WIKIPEDIA

Elo rating system

The **Elo**[a] **rating system** is a method for calculating the relative skill levels of players in zero-sum games such as chess or esports. It is named after its creator Arpad Elo, a Hungarian-American physics professor.


Arpad Elo, the inventor of the Elo rating system

The Elo system was invented as an improved chess-rating system over the previously used Harkness system,[1] but is also used as a rating system in association football (soccer), American football, baseball, basketball, pool, various board games and esports, and, more recently, large language models.

The difference in the ratings between two players serves as a predictor of the outcome of a match. Two players with equal ratings who play against each other are expected to score an equal number of wins. A player whose rating is 100 points greater than their opponent's is expected to score 64%; if the difference is 200 points, then the expected score for the stronger player is 76%.[2]

A player's Elo rating is a number that may change depending on the outcome of rated games played. After every game, the winning player takes points from the losing one. The difference between the ratings of the winner and loser determines the total number of points gained or lost after a game. If the higher-rated player wins, then only a few rating points will be taken from the lower-rated player. However, if the lower-rated player scores an upset win, many rating points will be transferred. The lower-rated player will also gain a few points from the higher rated player in the event of a draw. This means that this rating system is self-correcting. Players whose ratings are too low or too high should, in the long run, do better or worse correspondingly than the rating system predicts and thus gain or lose rating points until the ratings reflect their true playing strength.

Elo ratings are comparative only, and are valid only within the rating pool in which they were calculated, rather than being an absolute measure of a player's strength.

While Elo-like systems are widely used in two-player settings, variations have also been applied to multiplayer competitions.[3]

History

Arpad Elo was a chess master and an active participant in the United States Chess Federation (USCF) from its founding in 1939.[4] The USCF used a numerical ratings system devised by Kenneth Harkness to enable members to track their individual progress in terms other than tournament wins and losses. The Harkness system was reasonably fair, but in some circumstances gave rise to ratings many observers considered inaccurate.

On behalf of the USCF, Elo devised a new system with a more sound statistical basis.[5] At about the same time, György Karoly and Roger Cook independently developed a system based on the same principles for the New South Wales Chess Association.[6]

Elo's system replaced earlier systems of competitive rewards with a system based on statistical estimation. Rating systems for many sports award points in accordance with subjective evaluations of the 'greatness' of certain achievements. For example, winning an important golf tournament might be worth an arbitrarily chosen five times as many points as winning a lesser tournament.

A statistical endeavor, by contrast, uses a model that relates the game results to underlying variables representing the ability of each player.

Elo's central assumption was that the chess performance of each player in each game is a normally distributed random variable. Although a player might perform significantly better or worse from one game to the next, Elo assumed that the mean value of the performances of any given player changes only slowly over time. Elo thought of a player's true skill as the mean of that player's performance random variable.

A further assumption is necessary because chess performance in the above sense is still not measurable. One cannot look at a sequence of moves and derive a number to represent that player's skill. Performance can only be inferred from wins, draws, and losses. Therefore, a player who wins a game is assumed to have performed at a higher level than the opponent for that game. Conversely, a losing player is assumed to have performed at a lower level. If the game ends in a draw, the two players are assumed to have performed at nearly the same level.

Elo did not specify exactly how close two performances ought to be to result in a draw as opposed to a win or loss. Actually, there is a probability of a draw that is dependent on the performance differential, so this latter is more of a confidence interval than any deterministic frontier. And while he thought it was likely that players might have different standard deviations to their performances, he made a simplifying assumption to the contrary.

To simplify computation even further, Elo proposed a straightforward method of estimating the variables in his model (i.e., the true skill of each player). One could calculate relatively easily from tables how many games players would be expected to win based on comparisons of their ratings to those of their opponents. The ratings of a player who won more games than expected would be adjusted upward, while those of a player who won fewer than expected would be adjusted downward. Moreover, that adjustment was to be in linear proportion to the number of wins by which the player had exceeded or fallen short of their expected number.[7]

From a modern perspective, Elo's simplifying assumptions are not necessary because computing power is inexpensive and widely available. Several people, most notably Mark Glickman, have proposed using more sophisticated statistical machinery to estimate the same variables. On the other hand, the computational simplicity of the Elo system has proven to be one of its greatest assets. With the aid of a pocket calculator, an informed chess competitor can calculate to within one point what their next officially published rating will be, which helps promote a perception that the ratings are fair.

Implementing Elo's scheme

The USCF implemented Elo's suggestions in 1960,[8] and the system quickly gained recognition as being both fairer and more accurate than the Harkness rating system. Elo's system was adopted by the World Chess Federation (FIDE) in 1970.[9] Elo described his work in detail in *The Rating of Chessplayers, Past and Present*, first published in 1978.[10]

Subsequent statistical tests have suggested that chess performance is almost certainly not distributed as a normal distribution, as weaker players have greater winning chances than Elo's model predicts.[11][12] In paired comparison data, there is often very little practical difference in whether it is assumed that the differences in players' strengths are normally or logistically distributed. Mathematically, however, the logistic function is more convenient to work with than the normal distribution.[13] FIDE continues to use the rating difference table as proposed by Elo.[14]:table 8.1b

The development of the Percentage Expectancy Table (table 2.11) is described in more detail by Elo as follows:[15]

> The normal probabilities may be taken directly from the standard tables of the areas under the normal curve when the difference in rating is expressed as a z score. Since the standard deviation σ of individual performances is defined as 200 points, the standard deviation σ' of the differences in performances becomes σ√2 or 282.84. The z value of a difference then is D / 282.84. This will then divide the area under the curve into two parts, the larger giving P for the higher rated player and the smaller giving P for the lower rated player.
>
> For example, let $D = 160$. Then $z = 160 / 282.84 = .566$. The table gives .7143 and .2857 as the areas of the two portions under the curve. These probabilities are rounded to two figures in table 2.11.

The table is actually built with standard deviation $200(10/7)$ as an approximation for $200\sqrt{2}$.

The normal and logistic distributions are, in a way, arbitrary points in a spectrum of distributions which would work well. In practice, both of these distributions work very well for a number of different games.

Different ratings systems

The phrase "Elo rating" is often used to mean a player's chess rating as calculated by FIDE. However, this usage may be confusing or misleading because Elo's general ideas have been adopted by many organizations, including the USCF (before FIDE), many other national chess federations, the short-lived Professional Chess Association (PCA), and online chess servers including the Internet Chess Club (ICC), Free Internet Chess Server (FICS), Lichess, Chess.com, and Yahoo! Games. Each organization has a unique implementation, and none of them follows Elo's original suggestions precisely.

Instead one may refer to the organization granting the rating. For example: "As of April 2018, Tatev Abrahamyan had a FIDE rating of 2366 and a USCF rating of 2473." The Elo ratings of these various organizations are not always directly comparable, since Elo ratings measure the results within a closed pool of players rather than absolute skill.

FIDE ratings

For top players, the most important rating is their FIDE rating. FIDE has issued the following lists:

- From 1971 to 1980, one list a year was issued.
- From 1981 to 2000, two lists a year were issued, in January and July.
- From July 2000 to July 2009, four lists a year were issued, at the start of January, April, July and October.
- From July 2009 to July 2012, six lists a year were issued, at the start of January, March, May, July, September and November.
- Since July 2012, the list has been updated monthly.

The following analysis of the July 2015 FIDE rating list gives a rough impression of what a given FIDE rating means in terms of world ranking:

- 5,323 players had an active rating in the range 2200 to 2299, which is usually associated with the Candidate Master title.
- 2,869 players had an active rating in the range 2300 to 2399, which is usually associated with the FIDE Master title.
- 1,420 players had an active rating between 2400 and 2499, most of whom had either the International Master or the International Grandmaster title.
- 542 players had an active rating between 2500 and 2599, most of whom had the International Grandmaster title.
- 187 players had an active rating between 2600 and 2699, all of whom had the International Grandmaster title.
- 40 players had an active rating between 2700 and 2799.
- 4 players had an active rating of over 2800. (Magnus Carlsen was rated 2853, and 3 players were rated between 2814 and 2816).

The highest ever FIDE rating was 2882, which Magnus Carlsen had on the May 2014 list. A list of the highest-rated players ever is at Comparison of top chess players throughout history.

Performance rating

Performance rating or special rating is a hypothetical rating that would result from the games of a single event only. Some chess organizations [16]:p. 8 use the "algorithm of 400" to calculate performance rating. According to this algorithm, performance rating for an event is calculated in the following way:

1. For each win, add your opponent's rating plus 400,
2. For each loss, add your opponent's rating minus 400,
3. And divide this sum by the number of played games.

Example: 2 wins (opponents w & x), 2 losses (opponents y & z)

$$\frac{w+400+x+400+y-400+z-400}{4}$$

$$\frac{w+x+y+z+400(2)-400(2)}{4}$$

This can be expressed by the following formula:

$$\textbf{performance rating} = \frac{\textbf{total of opponents' ratings } + 400 \times (\textbf{wins} - \textbf{losses})}{\textbf{games}}$$

Example: If you beat a player with an Elo rating of 1000,

$$\textbf{performance rating} = \frac{1000 + 400 \times (1)}{1} = 1400$$

If you beat two players with Elo ratings of 1000,

$$\textbf{performance rating} = \frac{2000 + 400 \times (2)}{2} = 1400$$

If you draw,

$$\textbf{performance rating} = \frac{1000 + 400 \times (0)}{1} = 1000$$

This is a simplification, but it offers an easy way to get an estimate of PR (performance rating).

FIDE, however, calculates performance rating by means of the formula

$$\textbf{performance rating} = \textbf{average of opponents' ratings} + d_p,$$

where "rating difference" d_p is based on a player's tournament percentage score p, which is then used as the key in a lookup table where p is simply the number of points scored divided by the number of games played. Note that, in case of a perfect or no score d_p is 800.

p	d_p
1.00	+800
0.99	+677
0.9	+366
0.8	+240
0.7	+149
0.6	+72
0.5	0
0.4	−72
0.3	−149
0.2	−240
0.1	−366
0.01	−677
0.00	−800

Live ratings

FIDE updates its ratings list at the beginning of each month. In contrast, the unofficial "Live ratings" calculate the change in players' ratings after every game. These Live ratings are based on the previously published FIDE ratings, so a player's Live rating is intended to correspond to what the FIDE rating would be if FIDE were to issue a new list that day.

Although Live ratings are unofficial, interest arose in Live ratings in August/September 2008 when five different players took the "Live" No. 1 ranking.[17]

The unofficial live ratings of players over 2700 were published and maintained by Hans Arild Runde at the Live Rating website (https://web.archive.org/web/20080603001814/http://chess.liverating.org/) until August 2011. Another website, 2700chess.com (http://www.2700chess.com), has been maintained since May 2011 by Artiom Tsepotan, which covers the top 100 players as well as the top 50 female players.

Rating changes can be calculated manually by using the FIDE ratings change calculator.[18] All top players have a K-factor of 10, which means that the maximum ratings change from a single game is a little less than 10 points.

United States Chess Federation ratings

The United States Chess Federation (USCF) uses its own classification of players:[19]

- 2400 and above: Senior Master
- 2200–2399: National Master
  - 2200–2399 plus 300 games above 2200: Original Life Master[20]
- 2000–2199: Expert or Candidate Master
- 1800–1999: Class A
- 1600–1799: Class B
- 1400–1599: Class C
- 1200–1399: Class D

- 1000–1199: Class E
- 800–999: Class F
- 600–799: Class G
- 400–599: Class H
- 200–399: Class I
- 100–199: Class J

The K-factor used by the USCF

The *K-factor*, in the USCF rating system, can be estimated by dividing 800 by the effective number of games a player's rating is based on (N_e) plus the number of games the player completed in a tournament (m).[21]

$$K = \frac{800}{N_e + m}$$

Rating floors

The USCF maintains an absolute rating floor of 100 for all ratings. Thus, no member can have a rating below 100, no matter their performance at USCF-sanctioned events. However, players can have higher individual absolute rating floors, calculated using the following formula:

$$AF = \min\{100 + 4N_W + 2N_D + N_R, 150\}$$

where N_W is the number of rated games won, N_D is the number of rated games drawn, and N_R is the number of events in which the player completed three or more rated games.

Higher rating floors exist for experienced players who have achieved significant ratings. Such higher rating floors exist, starting at ratings of 1200 in 100-point increments up to 2100 (1200, 1300, 1400, ..., 2100). A rating floor is calculated by taking the player's peak established rating, subtracting 200 points, and then rounding down to the nearest rating floor. For example, a player who has reached a peak rating of 1464 would have a rating floor of $1464 - 200 = 1264$, which would be rounded down to 1200. Under this scheme, only Class C players and above are capable of having a higher rating floor than their absolute player rating. All other players would have a floor of at most 150.

There are two ways to achieve higher rating floors other than under the standard scheme presented above. If a player has achieved the rating of Original Life Master, their rating floor is set at 2200. The achievement of this title is unique in that no other recognized USCF title will result in a new floor. For players with ratings below 2000, winning a cash prize of $2,000 or more raises that player's rating floor to the closest 100-point level that would have disqualified the player for participation in the tournament. For example, if a player won $4,000 in a 1750-and-under tournament, they would now have a rating floor of 1800.

Theory

Pairwise comparisons form the basis of the Elo rating methodology.[22] Elo made references to the papers of Good,[23] David,[24] Trawinski and David,[25] and Buhlman and Huber.[26]

Mathematical details

Performance is not measured absolutely; it is inferred from wins, losses, and draws against other players. Players' ratings depend on the ratings of their opponents and the results scored against them. The difference in rating between two players determines an estimate for the expected score between them. Both the average and the spread of ratings can be arbitrarily chosen. The USCF initially aimed for an average club player to have a rating of 1500 and Elo suggested scaling ratings so that a difference of 200 rating points in chess would mean that the stronger player has an *expected score* of approximately 0.75.

A player's *expected score* is their probability of winning plus half their probability of drawing. Thus, an expected score of 0.75 could represent a 75% chance of winning, 25% chance of losing, and 0% chance of drawing. On the other extreme it could represent a 50% chance of winning, 0% chance of losing, and 50% chance of drawing. The probability of drawing, as opposed to having a decisive result, is not specified in the Elo system. Instead, a draw is considered half a win and half a loss. In practice, since the true strength of each player is unknown, the expected scores are calculated using the player's current ratings as follows.

If player A has a rating of R_A and player B a rating of R_B, the exact formula (using the logistic curve with base 10)[27] for the expected score of player A is

$$E_\mathrm{A} = \frac{1}{1 + 10^{(R_\mathrm{B} - R_\mathrm{A})/400}}\ .$$

Similarly, the expected score for player B is

$$E_\mathrm{B} = \frac{1}{1 + 10^{(R_\mathrm{A} - R_\mathrm{B})/400}}\ .$$

This could also be expressed by

$$E_\mathrm{A} = \frac{Q_\mathrm{A}}{Q_\mathrm{A} + Q_\mathrm{B}}$$

and

$$E_\mathrm{B} = \frac{Q_\mathrm{B}}{Q_\mathrm{A} + Q_\mathrm{B}}\ ,$$

where $Q_\mathrm{A} = 10^{R_\mathrm{A}/400}$, and $Q_\mathrm{B} = 10^{R_\mathrm{B}/400}$. Note that in the latter case, the same denominator applies to both expressions, and it is plain that $E_\mathrm{A} + E_\mathrm{B} = 1$. This means that by studying only the numerators, we find out that the expected score for player A is $Q_\mathrm{A}/Q_\mathrm{B}$ times the expected score for player B. It then follows that for each 400 rating points of advantage over the opponent, the expected score is magnified ten times in comparison to the opponent's expected score.

When a player's actual tournament scores exceed their expected scores, the Elo system takes this as evidence that player's rating is too low, and needs to be adjusted upward. Similarly, when a player's actual tournament scores fall short of their expected scores, that player's rating is adjusted downward. Elo's original suggestion, which is still widely used, was a simple linear adjustment proportional to the amount by which a player over-performed or under-performed their expected score. The maximum possible adjustment per game, called the K-factor, was set at $K = 16$ for masters and $K = 32$ for weaker players.

Suppose player A (again with rating R_A) was expected to score E_A points but actually scored S_A points. The formula for updating that player's rating is

$$R'_\mathrm{A} = R_\mathrm{A} + K \cdot (S_\mathrm{A} - E_\mathrm{A})\ .$$[1]

This update can be performed after each game or each tournament, or after any suitable rating period.

An example may help to clarify:

> Suppose player A has a rating of 1613 and plays in a five-round tournament. They lose to a player rated 1609, draw with a player rated 1477, defeat a player rated 1388, defeat a player rated 1586, and lose to a player rated 1720. The player's actual score is $(0 + 0.5 + 1 + 1 + 0) = 2.5$. The expected score, calculated according to the formula above, was $(0.51 + 0.69 + 0.79 + 0.54 + 0.35) = 2.88$.
>
> Therefore, the player's new rating is $[1613 + 32 \cdot (2.5 - 2.88)] = 1601$, assuming that a K-factor of 32 is used. Equivalently, each game the player can be said to have put an ante of K times their expected score for the game into a pot, the opposing player does likewise, and the winner collects the full pot of value K; in the event of a draw, the players split the pot and receive $\frac{1}{2}K$ points each.

> Note that while two wins, two losses, and one draw may seem like a par score, it is worse than expected for player A because their opponents were lower rated on average. Therefore, player A is slightly penalized. If player A had scored two wins, one loss, and two draws, for a total score of three points, that would have been slightly better than expected, and the player's new rating would have been $[1613 + 32 \cdot (3 - 2.88)] = 1617$.

This updating procedure is at the core of the ratings used by FIDE, USCF, Yahoo! Games, the Internet Chess Club (ICC) and the Free Internet Chess Server (FICS). However, each organization has taken a different approach to dealing with the uncertainty inherent in the ratings, particularly the ratings of newcomers, and to dealing with the problem of ratings inflation/deflation. New players are assigned provisional ratings, which are adjusted more drastically than established ratings.

The principles used in these rating systems can be used for rating other competitions—for instance, international football matches.

Elo ratings have also been applied to games without the possibility of draws, and to games in which the result can also have a quantity (small/big margin) in addition to the quality (win/loss). See Go rating with Elo for more.

Suggested modification

In 2011 after analyzing 1.5 million FIDE rated games, Jeff Sonas demonstrated according to the Elo formula, two players having a rating difference of X actually have a true difference of around $X(5/6)$. Likewise, one can leave the rating difference alone and divide by 480 instead of 400. Since the Elo formula is overestimating the stronger player's win probability, stronger players are losing points against weaker players despite playing at their true strength. Likewise, weaker players gain points against stronger players. When the modification is applied, observed win rates deviate by less than 0.1% away from prediction, while traditional Elo can be 4% off the predicted rate. [28]

Most accurate distribution model

The first mathematical concern addressed by the USCF was the use of the normal distribution. They found that this did not accurately represent the actual results achieved, particularly by the lower rated players. Instead they switched to a logistic distribution model, which the USCF found provided a better fit for the actual results achieved.[29] FIDE also uses an approximation to the logistic distribution.[14]

Most accurate K-factor

The second major concern is the correct "K-factor" used. The chess statistician Jeff Sonas believes that the original $K = 10$ value (for players rated above 2400) is inaccurate in Elo's work. If the K-factor coefficient is set too large, there will be too much sensitivity to just a few, recent events, in terms of a large number of points exchanged in each game. And if the K-value is too low, the sensitivity will be minimal, and the system will not respond quickly enough to changes in a player's actual level of performance.

Elo's original K-factor estimation was made without the benefit of huge databases and statistical evidence. Sonas indicates that a K-factor of 24 (for players rated above 2400) may be both more accurate as a predictive tool of future performance and be more sensitive to performance.[30]

Certain Internet chess sites seem to avoid a three-level K-factor staggering based on rating range. For example, the ICC seems to adopt a global $K = 32$ except when playing against provisionally rated players.

The USCF (which makes use of a logistic distribution as opposed to a normal distribution) formerly staggered the K-factor according to three main rating ranges:

K-factor	Used for players with ratings ...
$K = 32$	below 2100
$K = 24$	between 2100 and 2400
$K = 16$	above 2400

Currently, the USCF uses a formula that calculates the K-factor based on factors including the number of games played and the player's rating. The K-factor is also reduced for high rated players if the event has shorter time controls.[16]

FIDE uses the following ranges:[31]

K-factor	Used for players with ratings ...
$K = 40$	for a player new to the rating list until the completion of events with a total of 30 games, and for all players until their 18th birthday, as long as their rating remains under 2300.
$K = 20$	for players who have always been rated under 2400.
$K = 10$	for players with any published rating of at least 2400 and at least 30 games played in previous events. Thereafter it remains permanently at 10.

FIDE used the following ranges before July 2014:[32]

K-factor	Used for players with ratings ...
$K = 30$ (was 25)	for a player new to the rating list until the completion of events with a total of 30 games.[33]
$K = 15$	for players who have always been rated under 2400.
$K = 10$	for players with any published rating of at least 2400 and at least 30 games played in previous events. Thereafter it remains permanently at 10.

The gradation of the K-factor reduces rating change at the top end of the rating range, reducing the possibility for rapid rise or fall of rating for those with a rating high enough to reach a low K-factor.

In theory, this might apply equally to online chess players and over-the-board players, since it is more difficult for all players to raise their rating after their rating has become high and their K-factor consequently reduced. However, when playing online, 2800+ players can more easily raise their rating by simply selecting opponents with high ratings – on the ICC playing site, a grandmaster may play a string of different opponents who are all rated over 2700.[34] In over-the-board events, it would only be in very high level all-play-all events that a player would be able to engage that number of 2700+ opponents. In a normal, open, Swiss-paired chess tournament, frequently there would be many opponents rated less than 2500, reducing the ratings gains possible from a single contest for a high-rated player.

Formal derivation for win/loss games

The above expressions can be now formally derived by exploiting the link between the Elo rating and the stochastic gradient update in the logistic regression.[35][36]

If we assume that the game results are binary, that is, only a win or a loss can be observed, the problem can be addressed via logistic regression, where the games results are dependent variables, the players' ratings are independent variables, and the model relating both is probabilistic: the probability of the player A winning the game is modeled as

$$\Pr\{\mathsf{A\ wins}\} = \sigma(r_{\mathsf{A,B}}), \quad \sigma(r) = \frac{1}{1 + 10^{-r/s}},$$

where

$$r_{\mathsf{A,B}} = (R_{\mathsf{A}} - R_{\mathsf{B}})$$

denotes the difference of the players' ratings, and we use a scaling factor $s = 400$, and, by law of total probability

$$\Pr\{\mathsf{B\ wins}\} = 1 - \sigma(r_{\mathsf{A,B}}) = \sigma(-r_{\mathsf{A,B}}).$$

The log loss is then calculated as

$$\ell = \begin{cases} -\log \sigma(r_{\mathsf{A,B}}) & \text{if A wins,} \\ -\log \sigma(-r_{\mathsf{A,B}}) & \text{if B wins,} \end{cases}$$

and, using the stochastic gradient descent the log loss is minimized as follows:

$$R_\mathsf{A} \leftarrow R_\mathsf{A} - \eta\frac{\mathrm{d}\ell}{\mathrm{d}R_\mathsf{A}},$$

$$R_\mathsf{B} \leftarrow R_\mathsf{B} - \eta\frac{\mathrm{d}\ell}{\mathrm{d}R_\mathsf{B}}.$$

where η is the adaptation step.

Since $\frac{\mathrm{d}}{\mathrm{d}r}\log\sigma(r) = \frac{\log 10}{s}\sigma(-r)$, $\frac{\mathrm{d}r_{\mathsf{A},\mathsf{B}}}{\mathrm{d}R_\mathsf{A}} = 1$, and $\frac{\mathrm{d}r_{\mathsf{A},\mathsf{B}}}{\mathrm{d}R_\mathsf{B}} = -1$, the adaptation is then written as follows

$$R_\mathsf{A} \leftarrow \begin{cases} R_\mathsf{A} + K\sigma(-r_{\mathsf{A},\mathsf{B}}) & \textbf{if A wins} \\ R_\mathsf{A} - K\sigma(r_{\mathsf{A},\mathsf{B}}) & \textbf{if B wins,} \end{cases}$$

which may be compactly written as

$$R_\mathsf{A} \leftarrow R_\mathsf{A} + K(S_\mathsf{A} - E_\mathsf{A})$$

where $K = \eta\log 10/s$ is the new adaptation step which absorbs η and s, $S_\mathsf{A} = 1$ if A wins and $S_\mathsf{A} = 0$ if B wins, and the expected score is given by $E_\mathsf{A} = \sigma(r_{\mathsf{A},\mathsf{B}})$.

Analogously, the update for the rating R_B is

$$R_\mathsf{B} \leftarrow R_\mathsf{B} + K(S_\mathsf{B} - E_\mathsf{B}).$$

Formal derivation for win/draw/loss games

Since the very beginning, the Elo rating has been also used in chess where we observe wins, losses or draws and, to deal with the latter a fractional score value, $S_\mathsf{A} = 0.5$, is introduced. We note, however, that the scores $S_\mathsf{A} = 1$ and $S_\mathsf{A} = 0$ are merely indicators to the events when the player A wins or loses the game. It is, therefore, not immediately clear what is the meaning of the fractional score. Moreover, since we do not specify explicitly the model relating the rating values R_A and R_B to the probability of the game outcome, we cannot say what the probability of the win, the loss, or the draw is.

To address these difficulties, and to derive the Elo rating in the ternary games, we will define the explicit probabilistic model of the outcomes. Next, we will minimize the log loss via stochastic gradient.

Since the loss, the draw, and the win are ordinal variables, we should adopt the model which takes their ordinal nature into account, and we use the so-called adjacent categories model which may be traced to the Davidson's work[37]

$$\Pr\{\textbf{A wins}\} = \sigma(r_{\mathsf{A},\mathsf{B}}; \kappa),$$

$$\Pr\{\textbf{B wins}\} = \sigma(-r_{\mathsf{A},\mathsf{B}}; \kappa),$$

$$\Pr\{\textbf{A draws}\} = \kappa\sqrt{\sigma(r_{\mathsf{A},\mathsf{B}}; \kappa)\sigma(-r_{\mathsf{A},\mathsf{B}}; \kappa)},$$

where

$$\sigma(r;\kappa) = \frac{10^{r/s}}{10^{-r/s} + \kappa + 10^{r/s}}$$

and $\kappa \geq 0$ is a parameter. Introduction of a free parameter should not be surprising as we have three possible outcomes and thus, an additional degree of freedom should appear in the model. In particular, with $\kappa = 0$ we recover the model underlying the logistic regression

$$\Pr\{\textbf{A wins}\} = \sigma(r_{\mathsf{A},\mathsf{B}}; 0) = \frac{10^{r_{\mathsf{A},\mathsf{B}}/s}}{10^{-r_{\mathsf{A},\mathsf{B}}/s} + 10^{r_{\mathsf{A},\mathsf{B}}/s}} = \frac{1}{1 + 10^{-r_{\mathsf{A},\mathsf{B}}/s'}},$$

where $s' = s/2$.

Using the ordinal model defined above, the log loss is now calculated as

$$\ell = \begin{cases} -\log\sigma(r_{\mathsf{A},\mathsf{B}}; \kappa) & \textbf{if A wins,} \\ -\log\sigma(-r_{\mathsf{A},\mathsf{B}}; \kappa) & \textbf{if B wins,} \\ -\log\kappa - \frac{1}{2}\log\sigma(r_{\mathsf{A},\mathsf{B}}; \kappa) - \frac{1}{2}\log\sigma(-r_{\mathsf{A},\mathsf{B}}; \kappa) & \textbf{if A draw,} \end{cases}$$

which may be compactly written as

$$\ell = -(S_\mathsf{A} + \frac{1}{2}D)\log\sigma(r_{\mathsf{A},\mathsf{B}}; \kappa) - (S_\mathsf{B} + \frac{1}{2}D)\log\sigma(-r_{\mathsf{A},\mathsf{B}}; \kappa) - D\log\kappa$$

where $S_\mathsf{A} = 1$ iff A wins, $S_\mathsf{B} = 1$ iff B wins, and $D = 1$ iff A draws.

As before, we need the derivative of $\log\sigma(r;\kappa)$ which is given by

$$\frac{\mathrm{d}}{\mathrm{d}r}\log\sigma(r;\kappa) = \frac{2\log 10}{s}[1 - g(r;\kappa)],$$

where

$$g(r;\kappa) = \frac{10^{r/s} + \kappa/2}{10^{-r/s} + \kappa + 10^{r/s}}.$$

Thus, the derivative of the log loss with respect to the rating R_A is given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}R_\mathsf{A}}\ell &= -\frac{2\log 10}{s}\left((S_\mathsf{A} + 0.5D)[1 - g(r_{\mathsf{A},\mathsf{B}};\kappa)] - (S_\mathsf{B} + 0.5D)g(r_{\mathsf{A},\mathsf{B}};\kappa)\right) \\ &= -\frac{2\log 10}{s}\left(S_\mathsf{A} + 0.5D - g(r_{\mathsf{A},\mathsf{B}};\kappa)\right),\end{aligned}$$

where we used the relationships $S_\mathsf{A} + S_\mathsf{B} + D = 1$ and $g(-r;\kappa) = 1 - g(r;\kappa)$.

Then, the stochastic gradient descent applied to minimize the log loss yields the following update for the rating R_A

$$R_\mathsf{A} \leftarrow R_\mathsf{A} + K(\hat{S}_\mathsf{A} - g(r_{\mathsf{A},\mathsf{B}};\kappa))$$

where $K = 2\eta\log 10/s$ and $\hat{S}_\mathsf{A} = S_\mathsf{A} + 0.5D$. Of course, $\hat{S}_\mathsf{A} = 1$ if A wins, $\hat{S}_\mathsf{A} = 0.5$ if A draws, and $\hat{S}_\mathsf{A} = 0$ if A loses. To recognize the origin in the model proposed by Davidson, this update is called an Elo-Davidson rating.[36]

The update for R_B is derived in the same manner as

$$R_{\mathsf{B}} \leftarrow R_{\mathsf{B}} + K(\hat{S}_{\mathsf{B}} - g(r_{\mathsf{B,A}}; \kappa)),$$

where $r_{\mathsf{B,A}} = R_{\mathsf{B}} - R_{\mathsf{A}} = -r_{\mathsf{A,B}}$.

We note that

$$\begin{aligned} E[\hat{S}_{\mathsf{A}}] &= \mathbf{Pr}\{\mathbf{A\ wins}\} + 0.5\,\mathbf{Pr}\{\mathbf{A\ draws}\} \\ &= \sigma(r_{\mathsf{A,B}}; \kappa) + 0.5\kappa\sqrt{\sigma(r_{\mathsf{A,B}}; \kappa)\sigma(-r_{\mathsf{A,B}}; \kappa)} \\ &= g(r_{\mathsf{A,B}}; \kappa) \end{aligned}$$

and thus, we obtain the rating update may be written as

$$R_{\mathsf{A}} \leftarrow R_{\mathsf{A}} + K(\hat{S}_{\mathsf{A}} - E_{\mathsf{A}}),$$

where $E_{\mathsf{A}} = E[\hat{S}_{\mathsf{A}}]$ and we obtained practically the same equation as in the Elo rating except that the expected score is given by $E_{\mathsf{A}} = g(r_{\mathsf{A,B}}; \kappa)$ instead of $E_{\mathsf{A}} = \sigma(r_{\mathsf{A,B}})$.

Of course, as noted above, for $\kappa = 0$, we have $g(r; 0) = \sigma(r)$ and thus, the Elo-Davidson rating is exactly the same as the Elo rating. However, this is of no help to understand the case when the draws are observed (we cannot use $\kappa = 0$ which would mean that the probability of draw is null). On the other hand, if we use $\kappa = 2$, we have

$$g(r; 2) = \frac{10^{r/s} + 1}{10^{-r/s} + 2 + 10^{r/s}} = \frac{1}{1 + 10^{-r/s}} = \sigma(r)$$

which means that, using $\kappa = 2$, the Elo-Davidson rating is exactly the same as the Elo rating.[36]

Practical issues

Game activity versus protecting one's rating

In some cases the rating system can discourage game activity for players who wish to protect their rating.[38] In order to discourage players from sitting on a high rating, a 2012 proposal by British Grandmaster John Nunn for choosing qualifiers to the chess world championship included an activity bonus, to be combined with the rating.[39]

Beyond the chess world, concerns over players avoiding competitive play to protect their ratings caused Wizards of the Coast to abandon the Elo system for *Magic: the Gathering* tournaments in favour of a system of their own devising called "Planeswalker Points".[40][41]

Selective pairing

A more subtle issue is related to pairing. When players can choose their own opponents, they can choose opponents with minimal risk of losing, and maximum reward for winning. Particular examples of players rated 2800+ choosing opponents with minimal risk and maximum possibility of rating gain include: choosing opponents that they know they can beat with a certain strategy; choosing opponents that they think are overrated; or avoiding playing strong players who are rated several hundred points below them, but may hold chess titles such as IM or GM. In the category of choosing overrated opponents, new entrants to the rating system who have played fewer than 50 games are in theory a convenient target as they may be overrated in their provisional rating. The ICC compensates for this issue by assigning a lower K-factor to the established player if they do win against a new rating entrant. The K-factor is actually a function of the number of rated games played by the new entrant.

Therefore, Elo ratings online still provide a useful mechanism for providing a rating based on the opponent's rating. Its overall credibility, however, needs to be seen in the context of at least the above two major issues described—engine abuse, and selective pairing of opponents.

The ICC has also recently introduced "auto-pairing" ratings which are based on random pairings, but with each win in a row ensuring a statistically much harder opponent who has also won x games in a row. With potentially hundreds of players involved, this creates some of the challenges of a major large Swiss event which is being fiercely contested, with round winners meeting round winners. This approach to pairing certainly maximizes the rating risk of the higher-rated participants, who may face very stiff opposition from players below 3000, for example. This is a separate rating in itself, and is under "1-minute" and "5-minute" rating categories. Maximum ratings achieved over 2500 are exceptionally rare.

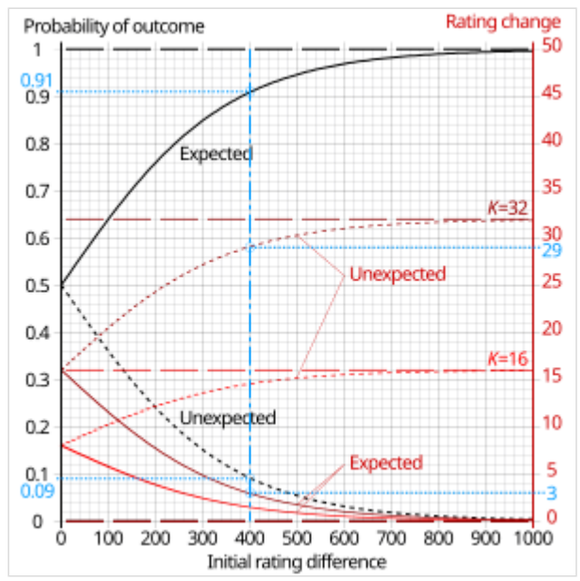
Ratings inflation and deflation

The term "inflation", applied to ratings, is meant to suggest that the level of playing strength demonstrated by the rated player is decreasing over time; conversely, "deflation" suggests that the level is advancing. For example, if there is inflation, a modern rating of 2500 means less than a historical rating of 2500, while the reverse is true if there is deflation. Using ratings to compare players between different eras is made more difficult when inflation or deflation are present. (See also Comparison of top chess players throughout history.)

Analyzing FIDE rating lists over time, Jeff Sonas suggests that inflation may have taken place since about 1985.[42] Sonas looks at the highest-rated players, rather than all rated players, and acknowledges that the changes in the distribution of ratings could have been caused by an increase of the standard of play at the highest levels, but looks for other causes as well.

The number of people with ratings over 2700 has increased. Around 1979 there was only one active player (Anatoly Karpov) with a rating this high. In 1992 Viswanathan Anand was only the 8th player in chess history to reach the 2700 mark at that point of time.[43] This increased to 15 players by 1994. 33 players had a 2700+ rating in 2009 and 44 as of September 2012. Only 14 players have ever broken a rating of 2800.

One possible cause for this inflation was the rating floor, which for a long time was at 2200, and if a player dropped below this they were struck from the rating list. As a consequence, players at a skill level just below the floor would only be on the rating list if they were overrated, and this would cause them to feed points into the rating pool.[42] In July 2000 the average rating of the top 100 was 2644. By July 2012 it had increased to 2703.[43]



Graphs of probabilities and Elo rating changes (for K=16 and 32) of expected outcome (solid curve) and unexpected outcome (dotted curve) vs initial rating difference. For example, player A starts with a 1400 rating and B with 1800 in a tournament using K = 32 (brown curves). The blue dash-dot line denotes the initial rating difference of 400 (1800 − 1400). The probability of B winning, the expected outcome, is 0.91 (intersection of black solid curve and blue line); if this happens, A's rating decreases by 3 (intersection of brown solid curve and blue line) to 1397 and B's increases by the same amount to 1803. Conversely, the probability of A winning, the unexpected outcome, is 0.09 (intersection of black dotted curve and blue line); if this happens, A's rating increases by 29 (intersection of brown dotted curve and blue line) to 1429 and B's decreases by the same amount to 1771.

Using a strong chess engine to evaluate moves played in games between rated players, Regan and Haworth analyze sets of games from FIDE-rated tournaments, and draw the conclusion that there had been little or no inflation from 1976 to 2009.[44]

In a pure Elo system, each game ends in an equal transaction of rating points. If the winner gains N rating points, the loser will drop by N rating points. This prevents points from entering or leaving the system when games are played and rated. However, players tend to enter the system as novices with a low rating and retire from the system as experienced players with a high rating. Therefore, in the long run a system with strictly equal transactions tends to result in rating deflation.[45]

In 1995, the USCF acknowledged that several young scholastic players were improving faster than the rating system was able to track. As a result, established players with stable ratings started to lose rating points to the young and underrated players. Several of the older established players were frustrated over what they considered an unfair rating decline, and some even quit chess over it.[46]

Combating deflation

Because of the significant difference in timing of when inflation and deflation occur, and in order to combat deflation, most implementations of Elo ratings have a mechanism for injecting points into the system in order to maintain relative ratings over time. FIDE has two inflationary mechanisms. First, performances below a "ratings floor" are not tracked, so a player with true skill below the floor can only be unrated or overrated, never correctly rated. Second, established and higher-rated players have a lower K-factor. New players have a $K = 40$, which drops to $K = 20$ after 30 played games, and to $K = 10$ when the player reaches 2400.[31] The current system in the United States includes a bonus point scheme which feeds rating points into the system in order to track improving players, and different K-values for different players.[46] Some methods, used in Norway for example, differentiate between juniors and seniors, and use a larger K-factor for the young players, even boosting the rating progress by 100% for when they score well above their predicted performance.[47]

Rating floors in the United States work by guaranteeing that a player will never drop below a certain limit. This also combats deflation, but the chairman of the USCF Ratings Committee has been critical of this method because it does not feed the extra points to the improving players. A possible motive for these rating floors is to combat sandbagging, i.e., deliberate lowering of ratings to be eligible for lower rating class sections and prizes.[46]

Ratings of computers

Human–computer chess matches between 1997 (Deep Blue versus Garry Kasparov) and 2006 demonstrated that chess computers are capable of defeating even the strongest human players. However, chess engine ratings are difficult to quantify, due to variable factors such as the time control and the hardware the program runs on, and also the fact that chess is not a fair game. The existence and magnitude of the first-move advantage in chess becomes very important at the computer level. Beyond some skill threshold, an engine with White should be able to force a draw on demand from the starting position even against perfect play, simply because White begins with too big an advantage to lose compared to the small magnitude of the errors it is likely to make. Consequently, such an engine is more or less guaranteed to score at least 25% even against perfect play. Differences in skill beyond a certain point could only be picked up if one does not begin from the usual starting position, but instead chooses a starting position that is only barely not lost for one side. Because of these factors, ratings depend on pairings and the openings selected.[48] Published engine rating lists such as CCRL are based on engine-only games on standard hardware configurations and are not directly comparable to FIDE ratings.

For some ratings estimates, see Chess engine § Ratings.

Use outside of chess

Other board and card games

- *Go*: The European Go Federation adopted an Elo-based rating system initially pioneered by the Czech Go Federation.
- *Backgammon*: The popular First Internet Backgammon Server (FIBS) calculates ratings based on a modified Elo system. New players are assigned a rating of 1500, with the best humans and bots rating over 2000. The same formula has been adopted by several other backgammon sites, such as Play65, DailyGammon, GoldToken and VogClub. VogClub sets a new player's rating at 1600. The UK Backgammon Federation uses the FIBS formula for its UK national ratings.[49]
- *Scrabble*: National Scrabble organizations compute normally distributed Elo ratings except in the United Kingdom, where a different system is used. The North American Scrabble Players Association has the largest rated population of active members, numbering about 2,000 as of early 2011. Lexulous also uses the Elo system.
- Despite questions of the appropriateness of using the Elo system to rate games in which luck is a factor, trading-card game manufacturers often use Elo ratings for their organized play efforts. The DCI (formerly Duelists' Convocation International) used Elo ratings for tournaments of *Magic: The Gathering* and other Wizards of the Coast games. However, the DCI abandoned this system in 2012 in favor of a new cumulative system of "Planeswalker Points", chiefly because of the above-noted concern that Elo encourages highly rated players to avoid playing to "protect their rating".[40][41] Pokémon USA uses the Elo system to rank its TCG organized play competitors.[50] Prizes for the top players in various regions included holidays and world championships invites until the 2011–2012 season, where awards were based on a system of Championship Points, their rationale being the same as the DCI's for *Magic: The Gathering*. Similarly, Decipher, Inc. used the Elo system for its ranked games such as *Star Trek Customizable Card Game* and *Star Wars Customizable Card Game*.

Athletic sports

The Elo rating system is used in the chess portion of chess boxing. In order to be eligible for professional chess boxing, one must have an Elo rating of at least 1600, as well as competing in 50 or more matches of amateur boxing or martial arts.

American college football used the Elo method as a portion of its Bowl Championship Series rating systems from 1998 to 2013 after which the BCS was replaced by the College Football Playoff. Jeff Sagarin of *USA Today* publishes team rankings for most American sports, which includes Elo system ratings for college football. The use of rating systems was effectively scrapped with the creation of the College Football Playoff in 2014.

In other sports, individuals maintain rankings based on the Elo algorithm. These are usually unofficial, not endorsed by the sport's governing body. The World Football Elo Ratings is an example of the method applied to men's football.[51] In 2006, Elo ratings were adapted for Major League Baseball teams by Nate Silver, then of Baseball Prospectus.[52] Based on this adaptation, both also made Elo-based Monte Carlo simulations of the odds of whether teams will make the playoffs.[53] In 2014, Beyond the Box Score, an SB Nation site, introduced an Elo ranking system for international baseball.[54]

In tennis, the Elo-based Universal Tennis Rating (UTR) rates players on a global scale, regardless of age, gender, or nationality. It is the official rating system of major organizations such as the Intercollegiate Tennis Association and World TeamTennis and is frequently used in segments on the Tennis Channel. The algorithm analyzes more than 8 million match results from over 800,000 tennis players worldwide. On May 8, 2018, Rafael Nadal—having won 46 consecutive sets in clay court matches—had a near-perfect clay UTR of 16.42.[55]

In pool, an Elo-based system called Fargo Rate is used to rank players in organized amateur and professional competitions.[56]

One of the few Elo-based rankings endorsed by a sport's governing body is the FIFA Women's World Rankings, based on a simplified version of the Elo algorithm, which FIFA uses as its official ranking system for national teams in women's football.

From the first ranking list after the 2018 FIFA World Cup, FIFA has used Elo for their FIFA World Rankings.[57]

In 2015, Nate Silver, editor-in-chief of the statistical commentary website FiveThirtyEight, and Reuben Fischer-Baum produced Elo ratings for every National Basketball Association team and season through the 2014 season.[58][59] In 2014 FiveThirtyEight created Elo-based ratings and win-projections for the American professional National Football League.[60]

The English Korfball Association rated teams based on Elo ratings, to determine handicaps for their cup competition for the 2011/12 season.

An Elo-based ranking of National Hockey League players has been developed.[61] The hockey-Elo metric evaluates a player's overall two-way play: scoring AND defense in both even strength and power-play/penalty-kill situations.

Rugbyleagueratings.com uses the Elo rating system to rank international and club rugby league teams.

Hemaratings.com was started in 2017 and uses a Glicko-2 algorithm to rank individual Historical European martial arts fencers worldwide in different categories such as Longsword,Rapier, historical Sabre and Sword & Buckler.[62]

Video games and online games

Many video games use modified Elo systems in competitive gameplay. The MOBA game League of Legends used an Elo rating system prior to the second season of competitive play.[63] The Esports game *Overwatch*, the basis of the unique Overwatch League professional sports organization, uses a derivative of the Elo system to rank competitive players with various adjustments made between competitive seasons.[64] *World of Warcraft* also previously used the Glicko-2 system to team up and compare Arena players, but now uses a system similar to Microsoft's TrueSkill.[65] The game *Puzzle Pirates* uses the Elo rating system to determine the standings in the various puzzles. This system is also used in FIFA Mobile for the Division Rivals modes. Another recent game to start using the Elo rating system is *AirMech*, using Elo[66] ratings for 1v1, 2v2, and 3v3 random/team matchmaking. *RuneScape 3* used the Elo system in the rerelease of the bounty hunter minigame in 2016.[67] *Mechwarrior Online* instituted an Elo system for its new "Comp Queue" mode, effective with the Jun 20, 2017 patch.[68] *Age of Empires II DE* and *Age of Empires III DE* are using the Elo system for their Leaderboard and matchmaking, with new players starting at Elo 1000.[69] Competitive Classic Tetris (Tetris played on the Nintendo Entertainment System) derives its ratings using a combination of players' personal best scores and a highly modified Elo system.[70]

Few video games use the original Elo rating system. According to Lichess, an online chess server, the Elo system is outdated, with Glicko-2 now being used by many chess organizations.[71] *PlayerUnknown's Battlegrounds* is one of the few video games that utilizes the very first Elo system. In *Guild Wars*, Elo ratings are used to record guild rating gained and lost through guild-versus-guild battles. In 1998, an online gaming ladder called *Clanbase*[72] was launched, which used the Elo scoring system to rank teams. The initial K-value was 30, but was changed to 5 in January 2007, then changed to 15 in July 2009.[73] The site later went offline in 2013.[74] A similar alternative site was launched in 2016 under the name *Scrimbase*,[75] which also used the Elo scoring system for ranking teams. Since 2005, *Golden Tee Live* has rated players based on the Elo system. New players start at 2100, with top players rating over 3000.[76]

Despite many video games using different systems for matchmaking, it is common for players of ranked video games to refer to all matchmaking ratings as *Elo*.

Other usage

The Elo rating system has been used in soft biometrics,[77] which concerns the identification of individuals using human descriptions. Comparative descriptions were utilized alongside the Elo rating system to provide robust and discriminative 'relative measurements', permitting accurate identification.

The Elo rating system has also been used in biology for assessing male dominance hierarchies,[78] and in automation and computer vision for fabric inspection.[79]

Moreover, online judge sites are also using Elo rating system or its derivatives. For example, Topcoder is using a modified version based on normal distribution,[80] while Codeforces is using another version based on logistic distribution.[81][82][83]

The Elo rating system has also been noted in dating apps, such as in the matchmaking app Tinder, which uses a variant of the Elo rating system.[84]

The YouTuber Marques Brownlee and his team used Elo rating system when they let people to vote between digital photos taken with different smartphone models launched in 2022.[85]

The Elo rating system has also been used in U.S. revealed preference college rankings, such as those by the digital credential firm Parchment.[86][87][88]

The Elo rating system has also been adopted to evaluate AI models. In 2021, Anthropic utilized the Elo system for ranking AI models in their research.[89] The LMSYS leaderboard briefly employed the Elo rating system to rank AI models[90] before transitioning to Bradley–Terry model.[91]

References in the media

The Elo rating system was featured prominently in the 2010 film *The Social Network* during the algorithm scene where Mark Zuckerberg released Facemash. In the scene Eduardo Saverin writes mathematical formulas for the Elo rating system on Zuckerberg's dormitory room window. Behind the scenes, the movie claims, the Elo system is employed to rank girls by their attractiveness. The equations driving the algorithm are shown briefly, written on the window;[92] however, they are slightly incorrect.

See also

- Elo hell
- Rating percentage index (RPI), another system that incorporates strength of opponents

Notes

a. This is written as "Elo", not "ELO", and is usually pronounced as /ˈiːloʊ/ or /ˈɛloʊ/ in English. The original name *Élő* is pronounced [ˈeːløː] in Hungarian.

References

Notes

1. Elo, Arpad E. (August 1967). "The Proposed USCF Rating System, Its Development, Theory, and Applications" (http://uscf1-nyc1.aodhosting.com/CL-AND-CR-ALL/CL-ALL/1967/1967_08.pdf#page=26) (PDF). *Chess Life*. **XXII** (8): 242–247.
2. Using the formula $100\% \,/\, (1 + 10^{-D/400})$ for D equal to 100 or 200.
3. Elo-MMR: A Rating System for Massive Multiplayer Competitions (https://cs.stanford.edu/people/paulliu/files/www-2021-elor.pdf)
4. Redman, Tim (July 2002). "Remembering Richard, Part II" (http://www.springfieldchessclub.com/icbarchive/ICB_2002_07.pdf) (PDF). Illinois Chess Bulletin. Archived (https://web.archive.org/web/20200630040943/http://www.springfieldchessclub.com/icbarchive/ICB_2002_07.pdf) (PDF) from the original on 2020-06-30. Retrieved 2020-06-30.
5. Elo, Arpad E. (March 5, 1960). "The USCF Rating System" (http://uscf1-nyc1.aodhosting.com/CL-AND-CR-ALL/CL-ALL/1960/1960_03_1.pdf) (PDF). *Chess Life*. **XIV** (13). USCF: 2.
6. Elo 1986, p. 4
7. Elo, Arpad E. (June 1961). "The USCF Rating System - A Scientific Achievement" (http://uscf1-nyc1.aodhosting.com/CL-AND-CR-ALL/CL-ALL/1961/1961_06.pdf#page=8) (PDF). *Chess Life*. **XVI** (6). USCF: 160–161.
8. "About the USCF" (http://www.uschess.org/about/about.php). United States Chess Federation. Archived (https://web.archive.org/web/20080926015601/http://www.uschess.org/about/about.php) from the original on 2008-09-26. Retrieved 2008-11-10.
9. Elo 1986, Preface to the First Edition
10. Elo 1986.
11. Elo 1986, ch. 8.73.
12. Glickman, Mark E., and Jones, Albyn C., "Rating the chess rating system" (http://www.glicko.net/research/chance.pdf) (1999), Chance, 12, 2, 21-28.
13. Glickman, Mark E. (1995), "A Comprehensive Guide to Chess Ratings". (http://www.glicko.net/research/acjpaper.pdf) A subsequent version of this paper appeared in the *American Chess Journal*, 3, pp. 59–102.
14. FIDE Rating Regulations effective from 1 July 2017 (https://handbook.fide.com/chapter/B022017). *FIDE Online (fide.com)* (Report). FIDE. Archived (https://web.archive.org/web/20191127231614/https://handbook.fide.com/chapter/B022017) from the original on 2019-11-27. Retrieved 2017-09-09.
15. Elo 1986, p159.
16. The US Chess Rating system (http://www.glicko.net/ratings/rating.system.pdf) (PDF) (Report). April 24, 2017. Archived (https://web.archive.org/web/20200207072639/http://www.glicko.net/ratings/rating.system.pdf) (PDF) from the original on 7 February 2020. Retrieved 16 February 2020 – via glicko.net.
17. Anand lost No. 1 to Morozevich (Chessbase, August 24 2008 (http://www.chessbase.com/newsdetail.asp?newsid=4860) Archived (https://web.archive.org/web/20080910150925/http://www.chessbase.com/newsdetail.asp?newsid=4860) 2008-09-10 at the Wayback Machine), then regained it, then Carlsen took No. 1 (Chessbase, September 5 2008 (http://www.chessbase.com/newsdetail.asp?newsid=4892) Archived (https://web.archive.org/web/20121109045914/http://chessbase.com/newsdetail.asp?newsid=4892) 2012-11-09 at the Wayback Machine), then Ivanchuk (Chessbase, September 11 2008 (http://www.chessbase.com/newsdetail.asp?newsid=4901) Archived (https://web.archive.org/web/20080913210432/http://www.chessbase.com/newsdetail.asp?newsid=4901) 2008-09-13 at the Wayback Machine), and finally Topalov (Chessbase, September 13 2008 (http://www.chessbase.com/newsdetail.asp?newsid=4908) Archived (https://web.archive.org/web/20080915230303/http://www.chessbase.com/newsdetail.asp?newsid=4908) 2008-09-15 at the Wayback Machine)
18. Administrator. "FIDE Chess Rating calculators: Chess Rating change calculator" (https://ratings.fide.com/calculator_rtd.phtml). *ratings.fide.com*. Archived (https://web.archive.org/web/20170928150121/https://ratings.fide.com/calculator_rtd.phtml) from the original on 2017-09-28. Retrieved 2017-09-28.
19. US Chess Federation (http://archive.uschess.org/ratings/ratedist.php) Archived (https://web.archive.org/web/20120618103954/http://archive.uschess.org/ratings/ratedist.php) 2012-06-18 at the Wayback Machine
20. USCF Glossary Quote:"a player who competes in over 300 games with a rating over 2200" (http://main.uschess.org/content/view/7327#Master) Archived (https://web.archive.org/web/20130308182917/http://main.uschess.org/content/view/7327#Master) 2013-03-08 at the Wayback Machine from The United States Chess Federation
21. "Approximating Formulas for the US Chess Rating System" (http://www.glicko.net/ratings/approx.pdf) Archived (https://web.archive.org/web/20191104083935/http://www.glicko.net/ratings/approx.pdf) 2019-11-04 at the Wayback Machine, United States Chess Federation, Mark Glickman, April 2017
22. Elo 1986, ch. 1.12.
23. Good, I.J. (1955). "On the Marking of Chessplayers". *The Mathematical Gazette*. **39** (330): 292–296. doi:10.2307/3608567 (https://doi.org/10.2307%2F3608567). JSTOR 3608567 (https://www.jstor.org/stable/3608567). S2CID 158885108 (https://api.semanticscholar.org/CorpusID:158885108).
24. David, H. A. (1959). "Tournaments and Paired Comparisons". *Biometrika*. **46** (1/2): 139–149. doi:10.2307/2332816 (https://doi.org/10.2307%2F2332816). JSTOR 2332816 (https://www.jstor.org/stable/2332816).
25. Trawinski, B.J.; David, H.A. (1963). "Selection of the Best Treatment in a Paired-Comparison Experiment" (https://doi.org/10.1214%2Faoms%2F1177704243). *Annals of Mathematical Statistics*. **34** (1): 75–91. doi:10.1214/aoms/1177704243 (https://doi.org/10.1214%2Faoms%2F1177704243).
26. Buhlmann, Hans; Huber, Peter J. (1963). "Pairwise Comparison and Ranking in Tournaments" (https://doi.org/10.1214%2Faoms%2F1177704161). *The Annals of Mathematical Statistics*. **34** (2): 501–510. doi:10.1214/aoms/1177704161 (https://doi.org/10.1214%2Faoms%2F1177704161).
27. Elo 1986, p. 141, ch. 8.4& Logistic probability as a rating basis
28. "The Elo rating system – correcting the expectancy tables" (https://en.chessbase.com/post/the-elo-rating-system-correcting-the-expectancy-tables). 30 March 2011.
29. Elo 1986, ch. 8.73
30. A key Sonas article is Sonas, Jeff. "The Sonas rating formula — better than Elo?" (http://www.chessbase.com/newsdetail.asp?newsid=562). *chessbase.com*. Archived (https://web.archive.org/web/20050305143008/http://www.chessbase.com/newsdetail.asp?newsid=562) from the original on 2005-03-05. Retrieved 2005-05-01.
31. FIDE Rating Regulations effective from 1 July 2014 (http://www.fide.com/fide/handbook.html?id=172&view=article). *FIDE Online (fide.com)* (Report). FIDE. 2014-07-01. Archived (https://web.archive.org/web/20140701031750/http://www.fide.com/fide/handbook.html?id=172&view=article) from the original on 2014-07-01. Retrieved 2014-07-01.
32. FIDE Rating Regulations valid from 1 July 2013 till 1 July 2014 (http://www.fide.com/fide/handbook.html?id=161&view=article). *FIDE Online (fide.com)* (Report). 2013-07-01. Archived (https://web.archive.org/web/20140715002648/http://www.fide.com/fide/handbook.html?id=161&view=article) from the original on 2014-07-15. Retrieved 2014-07-01.
33. "Changes to Rating Regulations" (https://web.archive.org/web/20120513170023/http://www.fide.com/component/content/article/1-fide-news/5421-changes-to-rating-regulations.html). *FIDE Online (fide.com)* (Press release). FIDE. 2011-07-21. Archived from the original (http://www.fide.com/component/content/article/1-fide-news/5421-changes-to-rating-regulations.html) on 2012-05-13. Retrieved 2012-02-19.
34. "*K*-factor" (https://web.archive.org/web/20120313023307/http://www.chessclub.com/help/k-factor). *Chessclub.com*. ICC Help. 2002-10-18. Archived from the original (http://www.chessclub.com/help/k-factor) on 2012-03-13. Retrieved 2012-02-19.
35. Kiraly, F.; Qian, Z. (2017). "Modelling Competitive Sports: Bradley-Terry-Elo Models for Supervised and On-Line Learning of Paired Competition Outcomes". arXiv:1701.08055 (https://arxiv.org/abs/1701.08055) [stat.ML (https://arxiv.org/archive/stat.ML)].

36. Szczecinski, Leszek; Djebbi, Aymen (2020-09-01). "Understanding draws in Elo rating algorithm" (https://www.degruyter.com/document/doi/10.1515/jqas-2019-0102/html?lang=en). *Journal of Quantitative Analysis in Sports*. **16** (3): 211–220. doi:10.1515/jqas-2019-0102 (https://doi.org/10.1515%2Fjqas-2019-0102). ISSN 1559-0410 (https://search.worldcat.org/issn/1559-0410). S2CID 219784913 (https://api.semanticscholar.org/CorpusID:219784913).
37. Davidson, Roger R. (1970). "On Extending the Bradley-Terry Model to Accommodate Ties in Paired Comparison Experiments" (https://www.jstor.org/stable/2283595). *Journal of the American Statistical Association*. **65** (329): 317–328. doi:10.2307/2283595 (https://doi.org/10.2307%2F2283595). ISSN 0162-1459 (https://search.worldcat.org/issn/0162-1459). JSTOR 2283595 (https://www.jstor.org/stable/2283595).
38. A Parent's Guide to Chess (http://www.chesscafe.com/text/skittles176.pdf) Archived (https://web.archive.org/web/20080528195206/http://www.chesscafe.com/text/skittles176.pdf) 2008-05-28 at the Wayback Machine *Skittles*, Don Heisman, Chesscafe.com, August 4, 2002
39. "Chess News – The Nunn Plan for the World Chess Championship" (http://www.chessbase.com/newsdetail.asp?newsid=2440). ChessBase.com. 8 June 2005. Archived (https://web.archive.org/web/20111119132830/http://www.chessbase.com/newsdetail.asp?newsid=2440) from the original on 2011-11-19. Retrieved 2012-02-19.
40. "Introducing Planeswalker Points" (https://web.archive.org/web/20110930202112/http://www.wizards.com/Magic/Magazine/Article.aspx?x=mtg%2Fdaily%2Ffeature%2F159b). September 6, 2011. Archived from the original (http://www.wizards.com/Magic/Magazine/Article.aspx?x=mtg/daily/feature/159b) on September 30, 2011. Retrieved September 9, 2011.
41. "Getting to the Points" (http://magic.wizards.com/en/articles/archive/week-was/getting-points-aaron-and-scott-2011-09-09). September 9, 2011. Archived (https://web.archive.org/web/20161018204636/http://magic.wizards.com/en/articles/archive/week-was/getting-points-aaron-and-scott-2011-09-09) from the original on October 18, 2016. Retrieved September 9, 2011.
42. Jeff Sonas (27 July 2009). "Rating inflation – its causes and possible cures" (https://en.chessbase.com/post/rating-inflation-its-causes-and-poible-cures). *chessbase.com*. Archived (https://web.archive.org/web/20131123073932/http://en.chessbase.com/post/rating-inflation-its-causes-and-poible-cures) from the original on 23 November 2013. Retrieved 27 August 2009.
43. "Viswanathan Anand" (http://www.chessgames.com/perl/chessplayer?pid=12088). Chessgames.com. Archived (https://web.archive.org/web/20130328031126/http://www.chessgames.com/perl/chessplayer?pid=12088) from the original on 2013-03-28. Retrieved 2012-08-14.
44. Regan, Kenneth; Haworth, Guy (2011-08-04). "Intrinsic Chess Ratings" (https://ojs.aaai.org/index.php/AAAI/article/view/7951). *Proceedings of the AAAI Conference on Artificial Intelligence*. **25** (1): 834–839. doi:10.1609/aaai.v25i1.7951 (https://doi.org/10.1609%2Faaai.v25i1.7951). ISSN 2374-3468 (https://search.worldcat.org/issn/2374-3468). S2CID 15489049 (https://api.semanticscholar.org/CorpusID:15489049). Archived (https://web.archive.org/web/20210420130735/https://ojs.aaai.org/index.php/AAAI/article/view/7951) from the original on 2021-04-20. Retrieved 2021-09-01.
45. Bergersen, Per A. "ELO-SYSTEMET" (https://wayback.archive-it.org/all/20130308184326/http://www.sjakk.no/nsf/elosystem_main.html) (in Norwegian). Norwegian Chess Federation. Archived from the original (http://www.sjakk.no/nsf/elosystem_main.html) on 8 March 2013. Retrieved 21 October 2013.
46. A conversation with Mark Glickman [1] (http://www.glicko.net/ratings/cl-article.pdf) Archived (https://web.archive.org/web/20110807205021/http://www.glicko.net/ratings/cl-article.pdf) 2011-08-07 at the Wayback Machine, Published in *Chess Life* October 2006 issue
47. "Elo-systemet" (https://web.archive.org/web/20131205025157/http://www.sjakk.no/nsf/elosystem_index.html). *Norges Sjakkforbund*. Archived from the original (http://www.sjakk.no/nsf/elosystem_index.html) on December 5, 2013. Retrieved 2009-08-23.
48. Larry Kaufman, Chess Board Options (2021), p. 179
49. "Backgammon Ratings Explained" (https://web.archive.org/web/20191114175326/http://results.ukbgf.com/explain-ratings). *results.ukbgf.com*. Archived from the original (https://results.ukbgf.com/explain-ratings) on 2019-11-14. Retrieved 2020-06-01.
50. "Play! Pokémon Glossary: Elo" (http://www.pokemon.com/us/play-pokemon/about/tournaments-glossary/#elo). Archived (https://web.archive.org/web/20150115173351/http://www.pokemon.com/us/play-pokemon/about/tournaments-glossary/#elo) from the original on January 15, 2015. Retrieved January 15, 2015.
51. Lyons, Keith (10 June 2014). "What are the World Football Elo Ratings?" (https://theconversation.com/what-are-the-world-football-elo-ratings-27851). *The Conversation*. Archived (https://web.archive.org/web/20190615101903/http://theconversation.com/what-are-the-world-football-elo-ratings-27851) from the original on 15 June 2019. Retrieved 3 July 2019.
52. Silver, Nate (2006-06-28). "Lies, Damned Lies: We are Elo?" (https://web.archive.org/web/20060822122806/http://baseballprospectus.com/article.php?articleid=5247). Archived from the original (https://www.baseballprospectus.com/news/article/5247/lies-damned-lies-we-are-elo/) on 2006-08-22. Retrieved 2023-01-13.
53. "Postseason Odds, ELO version" (http://www.baseballprospectus.com/statistics/ps_oddselo.php). Baseballprospectus.com. Archived (https://web.archive.org/web/20120307174118/http://www.baseballprospectus.com/statistics/ps_oddselo.php) from the original on 2012-03-07. Retrieved 2012-02-19.
54. Cole, Bryan (August 15, 2014). "Elo rankings for international baseball" (http://www.beyondtheboxscore.com/2014/8/15/5989787/elo-rankings-for-international-baseball). *Beyond the Box Score*. SB Nation. Archived (https://web.archive.org/web/20160102111843/http://www.beyondtheboxscore.com/2014/8/15/5989787/elo-rankings-for-international-baseball) from the original on 2 January 2016. Retrieved 4 November 2015.
55. "Is Rafa the GOAT of Clay?" (https://blog.universaltennis.com/2018/05/08/is-rafa-the-goat-of-clay/). 8 May 2018. Archived (https://web.archive.org/web/20210227064453/https://blog.universaltennis.com/2018/05/08/is-rafa-the-goat-of-clay/) from the original on 27 February 2021. Retrieved 22 August 2018.
56. "Fargo Rate" (https://fargorate.com/). Retrieved 31 March 2022.
57. "Revision of the FIFA/Coca-Cola World Ranking" (https://web.archive.org/web/20180612141237/https://resources.fifa.com/image/upload/revision-of-the-fifa-coca-cola-world-ranking.pdf?cloudid=jgxjkdrj1jfwyunjbkha) (PDF). FIFA. June 2018. Archived from the original (https://resources.fifa.com/image/upload/revision-of-the-fifa-coca-cola-world-ranking.pdf?cloudid=jgxjkdrj1jfwyunjbkha) (PDF) on 2018-06-12. Retrieved 2020-06-30.
58. Silver, Nate; Fischer-Baum, Reuben (May 21, 2015). "How We Calculate NBA Elo Ratings" (https://web.archive.org/web/20150523164940/http://fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/). FiveThirtyEight. Archived from the original (https://fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/) on 2015-05-23.
59. Reuben Fischer-Baum and Nate Silver, "The Complete History of the NBA," *FiveThirtyEight,* May 21, 2015.[2] (https://fivethirtyeight.com/interactives/the-complete-history-of-every-nba-team-by-elo/#bulls) Archived (https://web.archive.org/web/20150523115345/http://fivethirtyeight.com/interactives/the-complete-history-of-every-nba-team-by-elo/#bulls) 2015-05-23 at the Wayback Machine
60. Silver, Nate (September 4, 2014). "Introducing NFL Elo Ratings" (https://web.archive.org/web/20150912084549/https://fivethirtyeight.com/datalab/introducing-nfl-elo-ratings/). FiveThirtyEight. Archived from the original (https://fivethirtyeight.com/features/introducing-nfl-elo-ratings/) on September 12, 2015.

    Paine, Neil (September 10, 2015). "NFL Elo Ratings Are Back" (https://web.archive.org/web/20150911100901/http://fivethirtyeight.com/datalab/nfl-elo-ratings-are-back/). FiveThirtyEight. Archived from the original (https://fivethirtyeight.com/datalab/nfl-elo-ratings-are-back/) on September 11, 2015..
61. "Hockey Stats Revolution – How do teams pick players?" (http://www.hockeystatsrevolution.com/). *Hockey Stats Revolution*. Archived (https://web.archive.org/web/20161002040358/http://www.hockeystatsrevolution.com/) from the original on 2016-10-02. Retrieved 2016-09-29.
62. "About the Ratings - Hema Ratings" (https://hemaratings.com/about-ratings/). *Hemaratings*. Retrieved 2024-01-30.
63. "Matchmaking | LoL – League of Legends" (http://na.leagueoflegends.com/learn/gameplay/matchmaking). Na.leagueoflegends.com. 2010-07-06. Archived (https://web.archive.org/web/20120226001928/http://na.leagueoflegends.com/learn/gameplay/matchmaking) from the original on 2012-02-26. Retrieved 2012-02-19.
64. "Welcome to Season 8 of competitive play" (https://playoverwatch.com/en-us/blog/21363037). *PlayOverwatch.com*. Blizzard Entertainment. Archived (https://web.archive.org/web/20180312144148/https://playoverwatch.com/en-us/blog/21363037) from the original on 12 March 2018. Retrieved 11 March 2018.
65. "World of Warcraft Europe -> The Arena" (https://web.archive.org/web/20100923213409/http://www.wow-europe.com/en/info/basics/arena/index.html). Wow-europe.com. 2011-12-14. Archived from the original (http://www.wow-europe.com/en/info/basics/arena/index.html) on 2010-09-23. Retrieved 2012-02-19.
66. "AirMech developer explains why they use Elo" (https://www.carbongames.com/forums/viewtopic.php?p=83043#p83043). Archived (https://web.archive.org/web/20150217215610/https://www.carbongames.com/forums/viewtopic.php?p=83043#p83043) from the original on February 17, 2015. Retrieved January 15, 2015.
67. [3] (http://services.runescape.com/m=news/design-doc--bounty-hunter--deathmatch-pvp?acq_id=3001)
68. "MWO: News" (https://mwomercs.com/news/2017/06/1839-patch-notes-14120-20jun2017). *mwomercs.com*. Archived (https://web.archive.org/web/20180827174110/https://mwomercs.com/news/2017/06/1839-patch-notes-14120-20jun2017) from the original on 2018-08-27. Retrieved 2017-06-27.
69. "Age of Empires II: DE Leaderboards - Age of Empires" (https://www.ageofempires.com/stats/ageiide/). 14 November 2019. Archived (https://web.archive.org/web/20220127170906/https://www.ageofempires.com/stats/ageiide/) from the original on 27 January 2022. Retrieved 27 January 2022.
70. "List of the Best Tetris Players in the World (NES NTSC)" (https://listfist.com/list-of-the-best-tetris-players-in-the-world-nes-ntsc). 27 October 2020. Retrieved July 15, 2024.
71. "Frequently Asked Questions: ratings" (https://lichess.org/faq#ratings). *lichess.org*. Archived (https://web.archive.org/web/20190402042835/https://lichess.org/qa/55/how-old-is-lichess#ratings) from the original on 2019-04-02. Retrieved 2020-11-11.
72. "Wayback Machine record of Clanbase.com" (https://web.archive.org/web/20171105073119/http://www.clanbase.com/). Archived from the original (http://clanbase.com/) on 2017-11-05. Retrieved 2017-10-29.
73. "Guild ladder" (https://web.archive.org/web/20120301192358/http://wiki.guildwars.com/wiki/Guild_ladder). Wiki.guildwars.com. Archived from the original (http://wiki.guildwars.com/wiki/Guild_ladder) on 2012-03-01. Retrieved 2012-02-19.
74. "Clanbase farewell message" (http://clanbase.org/). Archived (https://web.archive.org/web/20131224094122/http://clanbase.org/) from the original on 2013-12-24. Retrieved 2017-10-29.
75. "Scrimbase Gaming Ladder" (https://scrimbase.com/about). Archived (https://web.archive.org/web/20171030003609/https://scrimbase.com/about) from the original on 2017-10-30. Retrieved 2017-10-29.
76. "Golden Tee Fan Player Rating Page" (http://www.goldenteefan.com/statistics/player-rating-handicap/). 26 December 2007. Archived (https://web.archive.org/web/20140101102859/http://www.goldenteefan.com/statistics/player-rating-handicap/) from the original on 2014-01-01. Retrieved 2013-12-31.
77. "Using Comparative Human Descriptions for Soft Biometrics" (http://eprints.soton.ac.uk/272922/1/IJCB%20(3).pdf) Archived (https://web.archive.org/web/20130308190221/http://eprints.soton.ac.uk/272922/1/IJCB%20%283%29.pdf) 2013-03-08 at the Wayback Machine, D.A. Reid and M.S. Nixon, International Joint Conference on Biometrics (IJCB), 2011
78. Pörschmann; et al. (2010). "Male reproductive success and its behavioural correlates in a polygynous mammal, the Galápagos sea lion (Zalophus wollebaeki)". *Molecular Ecology*. **19** (12): 2574–86. doi:10.1111/j.1365-294X.2010.04665.x (https://doi.org/10.1111%2Fj.1365-294X.2010.04665.x). PMID 20497325 (https://pubmed.ncbi.nlm.nih.gov/20497325). S2CID 19595719 (https://api.semanticscholar.org/CorpusID:19595719).
79. Tsang; et al. (2016). "Fabric inspection based on the Elo rating method" (https://web.archive.org/web/20201105011616/https://repository.hkbu.edu.hk/cgi/viewcontent.cgi?article=7236&context=hkbu_staff_publication). *Pattern Recognition*. **51**: 378–394. Bibcode:2016PatRe..51..378T (https://ui.adsabs.harvard.edu/abs/2016PatRe..51..378T). doi:10.1016/j.patcog.2015.09.022 (https://doi.org/10.1016%2Fj.patcog.2015.09.022). hdl:10722/229176 (https://hdl.handle.net/10722%2F229176). Archived from the original (https://repository.hkbu.edu.hk/cgi/viewcontent.cgi?article=7236&context=hkbu_staff_publication) on 2020-11-05. Retrieved 2020-05-05.
80. "Algorithm Competition Rating System" (https://web.archive.org/web/20110902110117/http://apps.topcoder.com/wiki/display/tc/Algorithm+Competition+Rating+System). December 23, 2009. Archived from the original (http://apps.topcoder.com/wiki/display/tc/Algorithm+Competition+Rating+System) on September 2, 2011. Retrieved September 16, 2011.
81. "FAQ: What are the rating and the divisions?" (http://www.codeforces.com/help). Archived (https://web.archive.org/web/20110925191346/http://codeforces.com/help) from the original on September 25, 2011. Retrieved September 16, 2011.
82. "Rating Distribution" (http://www.codeforces.com/blog/entry/870). Archived (https://web.archive.org/web/20111013103750/http://codeforces.com/blog/entry/870) from the original on October 13, 2011. Retrieved September 16, 2011.
83. "Regarding rating: Part 2" (http://www.codeforces.com/blog/entry/893). Archived (https://web.archive.org/web/20111013063622/http://codeforces.com/blog/entry/893) from the original on October 13, 2011. Retrieved September 16, 2011.
84. "Tinder matchmaking is more like Warcraft than you might think – Kill Screen" (https://web.archive.org/web/20170819190821/https://killscreen.com/articles/tinder-matchmaking-is-more-like-warcraft-than-you-might-think/). *Kill Screen*. 2016-01-14. Archived from the original (https://killscreen.com/articles/tinder-matchmaking-is-more-like-warcraft-than-you-might-think/) on 2017-08-19. Retrieved 2017-08-19.
85. "The Best Smartphone Camera 2022!" (https://www.youtube.com/watch?v=LQdjmGimh04). *YouTube*. 2022-12-22. Retrieved 2023-01-07.

86. Avery, Christopher N.; Glickman, Mark E.; Hoxby, Caroline M.; Metrick, Andrew (2013-02-01). "A Revealed Preference Ranking of U.S. Colleges and Universities". *The Quarterly Journal of Economics*. **128** (1): 425–467. doi:10.1093/qje/qjs043 (https://doi.org/10.1093%2Fqje%2Fqjs043).
87. Irwin, Neil (4 September 2014). "Why Colleges With a Distinct Focus Have a Hidden Advantage" (https://www.nytimes.com/2014/09/04/upshot/why-colleges-with-a-distinct-focus-have-a-hidden-advantage.html). The Upshot. *The New York Times*. Retrieved 9 May 2023.
88. Selingo, Jeffrey J. (September 23, 2015). "When students have choices among top colleges, which one do they choose?" (https://www.washingtonpost.com/news/grade-point/wp/2015/09/23/when-students-have-choices-among-top-colleges-which-one-do-they-choose/). *The Washington Post*. Retrieved 9 May 2023.
89. Askell, Amanda; Bai, Yuntao; Chen, Anna; Drain, Dawn; Ganguli, Deep; Henighan, Tom; Jones, Andy; Joseph, Nicholas; Mann, Ben (2021-12-09). "A General Language Assistant as a Laboratory for Alignment". arXiv:2112.00861 (https://arxiv.org/abs/2112.00861) [cs.CL (https://arxiv.org/archive/cs.CL)].
90. "Chatbot Arena Leaderboard Week 8: Introducing MT-Bench and Vicuna-33B | LMSYS Org" (https://lmsys.org/blog/2023-06-22-leaderboard). *lmsys.org*. Retrieved 2024-02-28.
91. "Chatbot Arena: New models & Elo system update | LMSYS Org" (https://lmsys.org/blog/2023-12-07-leaderboard). *lmsys.org*. Retrieved 2024-02-28.
92. Screenplay for *The Social Network*, Sony Pictures (http://flash.sonypictures.com/video/movies/thesocialnetwork/awards/thesocialnetwork_screenplay.pdf) Archived (https://web.archive.org/web/20120904142706/http://flash.sonypictures.com/video/movies/thesocialnetwork/awards/thesocialnetwork_screenplay.pdf) 2012-09-04 at the Wayback Machine, p. 16

Sources

- Elo, Arpad (1986) [1st pub. 1978]. *The Rating of Chessplayers, Past and Present* (Second ed.). New York: Arco Publishing, Inc. ISBN 978-0-668-04721-0.

Further reading

- Harkness, Kenneth (1967). *Official Chess Handbook*. McKay.

External links

- Mark Glickman's research page, with a number of links to technical papers on chess rating systems (http://www.glicko.net/research.html)