Beginner's guide to the revised ICCF rating system

Mark Glickman, Department of Statistics, Harvard University

Starting 2023, the ICCF will adopt a new rating system to rate game outcomes played in ICCF events. The decision to revise the rating system was the result of realizing that the old system was no longer able to adequately calculate accurate ratings. This was particularly evident at the top levels of play, where the frequency of drawn games had increased to a point where the ratings of high-rated players would barely change.

In 2021, the ICCF delegates recognized the need for a new system that would address the issue of an increased frequency of drawn games. To this end, they commissioned the development of a bespoke system by Dr. Mark Glickman, a respected expert in rating systems and a Senior Lecturer at Harvard University. Dr. Glickman is well-known for his work on the Glicko and Glicko-2 rating systems, which have been implemented in various chess organizations such as chess.com and lichess.org. He was also one of the co-developers of the Universal Rating System used by Grand Chess Tour.

Dr. Glickman developed the revised ICCF system from August 2021 through May 2022, using over six years' worth of ICCF game results and current rating information to fine-tune the formulas. The details of the new system were presented to the ICCF delegates in June 2022, and were accepted in August 2022. The implementation of the formulas for use by the ICCF began shortly afterward.

This document describes the basic features of the new rating system, and is intended for a non-technical audience.

Ratings and Rating Deviations:

One of the significant changes to the old rating system is the addition of a rating deviation (abbreviated RD) for every player. This is a concept that was borrowed from

the Glicko rating system[1] developed roughly 30 years ago. The rating, as in the old ICCF system, is a measure of one's playing strength. An RD is a measure of the unreliability of the rating. The larger a player's RD, the less reliable the player's rating. An RD of around 75 or lower indicates that a player's rating is reliable and trustworthy.

Players who are unrated (assuming no other external rating, such as a FIDE rating) start out with a rating of 1800 and an RD of 250, the largest possible RD in the new system. This assignment means that our best guess of the player's rating, without any extra information, is 1800. However, the RD of 250 indicates that the 1800 rating is very uncertain. For FIDE players who are ICCF-unrated, the starting RD is set to 150. Typically, players who have high RDs are either provisional players, or those who have not competed recently and have stale ratings. Conversely, players who have established ratings and who compete frequently tend to have low RDs. The inclusion of RDs into the revised ICCF rating system has some important implications for rating games. RDs can affect how much a player's rating changes from game outcomes and how much their opponent's rating changes as well.

For instance, let's say two players, Sarah and Joe, with the same rating, play a game, and Sarah wins. If both players have low RDs, meaning that their ratings are reliable reflections of their playing strength, then the new rating system would likely not change Sarah and Joe's rating much because their ratings were already reliable before they played. But if Sarah's RD was low and Joe's RD is large, indicating his rating is unreliable, then Sarah's rating would barely increase because she defeated an opponent with an unreliable rating. However, Joe's rating would likely decrease significantly because he lost to an opponent with a precise rating, and his own rating was unreliable to begin with. Finally, if Sarah's RD was high and Joe's RD was low, then Sarah's rating would increase by a large amount given that her rating was unreliable and Joe's was precise.

In general, when a player's RD is large, their rating changes can be expected to be more significant. Conversely, when a player's RD is low, their rating changes tend to be

[1] http://www.glicko.net/glicko/glicko.pdf

smaller. Moreover, when competing against an opponent with a large RD, the result of the game usually has little impact on one's own rating. But when competing against an opponent with a small RD, the result of the game can cause more substantial changes in the player's rating.

The revised ICCF rating system has been designed to provide players with high RDs an opportunity to improve their ratings at a faster rate than before. Because players with high RDs tend to receive larger rating increases (in absolute terms) than those with low RDs, a player with a high RD can play multiple games and quickly improve their rating in a relatively short amount of time. This approach aims to incentivize players who are new to ICCF events to start playing rated games, and to reward provisional players who consistently compete and participate in more rated games.

Rating procedure:

The revised ICCF rating system updates player ratings and RDs every three months, just like the old system. Here are the steps involved in the rating process, which is repeated every quarter:

1. Determine every player's rating and RD at the start of the 3-month rating period.
2. Update each player's rating and RD using the new rating formulas, based on games completed during the 3-month period. This step will result in an RD decrease for each player.
3. Use the formulas to increase each player's RD due to the passage of time to produce a new RD that is used at the start of the next 3-month period.

A few comments are worth mentioning to elaborate on this process. First, at the start of each rating period, every player is assigned a rating and RD, even if they are unrated. Unrated players are assigned a rating and RD (typically a high value) that are not based on ICCF game outcomes. They could be based on known rating information from another rating system (e.g., FIDE), or as mentioned above they would be assigned a rating of 1800 and an RD of 250 if no other information is available.

Second, all game outcomes in a 3-month period are rated simultaneously. This means that each player's rating and RD are updated simultaneously based on all the games completed during the rating period. While a player's rating could go up or down, the RD based on Step 2 always decreases. This reflects the idea that as more game results recorded, the player's updated rating becomes a more reliable measure of their playing strength.

Finally, it's important to note that after the RD decrease due to game outcomes in Step 2, the RD is then increased in Step 3 to account for the passage of time, and this new value is used at the start of the next 3-month rating period. Even if a player doesn't finish any games during the period, their RD will still increase. This is because as time passes, there is more uncertainty about a player's ability - they could be actively improving their play, or they could be getting rusty by neglecting chess. The increase in RD reflects the idea that we are slightly less sure about the player's strength after several months elapsed. However, it is worth noting that RDs above 120 do not increase due to the passage of time; they can only increase once RDs drop below 120.

Separate probabilities for wins, losses and draws:

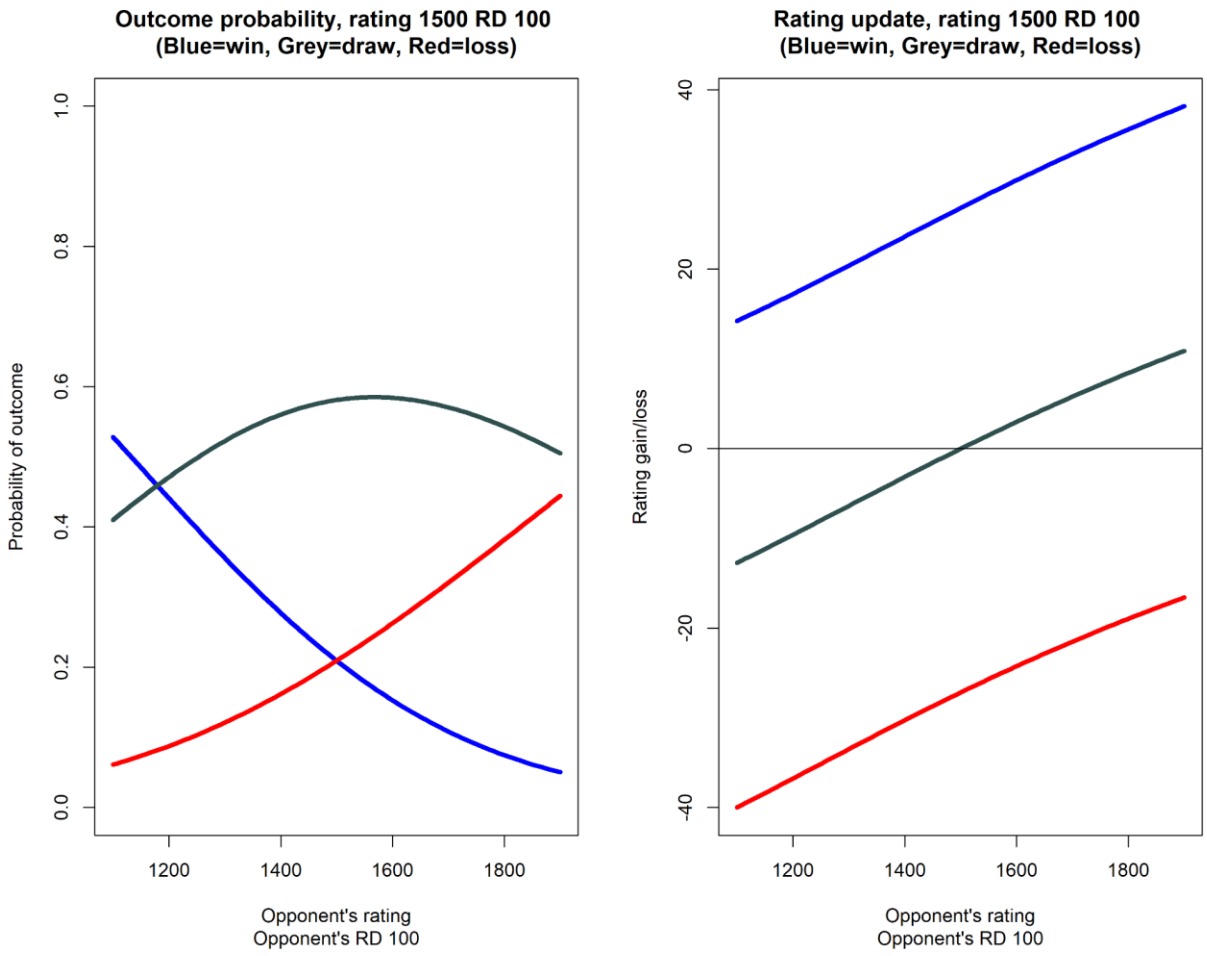
One of the main motivations for revising the ICCF rating system was the substantially higher frequency of drawn games among top-rated players relative to lower-rated players. This resulted in typically smaller rating changes for top players who rarely have decisive games against opponents of similar strength. One limitation of the previous system was that it was based on Arpad Elo's formulas from the 1950s, which did not account for the probability of a draw based on the players' ability levels. The system relied solely on a formula to determine the expected outcome or winning expectancy of a game, which was based purely on the rating difference between two players. As a result, the system was unable to address the issue of the high frequency of drawn games among top-rated players.

The revised ICCF rating system includes three separate probabilities for each type of game outcome (win, loss, or draw), instead of a single winning expectancy based solely on the rating difference between the two players involved. The probabilities of each

outcome are calculated based on the ratings and RDs of the players. The closer two ratings are to each other, the higher the probability of a draw. Also, the higher the RDs are for two players, indicating the unreliability of the players' ratings, the higher the probability of a draw. Additionally, the revised system recognizes that high-rated players tend to draw against each other at a higher rate than lower-rated players. A consequence of acknowledging that high-rated players draw against each other at a high rate is that rating gains and losses based on decisive games can be more pronounced. For example, a 1500-rated player defeating a 1300-rated player results in a rating gain of 20 points (assuming RDs of 100 for both players), while a 2500-rated player defeating a 2300-rated player would have their rating increase by 24 points, more than the rating gain for the lower-rated player pair.

The formulas to calculate the probabilities of each game outcome, and the formulas for rating changes based on these probabilities, were derived from analyzing over six years of previous ICCF game results. Optimizing the rating system with the only goal of producing accurate probabilities would cause ratings to change by amounts that are too large, and the ratings of top players would barely change based on drawn games. Instead, the derivation of the formulas was a compromise between three factors: (1) producing accurate probability calculations of game results, (2) avoiding rating inflation, and (3) maintaining the distribution of ratings over time. While the rate of drawn games among top players in recent years is well over 95%, the final formulas compute probabilities that are around 80%. This slight inaccuracy allows greater movement of ratings at the top levels, and prevents unintended drifts in the average of all ratings.

Below are graphs that illustrate the probabilities of a win, draw and loss for a player with a given rating and RD, and the corresponding rating change for each game result.

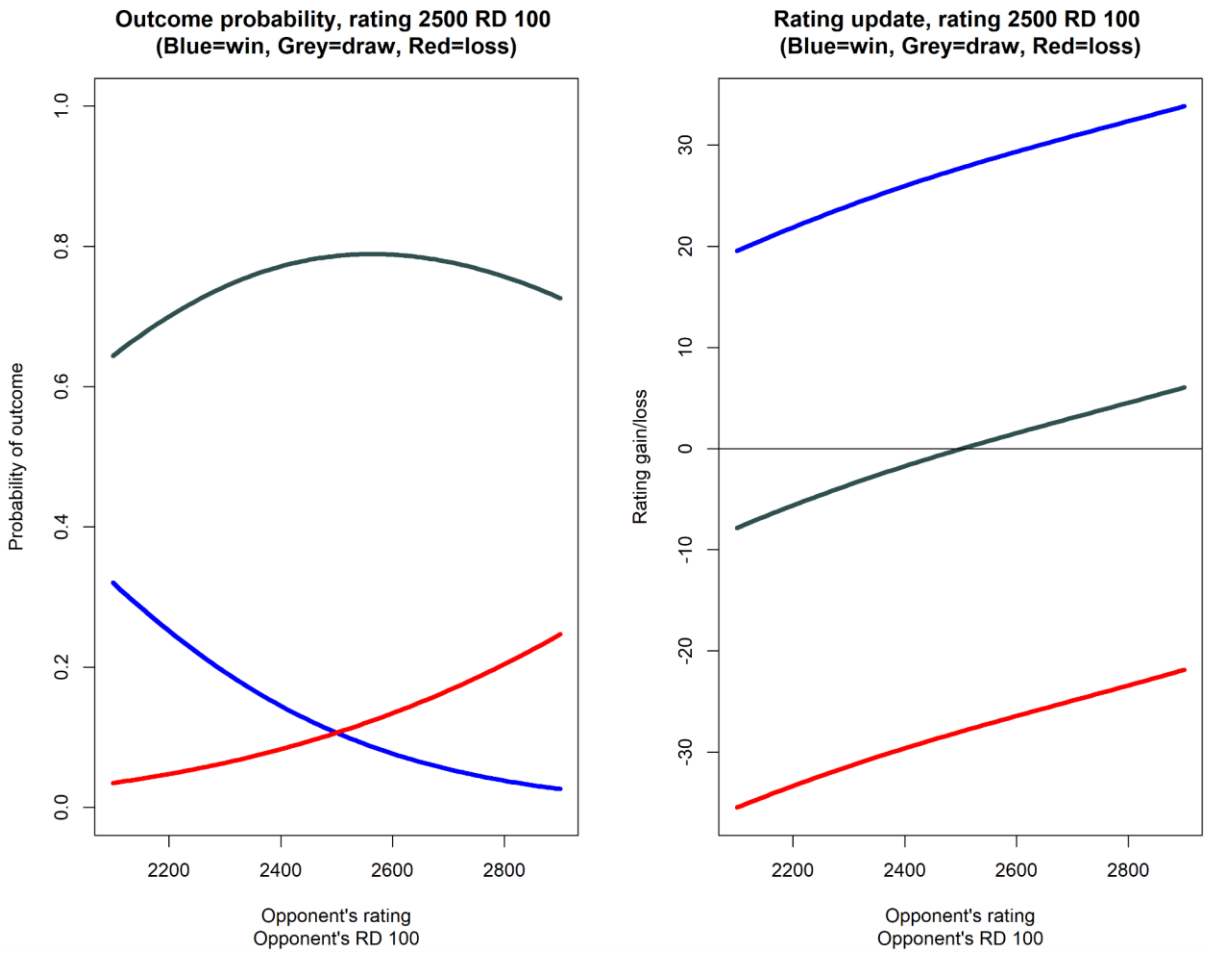


The figure above corresponds to a player with a rating of 1500 and RD of 100. The graph on the left shows the probabilities of a win (blue curve), draw (grey curve) and loss (red curve) against an opponent with a rating indicated on the horizontal axis (and with an RD of 100). For example, when competing against an opponent with a rating of 1400 and RD of 100, the probability of losing to this opponent is about 0.16, the probability of drawing is about 0.565, and the probability of winning is about 0.28. The probability of a draw is largest (highest point on the grey curve in the left graph) when the opponent's rating is also 1500. It is worth noting that the probability of a draw against an opponent rated 1900 (400 points higher than the 1500-rated player) is a bit higher than the probability of a draw against an opponent rated 1100 (400 points lower than the 1500-rated player), as can be seen by comparing the height of the grey curve on the right and left sides of the graph. This is because the revised rating system recognizes that higher-rated player pairs tend to draw games more frequently than

lower-rated player pairs. In this case, the probability of a draw between players rated 1500 and 1900 is higher than the probability of a draw between players rated 1500 and 1100 because the former set of ratings is, on average, higher than the latter.

The graph on the right shows the rating change based on defeating (blue), drawing (grey) and losing to (red) an opponent with a rating indicated on the horizontal axis. If the opponent's rating is 1500 (and their RD is 100), the player's rating does not change if the result is a draw, but the rating gain is about 27 points for a win, and a drop of 27 points for a loss. Again, it is worth noting the asymmetry in the rating change when playing an opponent with a rating 400 points higher versus 400 points lower. For example, when drawing an opponent rated 400 points higher, the rating gain for the 1500-player is 11 points. But when drawing an opponent rated 400 points lower, the rating loss for the 1500-player is 13 points. This asymmetry reflects that drawing a higher-rated opponent is more likely than drawing a lower-rated opponent, because the chance of a decisive outcome is higher.

The same analysis can be illustrated for player with a rating of 2500 and RD of 100.



There are several differences to point out in comparison to the analysis of the 1500-rated player. Based on the graph on the left, the probability of a draw is well over 0.6 when a 2500-rated player is playing against an opponent with a rating between 2100 and 2900 (and an RD of 100), with a peak of about 0.8 when the opponent's rating is also 2500. This is much higher than the probability of a draw by an 1500-rated player playing against an opponent within 400 rating points. The graph on the right, showing the rating change as a result of a win, draw or loss, has some important features worth noting. While the rating increase for a win against an opponent with the same rating is about the same (27 points for 1500-rated players, and 28 points for 2500-rated players), the rating increase for a win by a 2500-rated player does not depend as much on the opponent's rating as it does for an 1500-rated player. In other words, a 2500-rated player defeating an opponent rated 200 points lower (with an RD of 100) results in a 24 point rating increase, whereas an 1500-rated player defeating an opponent rated 200

points lower would experience only a 20-point increase. The curves on the right graph for the 2500-player are “flatter” than the curves on the right graph for the 1500-rated player. This means that, for higher-rated players, decisive game outcomes can lead to more significant rating changes.

Wrap-up:

The revised rating system recognizes that higher-rated player pairs tend to draw games more frequently than lower-rated player pairs, resulting in somewhat larger rating changes for decisive game outcomes for higher-rated players. Additionally, incorporating an RD as a measure of rating unreliability benefits players of all ratings, resulting in more accurate ratings that better reflect a player's skill level.

To convert to the new rating system, the ICCF will use existing ratings from several years ago as the starting point. When the new system is started, a player’s RD will be initiated based on the number of recent games completed. More details about the rollout of the new system are forthcoming.

Players who wish to learn more about the technical details of the revised rating algorithm are invited to read the technical specifications at [insert link here]. An online calculator that implements the new rating formulas is available at https://ratingscalculator.azurewebsites.net/, allowing players to calculate their own rating changes.