

## AGAGoR ratings — Explanatory Note

These ratings are calculated according to a slight variant of the European Go Federation's GoR algorithm, which in turn is basically an Elo rating scheme: each player is assigned a numerical rating; the probability of a player winning any given game is assumed to depend only on the difference in the ratings of the two players (allowing for handicaps changing rating by 100/stone; games on more than 6 stones are ignored). The ratings of the two players are adjusted slightly according to the outcome of each game in a tournament (the rating is assumed constant throughout a single tournament); the adjustment is larger for lower ranked players to allow for the greater variation of weaker players. The player's initial rating is determined by their initial claimed rank — 2100 for 1 dan, 2000 for 1 kyu etc. The adjustments are set so that the ranks correspond to a range of +/- 50 points; thus a 1 dan would expect to have a rating in the range 2050-2150.

In order for the algorithm not to inflate ratings over time, and to prevent wild oscillations for kyu players, some (ad hoc) limitations on a single game rating change are imposed (see the EGF web page for a complete description of the statistical theory and the parameters used). To allow for the fact that players strength may change radically between appearances at a tournament, if a player at a tournament claims a ranking difference of 2 stones or more (200 rating points) from their current rating, they are given a new rating based on their claim (which makes the mathematical theory behind the scheme dubious, but does make the final numbers seem more reasonable — but note that it makes them reasonable in the sense of corresponding with Australian ratings, which for middle dan players seem to be about 1 stone stronger than Japanese rankings, 1 to 2 stones weaker than EGF, and about the same as American). Because we have far fewer tournament, some other (ad hoc) variations have been introduced, the main one being that if the indicated change for one more players as calculated over a tournament is larger than a threshold (50) then the ratings for that tournament are recalculated with the indicated players being assumed to have changed their ratings initially; this is actually done in a number of steps with the players with very large indicated changes being modified first. This is iterated until the maximum change is less than the threshold (but in order to ensure convergence the maximum change for any individual player is decreased on each iteration). Secondly, a maximum initial rating of 2600 (6 dan) is assigned; strong players must earn their higher ratings (ideally all players should start with the same initial rating, and allow the system to work out relative levels; but we do not have enough games for this).

Two effects make the raw ratings conservative. The first is the maximum initial rank mentioned above. Secondly, there is a smaller difference in playing strengths at the top of the table (the difference between a 6d and a 7d player is less than 1 handicap stone). These factors tend to reduce the ratings of these players and thereby push lower ranked players down. To counter this effect, the Calibrated Ratings columns adjusts all ratings upwards so that one anchor player (Andrew Chi, a very strong player who tragically died) is set to 2800; this seems to give reasonable results. (An alternative view would be that the raw ranks correspond approximately to EGF ratings.)

Note the Rank column is the rank as claimed at the most recent AGA tournament. Ratings calculated over a small number of games are not meaningful. Ratings reflect playing strength at the last tournament not in current club play; and players may well have quite different strengths under different conditions such as club or internet games.