DETERMINATION OF THE McMAHON BAR

Geoff Kaniuk *geoff@kaniuk.co.uk*

August 2015

Contents

1 INTRODUCTION **5**

2 FLAWED TOURNAMENTS **7**

3 RANDOM PAIRING **10**

3.1 Player Opponents 10

3.2 Score 11

3.3 Correlation between crossing points for ΔS and ΔG 12

3.4 The equilibrium grade 14

4 SWISS PAIRING **15**

4.1 Pairing Method 15

4.2 Crossing point correlation 16

5 McMAHON PAIRING **17**

6 TOP GROUP PEER GAMES **19**

7 OPPONENT GRADES **20**

8 GROUP SCORES **22**

9 SCORE DISTRIBUTIONS AND THE BAR **24**

9.1 Player score and Group score 24

9.2 Winning chances above the bar 25

9.3 Bar separation 26

10 BAR DETERMINATION METHOD **28**

10.1 Statement of the method 28

10.2 Scope of the Monte Carlo trial 29

10.3 Solution failure 29

10.4 Variation with number of rounds 30

10.5 Tournament entry 30

10.6 Pairing sample rate 31

10.7 Software 32

11 MONTE-CARLO TRIAL RESULTS **32**
11.1 Solution failures . . . 32
11.2 Solution correlation . . . 32
11.3 Bar Depth . . . 33
11.4 Bar Separation . . . 34
11.5 Bar Population . . . 36
11.6 Uniqueness and ranking . . . 37

12 GUIDELINES **38**
12.1 Classical guidelines and maxims . . . 38
12.2 Additional guidelines . . . 39

13 SUMMARY AND CONCLUSION **40**
13.1 Assumptions . . . 40
13.2 Key quantities . . . 40
13.3 Conclusion . . . 41

A MODEL FOR BAR SEPARATION **42**

B PROBABILITY OF WIN **42**

C GLOSSARY **43**

D NOTATION **45**

E ALGORITHM INDEX **45**

List of Tables

1 McMahon bar settings . . . 5
2 Probability for flawed tournaments . . . 10
3 Entry for $T_{\text{r-gap}}$ . . . 13
4 Entry for $T_{\text{s-gaps}}$ . . . 16
5 Limits for the tournament entry . . . 30
6 $T_{\text{m-gap}}$ entry 5 rounds 42 players . . . 31
7 Parameter accuracy dependent on number of pairings . . . 32
8 Outlier features . . . 32
9 T_{mcmahon} entry 7 rounds 50 players . . . 35
10 Coefficients for winning probability . . . 43

List of Figures

1 Four players - pairing for round one . . . 7
2 Four players - start of round two . . . 8
3 Possible graphs for 6 players at start of round 3 . . . 9
4 Opponent grade difference for $\mathrm{T}_{\mathrm{ideal}}$ . . . 10
5 Mean group score for tournament $\mathrm{T}_{\mathrm{ideal}}$ . . . 12
6 Scatter plot for Z_S vs Z_G in random-pairing . . . 13
7 Solution process for $\mathrm{T}_{\text{r-gap}}$ . . . 14
8 Swiss forms for ΔG and ΔS in $\mathrm{T}_{\mathrm{ideal}}$ . . . 15
9 Scatter diagram for Z_S vs Z_G in swiss-pairing . . . 16
10 Solution process for $\mathrm{T}_{\text{s-gaps}}$ . . . 17
11 McMahon pairing quality in $\mathrm{T}_{\mathrm{ideal}}$ . . . 18
12 Peer Games . . . 19
13 McMahon opponent grade difference . . . 21
14 ΔG on the bar-layer . . . 22
15 Mean group-score for McMahon . . . 23
16 Group-score and differences on the bar-layer . . . 23
17 Shodan player-score and group-score histograms . . . 25
18 Tail distributions $T(g, 0, s)$ for bar at shodan . . . 26
19 The mid-score tail distributions $T_{\mathrm{mid}}(g, b)$ . . . 27
20 Mid score tail distribution $T_{\mathrm{mid}}(g, b)$ . . . 27
21 Sampling at 10 and 100 pairings for $\mathrm{T}_{\text{m-gap}}$ . . . 31
22 Scatter plot for Z_S vs Z_G in mcmahon-pairing . . . 33
23 The $\mathrm{T}_{\mathrm{outlier}}$ models . . . 33
24 Bar depth statistics . . . 34
25 Solution process and ramp models for $\mathrm{T}_{\mathrm{mcmahon}}$ . . . 35
26 Ramp statistics . . . 36
27 Cumulative distribution for the bar population . . . 36
28 Bar table statistics . . . 37
29 Unique winner probability and rank deviation for $\mathrm{T}_{\text{m-gap}}$ . . . 38

References

[1] http://senseis.xmp.net/?McMahonPairing/BarTheory

[2] http://www.kaniuk.co.uk/articles/pwin/prob-win.pdf

[3] http://www.kaniuk.co.uk/articles/tourstats/tourstats.pdf

[4] https://en.wikipedia.org/wiki/Round-robin_tournament

[5] Harold.N Gabow, Implementation of Algorithms for maximum matching on nonbipartite graphs(chapters 1 and 4),PhD Thesis,Stanford University,1974

[6] https://en.wikipedia.org/wiki/Table_of_simple_cubic_graphs#6_nodes

[7] http://www.kaniuk.co.uk/articles/pairing/mcmahon-weights-revised.pdf

[8] http://vannier.info/jeux/download/GothaHelp_en.pdf

[9] Paul L Meyer, Introductory Probability and Statistical Applications, Addison-Wesley, 1972

[10] https://en.wikipedia.org/wiki/Cumulative_distribution_function

[11] https://en.wikipedia.org/wiki/Ramp_function

[12] http://www.kaniuk.co.uk/source/

[13] http://www.kaniuk.co.uk/articles/pairing/mcmahon-bar-data.tar.bz2

[14] http://www.britgo.org/organisers/handbook/tournament4.html

1 INTRODUCTION

In a McMahon tournament [1], all players with grades above a certain grade called the bar, are given the same initial *mcmahon-score*. The bar is designed to satisfy three criteria:

1. There should be a unique winner.
2. Every player who has a reasonable chance of winning should be above the bar.
3. Top players should not run out of evenly matched opponents before the end of the tournament.

Since the McMahon system pairs players on the same mcmahon-score, the *unique winner* requirement can lead to an excessively large pool of players above the bar. For example, in an 8 round tournament, the pool would be 256 players - much larger than the usual total entry! For a large pool we generally find that there is no unique winner based on pure mcmahon-score, and indeed the top players may never meet. Tie-breaks are then needed to determine the winner.

In order to meet the second requirement, the bar is raised to include only those who have a reasonable chance of winning. However, if the pool is too small, the top players run out of opponents well before the end of the tournament, and will be playing low quality games against weaker players.

Most Go organisations have tables for setting the bar, defining the population range based on the number of rounds. Since it is not easy to pin down a value for 'a reasonable chance of winning', it comes as no surprise that there is a great deal of variety in the tables.

rounds	bga	macmahon	aga	egf
3	8	7	8	-
4	10	9	12	-
5	12	13	18	16
6	15	17	24	24
7	18	21	32	30
8	22	25	50	40
9	26	29	-	50
10	30	33	-	60

Table 1: McMahon bar settings

The McMahon Bar Theory article in Sensei's Library [1] provides references to various tables. Upper limits for the bar population are summarised in Table 1. For any number of rounds, a larger bar population will result in a wider range of player grades above the bar. The weakest players in this range will normally have little chance of winning the tournament.

One of the first issues to resolve is how do we quantify the requirement that players above the bar have a *'reasonable chance'* of winning the tournament. In the ideal case, the winner of the tournament wins all games played. This suggests that we start by examining the score distribution of players in the groups above the bar. Highest graded players inevitably meet lower graded players (also above the bar), and so have a greater probability of winning all their games.

The purpose of this study is to extract features of McMahon tournaments dependent on details of the player entry, which may be useful for setting the bar. The probability distribution of scores in *each* group is one such feature and is dependent on:

- The probability of win between players of differing grade.
- The population of players at each grade.
- The pairing algorithm.

The win probability between players is one of the most significant pieces of data collected in the European Go Database (*E.G.D*), and a suitable model for the representation of the win probability is presented in [2]. It is accurate in the range 12 kyu to 8 dan, and this model is used to simulate the results of games for all the statistics presented in this document. We assume at the outset that player strength is completely specified by player grade.

The distribution of player grades entering tournaments in Europe has been studied in tournament statistics [3] harvested from E.G.D. Models enabling us to generate plausible distributions have been derived, but in the first instance it will be useful to consider the case of purely random distributions of players in the range 8 kyu to 5 dan.

As far as the pairing algorithm is concerned, there are a variety of strategies in use. One of the simplest is the Swiss system, and this has a bearing on the behaviour of bar players in a McMahon tournament. There are however differences, as the Swiss tournament is closed, but in a McMahon tournament players above the bar continually meet winners from grades below the bar.

Perhaps the simplest *possible* pairing method, and one that is not normally considered, is pure *random-pairing*. There are of course good reasons why random-pairing is not used, but it should be possible to gain some insight into the behaviour of player scores in any group.

By pairing randomly there is the danger that the tournament may not be completed, at least without some players having repeat games or byes. This problem however is not restricted to random-pairing - the danger is there in any system especially if the ratio of players to number of rounds is below 2. In most cases it is definitely taboo for players to have repeat games, and there may not be enough time to arrange for the extra rounds that would be needed to pair players that received a bye.

A second issue with unstructured pairing is that the winner may well be decided by lottery. This again occurs in other pairing systems where tie-breaks need to be invoked to decide a winner. On both issues the question to be asked is: how bad is the effect? If random-pairing really is the worst kind of method, then at least we will have developed a sound basis for comparison.

The plan of this study is to start by examining the case of random-pairing in the context of random distributions of player grades. The purpose of this is simply to provide a framework for understanding how the player score distributions are affected by variations in entry grades. We next examine *swiss-pairing* and then *mcmahon-pairing* for small tournaments to gauge the effect of the feed from players below the bar.

In all the tournaments discussed throughout this study, it is assumed that there is an even number of players - so in principle there are no byes! Unless explicitly stated otherwise, *random* means that samples are drawn from a discrete uniform distribution over a finite range.

Technical terms are italicised when first introduced, and summarised in the Glossary - see Appendix C. It is convenient when referring to player grades to express these in *zero-shodan units*. So a shodan is represented by 0G and a 1 kyu player by −1G. A *group* refers to the set of all players of the same grade. The term *group g* denotes the set consisting of all players with grade g.

2 FLAWED TOURNAMENTS

Like *all-play-all*, random-pairing matches players blindly without regard to *any* game information, other than excluding repeat games. Figure 1 represents the pairing for round 1 in a 3 round all-play-all tournament with 4 players:

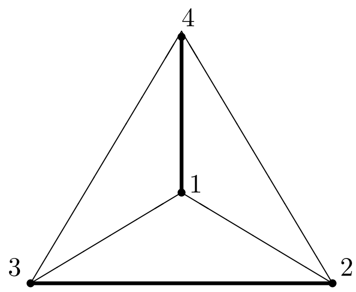


Figure 1: Four players - pairing for round one

The *vertices* in Figure 1 (numbered $1 \cdots 4$) identify the players. The *edges* joining the vertices indicate the allowed pairs. The heavy edges represent the pairing, so 1 plays 4 and 2 plays 3.

For the second round, we remove the edges for the pairing in round 1 to ensure that there are no repeat games. Then, following the standard Round-Robin

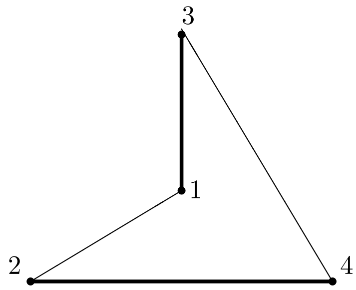


Figure 2: Four players - start of round two

algorithm [4], we rotate the diagram clockwise to produce Figure 2. The *graph* is a cycle of 4 edges, so there are two possible pairings for round 2: 1-3 and 2-4 (the standard round-robin pairing) or 1-2 and 3-4. Either way, the pairing is *perfect* (there are no byes) and we can complete all rounds of the tournament, even if we had chosen the pairing at random in each round.

All pairing algorithms operate along the following lines: we present the algorithm with a list of edges defining the allowed pairings between players. It then chooses a *matching* (a set of non-intersecting edges). This ensures that no player has two opponents in the same round.

For more rounds, it may happen some players would be given a bye. Let us call this a *flawed* tournament. Monte Carlo simulation is used to examine how frequently random-pairing might result in a flawed tournament :

Algorithm 1. `simulate-flawed-tournaments`

TS1. Choose an odd number of rounds r in the range 3 to 15. Then set the number N of players to be $r + 1$.

TS2. Set up a list of edges allowing each player to play *any* other player.

TS3. Pair all the players at random.

TS4. If there are any byes, the tournament is flawed, and we continue at TS1.

TS5. Remove paired players from the list of edges. If the list is not empty continue at TS3.

TS6. When the list is empty, the tournament completes without flaw. Continue at TS1.

A *maximum cardinality* matching algorithm *cmatch* [5] is used in the above simulation in step TS3. Such algorithms try to find a matching with the largest number of edges. Pairing at random is achieved by shuffling the list of edges presented to *cmatch* at each round.

Step TS1 is repeated 10 million times to ensure that the measured value for the probability of finding a flawed tournament is accurate to better than 0.1%. In most cases where the pairing was not perfect, the flaw occurs at the penultimate round (indicated by the column -1R) as shown in Table 2. There are a very few tournaments where the flaw does occur at 3 rounds prior to the last.

For six players it is not difficult to see why there is the possibility of a flawed tournament at round four. Let us go back to the beginning of round three.

At this point each player has just three possible opponents. There are exactly [6] two possible graphs[1] representing this situation as shown in Figure 3.

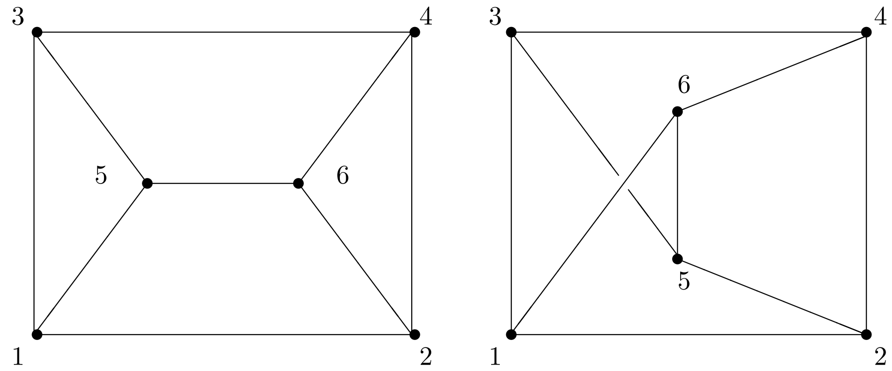


Figure 3: Possible graphs for 6 players at start of round 3

For the left hand diagram, one possible matching is 1–3, 2–4, 5–6, and when these edges are removed at the start of round 4, we are left with the cycle 1–5–3–4–6–2–1 of six edges. This means that there are two perfect pairings for rounds 4 and 5.

The other possible matching is 1–2, 3–4, 5–6, and now we see that this would be a bad choice. For at the start of round 4 we are left with two separated triangles 1–3–5 and 2–4–6, and only two players in each triangle can be paired .

In the right hand diagram choose the matching 1–3, 2–4, 5–6 and remove these edges from the graph. We are left with the cycle 1–2–5–3–4–6–1, which means that we can continue to pair the tournament for rounds 4 and 5. In fact, the right hand graph always has perfect matchings for the next three rounds, no matter which is chosen for round three.

The simulation covered a large number of matchings, but not every conceivable pairing method is encountered. I am indebted to Fred Holroyd and Tim Hunt for the example of a tournament with 10 players paired by the *split-and-cycle* method sometimes used in team tournaments.

Players are split into two equal teams A and B; one team remains seated and the other cycles through board positions at each round. This pairing method can continue happily for 5 rounds, but at round 6 every player in team A has played every player in team B.

Team A has 5 players and only 4 of them can now be paired. The same situation holds for team B, so we are left with two byes and the tournament is flawed.

[1] I thank Richard Parker for pointing me in the direction of cubic graphs.

players	-3R	-2R	-1R	0R	perfect
4	-	-	-	-	1.000
6	0	0	0.150	0	0.850
8	0	0	0.156	0	0.844
10	3.5×10^{-6}	0	0.213	0	0.787
12	0	0	0.247	0	0.753
14	0	0	0.277	0	0.723
16	7.0×10^{-7}	0	0.302	0	0.698

Table 2: Probability for flawed tournaments

3 RANDOM PAIRING

3.1 Player Opponents

The pairing algorithm decides each player's opponent, and in a random-pairing tournament the opponent can have *almost* any grade. There is a small bias in the choice of opponents, as players are not allowed repeated pairings, but in addition, players with the lowest grade $g_{\min}$ will play others of the same grade or *stronger*.

Let $G(g)$ be the average grade of the *opponents* of group g. It is highly likely that $G(g_{\min}) > g_{\min}$ in a randomly paired tournament, as the weakest players will sometimes meet stronger players. On the other hand, the strongest players will meet opponents of the same grade $g_{\max}$ or *weaker*, so we can expect $G(g_{\max}) < g_{\max}$.

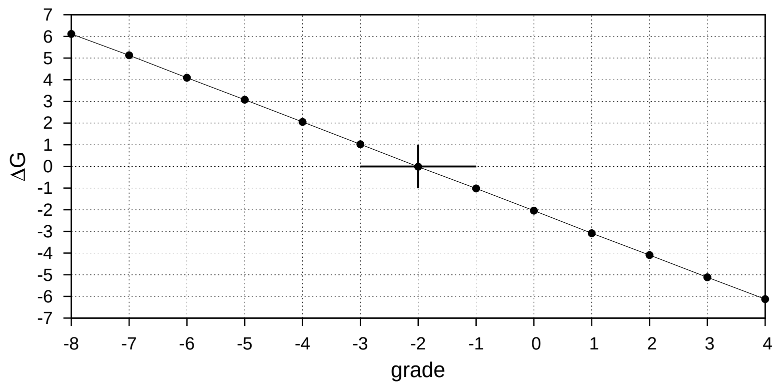


Figure 4: Opponent grade difference for $\mathrm{T}_{\text{ideal}}$

We define the *opponent-grade-difference* as:

$$\Delta G(g) = \overline{G(g) - g} \qquad (1)$$

Here $\overline{X}$ is the mean taken over simulations of pairings repeated many times on the same entry. From the discussion above, $\Delta G(g)$ varies from positive at $g = g_{\min}$ to negative at $g = g_{\max}$.

Consider the highly idealised 6 round tournament T_{ideal}, with 4 players in each grade varying from $-8\mathtt{G}$ to $4\mathtt{G}$. We simulate this tournament in the same way as described in Algorithm `simulate-flawed-tournaments`, from steps TS2 to TS6. Since there are many more players than number of rounds, we do not see any flawed tournaments, even with 1000 simulations of the same entry.

The behaviour of $\Delta G(g)$ for all trial pairings is shown in Figure 4. The linearity is expected[2] given the uniform entry. This clearly has a single *crossing point* at grade $G_c = -2\mathtt{G}$; these players enjoy a special status in that they play others on either side of their own grade to get a balanced set of opponents.

3.2 Score

Once the pairing for a round has been decided, it is up to the players to score as best they can. The outcome is statistically determined by the probability of win between players of given, possibly different, grades g_i, g_j. A game result is simulated by:

Algorithm 2. `simulate-result`
RS 1. Calculate the probability $P_{ij} = P_{\text{win}}(g_i, g_j)$ that player i beats j.

RS 2. Choose a real number f in the range $[0, 1]$.

RS 3. Assuming $g_i < g_j$, the result of the game is a win for i if $f < P_{ij}$.

Appendix B summarises the detail of the calculation for $P_{win}(g_i, g_j)$. At the end of each simulated tournament we obtain the total score for each player, and we define the *group-score* for group g to be:

$$s(g) = \frac{1}{n_g} \sum_{i=1}^{n_g} w_i$$

Here n_g is the number of players in the group, and player i has w_i wins. At the end of a randomly paired tournament with r rounds, the maximum score of any player is just r, so the group-score $s(g)$ lies in the range $0 \cdots r$.

The mean group-score taken over all simulations (with the same entry) is denoted $S(g)$ and increases with grade, as shown in Figure 5. The special group with grade G_c = -2$\mathtt{G}$ identified in the previous section, is seen to have a mean group-score of 2.9. Now in *any* random-pairing tournament with no byes, the average score of *all* players is given by:

$$S_{\text{mid}} \equiv \tfrac{1}{2} r$$

[2]Note that the tiny deviations visible at grades -8$\mathtt{G}$ and 4$\mathtt{G}$ arise from the rule: a player is never paired against self.

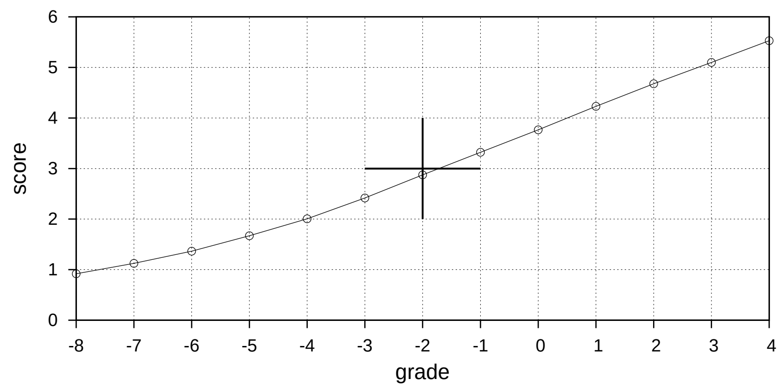


Figure 5: Mean group score for tournament T_{ideal}

We call this the *mid-score* for the tournament, which in the case of T_{ideal} is 3 points. It follows that the mean group-score of the special group is very close to S_{mid}. We define $\Delta S(g)$ as the difference of the mean group score from the mid-score:

$$\begin{aligned} \Delta S(g) &= S(g) - S_{\text{mid}} \\ &= \overline{s(g)} - S_{\text{mid}} \end{aligned} \tag{2}$$

The results demonstrate that for T_{ideal}, $\Delta S(G_c) \simeq 0$. Hence the crossing points of the two functions $\Delta G(g)$ and $\Delta S(g)$ are similar i.e. both are nearly equal to G_c. We next examine the relation between these two crossing points when the tournament is *not* ideal.

3.3 Correlation between crossing points for ΔS and ΔG

We can simulate the variety of entries in real tournaments. The number N of players generally increases with number of rounds, so we choose N from a range $[N_{\text{min}}(r), N_{\text{max}}(r)]$, where the extremes are linearly increasing functions of the number of rounds r. We fix the grade range to be the same as T_{ideal}.

Then, ignoring the complex details for the European player population per grade [3], we generate an entry by choosing the grades for the N players at random from the grade range [-8G, 4G]. Setting

$$n_{\text{min}} = 5 + 2.5r \quad n_{\text{max}} = 5 + 7.5r$$

we ensure that most grades are reasonably well populated for all rounds in the range 2 to 10.

We simulate each generated entry as described in section 3.1 to produce raw values for $\Delta G(g)$ and $\Delta S(g)$, for each grade g in the above range. These functions are less smooth than those produced for T_{ideal}, partly due to the reduced sampling rate from 1000 to 100 and partly due to variation in the population per grade. We represent $\Delta G(g)$ and $\Delta S(g)$ by linear models to smooth the raw data:

$$\begin{aligned} \Delta G(g) &= K_G\,(g - Z_G) \\ \Delta S(g) &= K_S\,(g - Z_S) \end{aligned} \tag{3}$$

The slopes K and zero-crossing points Z are obtained from a least squares fit using 4 points surrounding the changes of sign in the raw data.

The scatter plot in Figure 6 shows that the solutions Z_G and Z_S are well correlated for all rounds simulated. The correlation coefficient is 0.99, and there are no obvious outliers.

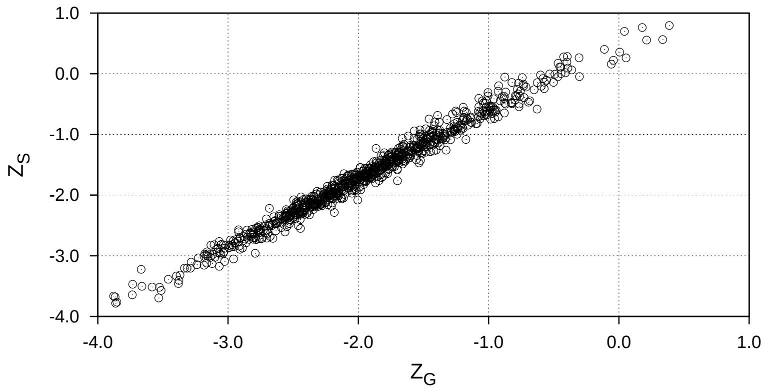


Figure 6: Scatter plot for Z_S vs Z_G in random-pairing

The solution process for a generated tournament $T_{\text{r-gap}}$ is illustrated in Figure 7. The rectangle encloses all the points used for fitting the $\Delta G(g)$ and $\Delta S(g)$ models.

This is a 3 round tournament of 28 players with an irregular entry including grade gaps specified in Table 3. The fit produced the coefficients:

$$\begin{aligned} Z_G &= -0.969 \quad K_G = -1.024 \\ Z_S &= -0.496 \quad K_S = 0.233 \end{aligned}$$

grade	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4
entry	1	3	2	1	0	4	0	2	3	3	1	6	2

Table 3: Entry for $T_{\text{r-gap}}$

Rounding the values of Z_G and Z_S gives us separated crossing-points at -1 and 0 respectively. In comparison, T_{ideal} produces only one crossing point.

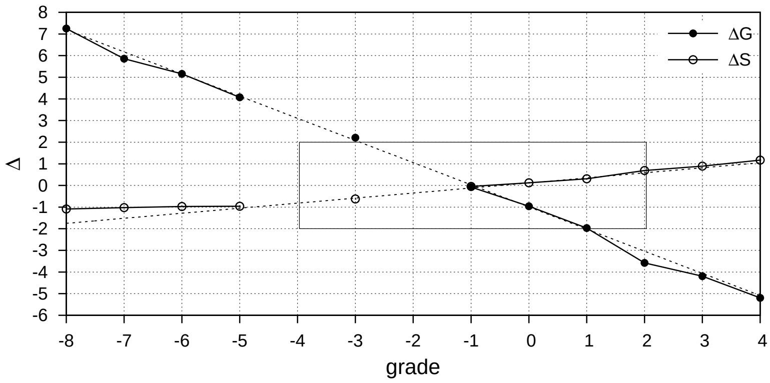


Figure 7: Solution process for $T_{\text{r-gap}}$

3.4 The equilibrium grade

The crossing-point Z_G is an *equilibrium-grade*, where players at that grade experience a balanced mix of opponents. There is another equilibrium-grade at Z_S, where players win half their games.

The maximum difference $|Z_G - Z_S|$ observed in the random-pairing trial is 0.82. This value, less than 1 grade apart, can be expected since the two conditions are closely related. For if the average grade of a player's opponents is very different from the player's own grade, then we would expect the player's score to be different from the average and vice versa.

Since the solution to $\Delta G(g) = 0$ is also the point at which $\Delta G(g)^2$ is minimised (and similarly for ΔS), we can obtain a unique equilibrium-grade by minimising the square sum of the models in Equation (3):

$$F(g) = K_G^2(g - Z_G)^2 + K_S^2(g - Z_S)^2$$

The minimisation of $F(g)$ produces the result:

$$\begin{aligned} G_\text{c} &= \lambda_G Z_G + \lambda_S Z_S \\ \lambda_G &= K_G^2/(K_G^2 + K_S^2) \\ \lambda_S &= K_S^2/(K_G^2 + K_S^2) \end{aligned} \tag{4}$$

Thus G_c is a weighted sum of the individual crossing points. Since the weights are positive and sum to unity, G_c always lies between Z_G and Z_S.

As we have seen, random-pairing leads to very simple statistics, and indeed the value of G_c for each tournament is very similar to the average grade for the tournament. Things are not so simple in Swiss tournaments.

4 SWISS PAIRING

4.1 Pairing Method

In the simplest application of swiss-pairing, all players start with zero wins, and pairing in the first round is random. Thereafter, players on the same number of wins are paired with each other. So the choice of opponents is no longer as free as it was in the random-pairing method discussed in the previous section. A player's score is the number of wins, and in a Swiss tournament the average of all the players' scores is $\frac{1}{2}r$, just as in random-pairing.

In Swiss tournaments, there is the possibility that the pairing matches players with different scores when the number of rounds is not exactly $\log_2(n)$, n being the number of players. However, with a reasonable turnout compared to the number of rounds, the incidence of these uneven games is normally much less than the number of evenly matched games. In the interests of simplicity we ignore the problem of uneven matching.

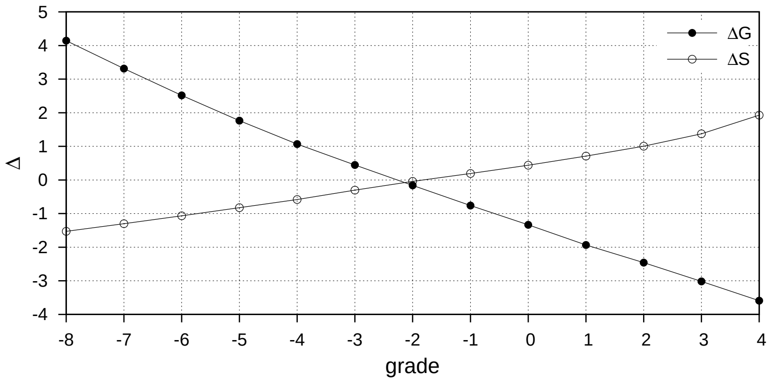


Figure 8: Swiss forms for ΔG and ΔS in $\mathrm{T_{ideal}}$

We use a *maximum weighted matching* algorithm `wmatch` [5] to carry out the pairing for Swiss tournaments. Following the discussion in [7], the simplest possible weight assignment for a Swiss tournament gives the weight for a pairing between players with scores s_i and s_j as:

$$W_{score}(i, j) = W\,\mathrm{sech}(\lambda(s_i - s_j)) + W_0 \tag{5}$$

The constant λ controls how severely uneven games are 'punished'. The value $\lambda = 0.1$ was found to be adequate for the simulations presented here. The values $W = 2^{33}$ and $W_0 = 2^{24}$ were set so that the weighted pairing is also maximum cardinality as discussed in [7]. The locally quadratic form for the weight avoids the degeneracy associated with a linear form [8], and ensures that the maximum weight for the pairing is obtained when the players have the same score. With the above assumptions, the pairing procedure for a single round is:

Algorithm 3. `swiss-pairing`

SP 1. Shuffle the list of players.

SP 2. For each possible pair ij, choose the weight as in Equation (5)

SP 3. Pair the players using `wmatch`.

SP 4. Simulate results as discussed in Algorithm `simulate-result`.

Applying swiss-pairing to T_{ideal} we obtain the forms for $\Delta G(g)$ and $\Delta S(g)$ shown in Figure 8. Again we see crossing points near to each other at $Z_G = -2.26$ and $Z_S = -1.79$, so in swiss-pairing we too have a unique equilibrium grade via Equation (4).

4.2 Crossing point correlation

The correlation between Z_G and Z_S for swiss-pairing is similar to that obtained for random-pairing. The correlation coefficient is 0.97, again high with no obvious outliers.

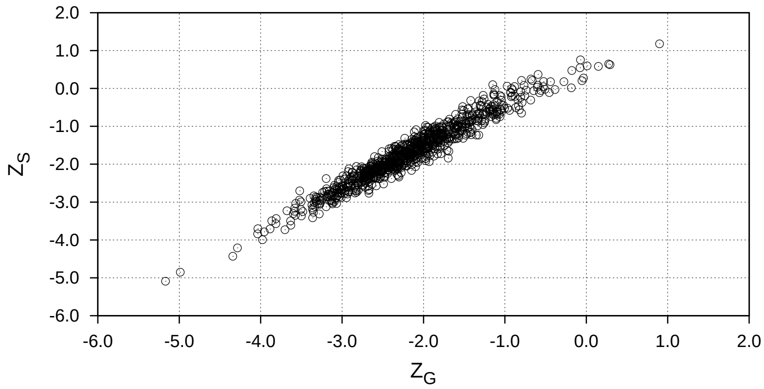


Figure 9: Scatter diagram for Z_S vs Z_G in swiss-pairing

The largest difference ΔZ in the crossing points is 1.25, occurring for a 2 round tournament with 11 players separated into two sections occupying the highest and lowest grades. Apart from this, there are 4 tournaments with $\Delta Z > 1$, the largest of which has $\Delta Z = 1.1$.

grade	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4
entry	1	4	0	4	1	2	0	3	4	2	0	10	1

Table 4: Entry for $T_{\text{s-gaps}}$

This is an 8 round tournament $T_{\text{s-gaps}}$ with 32 players and 3 grade gaps specified in Table 4. The raw data for the opponent-grade-difference $\Delta G(g)$ shows a change of sign between grades -3G and -1G as seen in Figure 10. So the 4 points used for the fit range from -4G to 0G. The mid-score $\Delta S(g)$ has raw values changing sign between -1G and 0G and uses a shifted set of 4 points from -3G to 1G for the least squares fit. The resulting combined solution obtained for $T_{\text{s-gaps}}$ is $G_c = -1.2$.

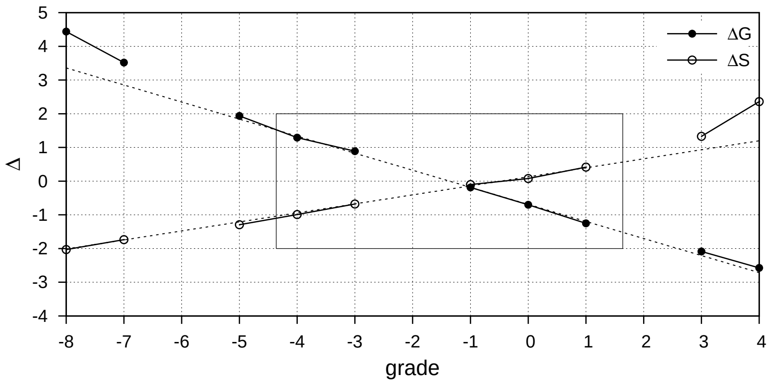


Figure 10: Solution process for $T_{\text{s-gaps}}$

Although there is no concept of a bar in random- or swiss- pairing, it seems that the equilibrium-grade takes on one of its rôles in the sense that it divides the entry into two parts. Players with grades above G_c win more than half their games on average, and so the tournament winner is likely to be found amongst these players.

The one dimensional models $\Delta G(g)$ and $\Delta S(g)$ developed for random- and swiss- pairing find a place in mcmahon-pairing, but the introduction of the bar adds a whole new dimension to the group behaviour.

5 McMAHON PAIRING

In the ideal McMahon tournament, the entry is well populated with no missing grades, and T_{ideal} provides us with a good example. Ignoring issues associated with the choice of player colour and player club or country, the pairing algorithm can be reduced to the simple form discussed for swiss-pairing in section 4.1, with the score there replaced by the mcmahon-score expressed in zero-shodan units.

The bar setting affects the pairing quality [7], and we expose this by simulating an entire tournament as described in swiss-pairing; then repeat the simulation for bar settings covering the entire range of grades in the entry.

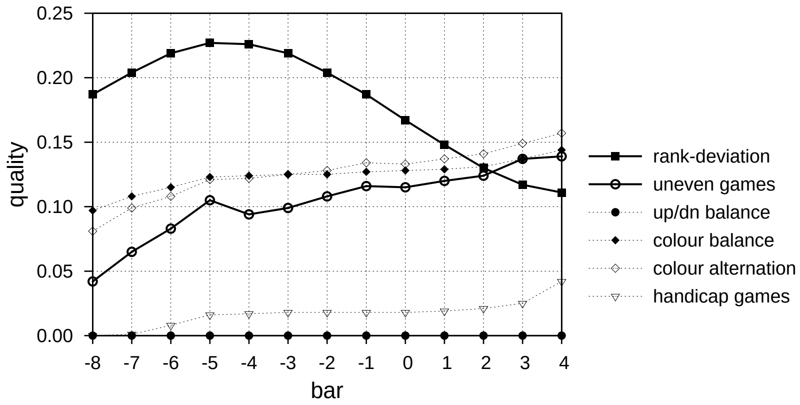


Figure 11: McMahon pairing quality in T_{ideal}

This process applied to T_{ideal} produces the results shown in Figure 11. Each quality item is normalised so that it has a range from best quality (0) to worst quality (1). The quality components shown are:

uneven games The number of uneven games for all players.

up/dn balance The magnitude of the difference between the number of games played up and games played down.

handicap games The number of games where handicap stones are awarded - i.e. when the mcmahon-score difference for a pair exceeds 1 point.

colour balance The magnitude of the difference between games played white and games played black.

colour alternation The magnitude of the number of colour reversals relative to the expected colour reversals.

rank-deviation The root-mean-square of the difference between the player's *mcmahon-rank* and *grade-rank*.

The only active weight in the pairing algorithm is the McMahon weight ensuring as many even games as possible, and this has its greatest direct effect on the first three qualities only.

We note firstly that there is a low incidence of handicap games for all bars below `4G`. There are a total of 156 games in the tournament, and the worst handicap quality of 0.042 translates to a mean of 6.6 handicap games per tournament. For bars from `-1G` and above, the uneven quality lies in the range 0.12 to 0.14, and this would produce a mean of 3.2 to 3.8 uneven games per round.

At the end of the tournament, players are ranked by mcmahon-score scaled to lie in the range $[0, 1]$. We also rank players by grade, also scaled to lie in the same range. We call these the mcmahon-rank and grade-rank respectively. The rank deviation shown in Figure 11 is the root-mean-square difference of the two ranks for each player. The rank deviation improves by a factor of two as the bar increases.

The quality analysis confirms that the simple pairing algorithm employed copes well with the uniform entry of T_{ideal} at all bar settings. The introduction of the bar is an independent variable affecting many characteristics of the McMahon and we take this up in the next sections.

6 TOP GROUP PEER GAMES

The bar can be set at any grade whatsoever, but the top groups will not be well served if it is set too high or too low. The top groups may include all the grades needed to ensure that the population exceeds the number of rounds. In the extreme case of a long tournament, where the entry is very thin at the higher grades, this could mean a huge grade range in the top groups. In such a situation, not all the players in these groups could be considered as peers.

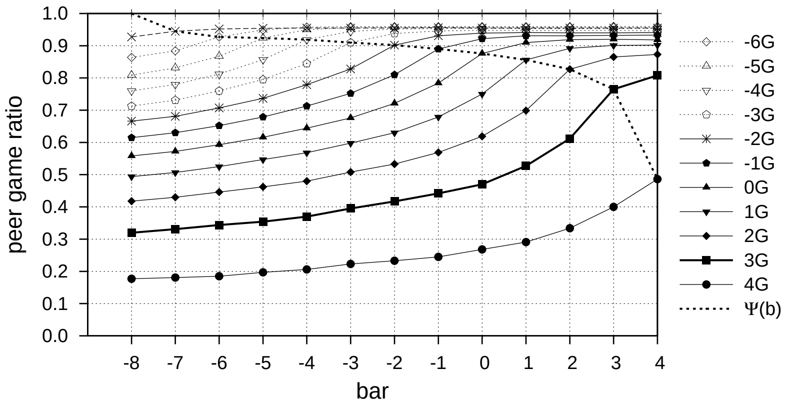


Figure 12: Peer Games

Define a *top-group* $Q(g)$ to consist of all players with grades from g to $g_{\max}$. We then examine how the bar setting b affects the proportion of *peer-games* in this top-group.

Suppose there are a total of N_g games played in the top-group $Q(g)$. The number of games *between* players in the top-group depends on the bar, and is denoted by $E_g(b)$. Such games are the peer-games, and we are interested in the behaviour of the ratio of peer-games to N_g for each grade g, as we vary the bar.

$$\Psi_g(b) = E_g(b)/N_g$$

The graphs in Figure 12 show the behaviour of the peer-games-ratio $\Psi_g(b)$ as the bar is varied. In particular we observe that for each grade g, $\Psi_g(b)$ rises steadily as the bar b is increased from its lowest value at $g_{\min}$. However, $\Psi_g(b)$ then abruptly levels off for bar values from $b = g$ to $b = g_{\max}$.

Consider the topmost group at 4G i.e. Q(4). If the bar is set to $g_{\min}$, the topmost group has a peer-games-ratio of 18.3%, so only 4.4 games out of the maximum of 24 is a peer-game. If the bar is set to $g_{\max}$ then the ratio rises to 49.2% - which is almost as high as it can be. At this bar, the 4 top players then do get to play each other.

The behaviour for the top group $Q(3)$ is quite different. Again at a bar -8G the peer-games-ratio is low. However, at a bar grade of 4G the ratio is a healthy 81.3%, and the bar can be reduced to 1G whilst still keeping the peer-games-ratio above 50%.

The *bar-group* consists of all players with grades in the range b to $g_{\max}$ where b is the bar setting. The peer-games-ratio for the bar-group is given by $\Psi_b(b)$. This is illustrated by the (dotted) graph $\Psi(b) = \Psi_b(b)$ shown in Figure 12. It is clear that for bars below the maximum grade, the peer-games-ratio for the bar-group is not very sensitive to the bar setting.

7 OPPONENT GRADES

The McMahon system pairs players on the same mcmahon-score, so if grades are fairly populated, we can expect that players well below the bar play others on similar grades. Above the bar, players may meet opponents of widely differing grades, especially in the early rounds. But this grade difference should narrow as the stronger players win through in the later rounds.

The average grade of opponents of players in group g depends on the bar b as well as on g, for the bar setting influences *whom* you get to play. We can extend the group opponent grade notion developed for random-pairing and swiss-pairing. Define $G_g(b)$ to be the average of the grades of opponents of group g, when the bar is set to b.

The incidence of equally graded games in McMahon tournaments is indicated similarly to (1), by the mean opponent-grade-difference ΔG_g defined by:

$$\Delta G_g(b) = \overline{G_g(b) - g} \tag{6}$$

The behaviour of $\Delta G_g(b)$ for the weakest group in $\mathrm{T}_{\text{ideal}}$ is shown highlighted in Figure 13. We see that ΔG settles down to a value of about 0.7 for all bars beyond -3G. As expected, the weakest players in the tournament are playing against stronger players, no matter where the bar is set (i.e. $\Delta G_{g_{\min}}(b) > 0$). The strongest group 4G in Figure 13, also highlighted, shows a grade difference steadily reducing in size as the bar increases from the weakest to the strongest grade. The strongest group always has opponents the same grade or weaker, so $\Delta G_{g_{\max}}(b) < 0$ for every bar.

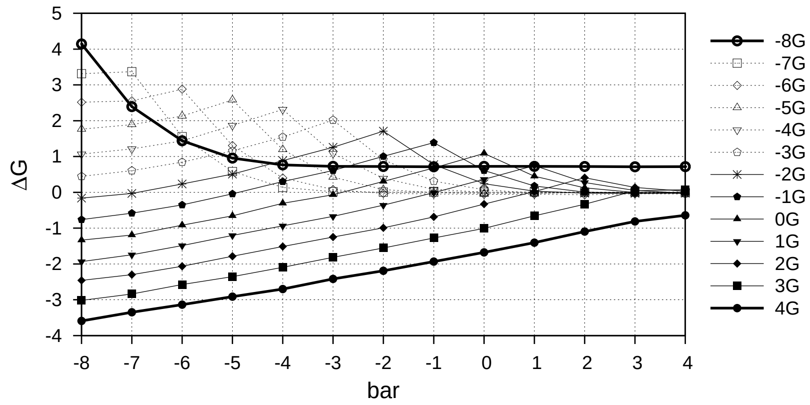


Figure 13: McMahon opponent grade difference

For the extreme strongest and weakest grades, we can see that the variation of opponent-grade-difference is very smooth. This however is not the case for any grade in between. Indeed we observe that for *every* grade between the weakest and the strongest, the group g has a peak value for ΔG_g when the bar $b = g$. To put this another way, set the bar at b somewhere in the range $g_{\min}+1$ to $g_{\max}-1$. Then $\Delta G_g(b) < \Delta G_b(b)$ for *all* grades g other than $g_{\min}$ or $g_{\max}$.

These results make it clear that the group at the bottom of the bar enjoys a special status, and really does stand between the weaker and stronger players in the tournament. We saw a similar phenomenon in the discussion on peer games, where the peer-game-ratio $\Psi_g(b)$ undergoes a sharp change when $b = g$.

In the (grade, bar) plane, the line $b = g$ provides a layer separating the weaker from the stronger players, and we call this the *bar-layer*. Define the opponent grade difference on the bar-layer:

$$\Delta G(b) = \Delta G_b(b) \tag{7}$$

We saw earlier that $\Delta G(g_{\min}) > 0$ and that $\Delta G(g_{\max}) < 0$. Therefore there is a zero crossing point for $\Delta G(b)$, somewhere between the minimum and maximum grades.

The graph in Figure 14 shows the values obtained for $\Delta G(b)$ extracted from the peak values of the individual grade graphs in Figure 13. The graph clearly shows a crossing point in the neighbourhood of 3G. This is an equilibrium-grade analogous to the one found for random-pairing and swiss-pairing.

When the bar is set to the equilibrium-grade players at the bottom of the bar meet roughly equal numbers of stronger or weaker players. At this bar setting there are also a healthy number of peer games as discussed in Section 6. The equilibrium-grade is a unique grade in McMahon tournaments just as it was in Random or Swiss tournaments.

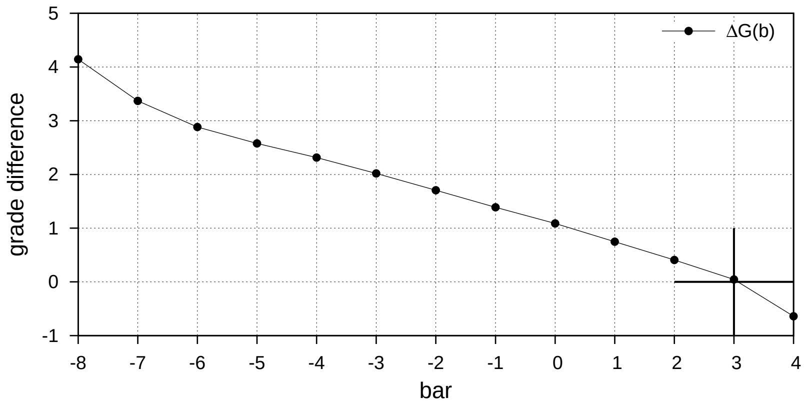


Figure 14: ΔG on the bar-layer

8 GROUP SCORES

We have so far discussed the influence of the bar on the *choice* of opponents, and now we consider how the bar setting affects the *score* of players. Extending the swiss-pairing discussion, we define the *group-score* $s(g, b)$ as the average of the mcmahon-scores in each group g when the bar is set to b.

We simulate the tournament $\mathrm{T}_{\mathrm{ideal}}$ to produce the mean group-score:

$$S_g(b) = \overline{s(g, b)} \tag{8}$$

for each grade g and bar b. The graphs in Figure 15 show that $S_g(b)$ levels off to a stable value $S_{\mathrm{lim}}(g)$ as the bar rises, for all grades g except g_{max}. The stable value for each grade can be defined as the value at the maximum possible bar setting:

$$S_{\mathrm{lim}}(g) = S_g(b{=}g_{\mathrm{max}}) \tag{9}$$

We observe that $S_{\mathrm{lim}}(g)$ changes in a regular manner: values are spaced very nearly 1 point apart at every grade. Moreover for $g = 0$, $S_{\mathrm{lim}}(g) = 2.96$, very nearly a perfect average score for the shodan group in a 6 round tournament. This means that to a good approximation the stable score can be represented by $S_{\mathrm{lim}}(g){=}g + \frac{1}{2}r$

In Section 7, we saw that when the bar is set to the maximum possible, any group g sees a balanced mix of opponent grades for $g_{\mathrm{min}} < g < g_{\mathrm{max}}$. We can then expect that most players will win roughly half their games. So players in group g will end up on the same final mcmahon-score which is the *initial-mcmahon-score* plus half the number of rounds[3] r.

[3] In the 'zero shodan' grading representation, the initial-mcmahon-score of a player is the grade of the player (or the bar grade) in zero-shodan units. Other systems may have a different origin for the mcmahon-score.

This is a universal mid-score for McMahon tournaments defined along the same lines as in random-pairing or swiss-pairing for each grade g by :

$$S_{\text{mid}}(g) = g + \frac{1}{2}r \tag{10}$$

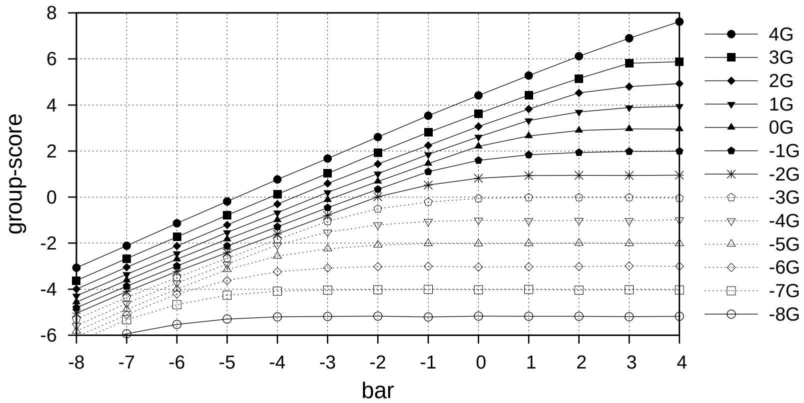


Figure 15: Mean group-score for McMahon

Therefore $S_{\text{lim}}(g) \simeq S_{\text{mid}}(g)$ for all grades apart from the highest and lowest. We present the difference as the graph labelled $S_{\text{lim}} - S_{\text{mid}}$ in Figure 16. When the bar is set to a value lower than g_{max} the group-score of every group above the bar is strongly affected, and for most groups it no longer tracks the mid-score value i.e. $S_g(b) \neq S_{\text{mid}}(g)$ when $b \neq g_{\text{max}}$.

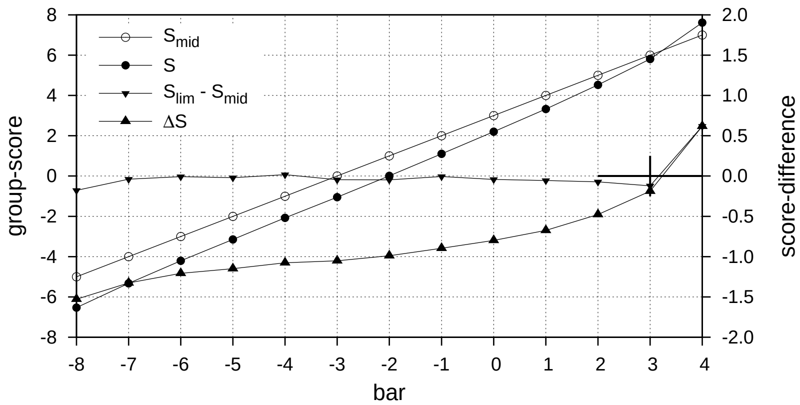


Figure 16: Group-score and differences on the bar-layer

Define the difference:

$$\Delta S_g(b) = S_g(b) - S_{\text{mid}}(g)$$

The behaviour of the group-score for players at the bottom of the bar i.e. on the bar-layer identified in the previous section is expressed by $S_b(b)$. We set $g = b$ in Equation (8) and define:

$$S(b) = S_b(b) \tag{11}$$

$$\begin{aligned} \Delta S(b) &= \Delta S_b(b) \\ &= S(b) - S_{\text{mid}}(b) \end{aligned} \tag{12}$$

It is apparent from Figure 16 that $S(b) < S_{\text{mid}}(b)$ up to the bar $b = 3\texttt{G}$. Beyond that point $S(b) > S_{\text{mid}}(b)$. Hence for our ideal tournament, we have the result that $\Delta S(b)$ changes sign near the value $b = 3\texttt{G}$.

This gives the same zero-crossing point as we found in the discussion of opponent grades in Section 7. The feature that we found in both random-pairing and swiss-pairing, namely that ΔG and ΔS have similar zero crossing points, now emerges in mcmahon-pairing for players in the bar-layer.

9 SCORE DISTRIBUTIONS AND THE BAR

9.1 Player score and Group score

A player's *individual score* is determined both by the probability of win and opponent choice. The group-score is determined by probability of win and the choice of all opponents in the group. Since all n players in the group have the same grade, they have the same score range $[s_{\min}, s_{\max}]$.

The player-score distributions in a given group are not necessarily independent, since the sum of the scores in a pair is 1. Nevertheless, since the players in the group have the same grade, we can assume that they have identical score distributions denoted by $p(s_i)$, where s_i is the score of the i^{th} player in the group.

The group-score s is proportional to the sum of the individual player scores:

$$s = k \sum_{i=1}^{n} s_i$$

Here n is the number of players in the group, and $k = 1/n$. Since the mean of a sum of random variables is the sum of the individual means [9], and since the individual means are the same, it follows that the mean player score is the same as the mean group-score.

The variances of the group score and the player score are however very different. For example, Figure 17 shows probability distributions for the shodan group when the bar is set to 0G. The player-score distribution is much broader than

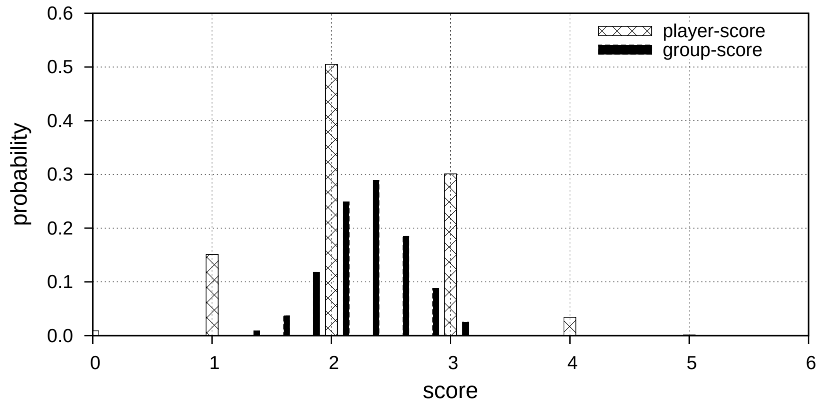


Figure 17: Shodan player-score and group-score histograms

the group-score distribution and its variance is 4.46 times the variance of the group-score distribution.

This result confirms that individual player scores are not independent. For if they were, the player-score variance would be exactly 4 times the group-score variance [9], since there are 4 players in the group. The excess shows that player scores in the group are positively correlated.

For the purpose of separating players above the bar in McMahon tournaments, the group-score provides a sharper instrument than player-score.

9.2 Winning chances above the bar

One of the commonly stated guidelines for McMahon tournaments is that players above the bar should have a reasonable chance of winning. It sometimes happens that the tournament winner does not win *every* game, but will win more games than any other player above the bar. Hence the probability that a group scores in *excess* of a particular value may help to quantify winning chances.

More precisely, we are concerned with the probability $T(g, b, s)$ that a group g scores more than s when the bar is set to b. Formally:

$$T(g, b, s) = \text{Prob}(\text{group-score} > s \mid \text{grade} = g, \text{bar} = b) \tag{13}$$

This probability is the complement of the cumulative distribution function, and is known as the *tail distribution* [10]. The tail distributions in T_{ideal} for higher grades at a bar setting of `0G` is shown in Figure 18. The vertical line in the plot at group-score 3 identifies the mid-score $S_{\text{mid}}(0)$, and its length covers the probability range $[0.05, 0.95]$. The highlighted graph is the shodan group at the bottom of the bar. It is apparent that the probability that it scores more than the mid-score is very low, and indeed it has no chance at all of winning.

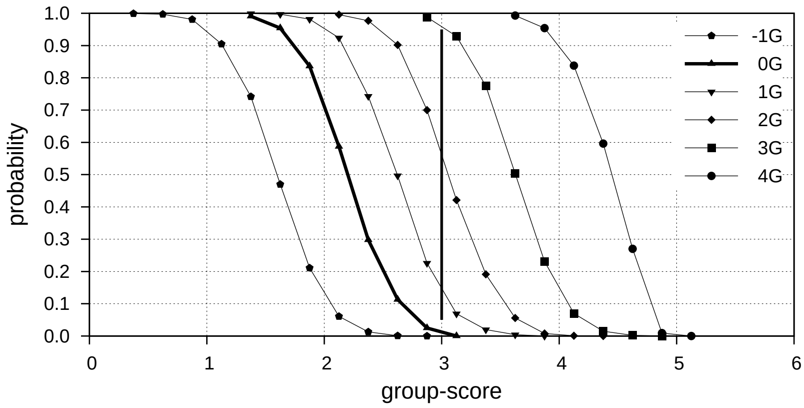


Figure 18: Tail distributions $T(g, 0, s)$ for bar at shodan

All groups below shodan have no chance of scoring more than $S_{\mathrm{mid}}(0)$. The groups 3G and 4G above shodan have *some* chance of scoring more than 5 and also have a *very good* chance of scoring more than $S_{\mathrm{mid}}(0)$. For reference, the 4G individual probability of scoring 6 is just 10.2%.

There is no acceptable value to be obtained for a 'reasonable chance' of winning. We can say however that the tail distribution provides a clear separation between grades: 0G and below have *no* chance of winning; 3G and above have *some* chance of winning.

9.3 Bar separation

The bar-group should clearly separate the performance of players above the bar from the performance of players below the bar.

In the tournament $\mathrm{T}_{\mathrm{ideal}}$ we observe in Figure 18 that when the bar is set to 0G, $T(g, 0, S_{\mathrm{mid}}(0))$ is very small for grades below grade 1G. This bar provides no distinction between -1G and 0G groups.

By setting the bar at a higher value, we might be able to improve the difference in performance of players at the bottom of the bar, from those below the bar. The tail distribution for each grade g at the mid-score for each bar b is obtained from Equation (13) with the group-score s set to $S_{\mathrm{mid}}(b)$. This gives us the restricted form $T_{\mathrm{mid}}(g, b)$:

$$T_{\mathrm{mid}}(g, b) = T(g, b, S_{\mathrm{mid}}(b)) \tag{14}$$

The mid-score performance for the shodan group is illustrated in the highlighted graph in Figure 19. Clearly we need to increase the bar to beyond the 2G setting to see any significant increase in performance of players at the bottom of the bar i.e. those in the bar-layer.

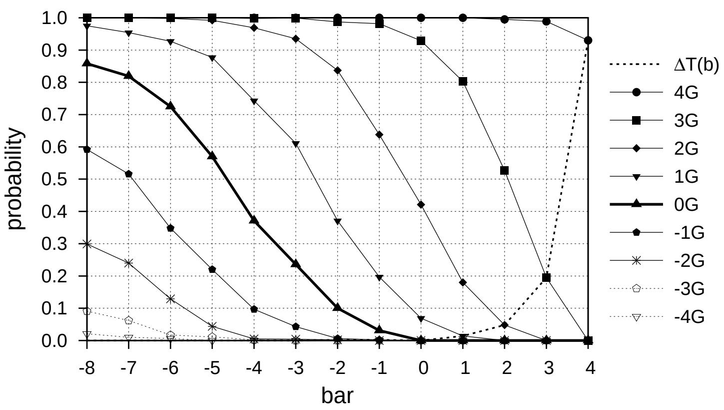


Figure 19: The mid-score tail distributions $T_{\text{mid}}(g, b)$

The mid-score performance of the players in the bar-layer is given by $T_{\text{mid}}(b, b)$ and illustrated by the dotted graph in Figure 19. The relation between the single variable $T_{\text{mid}}(b, b)$ and its two-variable parent $T_{\text{mid}}(g, b)$ for T_{ideal} is shown in Figure 20.

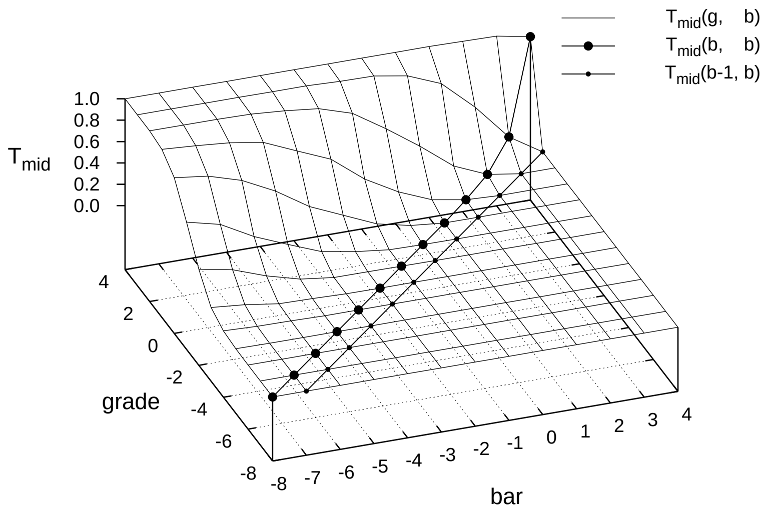


Figure 20: Mid score tail distribution $T_{\text{mid}}(g, b)$

The mid-score performance of players just below the bar is given by $T_{\text{mid}}(b, b{-}1)$, and is illustrated by the horizontal line in Figure 20. In T_{ideal}, the performance just below the bar is zero for all bars, but this is not always the case for tournaments with a more varied entry.

We formally define the *bar-separation* $\Delta T(b)$ as the difference in the performance across the bar-layer:

$$\begin{aligned} \Delta T(b) &= \gamma_{\mathrm{B}}(b) - \gamma_{\mathrm{L}}(b) \\ \gamma_{\mathrm{L}}(b) &= T_{\mathrm{mid}}(b-1, b) \\ \gamma_{\mathrm{B}}(b) &= T_{\mathrm{mid}}(b, b) \end{aligned} \tag{15}$$

The bar-separation $\Delta T(b)$ can be used as a means for judging the bar setting, independently of how it was arrived at. Thus, if $\Delta T(b) > 0$ it is good otherwise the setting may need further investigation. In Section 7 and Section 8 we saw that both ΔG and ΔS had similar crossing points near grade `3G`. A bar at `3G` is positioned well up the the curve $\Delta T(b)$, and so is judged to be good.

For non-ideal tournaments, the shape of $\Delta T(b)$ is much sharper than illustrated in Figure 19. It is usefully modelled by a ramp [11] function, discussed in Appendix A.

10 BAR DETERMINATION METHOD

The functions ΔG in Section 7 and ΔS in Section 8 provide us with the main tools for determining the bar from the tournament's entry. We firstly present a method for doing this, and then specify a Monte Carlo simulation trial to test the assumptions made in the method and to explore round-dependent features.

10.1 Statement of the method

We are given a tournament with a specified entry and number of rounds.

Algorithm 4. `find-bar`

B1. Set the bar at the highest grade.

B2. Simulate the tournament for N_p pairings.

B3. Collect the mean group-score and mean opponent-grade for the group at the bottom of the current bar.

B4. Lower the bar to the next non-empty group.

B5. Continue at B2 for 4 successive bars.

B6. Fit linear models to the mean group-score and mean opponent-grade data.

B7. Apply Equation (16) to obtain the bar setting.

In step **B6**, we fit linear models described in section 3.4 to find the zero crossing points and equilibrium grade. This step always succeeds. If there is just one group in the tournament it falls back to Swiss and there is no bar. This is equivalent to setting the bar in a McMahon tournament to $g_{\min}$. If there are two or more groups, then there is always a solution, because the variance of the sequence of bar values is non-zero.

In the above linear models, a higher value of K implies a more reliable crossing point. As we have seen in the graphs of ΔG in Figure 13 and ΔS in Figure 15, the zero crossing points are similar, but not identical. We take account of the crossing point reliability in step **B7**, by minimising the square sum of the linear models as in (4) to produce a weighted solution b_{sol} which always lies between the above crossing points:

$$b_{\text{sol}} = \frac{K_G^2 Z_G + K_S^2 Z_S}{K_G^2 + K_S^2} \tag{16}$$

The actual bar grade B_{sol} is obtained as the nearest non-empty grade to b_{sol}.

The simulation time for a single pairing increases as n^3, so in step **B2** we need to keep the entry in the simulated tournaments to the minimum possible. However we do not need to simulate the *entire* entry in a tournament. We certainly need the top 4 groups and all the groups below the top 4 which might possibly influence the performance of the top players. We thus only need to simulate the top $4 + r$ groups at most.

We have in effect seen the successful application of the above algorithm to the case of T_{ideal}.

10.2 Scope of the Monte Carlo trial

We specify a large-scale trial whose purpose is:

- Expose any entries preventing solution.
- Identify outliers.
- Identify features exhibiting clear dependency on the number of rounds.
- Generate a bar population table to compare with the traditional guidelines.

10.3 Solution failure

Some particular entries could prevent the above algorithm from producing an acceptable bar. Possible conditions for solution failure are:

There may be a flawed pairing. This would cause step **B2** to fail. A few such failures in repeated trials of the same entry would not be important, but a high rate of failure would mean that the entry is not capable of supporting a McMahon tournament.

Unacceptable solutions. For example, if the models for $\Delta G(b)$ and $\Delta S(b)$ have weak slope parameters, the solutions may be very far apart and the square sum *may* have a minimum which lies outside the range of grades in the tournament.

10.4 Variation with number of rounds

We now have the means to generate a classic table relating the bar population to the number of rounds. There are a number of other important features that may show a trend with number of rounds. One of these is the *bar-depth* defined as the difference between $g_{\max}$ and B_{sol}. Others include:

Solutions of Z_G and Z_S may differ. In $\mathrm{T}_{\mathrm{ideal}}$ we saw that the solutions were similar. We collect the coefficient for the correlation between zero crossing points.

Bar separation. We monitor the ramp hinge and slope parameters for ΔT .

Bar population. We obtain the population distribution and generate a bar table from the mean bar population as a function of the number of rounds.

Quality. As already mentioned, it is difficult to set the bar giving a population which guarantees a unique winner . We monitor the values for probability of a unique winner, as well as rank-deviation.

10.5 Tournament entry

The trial generates tournaments for rounds ranging sequentially from 2 to 10. For each number of rounds r, we first choose a total even entry n at random from a range $[n_{\min},\ n_{\max}]$ depending on r:

$$n_{\min} = N_{\min} + rR_{\min}$$
$$n_{\max} = N_{\max} + rR_{\max}$$

The variation in limits helps to provide a reasonably full population for the relevant higher grades in the tournament, but also allows some grades to have a zero entry. Once the total entry is known, we then assign to each of the n players a grade chosen at random from the range $[g_{\min},\ g_{\max}]$.

The limits for the entry generation are set to the values in Table 5.

	$N_{\min}$	$R_{\min}$	$N_{\max}$	$R_{\max}$	$g_{\min}$	$g_{\max}$
value	20	4	25	5	`-8G`	`4G`

Table 5: Limits for the tournament entry

For a 10 round tournament, the maximum number of players is 75. A `-8G` player could affect a bar set at `2G` by winning all games (but this is virtually impossible as we have seen). For a two round event, the minimum population is 28 and the `-8G` player would affect a bar set at the very unlikely grade `-6G`.

Once an entry is generated as described above, we find the bar according to Algorithm `find-bar`. Then repeat the process for 200 different entry samples, all with the same number of rounds.

10.6 Pairing sample rate

For each bar setting scanned by Algorithm `find-bar`, we need to simulate a number of pairings. All the results obtained for $\mathrm{T_{ideal}}$ are based on a sample of 1000 pairings. A sample rate lower than 100 can produce some severe distortions. For example, consider the 5 round tournament $\mathrm{T_{m\text{-}gap}}$ with entry:

grade	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4
entry	6	0	5	4	0	5	4	6	4	0	3	4	1

Table 6: $\mathrm{T_{m\text{-}gap}}$ entry 5 rounds 42 players

The leftmost plot in Figure 21 shows the Δ_G and Δ_S values for 10 pairings of $\mathrm{T_{m\text{-}gap}}$. The variations in the values are significantly reduced on simulating the tournament with 100 pairings as seen in the centre plot of Figure 21. Increasing the numbers of pairings beyond 1000 does not yield visible improvement.

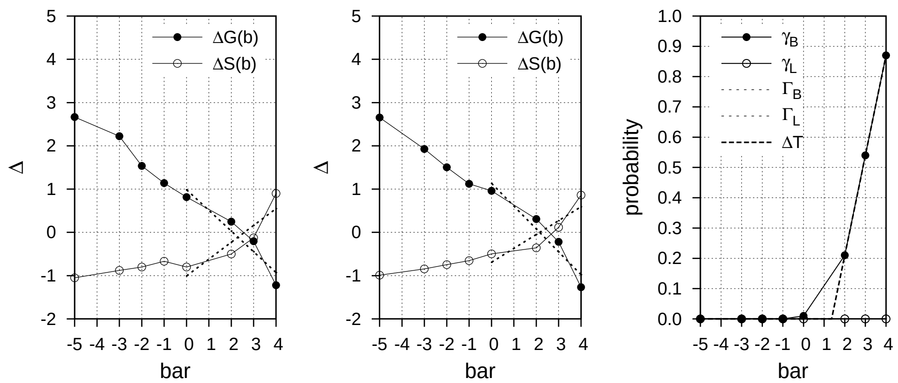


Figure 21: Sampling at 10 and 100 pairings for $\mathrm{T_{m\text{-}gap}}$

The rightmost plot shows values for the bar separation function ΔT discussed in section 9.3. A model for the bar-separation is presented in Appendix A. The term $\Gamma_{\mathrm{B}}(b)$ is the dominant component of the model for ΔT and this is shown in the rightmost plot of Figure 21.

Values for the derived parameters required in the bar solution and bar separation at increasing numbers of pairings are shown in Table 7

With 100 pairings per tournament the accuracy of b_{sol} at (± 0.1) is sufficient for our purposes. The steady state values for H_B and K_B are reached at 2000 pairings, beyond which the change is within (± 0.02)

pairings	b_{sol}	H_B	K_B
10	2.270	0.833	0.400
100	2.149	1.364	0.330
1000	2.126	1.465	0.338
2000	2.132	1.548	0.354

Table 7: Parameter accuracy dependent on number of pairings

10.7 Software

Software written to produce all the results in this document is called `toursim` [12]. It is a Linux command line application released under the GNU General Public License.

This program is a work in progress: there is no manual but the source is commented. The output files [13] required for results in this document contain extensive supporting detail.

11 MONTE-CARLO TRIAL RESULTS

11.1 Solution failures

There are 1800 different tournaments simulated over the entire round range and 100 complete pairings per entry generated. There are no flawed pairings in the 180,000 pairings sampled.

The solution b_{sol} lies above the maximum grade in 4 tournaments (all 2 rounds), but in all cases by less than 0.6 grade. In all these cases the maximum grade has an entry of 4 or more. Since there is an absolute limit on the bar population of 2^r players, there would be no need in practice to invoke Algorithm `find-bar`. The algorithm nevertheless does produce the value $B_{\text{sol}} = g_{\max}$ in these cases.

11.2 Solution correlation

The solution pair (Z_G, Z_S) for every entry in the trial is presented in the scatter plot Figure 22. The three points **A,B,C** identify clear outliers. The rectangle contains the 2-round tournaments discussed in the previous section.

point	rounds	$Z_S - Z_G$	-1G	0G	1G	2G	3G	4G	bar
A	7	0.90	6	7	1	0	1	6	3G
B	4	1.14	5	0	0	2	1	4	3G
C	4	1.37	6	6	3	0	1	5	4G

Table 8: Outlier features

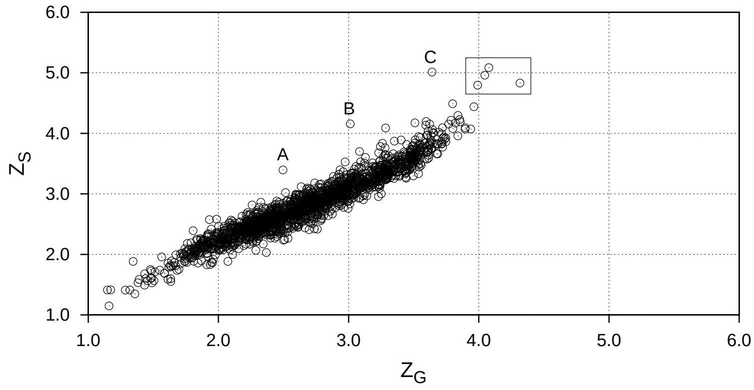


Figure 22: Scatter plot for Z_S vs Z_G in mcmahon-pairing

One common feature of the models for the outliers seen in Figure 23 is that that the fit to $\Delta G(b)$ and $\Delta S(b)$ seems to be poor. Even if we were to interpolate the solution for the crossing points *directly* from the raw data, we still obtain estimates for B_{sol} which agree with Algorithm `find-bar`.

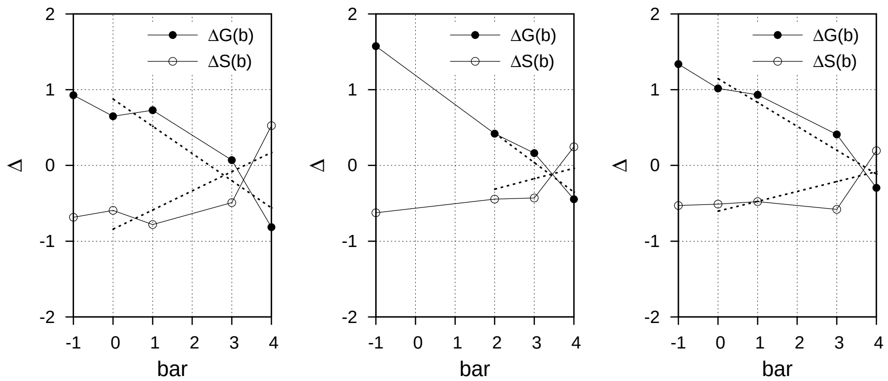


Figure 23: The $\text{T}_{\text{outlier}}$ models

Simple interpolation may fail however if the changes in sign for $\Delta G(b)$ and $\Delta S(b)$ lie at the ends a large grade gap. The entries for the higher grades in the outliers are given in Table 8, and they all show grade gaps 1 or more grades below the solution point.

11.3 Bar Depth

For any number of rounds, the observed bar-depth lies in the range $0 \cdots 4$. The left hand plot in Figure 24 shows histograms for the bar-depth distributions.

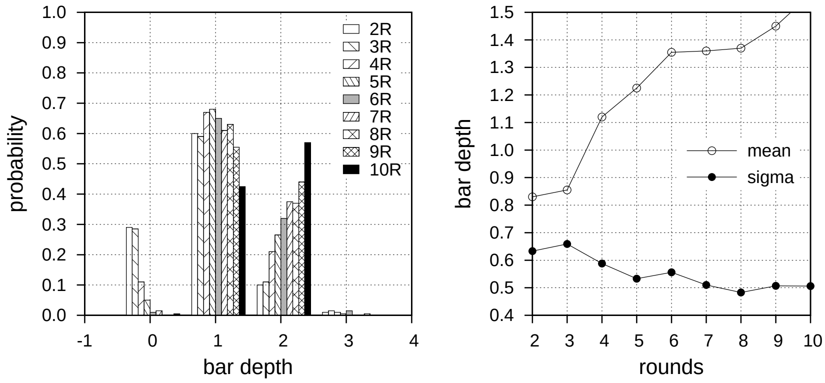


Figure 24: Bar depth statistics

The individual plots are displaced along the bar-depth axis to distinguish the different number of rounds.

For each number of rounds, the most common bar-depth is 1 grade, i.e. the bar is just 1 grade below the highest grade.

There is 1 case with a bar-depth of 4. This is a 6 round tournament, where the top five grades from `0G` to `4G` have entries 7,0,0,2,3, and the bar obtained is at grade `0G`.

It is very rare for the bar-depth to be as high as 3. In almost all cases when this happens, there is a gap in the entry at the higher grades. When there is a missing grade say g_{miss} near g_{max} but nevertheless a healthy population above g_{miss}, the bar sometimes does not cross the gap. This is illustrated in the tournament $\mathrm{T}_{\text{m-gap}}$ discussed in Section 10.6. It is a 5 round event with $B_{\text{sol}} = 2$ giving a population of 8 players. There is a gap at `1G` and then a group of 4 players at `0G`. Crossing the gap would induce large changes in the values of ΔG and ΔS, moving the solution well away from optimal.

There is *one* case where the bar-depth is 3, and there are no missing grades. This is also a 6 round tournament with top 5 grades 7,5,2,1,1. Although there are no missing grades, the top groups are very thin and the algorithm lowers the bar to `1G` giving a bar population of 9.

The right hand plot in Figure 24 shows that the mean bar-depth increases with the number of rounds, but the rate of change per round is quite small averaging 0.09 per round.

11.4 Bar Separation

As mentioned in section 10.5, we smooth the individual components of $\Delta T(b)$ via the ramp models $\Gamma_B(b)$ and $\Gamma_L(b)$ discussed in Appendix A. Figure 25 shows

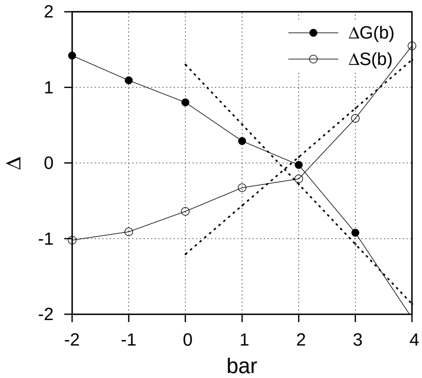
grade	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4
entry	7	3	4	5	5	4	2	3	3	9	1	3	1

Table 9: $\mathrm{T}_{\mathrm{mcmahon}}$ entry 7 rounds 50 players

the models arising from a 7 round tournament $\mathrm{T}_{\mathrm{mcmahon}}$ and Table 9 shows the entry generated in the trial.

The left hand plot indicates a bar solution $B_{\mathrm{sol}} = 2\mathtt{G}$. The right hand plot confirms that this bar is well up on the ramp Γ_{B}.

The *BGA* tables in Section 1 suggest a maximum bar population of 18 players for a 7 round tournament, and Table 9 would then allow a bar at 0G. At this grade ΔT is zero, meaning that there is no distinction in the performance of 0G from -1G players.



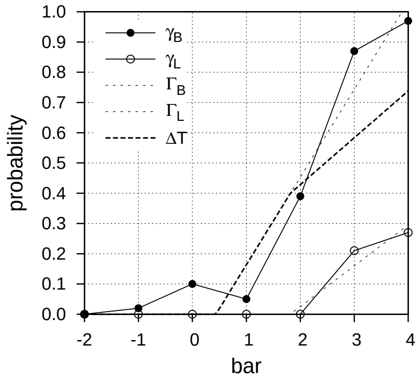


Figure 25: Solution process and ramp models for $\mathrm{T}_{\mathrm{mcmahon}}$

We are interested in the probability P_{sep} that the bar-separation is non-zero at the bar $b = B_{\mathrm{sol}}$. The slowly decreasing values in Figure 26 are consistent with the observed increase in bar-depth for increasing population found in section 11.3. For players at the bottom of a deep bar will meet an increased number of stronger opponents and so find it harder to beat the mid-score.

The dominant component in ΔT is the ramp function Γ_{B} with hinge point H_B, and it is of interest to see how far up this ramp the bar solution rises. To this end we consider a normalised value:

$$\Theta = (b_{\mathrm{sol}} - H_B)/(g_{\mathrm{max}} - H_B)$$

The value of Θ lies in the range $[0, 1]$ when b_{sol} is on the ramp, and becomes negative when it is below the ramp's hinge. The higher the value of Θ, the better is the separation in the performance of players at the bottom of the bar, compared to those below the bar.

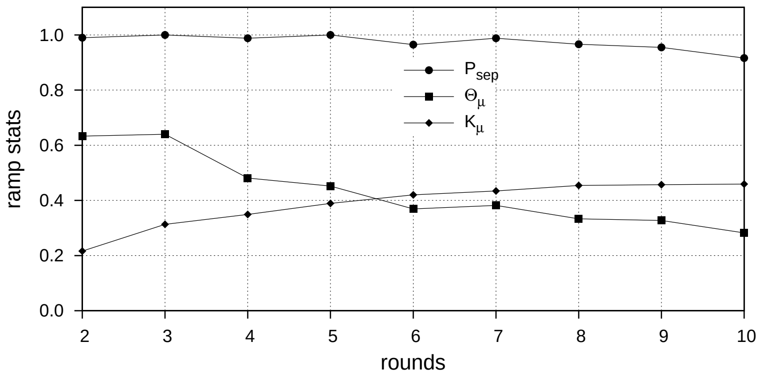


Figure 26: Ramp statistics

The slope K_B in the ramp model has a maximum value of 1 corresponding to the case where the bar is at the maximum grade and the value of Γ_B is zero at grade $g_{max} - 1$. The quantities Θ and K_B all have the same range and are presented along with P_{sep} in Figure 26. The mean normalised separation Θ_μ shows a decreasing trend with number of rounds. This means that b_{sol} is getting closer to the hinge point of the ramp, and this is consistent with the decrease in P_{sep} observed above.

11.5 Bar Population

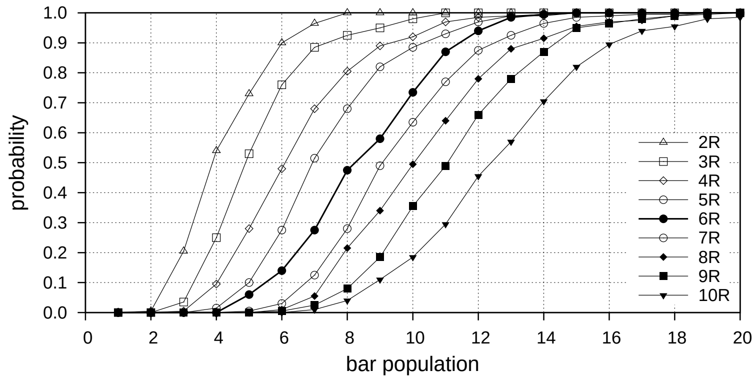


Figure 27: Cumulative distribution for the bar population

The underlying distribution for the bar population shown in Figure 27 reveals a fairly regular dependence on the number of rounds.
This regularity emerges in the bar table, which here we present graphically in Figure 28. The dotted graphs display the BGA minimum and maximum population [14] range. The actual obtained minimum in the trial is somewhat

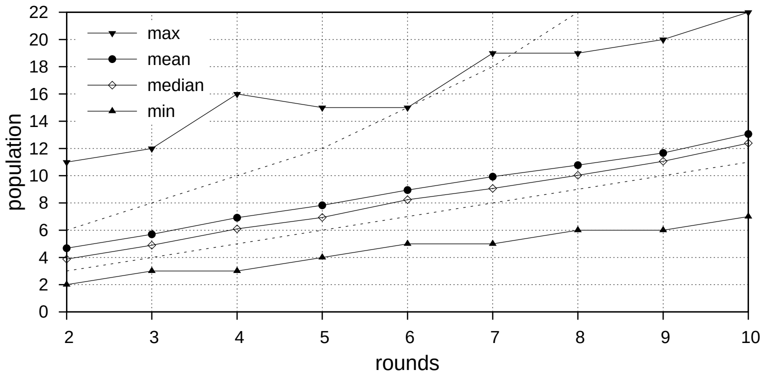


Figure 28: Bar table statistics

below the classical minimum, defined as one more than the number of rounds. The cdf plot in Figure 27 indicates for example, that in a 6 round tournament, there is a 14% chance that the bar population is lower than the number of rounds.

Note that the trends observed in this trial are to some extent associated with the assumptions made for the dependence of total entry size on the number of rounds, as defined in Table 5. The rough nature of the trial results for minimum and maximum populations reflects the low number (200) of different tournaments sampled for each round, but the trends are clear.

11.6 Uniqueness and ranking

It is desired that a tournament produces a unique winner without recourse to tie-breaks. There is also a desire that the tournament should produce a sensible ranking for runners up. These requirements are often incompatible.

In Figure 29 we present a normalised version of the rank deviation illustrated in Figure 11, along with the probability of a unique winner for all the possible bar settings in the tournament $\mathrm{T}_{\text{m-gap}}$.

The normalised grade-rank R_G and mcmahon-rank R_S have values lying in the range $[0, 1]$ and are defined by:

$$\begin{aligned} R_G(g) &= (g - g_{\min})/(g_{\max} - g_{\min}) \\ R_S(g, b) &= (s(g, b) - S_{\min})/(S_{\max} - S_{\min}) \end{aligned} \tag{17}$$

In zero-shodan units $S_{\min} = g_{\min}$, and $S_{\max} = b + w_{\max}$, where $w_{\max}$ is the maximum number of wins above the bar. The player's mcmahon-score is given by $s(g, b)$.

The graph labelled δ_{rank} shows the values of the rms-difference between $R_G(g)$ and $R_S(g, b)$ summed over all grades and 1000 pairings in the simulation of

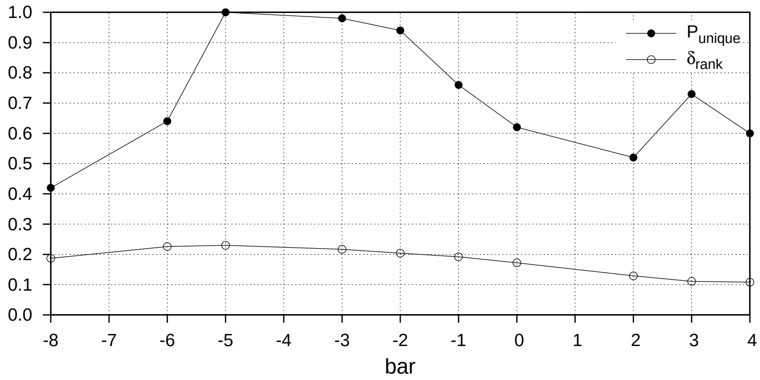


Figure 29: Unique winner probability and rank deviation for $T_{\text{m-gap}}$

$T_{\text{m-gap}}$. It decreases slowly with increasing bar, and the behaviour seen here is typical.

The graph P_{unique} shows a modest peak at bar 3G, but the behaviour near the bar solution takes on many different forms for different entries. $T_{\text{m-gap}}$ provides a unique winner at a bar grade -5G. Referring to Table 6, this gives a population of 31. A player is drawn up at round 1 to give a perfect population of 32 for a 5 round tournament. Such a bar is of course out of the question!

12 GUIDELINES

In the light of the results obtained in this study, we revisit the guidelines for setting the bar.

12.1 Classical guidelines and maxims

Where possible, we supply additional information to clarify the guideline.

Bar set too high. Top players run out of peer opponents early. For T_{ideal} the proportion of peer games decreases slowly as the bar is lowered from the maximum (Section 6).

Bar set too low. The top players may never meet. For T_{ideal} the proportion of peer games played by the **topmost** group drops from 50% to 26% when the bar drops to shodan.

Smallest bar population It is generally accepted that the population above the bar should be at least the same as the number of rounds. However the bar population distribution obtained in the Monte Carlo trial does not exclude values lower than the number of rounds.

Winning chances above the bar. The winning probability above the bar is very sensitive to grade, and players certainly do not have an *equal* chance of winning. There is no natural way of defining a 'reasonable' chance to win.

Unique winner For tournaments of 4 rounds or more, the simulations carried out in this study showed that on average, the probability of a unique winner is less than 60%

12.2 Additional guidelines

The emphasis throughout has been that the position of the bar is sensitive to the actual number of players in each grade group. There is no simple formula for setting the bar given the entry distribution, and in the end we require an algorithm such as the one discussed in Section 10 to produce a solution. Some simple rules of thumb have emerged though, and these might be useful when setting the bar manually.

Avoid the highest grade. It is unusual for the bar to be set at the highest grade for tournaments of more than 4 rounds. In the Monte Carlo trial, the probability of a bar at the maximum grade is less than 2%.

Use the highest grade where needed. For the smaller tournaments in the range 2 to 4 rounds, the probability of a bar at the top grade ranges from 10% to 30%. This may happen when there is a relatively large top group followed by a gap in the entry.

Bar depth. For most tournaments, the bar-depth should be 1 or 2. It is very unusual for the bar-depth to be 3, and the most common bar-depth is 1 for all number of rounds.

Mind the gap. Where there is a grade with zero or very few players within 4 grades of the maximum grade, be careful of crossing the gap to boost the bar population. This can lead to an excessive bar-depth and poor performance for players at the bottom of the bar. The tournament $T_{\text{m-gap}}$ is a case in point. It might be tempting to cross the gap at `1G` and set the bar at `0G` to give the populations of 12 according to Table 6. The graphs in the centre plot of Figure 21 show just how far the performance of the players at the bottom of the bar would move from the ideal.

13 SUMMARY AND CONCLUSION

13.1 Assumptions

The key assumptions made in this study are:

Strength. A player's grade defines the exact strength of the player used in simulating game results.

No repeats. There are no repeat games, irrespective of the pairing method.

Pairing. Apart from random-pairing, the pairing algorithm applies a maximum weight to players with equal scores.

Entry distribution. The number of players in a tournament is even, and we assume that the available pool of players increases linearly with the number of rounds. Player grades are chosen at random from a fixed range.

Winning probability. The probability of win between players of differing strength is derived from the E.G.D published statistics and is defined in Appendix B.

13.2 Key quantities

The study has identified key quantities applicable to any pairing method.

Mid score. In any single tournament, the mid-score for random- and swiss-pairing is half the number of rounds. For mcmahon-pairing it depends on the group grade, and is the initial mcmahon-score plus half the number of rounds.

Group score. The average of the scores of all players in the group.

Group opponent-grade. The average grade of all opponents of players in the group.

Mid Tail distribution. Probability that the mean group-score exceeds the mid-score.

Rank deviation. Each player is assigned a grade rank and a score rank as defined in Equation (17). The rank-deviation is the root mean square difference in the two rankings.

13.3 Conclusion

The main conclusions to be drawn from this study are:

Algorithm. A Monte Carlo algorithm for determining an equilibrium-grade from the player entry grades in Random, Swiss, or McMahon tournaments has been identified. It relies on minimising a quadratic form constructed from the mid-score, group-score, and opponent-grade.

Sensitivity. The equilibrium-grade is very sensitive to the exact distribution of player numbers in the top grades.

No formula. No simple formula in Swiss or McMahon tournaments has been found which relates the equilibrium-grade to the number of rounds or entry statistics, such as the first 4 moments or entropy of the populations per grade.

Unique winner. The probability of a unique winner is a function of the bar. No feature of this function has been found which is directly related to the equilibrium grade for arbitrary entries.

Rank deviation. The rank-deviation is a function of the bar. This function generally rises for low bar settings and then falls gently as the bar is increased to the maximum grade.

The bar. The equilibrium-grade is a prime candidate for setting the bar in any McMahon tournament.

A MODEL FOR BAR SEPARATION

The behaviour of $T_{\mathrm{mid}}(b,g)$ along the bar-layer line $g = b$ as illustrated in Figure 19, consists of a long horizontal flat section followed by a rapid and fairly linear rise. This behaviour can usefully be modelled by a ramp-function. The ramp turns at a point H_B called the hinge, and can be expressed in the form:

$$\begin{aligned} \Gamma_{\mathrm{B}}(b) &= K_B(b - H_B) \quad \text{for} \quad H_B \le b \le g_{\max} \\ &= 0 \qquad\qquad\quad\ \text{for} \quad b < H_B \end{aligned} \tag{18}$$

For our purposes the sloping part can be obtained by a least squares fit to $\gamma_{\mathrm{B}}(b) \equiv T_{\mathrm{mid}}(b,b)$ values for the 3 highest non-empty grades.

The mid-score performance for players just below the bar can also be represented by a ramp function. In this case the hinge is usually at a much higher grade than for the bar-layer so the ramp part is very short. The ramp is obtained by a least squares fit to the top three non-empty grades for $\gamma_{\mathrm{L}}(b) \equiv T_{\mathrm{mid}}(b, b-1)$.

If there are missing grades in the entry near $g_{\max}$, then there are some points where the raw data $\gamma_{\mathrm{B}}(b)$ is not defined. Such points are ignored in the least squares process, and we simply search for the three largest grades with non-zero populations.

Missing grades affect the data for $\gamma_{\mathrm{L}}(b)$ more severely, since even if the entry at grade b is non-zero, there might be a gap at $b-1$. Again, for each non-empty grade b, we search for the highest non-empty grade below b. This grade can be denoted by $b \ominus 1$. So the raw data required for modelling the performance of players below the bar takes the form:

$$\gamma_{\mathrm{L}}(b) = T_{\mathrm{mid}}(b, b \ominus 1), \quad n(b) > 0$$

This raw data leads to a ramp model $\Gamma_L(b)$ having a hinge at H_L, and slope K_{L}.

The model for the bar-separation $\Delta T(b)$ function is defined as the difference in the ramp models across the bar-layer:

$$\Delta T(b) = \Gamma_{\mathrm{B}}(b) - \Gamma_{\mathrm{L}}(b) \tag{19}$$

B PROBABILITY OF WIN

A model [2] for the probability of win between players with ratings r and s is based on published data provided by the European Go Database. The model can be expressed using the standard S-shaped error function erf as follows:

$$\begin{aligned} p(r,s) &= \tfrac{1}{2}[1 - erf(\Lambda(r,s))] \\ \Lambda(r,s) &= \sum_{n=0}^{3} h_n(s-r)e^{n\,K\,\min(r,s)} \\ h_n(x) &= u_n x + v_n x^3, \quad n = 0 \cdots 3 \end{aligned}$$

Here u_n and v_n are *positive* constants specifying monotonic increasing cubics, and the best-fit non-zero values for the coefficients are:

$\mathbf{u_0}$	$\mathbf{v_0}$	$\mathbf{u_1}$	$\mathbf{u_3}$	**K**
0.0351224	0.00445376	0.156777	0.0164481	0.18818

Table 10: Coefficients for winning probability

C GLOSSARY

all-play-all (p7): Each player has one game against each of the others.

bar-depth (p30): The difference between the bar grade and the maximum grade.

bar-group (p20): The collection of players with grades $\geq$ bar.

bar-layer (p21): A line in the (bar,grade) plane representing players at the bottom of a variable bar.

BGA (p35): British Go Association.

crossing-point (p11): A sequence $f(i)$ changes sign between i and $i+1$. Inverse interpolation provides the crossing point. .

edges (p7): An edge in a graph contains exactly two distinct vertices.

E.G.D (p6): European Go Database.

equilibrium-grade (p14): A grade minimising the square sum of ΔS and ΔG.

flawed (p8): A tournament is flawed if the pairing leaves two or more players unpaired in any round.

grade-rank (p18): An ordering by player grade.

graph (p8): A graph is formed from a set of vertices and a set of edges joining some of the vertices.

group (p7): For a given tournament, the set of all players of the same grade.

group g (p7): Each player in the group has grade g.

group-score (p11): The average score of all players in the group.

initial-mcmahon-score (p22): The player's grade if player is below the bar, otherwise the bar grade.

matching (p8): A matching is a set of edges with no vertex in common.

maximum cardinality (p8): The matching has the largest possible number of edges.

maximum weighted matching (p15): Each edge is given a weight, and the algorithm chooses a matching with the largest sum of weights.

mcmahon-score (p5): The initial-mcmahon-score plus number of wins.

mcmahon-pairing (p7): Players on the same mcmahon-score are paired.

mcmahon-rank (p18): An ordering by player mcmahon-score.

mid-score(Random or Swiss) (p12): Half the number of rounds.

mid-score(McMahon) (p22): Half the number of rounds incremented by the bar grade.

opponent-grade-difference (p10): The difference between the average grade of all opponents of a group and the group grade.

peer-games (p19): Games between players in a top-group.

perfect (p8): A pairing is perfect if every player is paired.

random (p7): Drawn from a discrete uniform distribution.

random-pairing (p6): Players are paired at random, subject to the condition that there are no repeat games.

split-and-cycle (p9): One or more players remain in fixed seats, the others cycle at each round.

swiss-pairing (p7): Players with the same number of wins are paired.

tail distribution (p25): Probability $X > x$.

top-group (p19): Players in the range of groups from a given grade to the maximum grade.

unique winner (p5): Only one player achieves the maximum score.

vertices (p7): Edges in a graph meet in its vertices.

zero-shodan units (p7): Kyu grades are negative, dan grades start at zero, and grades increase by one unit.

D NOTATION

$\overline{X}$	Mean of a random variable for all samples in a trial.	p11
b	An arbitrary bar.	p19
b_{sol}	The raw bar solution.	p29
B_{sol}	The non-empty group nearest b_{sol}.	p29
g	grade.	p10
$g_{\max}$	Highest grade in the entry.	p10
$g_{\min}$	Lowest grade in the entry.	p10
G_{c}	Equilibrium grade.	p11
$G(g)$	Average opponent-grade for players of grade g.	p10
$G_g(b)$	Average opponent grade for group g at bar b.	p20
$\Delta G_g(b)$	Mean difference $G_g(b)$ from g.	p20
ΔG	Mean difference of opponent-grade from group grade.	p11
H_B	Hinge point for ramp Γ_B.	p42
H_L	Hinge point for ramp Γ_L.	p42
K_B	Slope of ramp Γ_B.	p42
K_L	Slope of ramp Γ_L.	p42
$\Psi_g(b)$	Peer-games ratio for top-group g and bar b.	p19
P_{sep}	Probability$\Delta T(b) > 0$ at bar solution B_{sol}.	p35
r	Number of rounds.	p8
$s(g)$	Group score for group g in random- or swiss- pairing.	p11
$S(g)$	Mean group score for group g.	p11
$s(g, b)$	Group score for group g at bar b.	p22
$S_g(b)$	Mean group score for group g at bar b.	p22
S_{mid}	Mid score.	p12
$S_{\text{mid}}(b)$	Average scores for players at bottom of bar.	p23
ΔS	Difference mean group-score from mid-score.	p12
$T(g, b, s)$	Tail distribution for given bar and grade.	p25
$T_{\text{mid}}(g, b)$	Tail distribution for score $s = S_{\text{mid}}(b)$.	p26
$\Gamma_B(b)$	Linear model for $T_{\text{mid}}(b, b)$.	p31
$\Gamma_L(b)$	Linear model for $T_{\text{mid}}(b, b-1)$.	p42
$\Delta T(b)$	Bar separation.	p28
Θ	Relative ramp distance.	p35
Z_G	Crossing point for ΔG.	p13
Z_S	Crossing point for ΔS.	p13

E ALGORITHM INDEX

1. `simulate-flawed-tournaments` p8
2. `simulate-result` p11
3. `swiss-pairing` p16
4. `find-bar` p28