Report of the
USCF Ratings Committee
August 1996

Mark E. Glickman, Ph.D.
Chairman, USCF Ratings Committee

Contents

Preface – General Philosophy of Current Ratings Committee Work

Ratings Committee work currently is dominated by two goals:

- To produce a rating system that *predicts* performances as accurately as possible, and
- To produce separate measurement systems that may be based on the rating system which serve other functions, such as rewarding players for meritorious performances, or determining sectioning strategies for tournaments.

The current ratings system serves many functions, some of which conflict with each other. For example, while the current rating system tries to predict performances accurately, it also serves as a promotional tool where a player may be more encouraged to continue tournament participation if his or her rating increases, and conversely discouraged if a player's rating decreases. To relieve this and other burdens from the rating system, the rating system is being viewed solely a predictor of performances. Other measures (such as titles), whose computation may rely on ratings, are intended to enhance and encourage tournament participation. This year's proposal takes steps towards implementing this philosophy.

Ratings Committee Members

Christopher Avery Ph.D. in Business (Economics), Stanford University (1993). Diploma in Mathematical Statistics, Cambridge University (1989). B.A. in Applied Mathematics, Harvard University (1988). Currently holds an Assistant Professorship of Public Policy, Kennedy School of Government, Harvard University.

Harry Cohen Ph.D. in Operations Research, Massachusetts Institute of Technology (1975). Member of the USCF Ratings Committee 1988–present. Currently a principal in a transportation planning and management services consulting firm.

Dmitry Dakhnovsky B.S. in Computer Science, Economics and Mathematics, Carnegie Mellon University (expected 1997). M.S. in Software Engineering, Carnegie Mellon University and Software Engineering Institute (expected 1998). Programmed and implemented the "Glicko" rating system for the Internet Chess Club.

Thomas Doan M.A. in Economics, University of Minnesota (1980), M.A. in Math, University of Wisconsin (1977). Currently developer of statistical software.

Mark Glickman Ph.D. in Statistics, Harvard University (1993). B.A. in Statistics, Princeton University (1986). Member of the USCF Ratings Committee 1985–present; chairman 1992–present. Currently holds an Assistant Professorship of Mathematics at Boston University.

Bill Goichberg B.A. in Political Science, New York University (1963). USCF Policy Board member, 1975–1978, and 1989–1992. Professional chess organizer and director.

Albyn Jones Ph.D. in Statistics, Yale University (1986). B.A. in Mathematics, UCLA (1978). Member of the USCF Ratings Committee 1992–present. Currently holds an Associate Professorship in Statistics at Reed College, Portland, Oregon.

Alan Losoff B.S. in Mathematics, Illinois Institute of Technology (1969). Proofreader for "The Ratings of Chessplayers" by Arpad Elo. Member of the USCF Ratings Committee 1991–present. Computer programmer working on mathematical modeling for a financial derivatives firm.

Paul Matthews Ph.D. in Experimental Psychology, Stanford University (?). Designer of the current American Go Association (AGA) ratings algorithm, methodology, and software. Ratings Statistician for the AGA. Currently a Senior Consultant in Applied Research at Bellcore.

Andrew Metrick Ph.D. in Economics, Harvard University (1994). B.A. in Economics, Yale University (1989). Currently holds an Assistant Professorship of Economics, Harvard University.

Kenneth Sloan Ph.D. in Computer and Information Science, University of Pennsylvania (1977). Sc.B. in Applied Mathematics, Brown University (1970). Member of the USCF Ratings Committee 1992–present. Currently holds an Associate Professorship at the University of Alabama in Birmingham.

Ratings Committee Special Consultant:

Christopher Chabris B.A. in Computer Science, Harvard University (1988). Special Consultant to the USCF Ratings Committee 1993–present. Ph.D. Candidate, Department of Psychology, Harvard University. Editor in Chief, American Chess Journal.

Policy Board Liaison:

Frank Camaratta M.S. in Applied Mathematics, Applied Mechanics, and Aerospace Engineering, Drexel Institute of Technology (1968). Treasurer and former vice president of USCF. Chairman of the Computer Rating Agency, 1985–present. Ratings Committee chairman, 1986–1990.

Committee Motions

(Note: in the text, "the Committee Report" refers to this document)

ADM: USCF members with prison membership are precluded from appearing on any Top-50 list.

ADM: Ratings of USCF members with prison membership are to be treated as provisional based on one game once such players become regular members.

ADM: Lowering of rating floors: Effective immediately, a player's rating may not go 200 points below the highest 100-point boundary lower than a player's highest attained rating. In effect, this "drops" the rating floors by 100 points. If the player's highest rating, however, is lower than 1600, this rule is not in effect, and the player's rating "floor" is 0. Examples: (a) A player who has a highest rating of 1705 would have a rating floor of 1500; (b) A player who has a highest rating of 1695 would have a rating floor of 1400; (c) A player who has a highest rating of 1550 would have a rating floor of 0.

The Policy Board recognizes and authorizes the use of the FIDE-to-USCF and the CFC-to-USCF conversion as described in Section 3 of the Committee Report.

The Policy Board authorizes the use of the FIDE-to-USCF conversion, and the CFC-to-USCF conversion for the purpose of assigning ratings to USCF-unrated players with FIDE ratings or CFC ratings, as described in Section 3.3. The converted rating would be treated as a provisional USCF rating based on 10 games when updating ratings from an event.

The Policy Board recognizes that any proposed change to the USCF rating system or USCF Title System must be submitted to the Ratings Committee for study and recommendation.

The Policy Board authorizes the modification to the rating formula, described in Section 4.3 of the Committee Report, which, in effect, increases the value of K in the rating update formula for players with low ratings.

The Policy Board authorizes the use of the "delta" schedule for the USCF Title System as described in Section 5.2 of the Committee Report.

The Policy Board recognizes and authorizes the incorporation of "supernorms" into the USCF Title System, along with the proposed supernorm criteria specified in Section 5.3 of the Committee Report.

The Policy Board authorizes the proposed rule in the USCF Title System that a norm is only in effect for three years, after which it expires.

The Policy Board authorizes the proposed rule in the USCF Title System that a norm cannot be won in Quick chess, $\frac{1}{4}K$ and $\frac{1}{2}K$ events.

1 Summary of Ratings Committee Work

Below is the text that appears in the 1996 USCF Annual Report. Citations to the "Ratings Report" refer to this document.

The Ratings Committee has tackled a variety of issues this past year; several of them involved refining unfinished Committee proposals, and the rest involved new business. We summarize our work below. A detailed account of our work can be found in the "Report of the USCF Ratings Committee, August 1996" (henceforth, the "Committee Report"). This can be obtained at the Ratings Committee workshop, the Delegates' meeting, as well as through the World Wibe Web at http://math.bu.edu/people/mg.

The structure of the USCF Title System will be undergoing major changes, as described in the 1994 Ratings Committee Report. Three aspects of the system had been unresolved, so the Committee has addressed them. First, we proposed that titles will correspond to playing strengths at ratings of even multiples of 100 from 400 to 1800, and then at every 100 rating point up through 2400. A list of the title names are in the Committee Report. Second, in order to make the titles corresponding to 2200 and above more prestigious, we have proposed an extra criterion for earning norms for these titles. Specifically, we proposed that at least 2 of the required 5 norms for these titles must be earned in "supernorm" events. Supernorm events are either Grand Prix events with at least 20 total Grand Prix points available, or one of a list of prestigious or "heritage" events (such as any the U.S. Amateur Team Championships, the Pan-Am Intercollegiate Team Championships, and so on). The list of proposed events appears in the Committee Report. Finally, we proposed to add the restriction that norms generated from an event only last for three years, after which the norms expire if the title is not earned. This has the benefit of encouraging tournament participation, which we think will be helpful in promoting competitive chess.

A recurring issue, raised again this year, was whether to abolish rating floors. Most Committee members have strong feelings that the rating floors serve purposes which are more soundly implemented through other approaches, such as the title system or through information on maximum attained rating. Also, the rating floors are at odds with the notion that the rating system intends to predict game outcomes, and therefore their existence seem counter to purpose of the rating system. As a result of presumed rating inflation of lower rated players, the Policy Board has passed a motion in February to abolish rating floors below 1400 (except for the floor at 0). Several members of the Committee have made known their feelings that all the floors (except the floor at 0) should be removed as soon as possible. Recognizing that removing all the rating floors may result in unwanted rating deflation, as well as the possible anti-promotional effect that players may compete less often knowing that their rating has no floor, we proposed a gradual abolition of the rating floors while the new title system eventually takes hold. The details of the gradual floor abolition are summarized in the Committee Report. To counteract possible rating deflation, we are proposing to the Policy Board a temporary anti-inflationary system, the details of which can be found in the Committee Report.

It was brought to our attention that ratings of players in prison systems may be out of sync with the rest of the rating pool. For example, Claude Bloodgood, who, by his own admission, is probably not better than expert strength,[1] has a rating in the high 2600's because he is so much stronger than his chessplaying prisonmates. Thus Bloodgood's name appears on the USCF top-50 list when it is not clear his strength is as high as his rating indicates. The Committee proposes that players with prison memberships not be allowed to appear on any top-50 list, and to treat their ratings as provisional based on 1 game as soon as they no longer have a prison membership.

The Policy Board passed a motion in February that prevents opponents of players at their rating floor in round robin events to increase in rating. The Committee felt that this was an unnecessary motion, and that it may result in smaller attendance at Quads and other round robin events. We proposed at the May 1996 Policy Board meeting that this motion be reversed. The original motion was rescinded.

The Committee was charged early this year in responding to whether unplayed drawn games (assuming they became officially "recognized") should be rated. The Committee, with the exception of two members, answered by proposing that unplayed drawn games should not be rated, simply because an unplayed game provides no evidence of playing strength.

Finally, a change was proposed in the method to update USCF ratings based on FIDE events. The change, which is based on the method proposed in the "1994 Report of the Ratings Committee", downweights the impact of the FIDE events by a factor of 4. This change was suggested because reporting of FIDE events have been late by as much as 6 months relative to USCF events, so the information about current playing strength is not demonstrably trustworthy. We would like to thank Nick deFirmian for bringing this issue to our attention.

The Committee is considering the eventual transition to a two-parameter rating system. Rather than each player having just a rating, each player would have, under a two-parameter system, both a rating and a measure of its uncertainty. Such a change to the system would have profound implications on rating changes, which would reflect the precision in players' ratings. A particular two-parameter system, the "Glicko" system, is already well-established on the Free Internet Chess Server. We are considering extensions to this system that are better able to track players who improve quickly over time.

[1]Since the time of submitting this summary, a letter from Claude Bloodgood appeared which indicated that he believes his rating to be somewhere between 2200 and 2400.

2 Rating Floors

(Ken Sloan)

The establishment of 100-point Rating Floors has proven to be a very controversial element of the USCF Ratings system. The Policy Board has been drawn into this controversy, voting on several changes to the system. Proponents of ratings floor claim that they serve a promotional purpose – allowing players to maintain their near-peak ratings as a source of pride. Opponents claim that they detract from the predictive value of ratings, and are inflationary. Some claim that floors are an anti-sandbagging mechanism. A growing number of players (our current estimate is 9% of all active players) are on their ratings floors. See Figure 2.1 for the proportion of players having established ratings ending in the two-digits 00 over the last several years.

The sense of the Committee is that promotional, pride, and anti-sandbagging issues are better served by the Title system. Titles record a player's peak performance and remain even if the player's strength (and rating) subsequently fall back. We agree that floors disturb the predictive function of ratings and are inherently inflationary. If the primary purpose of the ratings system is to be to measure and predict performance at the chess board, then ratings floors are undesirable.

Perhaps the most potent effect of ratings floors has been the stimulation of clever ideas on how to "game" the system. The common theme has been that a player on his floor is decoupled from his losses. Some claim that this encourages people to play. An equal number claim that players lose interest. These debates might be dismissed as simple differences of opinion. However, it turns out that floors have provided a mechanism by which unscrupulous players or organizers can, without penalty, manipulate the ratings system (in particular, artificially raise the rating on one player without affecting the rating of the player on his floor). Many of the hard problems referred to the Committee this year can be traced directly to ratings floors.

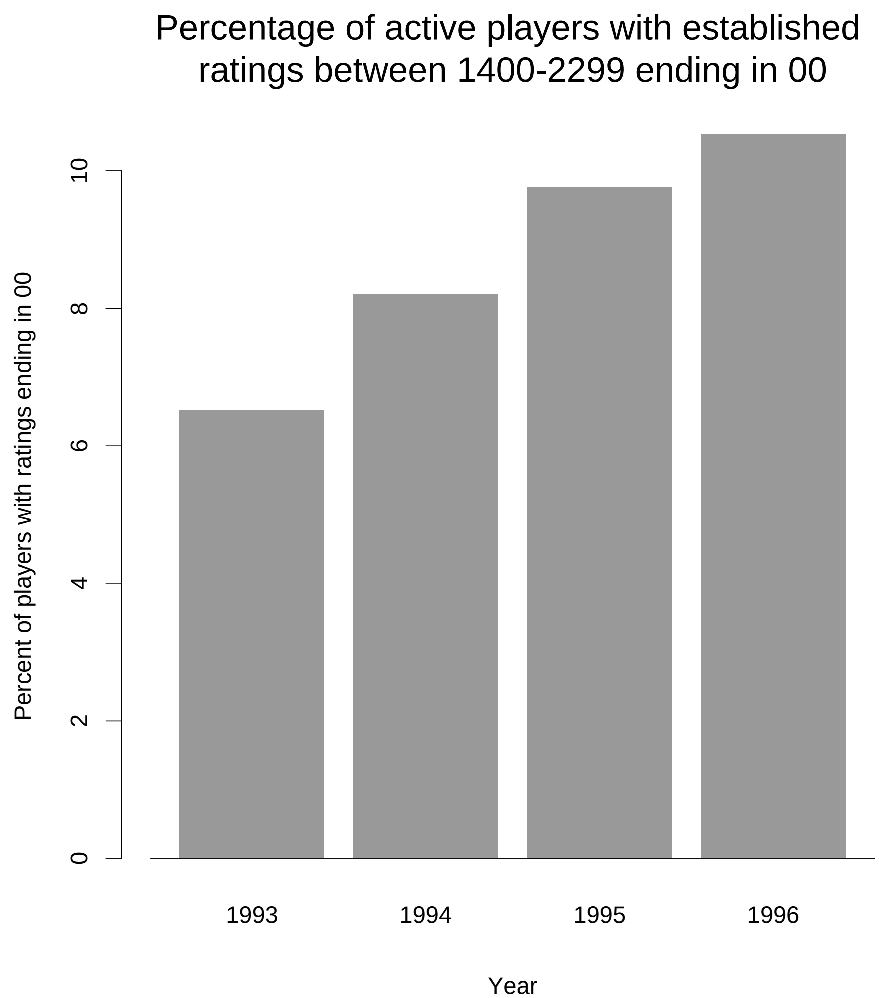

Figure 2.1: The barplot shows the percentage of established ratings between 1400–2299 among active players in the previous year that end with 00. In the December 1995 rating list, approximately 10.5% of such players had 00 as the last two digits of their rating. Because approximately 1% of players would naturally have 00 has the last two digits of their rating, we can conclude that roughly 9.5% of players are at their rating floor.

3 1996 Foreign Rating Conversions

If an unrated player has a USCF-unrated opponent with either a CFC (Canadian) rating or a FIDE rating, the FIDE-rating or the CFC-rating would be converted to a USCF rating given by a conversion table. The converted rating would then be used as if the player had an established USCF rating. This conversion is only used when an unrated player's opponent is unrated.

3.1 FIDE-to-USCF Rating Conversion

The Committee has determined a 1996 conversion of FIDE ratings to USCF ratings, as shown in Table 1. Details of the conversions are described in Appendix B.

Table 1: USCF rating conversion from FIDE rating

FIDE	USCF	FIDE	USCF	FIDE	USCF	FIDE	USCF	FIDE	USCF	FIDE	USCF
2100	NA	2200	2215	2300	2346	2400	2463	2500	2566	2600	2638
2105	NA	2205	2222	2305	2352	2405	2469	2505	2571	2605	2640
2110	NA	2210	2229	2310	2358	2410	2474	2510	2575	2610	2643
2115	2093	2215	2236	2315	2364	2415	2480	2515	2579	2615	2645
2120	2101	2220	2242	2320	2370	2420	2485	2520	2584	2620	2647
2125	2108	2225	2249	2325	2376	2425	2490	2525	2588	2625	2650
2130	2116	2230	2256	2330	2382	2430	2496	2530	2592	2630	2652
2135	2123	2235	2262	2335	2388	2435	2501	2535	2595	2635	2654
2140	2130	2240	2269	2340	2394	2440	2506	2540	2599	2640	2656
2145	2137	2245	2276	2345	2400	2445	2512	2545	2603	2645	2658
2150	2145	2250	2282	2350	2406	2450	2517	2550	2607	2650	2659
2155	2152	2255	2289	2355	2411	2455	2522	2555	2610	2655	2661
2160	2159	2260	2295	2360	2417	2460	2527	2560	2614	2660	2662
2165	2166	2265	2302	2365	2423	2465	2532	2565	2617	2665	NA
2170	2173	2270	2308	2370	2429	2470	2538	2570	2620	2670	NA
2175	2180	2275	2315	2375	2434	2475	2543	2575	2623	2675	NA
2180	2187	2280	2321	2380	2440	2480	2548	2580	2627	2680	NA
2185	2194	2285	2327	2385	2446	2485	2552	2585	2629	2685	NA
2190	2201	2290	2334	2390	2452	2490	2557	2590	2632	2690	NA
2195	2208	2295	2340	2395	2457	2495	2562	2595	2635	2695	NA

It is worth noting that the calculated conversions are not constant for all FIDE values. For example, to obtain the estimated USCF rating for a 2250 FIDE-rated player, 32 points need to be added; for a 2400 FIDE-rated player, 63 points need to be added; and for a 2660 FIDE-rated player, 2 points need to be added.

3.2 CFC-to-USCF Rating Conversion

The Committee has determined a 1996 conversion of CFC ratings to USCF ratings, as shown in Table 2. Details of the conversions are described in Appendix B.

Table 2: USCF rating conversion from CFC rating

CFC	USCF	CFC	USCF	CFC	USCF	CFC	USCF	CFC	USCF
1510	1421	1710	1632	1910	1847	2110	2065	2310	2288
1520	1433	1720	1641	1920	1860	2120	2074	2320	2299
1530	1445	1730	1651	1930	1872	2130	2083	2330	2311
1540	1457	1740	1661	1940	1885	2140	2092	2340	2323
1550	1469	1750	1671	1950	1897	2150	2102	2350	2335
1560	1480	1760	1681	1960	1909	2160	2113	2360	2347
1570	1492	1770	1691	1970	1921	2170	2124	2370	2359
1580	1503	1780	1702	1980	1932	2180	2135	2380	2371
1590	1514	1790	1712	1990	1943	2190	2147	2390	2383
1600	1525	1800	1723	2000	1954	2200	2158	2400	2396
1610	1535	1810	1733	2010	1966	2210	2169	2410	2408
1620	1545	1820	1744	2020	1977	2220	2181	2420	2420
1630	1555	1830	1754	2030	1988	2230	2193	2430	2432
1640	1565	1840	1765	2040	1998	2240	2205	2440	2445
1650	1575	1850	1776	2050	2009	2250	2216	2450	2457
1660	1585	1860	1787	2060	2019	2260	2228	2460	2469
1670	1594	1870	1798	2070	2029	2270	2240	2470	2482
1680	1604	1880	1810	2080	2039	2280	2252	2480	2494
1690	1613	1890	1822	2090	2048	2290	2264	2490	2507
1700	1622	1900	1835	2100	2056	2300	2276	2500	2520

As with the FIDE-to-USCF conversion, the calculated conversions are not constant for all CFC ratings. For example, to estimate a USCF rating for a player with a CFC rating of 1510, 89 points would be subtracted; to estimate a USCF rating for a player with a CFC rating of 2440, 5 points would be added.

3.3 Additional use of converted ratings in rating system

The Ratings Committee recommends the use of the FIDE-to-USCF or CFC-to-USCF conversions for the purpose of updating ratings from an event. We propose that prior to any other computations, USCF unrated players in an event with either CFC or FIDE established ratings have their ratings converted to USCF ratings by Tables 1 or 2. These ratings are then treated as provisional based on 10 games. The FIDE conversion takes precedence if a USCF-unrated player has ratings in both the CFC and FIDE systems.

4 Rating System Modifications

4.1 Introduction

In our 1994 Ratings Committee Report, we described a modification to the rating system that permits unrated and provisional rating calculations that are similar to established rating calculations, but with different values of the constant K. The value of K in the established rating formula determines the amount of weight to place on game outcomes relative to one's pre-event rating. If K is large, game outcomes from an event mostly determine a post-event rating; if K is small, then more trust is being placed on a pre-event rating so that the player's post-event rating will not likely differ greatly from the player's pre-event rating.

To the motions that have been passed by the Policy Board, we propose the addition of one new feature. When a player is low-rated (say, under 1000), the value of K for this player is made larger. The reason is that players with low ratings (e.g., scholastic players) tend to improve over time, so that assigning a large K for such players in essence accelerates their movement upwards. We describe the proposed implementation below.

4.2 Unrated and Provisional Rating Calculations

Below is a summary of the rating procedure described in the 1994:

<u>Procedure for a single event</u>

- Determine ratings for unrated players first
- Next, determine ratings for provisionally rated players
- Finally, compute ratings for established players.

<u>Unrated players</u>

- Compute "event performance rating" (EPR) to obtain first provisional rating. See the Report of the USCF Ratings Committee August 1994 for details.

- If player achieves either all wins or all losses, then save the game outcomes and the opponents' ratings. Compute a nominal rating (for publication purposes) using the EPR algorithm in the 1994 Ratings Committee Report. Repeat the unrated procedure for subsequent events, acting as if the events were a single tournament, until the player has either a win and a loss, or a draw. (Note – recent simulations show that this step may not be necessary. We may implement the system by just computing the EPR on the first event.)
- If an *opponent* is unrated, impute (in order of precedence)
  - Converted rating from non-USCF system (if one exists), or
  - Age-based rating

  as described in the 1994 Ratings Committee Report.

Provisional ratings

Use established rating formula with varying K:

- If a player with pre-event rating R has already played a total of N rated games ($N \geq 4$), and then competes in an m-round event, we propose that the value of K should be

$$K = \begin{cases} 600/(N+m-1) & \text{if } 4 \leq N+m \leq 20 \\ 32 & \text{if } N+m > 20 \text{ and } R < 2100 \\ 24 & \text{if } N+m > 20 \text{ and } 2100 \leq R < 2400 \\ 16 & \text{if } N+m > 20 \text{ and } R \geq 2400 \end{cases}$$

  (An alternate formula, which has been the topic of recent discussion, is to replace $600/(N+m-1)$ with $800/(N+m)$.)

4.3 Modification to the provisional rating formula

To recognize the variability of low-rated players, Ratings Committee member Tom Doan proposed the following modification:

Suppose a provisionally rated player with pre-event rating R, having played a total of N rated games ($N \geq 4$), competes in an m-round event. Define

$$N^* = \min(N,\ \max(8,\ 0.02R)),$$

the "adjusted" N.

The value of K is then determined by

$$K = \begin{cases} 600/(N^*+m-1) & \text{if } 4 \leq N^*+m \leq 20 \\ 32 & \text{if } N^*+m > 20 \text{ and } 1000 \leq R < 2100 \\ 24 & \text{if } N^*+m > 20 \text{ and } 2100 \leq R < 2400 \\ 16 & \text{if } N^*+m > 20 \text{ and } R \geq 2400 \end{cases}$$

For low-rated players who have competed in several tournaments, this formula guarantees that N^* will remain low (but not below 8), so that K will remain large when performing rating updates. This rating modification gives a value of K near 50 for an established rating of 400 or below, and around 38–40 for a 600-rated player.

5 Title System Modifications

5.1 Introduction

In the 1994 Ratings Committee report, we introduced a reconstruction of the USCF Title System which has been approved by the Policy Board. While this system has not yet been implemented, we have proposed three refinements to our original proposal which we describe below. We review the major ideas of the new Title system that were described in the 1994 Ratings Committee Report.

5.2 The Delta schedule

The 1994 Ratings Committee Report established the following principle upon which to revise the old Title system:

- A player possessing the ability of a Y-rated player would have approximately a 50% probability of obtaining the Y-rated title (via five norms) in 10 events.

This can be shown to be equivalent (see Appendix A) to

- A player possessing the ability of a Y-rated player would have approximately a 0.325 probability of obtaining a Y-rated norm in a single event.

Specifically, a player is awarded a norm or multiple norms if his or her attained result in an event exceeds the expected result of a Y-rated player by a certain threshold amount (which depends on the number of games in the event). This threshold amount is denoted Δ. Two norms in a single event are awarded with probability $(0.325)^2$, three norms with probability $(0.325)^3$, and so on. Note that this is a revision of the result that appeared in the 1994 Ratings Committee Report.

Using conservative estimates of the variability of game results given players' ratings, and using the normal distribution as a conservative approximation to the distribution of a player's total score in an event (as an approximation to the Binomial distribution), the calculation of Δ's corresponding to the above rule are straightforward. From a single event in which a competitor plays n games, the following modification of the 1994 norm schedule is proposed:

- To earn 1 norm, $\Delta = 0.227\sqrt{n}$.
- To earn 2 norms, $\Delta = 0.625\sqrt{n}$.
- To earn 3 norms, $\Delta = 0.910\sqrt{n}$.
- To earn 4 norms, $\Delta = 1.142\sqrt{n}$.
- To earn the title, $\Delta = 1.343\sqrt{n}$.

5.3 Supernorms

Norms earned at certain prestigious or heritage events are proposed to be denoted "supernorms." Supernorms can be earned either in

- Grand Prix events with guaranteed total of 20 Grand Prix points, or
- any of the following regular events:

  The U.S. Open
  The World Open
  The National Open
  The New York Open
  The U.S. Team Championships
  The U.S. Class Championships
  The U.S. Masters
  The MidWest Masters
  The U.S. Junior Open
  The U.S. Senior Open
  The U.S. Junior Chess Congress
  The National Junior H.S. Champs
  The National High School Champs
  The U.S. Junior Invitational
  The U.S. Men's and Women's Closed Championships
  The Denker
  The Pan-Am Intercollegiate

The Supernationals

The National Chess Congress

The Ratings Committee is continuing to review other possible criteria for a tournament to be deemed a supernorm event.

5.4 Title names

The following is the current proposed list of titles according to rating level, and the minimum number of supernorms required to earn the title.

Rating Level	Title	Supernorms required for Title
400	Category VIII	0
600	Category VII	0
800	Category VI	0
1000	Category V	0
1200	Category IV	0
1400	Category III	0
1600	Category II	0
1800	Category I	0
1900	Candidate Expert	0
2000	Expert	0
2100	Candidate Master	0
2200	Master	2
2300	Candidate Senior Master	2
2400	Senior Master	2

5.5 Proposed Norm and Title rules

The first three rules have already been approved by the Policy Board. The fourth and fifth rules, governing the duration that a norm is in effect and the type of event in which a norm can be won, are proposed for this year's meeting.

1. Norms can only be earned in events of 4 rounds or more.
2. A norm is earned, or multiple norms are earned, towards a Y-rated title when a player's total score in an event exceeds the expected total score of a Y-rated player by the value of Δ given in Section 5.2.

3. A player's results from an event apply simultaneously to every norm for titles not already earned. Thus, a player may be working on several titles at once.

4. A norm remains in effect for a total of three years, after which it expires.

5. A norm cannot be earned in Quick chess events, $\frac{1}{4}K$ events, or $\frac{1}{2}K$ events.

6 American Go Association's Two-Parameter Rating System

(Paul Matthews)

In 1989, the American Go Association (AGA) adopted a "two parameter" (i.e., rating with variance) rating system for estimating playing strengths. The new model provided a mathematically consistent framework to solve outstanding problems that years of ad hoc tinkering with the previous Elo-type rating system had not only failed to solve but had made worse by introducing anomalies.

In a nutshell, the second parameter (i.e., variance) represents the degree of uncertainty about a player's rating. Uncertainty occurs naturally. For example,

- Novice players may be very weak, but have the potential to grow stronger very quickly. Low ratings are highly uncertain.
- Tournament directors are very uncertain about the actual strengths of previously unrated players.
- Players who are rated often have more certain ratings than players who have not been rated for several years.
- Ambitious players who study can improve rapidly, and are likely to be stronger than their rating. When a player makes that claim, his/her rating is less certain.
- Hobby players who are satisfied to improve slowly, or not at all, have fairly certain ratings. These players are bedrock for a ratings system.

In a two parameter model, less certain ratings adjust more quickly in response to tournament results. In effect, the relatively certain ratings of stable players anchor the system, and the ratings of others slide into a consistent relationship.

But it's not just having the second parameter. In the AGA system, the two-parameter model is also used to express more subjective estimates of playing strength, such as an improved player who is believed to be "about two ranks stronger than his rating", more or less. When certified by a responsible party, such as a tournament director, such estimates are combined with official rating records to obtain a better basis for estimation. By incorporating all available information, suitably qualified by the degree of uncertainty, AGA ratings adjust very quickly to demonstrated changes in playing strength, which delights players, and ratings deflation is virtually eliminated.

Implementing this two-parameter system, allowing for subjective estimates of playing strength, was accomplished by using standard "Bayesian" methodology. The advantage of using the Bayesian approach is that, unlike the classical "frequentist" approach to statistical inference, the amount of certainty one has in a rating can be described using probabilities. For example, using Bayesian methods, one can make a statement like "with 95% probability, a player's true average ability falls between 1600 and 1800," whereas using classical methods, one cannot.

In the beginning, AGA ratings were bootstrapped by seeding all players at the same rating with a large variance, basically the average rating of the entire tournament player population. Several years of tournament data were used in the bootstrapping, but in general, given about 7 rounds in a Swiss tournament, it is possible to recover ratings fairly well without knowing any of the individual player ratings in advance.

The AGA ratings method is widely accepted, and has been adopted in Canada, Europe, and for playing go on the Internet.

The AGA estimation algorithm for the two-parameter system is currently implemented in the C++ programming language, and runs on a Windows PC with execution times on the order of one minute for multiple tournaments combined. The algorithm as it stands would handle USCF data volumes.

Although go and chess are different games, the ratings issues are largely the same. The two parameter Bayesian model used by the AGA is an example of promising technology that could be incorporated into a future USCF ratings system.

A Probability of earning titles

Under the revised title system, a player needs to earn five norms to obtain a title. Suppose in any event there is a p probability of earning at least one norm. For a particular choice of p, the table of "delta's" is chosen to ensure that the probability of earning at least two norms in one event is p^2, and so on up to earning the title in one event is p^5.

The goal here is to find the probability, $P_T(N)$, of earning a title in N events.

Derivation:

Suppose a player competes in N events. Let

$$\begin{aligned} X_0 &= \text{\# of events player earns no norms} \\ X_1 &= \text{\# of events player earns } \textit{exactly} \text{ one norm} \\ X_2 &= \text{\# of events player earns } \textit{exactly} \text{ two norms} \\ X_3 &= \text{\# of events player earns } \textit{exactly} \text{ three norms} \\ X_4 &= \text{\# of events player earns } \textit{exactly} \text{ four norms} \\ X_5 &= \text{\# of events player earns } \textit{exactly} \text{ five norms.} \end{aligned}$$

With events being independent, the joint probability distribution of $(X_0, X_1, X_2, X_3, X_4, X_5)$ is Multinomial with

$$(X_0, X_1, X_2, X_3, X_4, X_5) \sim \mathrm{M}(N, (1-p, p-p^2, p^2-p^3, p^3-p^4, p^4-p^5, p^5)).$$

The probability mass function, $f(x_0, x_1, x_2, x_3, x_4, x_5|N, p)$, that is, $\Pr(X_0 = x_0, X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4, X_5 = x_5|N, p)$, is given by

$$f(x_0, x_1, x_2, x_3, x_4, x_5|N, p) = \binom{N}{x_0, x_1, x_2, x_3, x_4, x_5} (1-p)^{x_0}(p-p^2)^{x_1}(p^2-p^3)^{x_2}(p^3-p^4)^{x_3}(p^4-p^5)^{x_4}(p^5)^{x_5}.$$

The event that a player wins a title in N events is equivalent to

$$0 \cdot X_0 + 1 \cdot X_1 + 2 \cdot X_2 + 3 \cdot X_3 + 4 \cdot X_4 + 5 \cdot X_5 \geq 5,$$

so that

$$P_T(N) = \sum_A f(x_0, x_1, x_2, x_3, x_4, x_5|N, p)$$

where

$$A = \{(x_0, x_1, x_2, x_3, x_4, x_5) : 0 \cdot x_0 + 1 \cdot x_1 + 2 \cdot x_2 + 3 \cdot x_3 + 4 \cdot x_4 + 5 \cdot x_5 \geq 5\}.$$

Computationally, this is more easily implemented as

$$P_T(N) = 1 - \sum_{A^c} f(x_0, x_1, x_2, x_3, x_4, x_5 | N, p)$$

where

$$A^c = \{(x_0, x_1, x_2, x_3, x_4, x_5) : 0 \cdot x_0 + 1 \cdot x_1 + 2 \cdot x_2 + 3 \cdot x_3 + 4 \cdot x_4 + 5 \cdot x_5 < 5\}$$

because the number of 6-tuples in A^c is much fewer than in A.

Examples:

- When $p = 0.5$ and $N = 10$, $P_T(10) = 0.9102173$.
- When $p = 0.1$ and $N = 10$, $P_T(10) = 0.009230212$.
- When $p = 0.32575115$ and $N = 10$, $P_T(10) = 0.5$.

B Foreign Conversions - Technical Details

The Committee has determined a conversion to predict a USCF rating from a FIDE ratings or a CFC rating for purposes of pairing FIDE-rated or CFC-rated USCF-unrated players into USCF-rated events. This is accomplished by identifying players common to both the active USCF and FIDE or CFC pool of players, and fitting a local regression model ("loess") to the data.

Among the 614 players who competed in both USCF and FIDE events in 1995, only players with established USCF ratings, FIDE ratings of at least 2200, and 10 or more FIDE-rated games in 1995, were included. This resulted in a total of 162 players used in the FIDE analysis. For the CFC analysis, 129 players were identified as being active in the first half of 1996 in both the US and Canada based on the annual USCF and CFC lists. We included only players who had established USCF ratings which resulted in a total of 101 players used in the analysis.

The loess fits were performed as robust procedures (not adversely affected by outliers) using a smoothness criterion based on 75% samples of the data at each point. The results of the fits appear on Table 1 in Chapter 3. The loess fits revealed non-linearity in the relationships between FIDE and USCF ratings. A plot of the conversion, and of the FIDE-USCF difference as a function of FIDE rating, is shown in Figures B.1 and B.2, respectively. The analogous plots for the CFC conversion is shown in Figures B.3 and B.4.

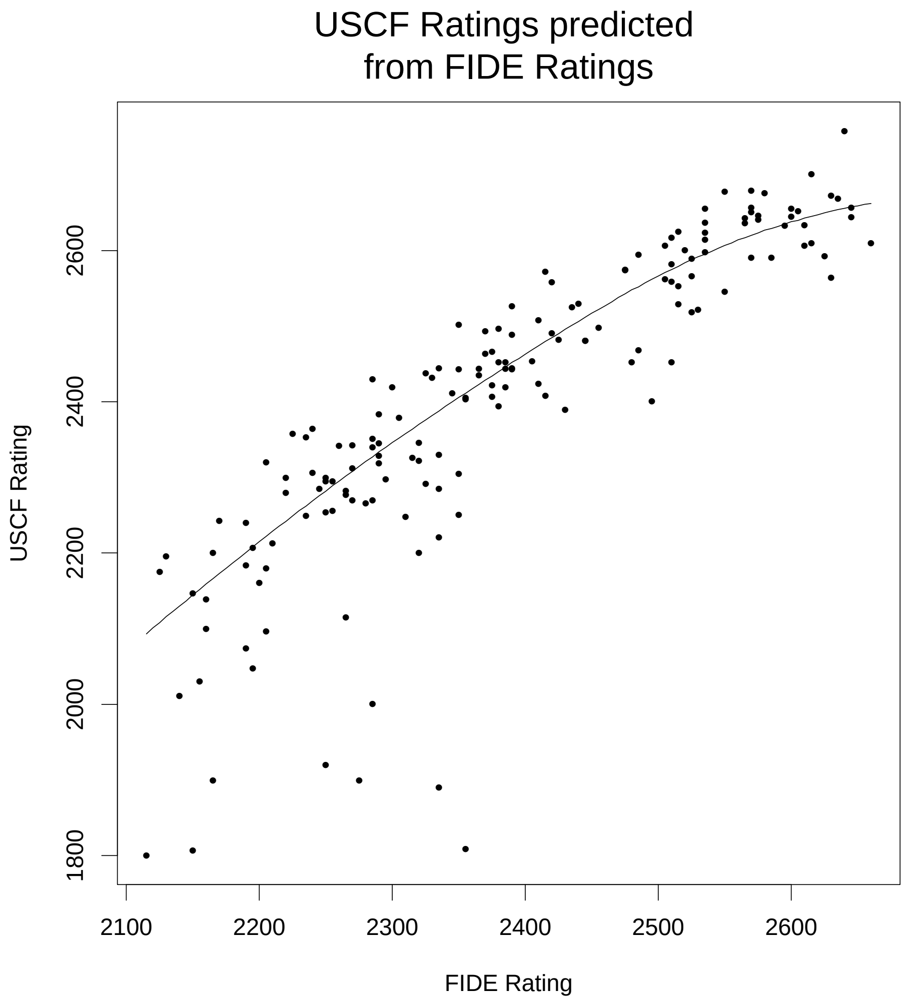

Figure B.1: Curve tracing the FIDE-to-USCF conversion. The observations corresponding to low USCF ratings have little impact on the conversion due to the "robust" properties of the procedure.

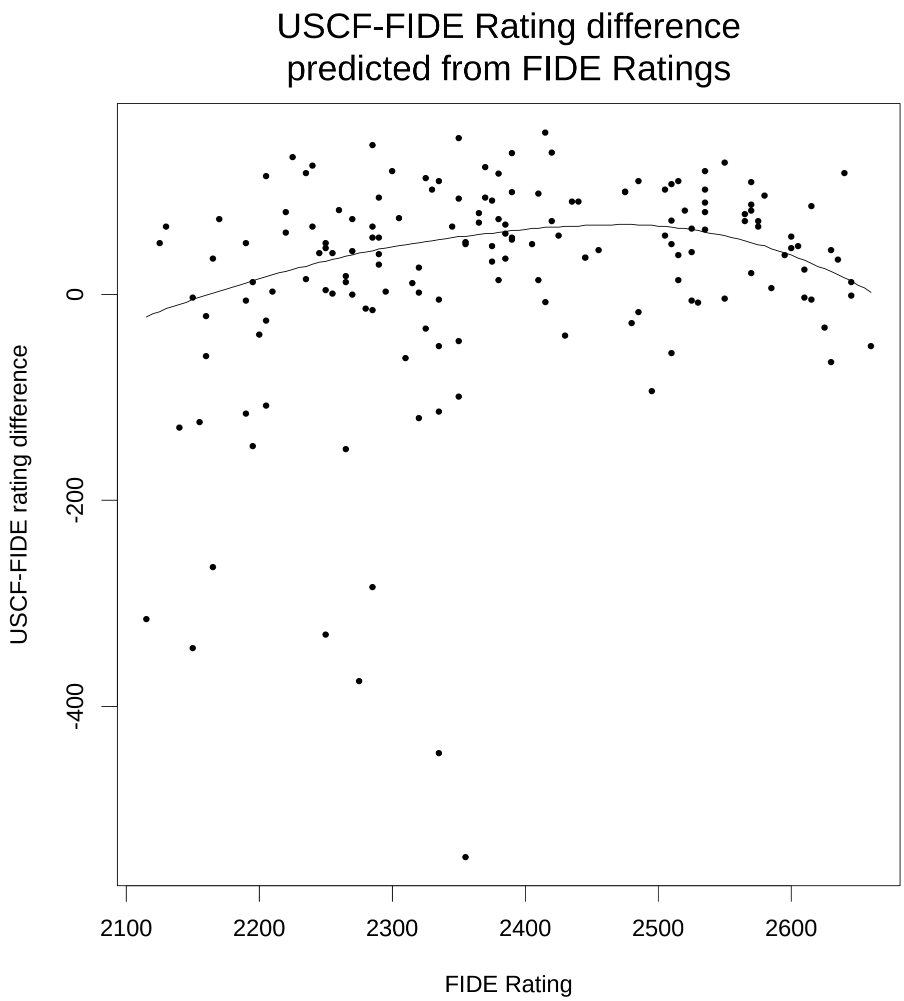

Figure B.2: Curve tracing the amount to add to a FIDE rating to convert it to a USCF rating. The curve demonstrates that low and high FIDE ratings require only a small adjustment to bring to the USCF scale, while medium FIDE require a large adjustment.

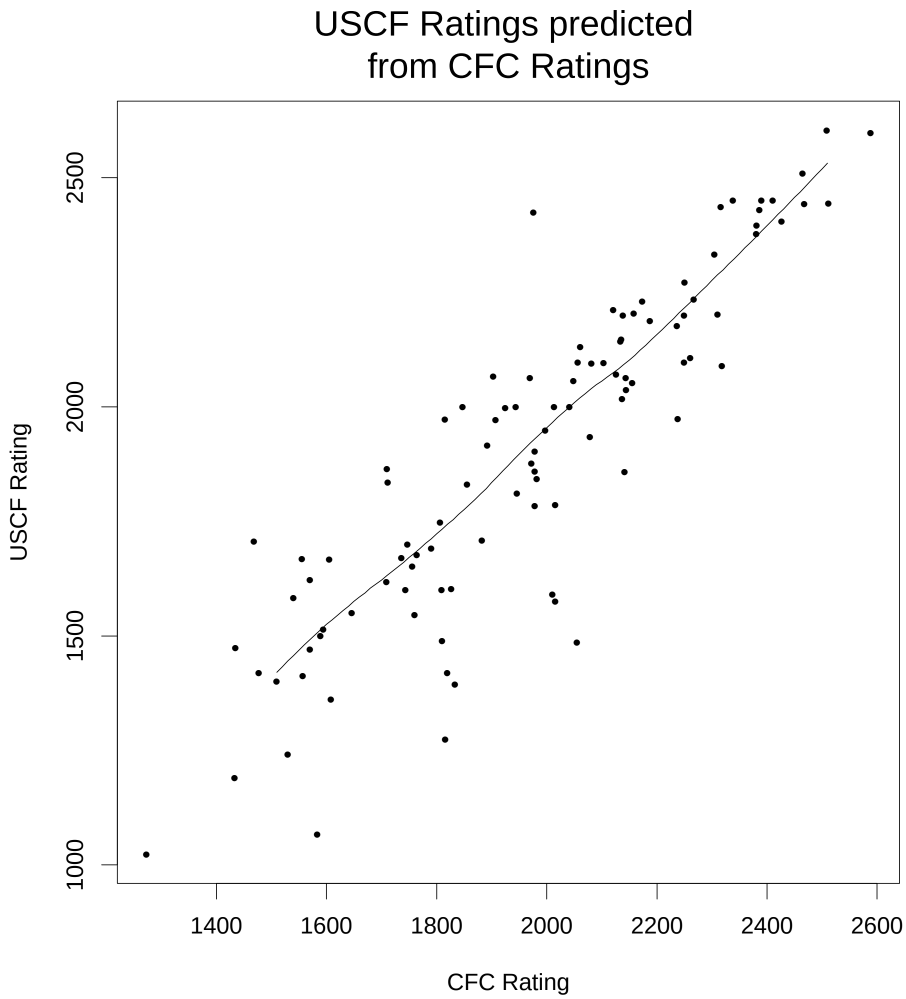

Figure B.3: Curve tracing the CFC-to-USCF conversion.

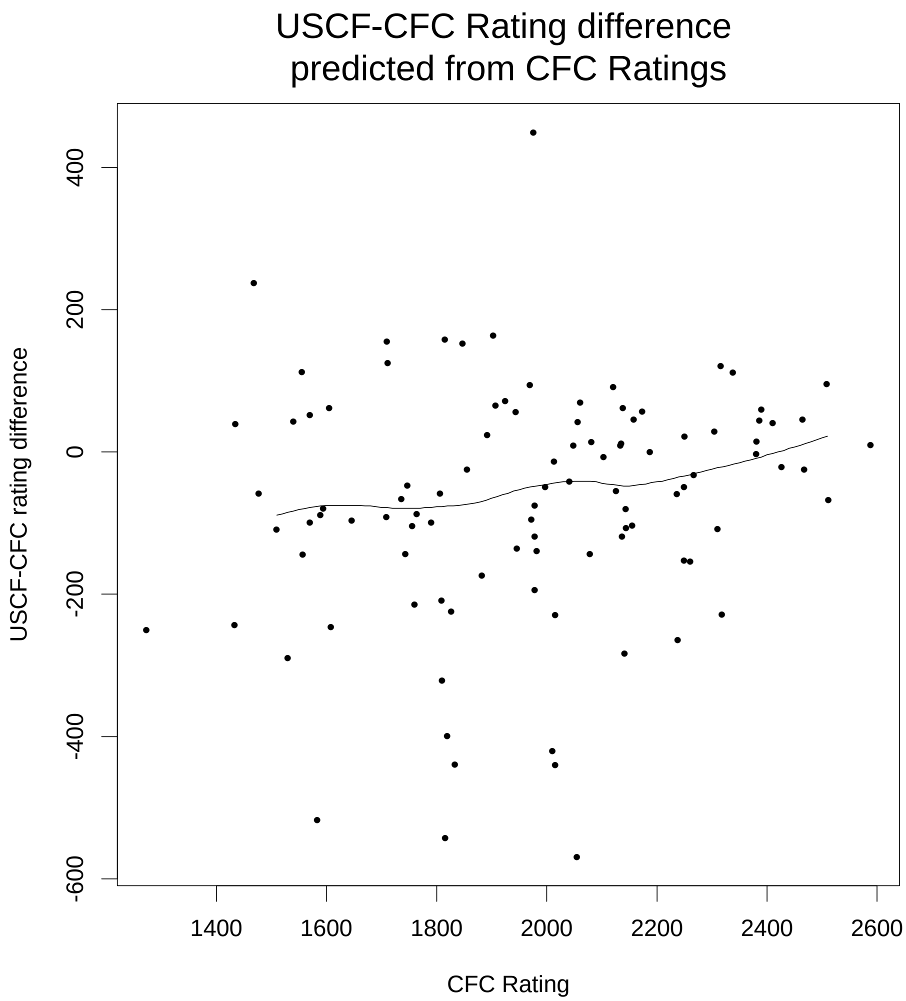

Figure B.4: Curve tracing the amount to add to a FIDE rating to convert it to a USCF rating. The curve demonstrates that for low CFC ratings, a subtraction is necessary to convert to a USCF rating. For higher CFC ratings, a small amount is to be added.