

Modul (1990)

Thomas Mally

Interview mit dem Interviewer

Göran Grotting über

die Schwedische Rating-Liste

Der Name Göran Grotting ist untrennbar mit der Schwedischen Rating-Liste verbunden, deren Einfluß auf die Schachcomputer-Welt in den letzten Jahren ständig gewachsen ist. Die Liste wird regelmäßig in Modul, dem ICCA Journal und anderen internationalen Fachzeitschriften veröffentlicht und wird allgemein wegen ihrer Zuverlässigkeit und Objektivität geschätzt, obwohl in letzter Zeit auch kritische Stimmen zu der auf Partien von Computern gegen Computer beruhenden Methode laut geworden sind.



Göran Grotting ist 42 Jahre alt und arbeitet als Journalist bei der Tageszeitung "Göteborgs-Posten." Er lebt in Lindome, einer kleinen Ortschaft 15 km südlich von Göteborg an der Schwedischen Westküste, mit seiner Frau Gunmari und drei Söhnen im Alter von 9, 10 und 11 Jahren. Außer für Schach und Computer interessiert er sich noch für Langlauf, Ahnenforschung, Briefmarken und Science Fiction (Lieblingsautor: Isaac Asimov).

Er fährt einen Saab-Turbo (nicht von Schätzle+Bsteh!) mit 175 PS und beschwert sich andauernd über die Schwedischen Tempolimits (umweltschonende 90 km/h auf der Autobahn). Modul-Leser kennen ihn nicht nur als Urheber und "Ziehvater" der Schwedischen Rating-Liste, sondern auch als Verfasser von zahlreichen Interviews mit Größen der Schachcomputer-Szene. Höchste Zeit, einmal den Interviewer zu interviewen!

Die Modul-Delegation traf Göran und seine Familie in dem kleinen Salzburger Ort Werfen, wo ein gemeinsamer Besuch von Burg Hohenwerfen auf dem Programm stand. Während die Junioren andächtig jene Stellen besichtigten, an denen Clint Eastwood vor gut 20 Jahren bei den Dreharbeiten zu dem Film "Agenten sterben einsam" über den Schloßhof geballert hatte, gab uns Grotting senior bereitwillig Auskunft über sein liebstes Thema.

Modul: Wie hat es damals mit der Schwedischen Rating-Liste begonnen?

Göran Grotting: Irgendwann im Jahre 1984 hatte ich die Idee, die gesammelten Computerresultate in eine für alle Schachspieler verständliche Rating-Liste umzusetzen. In Ply, der Zeitschrift der Schwedischen Schachcomputer-Vereinigung, hatten wir seit Jahren verschiedene Resultate von Spielen der Computer untereinander veröffentlicht und uns natürlich Gedanken darüber gemacht, wieviele Punkte der Unterschied zwischen Geräten wie z.B. dem Prestige und dem Mark V wirklich ausmachen würde. Professor Elo (er hat österreichische Vorfahren!) hat ja mit seinem Ratingsystem eine Methode entwickelt, wie man eine gewisse Anzahl von Gewinnprozenten in einen Rating-Unterschied umrechnen kann. Seine Tabelle diente mir als Arbeitsgrundlage.

Nachdem mir dieser Einfall gekommen war, verbrachte ich viele Abende damit, die erste Schwedische Elo-Liste mit Hilfe von Papier und Bleistift sowie mit einem Taschenrechner auszuarbeiten. Die Liste, die sich daraus ergab, wurde vom Prestige angeführt und schien mit meinen eigenen Erfahrungen recht gut übereinzustimmen. Ich rief unseren Präsidenten Thoralf Karlsson an und berichtete ihm von meinen Berechnungen. Er war begeistert, und die Liste wurde in der nächsten Nummer von Ply veröffentlicht.

Seitdem ist die Liste regelmäßig in Ply erschienen, in den letzten Jahren auch in mehreren ausländischen Zeitschriften. Während der ersten Jahre war ich ein wenig verärgert darüber, daß das Ausland so wenig Notiz von der großen Anzahl von Computerpartien nahm, die in Schweden gespielt wurde. Heute hingegen ist es mir ein wenig peinlich, wie ernst viele Leute in anderen Ländern unsere Liste nehmen. Ich kann noch hinzufügen, daß die einschlägigen Berechnungen schon bald einem Computer anvertraut wurden. Heute wird die Liste innerhalb von Sekunden mit Hilfe eines von Plymate-Programmierer Lars Hjorth geschriebenen Programms auf einem Atari ST erstellt.

Modul: Die Tester der Schwedischen Schachcomputer-Vereinigung spielen 300-400 Partien im Monat. Was für Leute sind das?

Göran Grotting: Alle Resultate kommen von unseren Mitgliedern, die mit eigenen oder von der Vereinigung zur Verfügung gestellten Geräten arbeiten. Dabei kommen gut 90 Prozent der Ergebnisse von einer kleinen Gruppe von etwa 10 Personen, die mit großer Hingabe Tag für Tag, Woche für Woche und Monat für Monat an den Geräten sitzen und testen. Die eifrigsten Tester sind Junggesellen und Pensionisten, in einigen Fällen auch Behinderte, die in dieser Testarbeit eine neue Lebensaufgabe gefunden haben. So ist z.B. unser Supertester Hans Hultqvist, der schon mehr als 2000 Computerpartien gespielt hat, nach einem Verkehrsunfall an den Rollstuhl gefesselt. Oft spielen unsere Leute zwei Partien gleichzeitig manchmal sogar drei!

Modul: Woher wißt ihr, daß ihr euch auf die Resultate eurer Tester verlassen könnt?

Göran Grotting: Unsere ganze Testarbeit beruht auf Vertrauen. Keiner von unseren Testern hat irgendwelche kommerziellen Interessen. Sie alle sind - wie auch ich - auf der Suche nach der sogenannten "Wahrheit". Sie wollen wissen, wie sich die Spielstärken der verschiedenen Computer wirklich zueinander verhalten. Wir haben über Telefon sehr engen Kontakt mit unseren Testern und erleben immer wieder, wie sie über ihre eigenen oder fremde Resultate begeistert oder traurig sind. So lernen wir unsere Tester kennen und merken schnell, ob ihr Engagement echt ist oder nicht.

Außerdem wäre es ja ganz und gar sinnlos, hunderte Stunden beim Testen zu verbringen, nur um nachher ein gefälschtes Ergebnis abzuliefern! Wenn irgendein Tester die Resultate ständig zugunsten eines bestimmten Computers verfälschen würde, würden wir das durch Vergleich mit den Ergebnissen anderer Tester für das gleiche Gerät rasch herausbekommen.

Modul: Habt ihr niemals den Verdacht gehabt, daß ein Tester nicht ganz aufrichtig war?

Göran Grotting: Doch, einmal. Es handelte sich dabei um eine Person, die das einzige Gerät eines bestimmten Typs in ganz Schweden besaß, sodaß wir keine Vergleichsmöglichkeiten hatten. Die für diesen Computer gemeldeten Ergebnisse waren außerordentlich gut, und sein Rating stieg von Mal zu Mal höher. Dennoch konnten wir nicht mit Sicherheit feststellen, ob der Tester die Wahrheit sagte oder nicht. Wir wissen ja, daß es große zufällige Abweichungen geben kann.

Schließlich ging das so weit, daß wir den Beschluß faßten, diese Person nicht mehr als Tester einzusetzen. Der fragliche Computer scheint heute nicht mehr in unserer Liste auf. Ich bin mir aber nach wie vor nicht klar darüber, ob die Resultate damals korrekt waren oder nicht. Es tut uns heute noch von Herzen leid, daß wir den Kontakt mit einem begeisterten Computerfan auf diese Weise abbrechen mußten. Eines ist sicher: Falls er damals tatsächlich gelogen hat, dann bestimmt nicht wegen irgendwelcher kommerzieller Interessen. Er hatte einfach seinen Computer so sehr ins Herz geschlossen daß er ihn nicht verlieren sehen konnte!

Modul: Erzähle uns etwas über die Regeln, nach denen die Testarbeit vor sich geht!

Göran Grotting: Wenn ein neuer Computer herauskommt, versuchen wir es so einzurichten, das er gegen 50 viele verschiedene Geräte wie möglich jeweils mindestens 20 Partien spielt. Natürlich müssen die Partien abwechselnd mit Weiß und mit Schwarz durchgeführt werden. Die Computer spielen, sofern vom Hersteller nichts anderes empfohlen wird, mit ihrer Standardeinstellung (z.B. bei der Selektivität) und, wenn vorhanden, mit ihrer Turnierbibliothek. Die Partien werden mit einer Zeitkontrolle von 40 Zügen in 2 Stunden gespielt, weil das eben der Modus ist, der auch den normalen Elozahlen für Menschen zugrunde liegt. Die Partien müssen nicht unbedingt bis zum Matt gespielt werden, sollten aber erst dann abgebrochen werden, wenn über den Ausgang kein Zweifel mehr besteht. Bei einem neuen Computer legen wir in erster Linie Wert darauf, ihn gegen Geräte der gleichen Spielstärke-Kategorie antreten zu lassen. Die Eröffnungen werden von den Computern selbst gewählt. Falls eine Partie eine frühere ganz oder teilweise wiederholt, wird sie trotzdem angerechnet.

Modul: Könnt ihr euch auf das Urteil eurer Tester verlassen?

Göran Grotting: Ich bin mir ziemlich sicher, daß man über manche Entscheidungen streiten könnte, wenn man die Partieprotokolle genau studieren würde. Es kann vorkommen, das ein Tester eine Partie fehlerhaft einschätzt oder voreilig abbricht. Aber der entscheidende Punkt ist, daß sich solche unregelmäßigkeiten auf die Dauer ausgleichen, solange der Tester keine bestimmte Absicht verfolgt. Einmal wird Computer A benachteiligt, dann wieder Computer B. Das Rating eines Computers beruht aber normalerweise auf mehreren Hunderten von Partien, manchmal sogar über tausend!

Modul: Nun, die Ehrlichkeit eurer Testarbeit wird ja auch kaum jemals in Zweifel gezogen...

Göran Grotting: Zumindest nicht von Leuten, die selber frei von kommerziellen Interessen sind. Ich war aber z.B. sehr verwundert, als ich unlängst in Modul einen Artikel von Hern Pordzik über den Super Expert B lesen mußte, in dem angedeutet wurde, das wir Novag-Geräte entweder absichtlich oder aus Unfähigkeit benachteiligen. Merkwürdigerweise stimmte dieser Artikel fast Wort für Wort mit einem Brief überein, den mir Novag-Chef Peter Auge kurz zuvor geschickt hatte! War sich so zum Sprachrohr einer Firma macht, darf natürlich nicht erwarten, mit seiner Kritik ernst genommen zu werden. Die Modul-Leser werden sich bestimmt ihr eigenes Urteil darüber gebildet haben.

Modul: Aber wie steht es mit der Frage, ob die Resultate der Computer untereinander die wahren Unterschiede in der spielstärke widerspiegeln, so wie sie von Menschen empfunden werden?

Göran Grotting: Ah, danke für diese Frage! Der Tag, an dem ich zu der Überzeugung käme, daß die Testarbeit Computer gegen Computer keine Relevanz für das Verhalten der Computer im Spiel gegen Menschen hat, wäre auch der Tag, an dem ich mich von unserer Liste, die ich seit ihrem Entstehen betreut habe, lossagen würde. Partien von Computern untereinander haben ja für sich allein keine Existenzberechtigung.

Sie sind lediglich ein Ersatz für Spiele von Computern gegen Menschen, die sich leider nicht in ausreichender Anzahl organisieren lassen. Die Schwedische Schachcomputer-Vereinigung versucht neben ihrer Testarbeit ständig, so viele Partien gegen Menschen zu spielen wie möglich. Leider gibt es von Seiten vieler Schachspieler wachsenden Widerstand gegen die Teilnahme von Computern an Turnieren, was ihr ja auch in Österreich schon habt feststellen müssen.



SSDF Platz 10: Dallas 68020 14 MHz mit Elo 1994 ...
(Bild: Heinz-Gerd Lehmann)

Unsere gesammelten Ergebnisse im Spiel gegen Menschen, die auch im letzten Modul veröffentlicht wurden, dienen uns als vergleichsgrundlage für unsere Liste. Ich meine, daß diese Ergebnisse Mensch/Computer mit unserer Computer/Computer-Liste recht gut übereinstimmen. Natürlich muß man dabei den riesigen unsicherheitsfaktor berücksichtigen, der sich für ein Rating ergibt, das nur auf 9 oder 15 gespielten Partien beruht. Wenn man diese Fehlerbreite in Rechnung stellt, kann man zumindest nicht sagen, daß zwischen den beiden Listen ein eklatanter Widerspruch bestünde.

Natürlich bleibt es jedem einzelnen überlassen, darüber Spekulationen anzustellen. Auch was die verschiedenen Prozessor-Geschwindigkeiten betrifft, finden sich in unserer Liste ziemlich genau die theoretisch zu erwartenden Rating-Unterschiede zwischen gleichen Programmen, die mit verschiedener Geschwindigkeit laufen. Wenn z.B. der Excellence 4 MHz schlechter platziert wäre als der Excellence 3 MHz, müßte man den Schluß ziehen, das mit unserer Testarbeit etwas nicht stimmt. Aber so ist es nicht!

Modul: Worauf beruhen eigentlich die Schätzungen, wieviel zusätzliche Elopunkte eine Geschwindigkeitssteigerung bringt?

Göran Grotting: Im Computerschach hat man lange Zeit angenommen daß eine Verdoppelung der Rechengeschwindigkeit eine Steigerung von etwa 100 Elopunkten mit sich bringt. Diese Zahl stammte aus einer älteren Untersuchung über Partien, die das Programm BELLE mit verschiedenen Einstellungen gegen sich selbst gespielt hatte. Wir halten eine Zahl von 70-80 Punkte für eine bessere Annäherung an die Verhältnisse, wie sie sich aus unserer Liste ergeben. Wenn wir alle Programme einbeziehen, also auch jene, die nicht mehr in der Liste aufscheinen, gibt es insgesamt neun Beispiele dafür, daß das gleiche Programm in zwei Versionen mit verschiedenen Geschwindigkeiten getestet wurde.

Aus diesen Zahlen läßt sich berechnen, daß der durchschnittliche unterschied bei einer Verdoppelung der Geschwindigkeit etwa 88 Punkte ausmacht. Es ist aber ziemlich wahrscheinlich, daß verschiedene Programme verschieden stark von einer Erhöhung der Geschwindigkeit profitieren. Wir vermuten z.B. daß selektive Programme davon größeren Nutzen haben als Brute-Force-Programme, aber konkrete Beweise sind infolge der Fehlerbreiten schwer zu finden.

Wenn man verschiedene Richard-Lang-Programme mit einander vergleichen will, dann muß man bedenken, daß es sich dabei um verschiedene Prozessor-Typen handelt (68000, 68020 und 68030). Das gleiche gilt für die neuesten Fidelity-Programme. Da spielen dann weit mehr Faktoren hinein als nur die in MHz gemessene Taktfrequenz, und die verschieden großen Hash Tables komplizieren die Sache noch weiter. Mißt man die Lösezeiten für gewisse Stellungen, dann kann man zu ganz unterschiedlichen Ergebnissen kommen. Der Mach IV ist z.B. in manchen Stellungen dreimal schneller als der Mach III, in anderen aber nur 1,8 mal! (Der durchschnittliche Wert ist 2,4 mal.)

Modul: Und wie steht es mit dem Irazoqui-Effekt, dem Novag-Effekt und dem Einfluß der Eröffnungsbibliothek?

Göran Grotting: Beginnen wir mit Herrn Irazoqui, einem ehemaligen Redakteur von Computer Chess Reports. Er war der Meinung, daß beim ständigen Spielen von Computern gegeneinander die objektiv vorhandenen Rating-Unterschiede um einen Faktor von 1:2 vergrößert werden. Wir haben dafür in unseren Ergebnissen keine Anhaltspunkte gefunden. Larry Kaufman, der jetzige Redakteur von CCR, beziffert den gleichen Effekt mit 1:1,25. Ich wage nicht zu behaupten, daß er da völlig unrecht hat. Ich könnte mir vorstellen, daß es einen solchen Effekt tatsächlich gibt, daß er sich aber weniger stark auswirkt, als Kaufman annimmt.

Der sogenannte "Novag-Effekt" soll darauf beruhen daß Novag-Computer gegen Menschen erfolgreicher spielen als gegen andere Computer. Ich vermute, daß diese Theorie hauptsächlich aus der Publicity-Abteilung von Novag stammt, denn in unseren Zahlen lassen sich keine überzeugenden Belege dafür finden. Ich kann durchaus der Auffassung zustimmen, daß der Super Constellation von 1984 gegen Menschen besser abgeschnitten hat als gegen spätere Computermodelle, die offenbar gerade gegen den Super Conny getestet worden waren. Aber ich kann in unseren Zahlen keine Anzeichen dafür finden, daß spätere Novag-Generationen wie Forte oder Super Forte gegen Menschen besser abgeschnitten hätten als gegen Computer.

In letzter Zeit haben wir viel über die Auswirkung von sogenanntem "booking" auf unsere Ergebnisse nachgedacht. Mein grundsätzlicher Eindruck ist, daß die Eröffnungsbibliothek eines Computers für die Spielstärke eine eher geringe Rolle spielt, so sonderbar das auch für Theoriefanatiker klingen mag. Sicher gibt es Beispiele dafür, daß ein Computer von seinen Eröffnungen her auf das Spiel gegen ein anderes Gerät getrimmt wurde. Aber wir haben mit Hilfe der Statistik festgestellt, daß eine Eröffnung, die gegen Computer A gut ist, sich gegen B und C schlecht auswirken kann und umgekehrt.

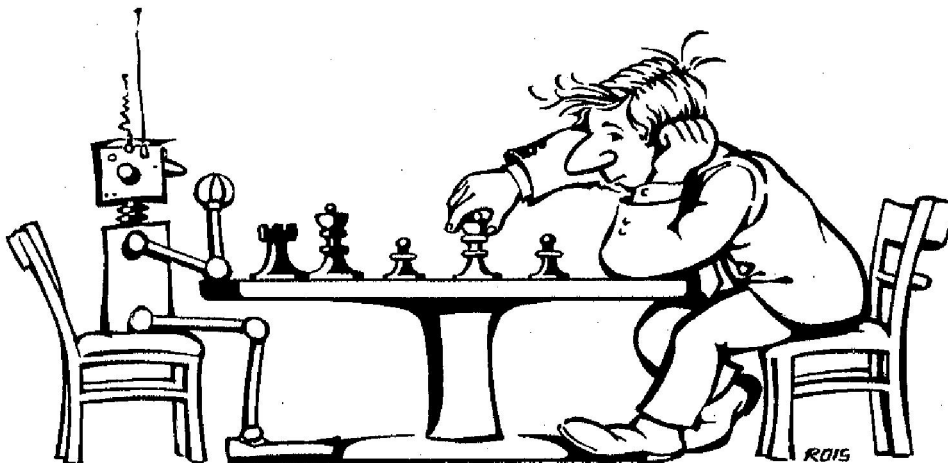
Der Effekt geht in der Menge der gespielten Computerpartien unter. Die Zeit reicht einfach nicht dazu aus, das "Buch" eines Computers auf 40 verschiedene Konkurrenzgeräte abzustimmen. Gott sei Dank!

Modul: Wie aus verschiedenen Modul-Beiträgen hervorgeht, interessieren sich viele Leser für sogenannte Stellungstests. Mehrmals wurde sogar der Wunsch geäußert, die Schwedische Liste ganz durch eine Batterie von Testpositionen zu ersetzen. Wie steht ihr dazu?

Göran Grotting: Ich persönlich habe viel Spaß an Testpositionen. Ich kann dabei selber eine Menge über Schach lernen, und es ist interessant festzustellen, ob mein Computer mit dieser oder jener Stellung fertig wird oder nicht. Aber leider habe ich bis jetzt keinen Test gefunden, der mir verlässlichere Auskunft über die Spielstärke eines Geräts hätte gehen können als unsere Rating-Liste.

Ich glaube nicht, daß man einen Test konstruieren kann, der mehr Aussagekraft besitzt als z.B. 100 Computerpartien (die wir oft innerhalb von wenigen Monaten spielen). Wenn man bedenkt, daß sich eigentlich jeder Zug in einer richtigen Partie mit einer Testposition vergleichen läßt, dann bedeutet das, daß 100 Partien etwa dasselbe leisten wie ca. 6000 Teststellungen!

Ich glaube aber durchaus, daß man mit Hilfe von Testpositionen ein Leistungsprofil eines Schachprogramms erstellen kann. Man kann z.B. feststellen, ob ein bestimmtes Programm taktisch stark ist oder das Endspiel besonders gut behandelt etc. Aber wie soll man diese einzelnen Faktoren gewichten, um die Gesamtspielstärke zu bekommen?



Modul: Manche Verfechter von Positionstests behaupten, daß die Schwedische Liste ihre Rolle bald ausgespielt haben wird, weil die Computerproduzenten ihre Geräte zunehmend auf das Spiel gegen andere Computer ausrichten. Wie steht's damit?

Göran Grotting: Denken diese Leute nicht daran, wie einfach es umgekehrt für einen Computerprogrammierer wäre, sein Programm an einen von vornherein bekannten Test anzupassen, sodaß es ein nahezu perfektes Resultat erreichen würde? Oder glauben sie vielleicht, daß sie ihre Tests vor Richard Lang und Co. geheimhalten können? Die Programmierer lesen ja auch die einschlägigen Fachzeitschriften und könnten z.B. die betreffenden Positionen ganz einfach in ihrer Eröffnungsbibliothek unterbringen.

Bei allen Tests, die darauf abzielen, die Spielstärke eines Computers zu messen, ist höchste Genauigkeit die Voraussetzung, sonst wären sie sinnlos. Ich kann ja z.B. über die Spielstärke eines neuen Saitek-Programms ganz brauchbare Vermutungen anstellen - wenn ich z.B. annehme, daß ein neues Saitek-Programm etwa 25-50 Punkte stärker sein wird als das vorige, so werde ich im allgemeinen nicht sehr weit daneben liegen. Dazu brauche ich keine Tests!

Schwedische Eloliste 4/1990

	ELO	+/-	Sp.	%	Geg.
Portorose 68030 36 MHz	2248	41	183	85	1955
Elite Nr.9 68030 32 MHz	2120	36	242	74	1942
Portorose 68020 12 MHz	2116	26	458	76	1916
Almeria 68020 12 MHz	2069	37	227	75	1879
Mach IV 68020 20 MHz	2066	26	442	70	1917
Polgar 10 MHz	2059	77	52	65	1947
Portorose 68000 12 MHz	2021	28	380	72	1860
Roma 68020 14 MHz	2002	22	646	68	1873
Elite Nr.5 (68000x2)	1998	56	99	58	1944
Dallas 68020 14 MHz	1994	25	489	72	1830
Almeria 68000 12 MHz	1993	25	476	71	1840
Mach III 68000 16 MHz	1981	16	1180	64	1880
Dallas 68000 12 MHz	1953	18	900	68	1818
Polgar 5 MHz	1949	25	483	58	1893
Roma 68000 12 MHz	1945	17	1080	62	1858
Academy 5 MHz	1912	21	702	54	1883
Amsterdam 68000 12 MHz	1898	18	940	62	1815
Super Expert C 6 MHz (sel 5)	1898	84	44	51	1890
Mach IIc 68000 12 MHz	1896	21	723	57	1848
Mega IV 4.9 MHz	1894	19	862	56	1852
Maestro D 10 MHz	1892	27	431	51	1886
Super Expert B 6 MHz (sel 4)	1878	29	369	58	1822
MM IV 5 MHz	1877	18	932	56	1836
Sphinx Galaxy 4 MHz	1859	21	667	48	1875
Psion Atari 8 MHz	1848	17	1031	50	1847
Excel Club 68000 12 MHz	1826	19	889	52	1811
Mega IV Brute Force	1822	43	168	54	1791
Avant Garde 5 MHz	1804	16	1156	46	1831
Super Expert/Forte 5 MHz	1799	20	797	45	1838
Par Excellence 5 MHz	1794	19	827	51	1786
Rebell 5 Mhz	1786	19	880	42	1840
Forte B 5 MHz	1785	19	835	44	1829
Forte A 5 MHz	1776	18	939	49	1784
Stratos/Analyst 6 MHz	1775	19	871	40	1848
Super Mondial 4 MHz	1775	19	825	41	1842
Conchess Plymate 5.5 MHz	1775	15	1441	46	1806
Leonardo Maestro A 6 MHz	1775	25	511	46	1803
Excellence 4 MHz	1763	15	1307	43	1813
Simultano 5 MHz	1760	24	527	36	1861
Expert 4 MHz	1756	21	683	46	1785
MM II 3.7 MHz	1736	38	208	52	1719
Turbostar 432 4 MHz	1730	18	998	45	1765
Excellence 3 MHz	1722	19	875	42	1778
Chessplayer 2150 Atari/Amiga	1708	76	53	42	1761
Super Constellation 4 MHz	1694	15	1352	37	1786
Final Chesscard	1686	78	51	40	1756
Super VIP	1661	52	112	36	1762
Superstar 36K 2 MHz	1646	19	854	33	1771
Europa	1645	51	117	52	1633
Elite A/S 3.2 MHz	1638	20	778	32	1771
Quattro 4 MHz	1627	23	560	33	1754
Conchess Glasgow 2 MHz	1621	21	719	34	1735
Chessmaster 2000 Atari	1620	37	225	29	1779
Constellation 3.6 MHz	1605	26	472	41	1669
Constellation Primo	1598	45	149	44	1644
Super Enterprise	1512	30	331	27	1687
Turbo S 24K	1432	64	75	21	1657

Modul: Apropos Genauigkeit! Wie war das mit dem ersten Rating für den Mephisto Polgar? Kann man da nicht von einem echten Flop sprechen?

Göran Grotting: Für einen außenstehenden Betrachter mag es wirklich so aussehen. Es Stimmt: der Polgar ist ein Musterbeispiel dafür, daß ein Computer auch außerhalb der 95%igen Fehlerbreite liegen kann, und das Rating des Polgar ist inzwischen ja auch drastisch gesunken. Das ist aber auch schon bei anderen Geräten vorgekommen wie zuletzt beim Portorose 68030. Was den Polgar betrifft, so habe ich in meinen Kommentaren zur Rating-Liste ausdrücklich darauf hingewiesen, daß seine ersten Ergebnisse einfach zu schön waren, um wahr zu sein, und daß sein Einstiegs-Rating viel zu hoch schien. Es kommt ganz darauf an, wie man die in der Rating-Liste aufgeführten Zahlen interpretiert.

Die meisten Leute nehmen die Rating-Liste viel zu wörtlich. Manchmal werden aus einem Unterschied von nur 10-20 Punkten zwischen zwei Computern die tollsten Schlußfolgerungen gezogen. Viele glauben auch, daß die angegebene Fehlerbreite eine absolute Grenze ist, was aber überhaupt nicht stimmt! Wir können mit einiger Sicherheit sagen, daß 19 von 20 Computern innerhalb dieser Fehlergrenzen liegen. Das bedeutet aber auch, daß auf unserer Liste im Augenblick wahrscheinlich zwei bis drei Computer aufscheinen, die außerhalb der Fehlerbreite liegen, wobei aber weder ich noch irgendjemand anderer sagen kann, um welche Geräte es sich handelt. Und die, die außerhalb der Fehlergrenzen liegen, können im Prinzip auch beliebig weit außerhalb liegen!

Wenn man nach etwa 500 gespielten Partien die Einzelresultate eines Computers genauer betrachtet, findet man immer wieder die eine oder andere Begegnung aus 20 Partien, die in negativer oder positiver Hinsicht ein überraschendes Ergebnis gebracht hat. Diese stark abweichenden Einzelresultate gleichen sich im Laufe der Zeit gegenseitig aus bzw. gehen in der großen Anzahl von gespielten Partien unter. Ich habe schon längst aufgehört, mich über solche Abweichungen zu wundern. Ich weiß einfach, daß so etwas vorkommen kann, und es besteht kein Anlaß, einen Tester deswegen der Unehrllichkeit zu verdächtigen.

Wenn wir bei unserem Testverfahren Pech haben, treten solche abweichenden Resultate zu Beginn der Testphase eines neuen Computers gehäuft auf. Das schlägt sich dann natürlich stark auf die Berechnung des ersten Ratings nieder. So etwas ist offenbar auch beim Mephisto Polgar passiert. Wenn man die Rating-Listen im Laufe der Zeit vergleicht, sieht man aber, daß es durchaus der Normalfall ist, daß ein Computer schon bei seinem ersten Auftreten ziemlich richtig liegt und sich sein Rating in der Folge nur wenig verändert.

Es ist faszinierend zu beobachten, wie neue Computer sich sozusagen immer mehr auf ihre richtige Position "einschießen" und ihr Rating sich von einer Liste zur anderen immer weniger ändert. In diesem Frühjahr haben wir z.B. beobachten können, wie die drei Portorose-Versionen sich immer besser in das erwartete Verhältnis zueinander eingeordnet haben. Der Portorose 68030 ist etwa 3,6 mal schneller als der 68020, der wiederum doppelt so schnell ist wie ein 68000. Theoretisch müßte der Unterschied zwischen den ersten beiden etwa 144 Punkte und zwischen den letzten beiden 80 Punkte betragen.

Wie ein Blick auf die neueste Liste verrät, liegt die Praxis gar nicht so weit von der Theorie entfernt! Nach dieser kleinen Lektion hoffe ich, daß die Modul-Leser die Bedeutung der Fehlerbreite besser verstehen und einen Rating-Unterschied von 20-30 Punkten nicht allzu ernst nehmen! Das gilt natürlich ganz besonders, solange wir nicht mehr als etwa 100 Partien gespielt haben. Bei dieser Gelegenheit möchte ich Modul für die Art danken, in der die Schwedische Rating-Liste präsentiert wird.

Ich halte es für sehr wichtig, daß man nicht nur die Rating-zahlen abdruckt, sondern auch die Fehlerbreite und die Anzahl der gespielten Partien. Leider tun CSS und andere Fachzeitschriften das nicht. Ich glaube auch, daß es für Modul-Leser interessant wäre, wenn die Redaktion ab und zu Platz dafür fände, alle Einzelresultate abzudrucken, auf denen die Liste beruht.

Modul: Um den Platz ist es bei uns leider schlecht bestellt! Wir weisen für Interessenten darauf hin, daß diese Resultate nicht nur ständig im Swedish Rating Service, sondern zu Beginn eines Jahres auch in Ply und im ICCA Journal abgedruckt werden. Eine andere Frage: Warum ist das Niveau der Schwedischen Rating-Liste im Lauf der Jahre mehrmals gesenkt worden?

Göran Grotting: Dafür gibt es mehrere mögliche Ursachen. Ein Grund kann einfach darin liegen, daß wir 1984 mit einem zu hohen Niveau begonnen haben. Ein anderer möglicher Grund ist das, was ich den "Zeit-Effekt" nenne. Ich glaube ganz einfach, daß sich inzwischen mehr Schachspieler an das Spiel gegen Computer gewöhnt haben, und daß es heute für einen Computer nicht mehr so leicht ist, gegen Menschen ein hohes Rating zu erreichen, wie das z.B. 1984 der Fall war. Das Niveau, das unsere Liste derzeit hat, halte ich für durchaus realistisch, zumindest für Schwedische Verhältnisse und - wohlgemerkt! - für Partien auf der Turnierstufe. Im Blitz- und Schnellschach haben die Computer eine viel höhere relative Spielstärke, während es im Fernschach gerade umgekehrt ist. Das Niveau der Schwedische Rating-Liste wird auch in Zukunft weiter angepaßt werden. Das machen wir einmal im Jahr, genauer gesagt jeden Herbst. Wie es derzeit aussieht, wird die Liste im nächsten Herbst - zum ersten Mal seit ihrem Bestehen! - ein wenig angehoben werden, und zwar um etwa 10 Punkte. Wir werden sehen!

Modul: Noch etwas?

Göran Grotting: Ach ja, wahrscheinlich könnten wir noch einmal so lange reden, wenn wir auf Details eingehen würden. Aber ich glaube, daß wir an dieser Stelle aufhören sollten. Ich nehme an, daß die Modul-Leser zu diesem Interview Fragen oder Kommentare haben werden. Ich werde darauf gerne in einer der nächsten Nummern von Modul antworten.

Modul: Dann herzlichen Dank im Namen unserer Leser und in unserem eigenen!

Göran Grotting: Ich habe für das gemütliche Plauderstündchen zu danken!

Göran Grotting und Thomas Mally.

P.S.: Nur einen Tag nachdem Göran mit seinen sehr lebhaften Sprößlingen die Eishöhlen bei Werfen besucht hatte, wurde die Gegend von einem heftigen tektonischen Beben erschüttert. Wie nicht anders zu erwarten, bestreitet Göran vehement jeden kausalen Zusammenhang und beruft sich wieder einmal auf den Zufall.

Erwerbsquelle: 06-1990, Das österreichische Schachcomputer-Magazin Modul 2/1990 S. 12-18, Interview mit dem Interviewer. Göran Grotting über die Schwedische Rating-Liste. (Leicht bearbeitet durch Hein Veldhuis.)

Internet

<http://ssdf.bosjo.net/list.htm> (The SSDF Ratinglist)

<http://home.interact.se/~w100107/historik.htm> (Ply/SSDF – the story)

Last Updated on February 2, 2010