Introduction

The purpose of this study is to find out how the strength of play has risen in the course of time. FIDE rating of 2008 serves as a basis of comparison. There exist two types of analyses: those that comprise players, and those that comprise certain time periods. The advantage of time period-based analyses lies in the fact that their results are not affected by style of play so much as it could sever the credibility of results. Tactical play by humans often contains moves that objectively are not optimal, but give good chances in practical play. If games are chosen randomly, the probability of either wild, or calmer and positional games happening to be within a selection is equal. Thus decades from 1860s till 2000s are taken as an object in the study. To ensure the coherence of data, players from the chessmetrics.com rating interval 2600-2700 were taken into consideration. The average rating stays therefore always near 2650. For the sake of interest, Carlsen's games from Nanking 2009 are also included where his performance rating was even 3002. It has been often described as the most impressive display of chess skill ever by human. Later on Fischer vs Larsen and Taimanov, and Karpov in Linares 1994 will be analyzed as well.

Some theoretical principles.

Every computerized analysis whose aim is to ascertain the playing strength of a player in comparison with today's FIDE-rated chessplayers, has to consider following theoretical principles:

1. All moves must be taken under scrutiny as long as one of the players does not have a won position, except opening moves.

It's crucially important that all moves after leaving theory are analyzed, provided the position is more or less equal. If one picks an arbitrary selection of moves from a game, the whole picture will not be seen like in the case of all other measurings in the world. As long as the position remains more or less equal, a player supposedly mostly retains the objective and neutral opinion. Things are different when the position inclines decisively towards one or other side. For example, if a player is in a clearly lost position, there would be no reason to expect him to make every effort due to psychological reasons, or he would resort to swindling, or just go into complete passivity. In a won position pursuit of accuracy, on the other hand, is not the most essential, if there exist other, more natural for humans, ways leading to victory. It must be admitted, though, that in the case of data size being large enough it is not so important, as it would be cancelled out statistically.

2. It must be considered that time control may be different by games, whereas in earlier times they were longer than nowadays.

Time controls in tournaments and matches have become shorter in the course of time, and due to the spread of computers, adjournments have disappeared. One of reasons must be making chess more attractive to the audience. Shorter time controls also decrease the number of number of draws via the lessened quality of play.

3. Longer time control leads to more accurate play.

By doubling thinking time, the playing strength of humans increases further than computers (cs 50-70 ELO). Also, the increase in accuracy is sharper at faster time controls.

4. Positions in a game of chess are of different degrees of difficulty that affect the accuracy of play.

Generally, positions occurring in a game can be classified into two categories: tactical and positional/strategical. The more tactical elements there are in a position, the more difficult it is for a

human to maintain the accuracy of play. I prefer the terms 'higher accuracy of play' or 'lower accuracy of play' over 'easier/harder to find best moves' because in more than a half of positions the notion of 'best move' matters extremely little. It is possible to determine the difficulty of position with the help of various parameters whose sum total makes up the general factor of difficulty.

5. The level of play or the rating of a player is indicated by what his expected average error is at a certain degree of difficulty.

The rating of a player is a reflection of his playing skills in a numerical way. Depending on the style of play, the expected average error of a player of the same level of play can be slightly different with respect to the extremes of the degree of difficulty, but the same altogether. The corresponding indicator of positionally playing chessplayer in simpler and more strategical positions is higher than that of a equistrong tactical player. The accuracy of play of a tactical player, nonetheless, is relativelt better in more difficult and tactical positions.

6. The expected error is a function of the actual error and the difficulty factor:

$$f\,(expected\ error) = \frac{actual\ error}{difficulty}$$

The general difficulty factor is equal to the joint effect of all separate difficulty factors. How much a factor influences the accuracy of play can be determined from measuring how much relatively the actual error changes within an equal sized amount of positions.

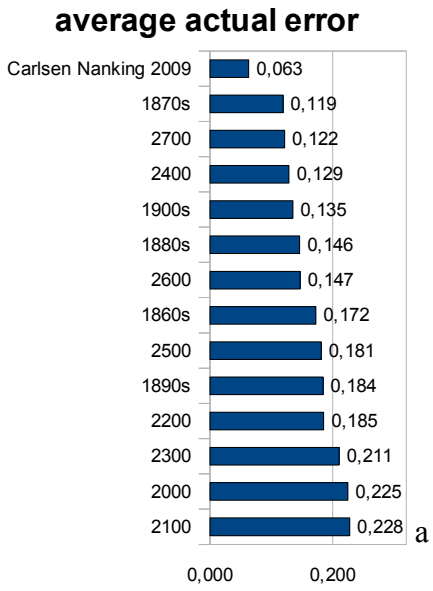
7. The outcome of a game of chess is determined by the actual accuracy of play.

Since the difficulty of positions is also influenced by moves players make, generally more threatening and aggressive moves tend to highen the difficulty for the opponent, it is theoretically possible that both equally good players with a similar style of play are playing at equal strength, but one of them loses due to his positions being more difficult. It demonstrates how important it is to put the opponent under pressure, to strive for the initiative. It makes move-choosing more complicated.

The methods of analysis used in this study.

4 phases can be distinguished:

1. Find out the actual average error. It is calculated by taking the eval of best move by Rybka and substracting the evaluation of the move made by a player from it.
2. Next find out various difficulty parameters, including thinking time.
3. Derive the average expected error on the basis of difficulty factors and actual average error.
4. Compare the results with that of the modern FIDE-rated players to get an overview of the playing level of players being analyzed.



In determining the average actual error Rybka 3 Default with default settings has been used. The GUI was Arena 2.0. Each position was analyzed for 5 minutes with 5 simultaneous PV-s on dual 2GHz Athlon. Since the amount of data was relatively a

scarce, stabilizing points have been used, otherwise the conclusions could have been misleading.

- The largest mistake possible is 2.00. In the case of large mistakes, it doesn't matter how large it exactly was, but the fact that a mistake was committed. An example can be instanced where player A in a game makes 5 mistakes 2.01; 2.68; 1.94, 2.08 and 2.44, but player B only one mistake withe the value of 13.34. If I didn't use the boundary value, we would arrive at the conclusion that both players had played at a similar level of accuracy, which would be obviously misleading. 2.00 is chosen arbitrarily and is not actually more valid than any other similar value.
- Only positions during which the situation of the game stays more or less equal are considered. Thus, positions where both the evaluation of move proposed by computer and the evaluation of move actually played are outside of the evaluation interval [2; -2] and are with the same sign, have been removed.
- Games whose blunder percentage exceeds 15% were removed, except when the number of valid moves is under 30.

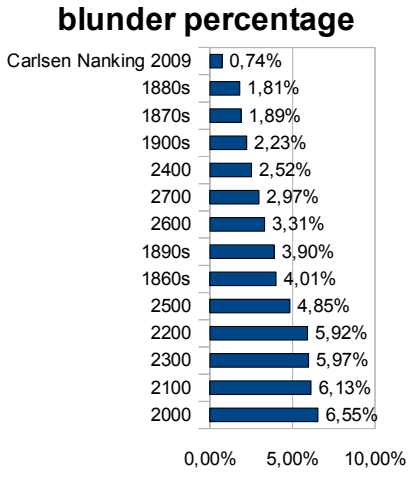
There is no point in including opening moves if purely chess skills are being analyzed, not how much theory one has learnt by heart. Because opening theory is advancing every year, the start point where the analysis begins depends on a time-period as shown in the following table:

1860-1879	1880-1899	1900-1919	1920-1939	1940-1959	1960-1979	1980-1999	2000s
8	9	10	11	12	13	14	15

The minimal length of games is 20 moves + the start point of the analysis. So, it ranges between 28-35 according to period.

Since Rybka 3 is quite untrustworthy in properly evaluating endgames, positions with less than 10 pieces or pawns have been left out of consideration. An exception has been made for mistakes that really turn the situation around: a won position to a drawn or a lost one and vice versa.

What actual average error indicates in the ideal situation is how much the accuracy of play deviates from the absolute point of view of perfect play. As mentioned previously, exactly this criterion plays a determining role in the outcome of a game; the fate of a game depends on to what extent one can bring his divergence from the absolute to a minimum. Unfortunately, a serious matter acts as an obstacle: how easily one is able to lower his error rate depends not only on his playing skills, but also on factors that make maintaining one's accuracy of play harder. This is why it is necessary to take those factors into account in the analysis. The more of them, the more trustworthy the calculation of the expected error will be.

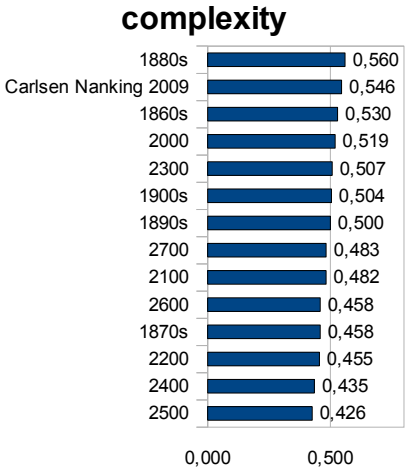


For the sake of interest, the frequency of blunders is displayed here. A blunder is a mistake valued 1.00 or more, but not over 2.00.

There are two types of difficulty factors: external and internal ones. External factors are thinking time, psychological pressure as a result of some external influences, feeling, conditions at playing venue etc. Only thinking time can be measured and considered. Internal factors develop in the course of game, and are dependent on the placement of pawns and pieces that in turn depends on the style of play and openings. Unlike external ones, internal factors are also valuable indetermining the playing style of chessplayers. Internal factors can be classified into two groups: direct (complexity, difference) and indirect ones (material, evaluation).

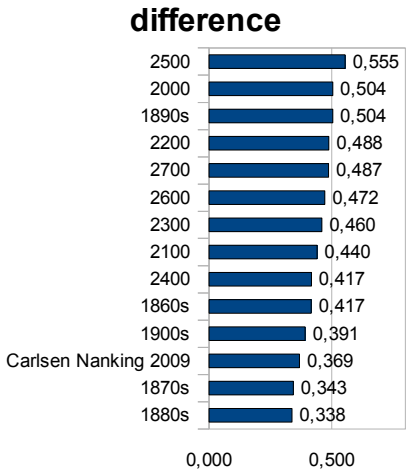
The difference between direct and indirect factors lies in the fact that direct factors determine the the tacticality or the positionality of a position directly, indirect ones do not. Higher complexity and difference occur chiefly in tactical positions, a large amount of material on the board and a high evaluation do not reveal anything about the type of positions.

A short description of the five difficulty parameters usen here in the study is as follows:



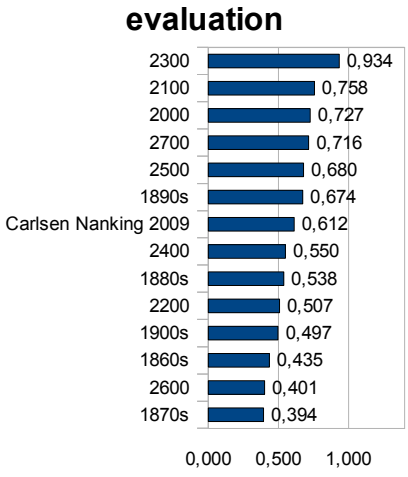
1. *Complexity.*

Complexity shows for what extent a position is tactically complicated, irrational and unclear where concrete calculation has greater prominence than positional evaluation. A high complexity can occur also in seemingly calm positions but where, however, some surprising and hard-to-see tactical shots are found. Another engine is used - Stockfish 1.4, because its evaluation figures are larger than Rybka 3, enabling more fine-grained distinguishing.
Complexity is calculated this way: all instances of Stockfish changing its best move at each depth from 2 to 15 are determined. After that all differences between first and second best move at each point of move change are added together. One has to take into account the fact that changes at greater depths are more difficult to see for players, so all differences within the interval 10-15ply were further multiplied by two.



2. *Difference*

Difference is simply the evaluation difference between the best and the second best move. Since all the players analyzed here generally play on a quite high level and do not make oversights so easily, the maximal difference is set at 3.00 and all higher values are regarded as 3.00. Positions where best moves are forced have greater difference. It shows how important it is for a player to choose best moves. It may seem a bit strange that higher difference is in correlation with lower accuracy of play, whereas the probability of choosing the best move increases. But that's precisely why chess is interesting, very often the best move is not obvious and clearly distinguished at all. In more peaceful and strategical positions difference factor is relatively small, and enabling a player to choose among many more-or-less equal moves. Although it has not been considered here, a smaller difference is presumably in correlation with how moany PV-s are within a certain evaluation window.

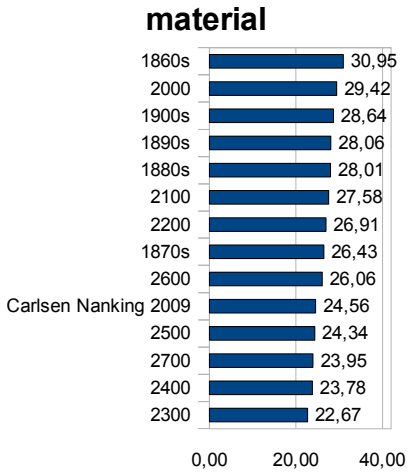


3. *Evaluation*

Evaluation describes how far a position has drifted from the complete equilibrium towards a decisive result. It is the same as the absolute value of the evaluation of an engine on the best variation. Also, it shows indirectly how tactical a position is, since the frequency of evaluation jumps in messed-up positions is certainly bigger either as a result of participants committing more errors or engines having troubles in orienting tin a thicket of variations. Secondly, in positions where evaluations are higher, deviations tend to have bigger values due to bigger proportions even if other factors remain same.

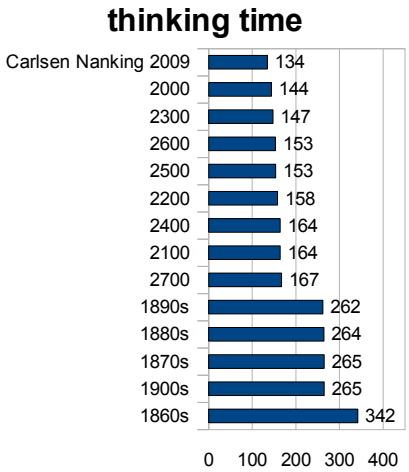
4. Material

The system devised by Larry Kaufman with the help of computer was used in describing the quantity of material. Queen 9.75, rook 5, bishop and knight 3.25; bishop pair adds 0.5. Material describes the 'endgameness' of a position. In general, there is a tendency for the accuracy of play to be greater towards material diminishing.



5. Thinking time

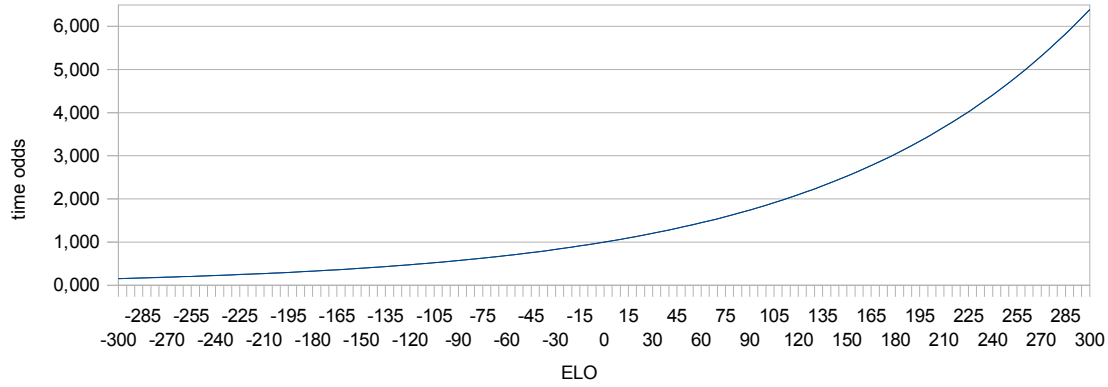
There are numerous sites on the internet where information on time controls in various times at various tournaments and matches are available. In case no data on a particular event was available, estimates given by Encyclopaedia Britannica were used: 1880-1925 4 min; 1926-1945 3 min 20 s; 1946- 1985 3 min 45 s; 1986-... 3 min. One of the most annoying problems were adjourned games. How to measure them? There is no some uniform criterion, in earlier times there even was an unwritten statute that adjourned games were not analyzed. It is known that soviet masters used to help each other in analyzing them, especially against foreign players. In any case, it seems impossible to find a decent solution, so I included an arbitrary 1 hour to every adjourned game.



Next, one of the most difficult problems needed to be solved: how changes in the thinking time affect the accuracy of play? Computer's playing strength increases by 50-70 ELO each time thinking time is doubled, slight diminishing returns can also be observed. In case of humans, these phenomena are more intensified. One possibility was to use clock simuls, as displayed below:

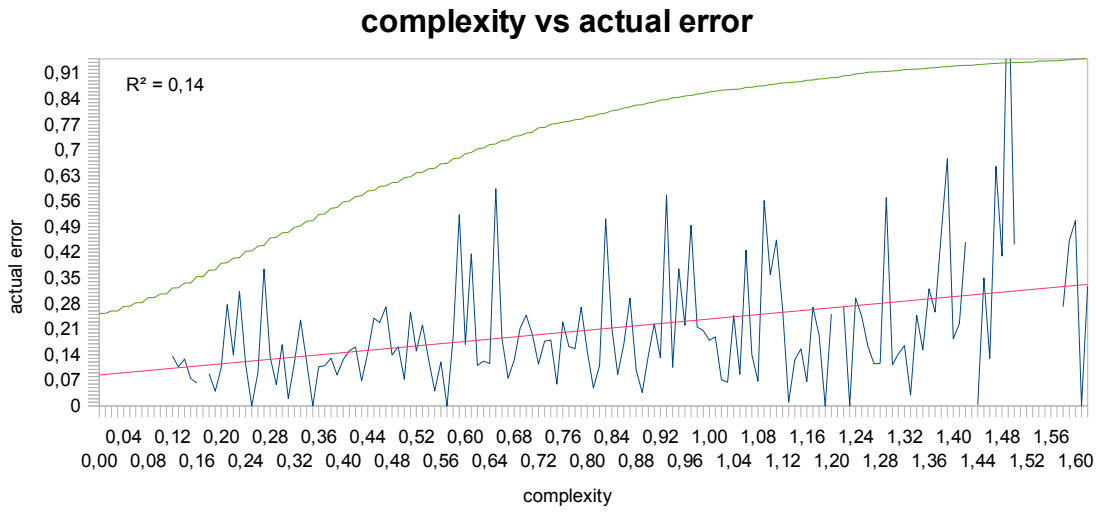
	player	opposition	year	time odds	opposition	performance	player	gap
1	Kasparov	Czech team	2001	3,3	2596	2609	2860	-251
2	Kasparov	Czech team	2001	3,3	2596	2621	2860	-239
3	Kasimdzhanov	Uzbek team	2007	5,0	2408	2446	2682	-236
4	Kramnik	German team	2004	4,0	2604	2617	2765	-148
5	Tal	Warsaw team	1966	8,0	2378	2441	2763	-322
6	Fischer	Greek team	1968	5,0	2348	2398	2795	-397
7	Kasparov	German team	1992	4,0	2632	2657	2865	-208
8	Kasparov	Israeli team	1998	4,0	2620	2658	2854	-196
9	Kasparov	Israeli team	1998	4,0	2620	2658	2854	-196
			average	4,511	2534	2567	2811	-243,7

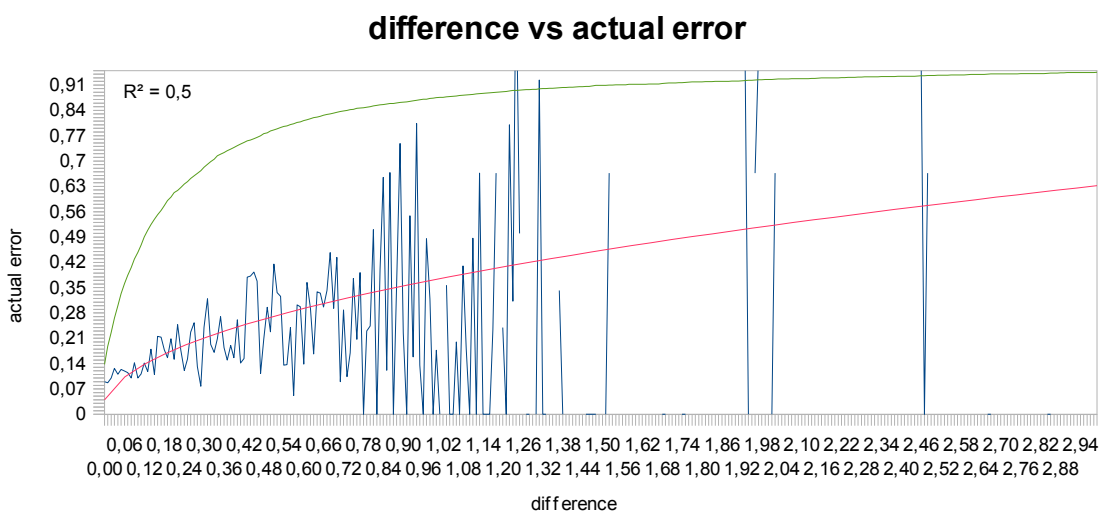
The method consists of comparing the performance rating of a player giving a simul to his own actual rating and dividing it by mean time odds. In current case it appears that 4.5-fold time odds causes a player to lose his playing strength by ca 250 ELO points. according to the following graph, doubling thinking time means a rise of playing level by about 110 ELO.

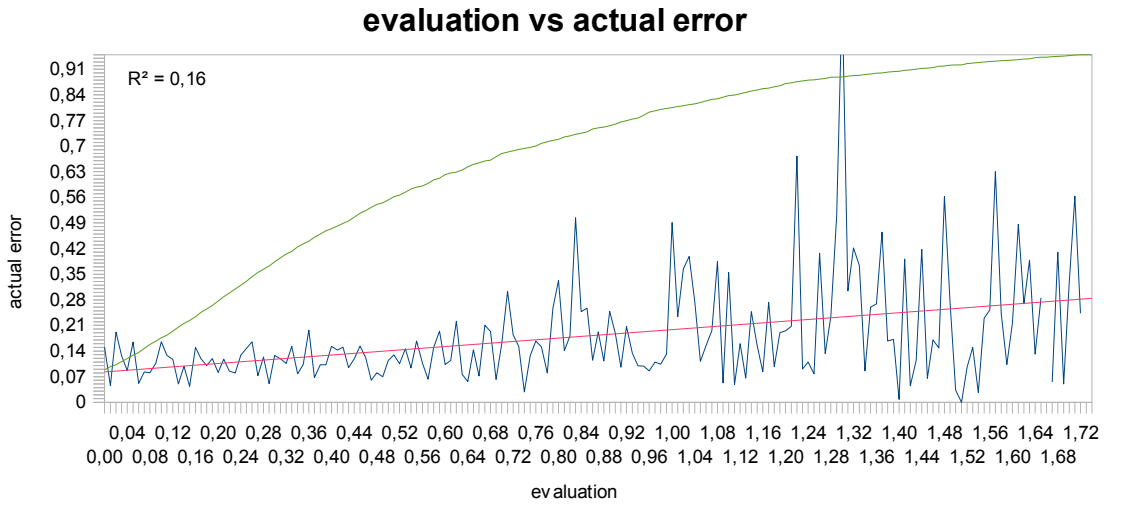


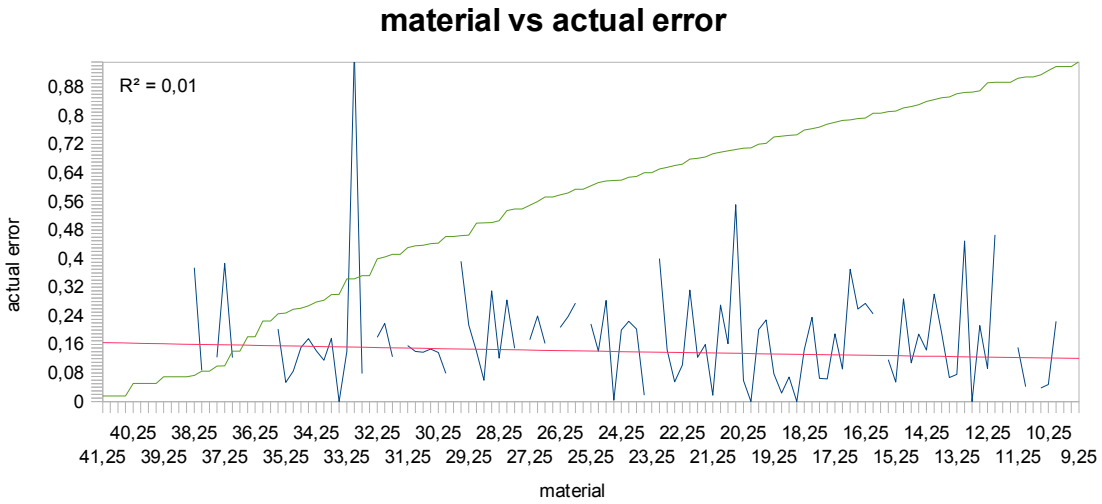
After determining the difficulty factors of every cohort, the average expected error has to be calculated on the basis of them. It shows hypothetically what the accuracy of play of chessplayers would be if each had identical types of positions and time controls. The average expected error is a function of playing accuracy in respect to all difficulty factors. Depending on the style of play and the nature of positions, the curve may be flatter or steeper. The accuracy of play of tactical players in positions that are more complicated and more familiar to them is relatively better compared to players inclined towards positional play. But it is lower in positions requiring strategical thinking and and generally in endgames. Their curve of the expected error is therefore flatter. On the other hand, the curve of positional players is steeper.

The influence of each difficulty factor on actual error is individual. It is strongest in the case of the difference, and the lowest in the case of the material.

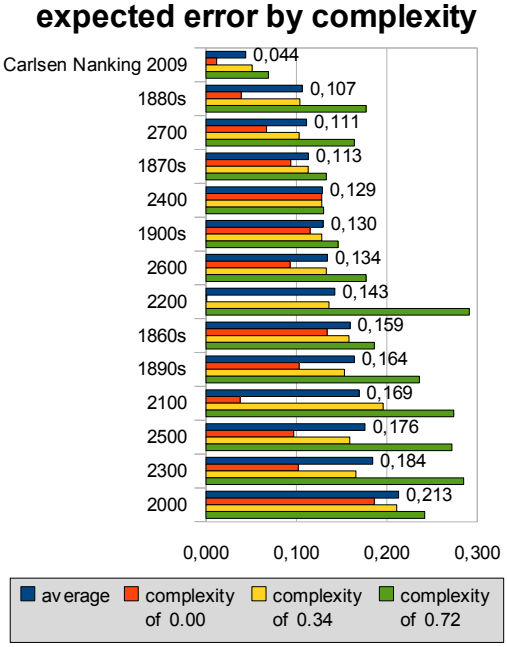


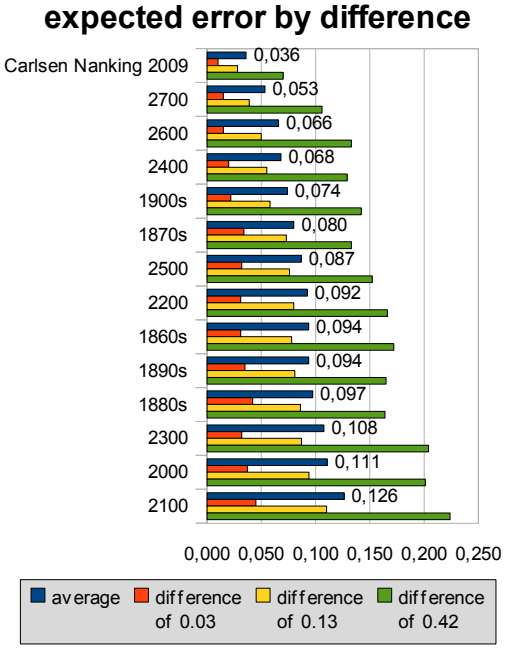


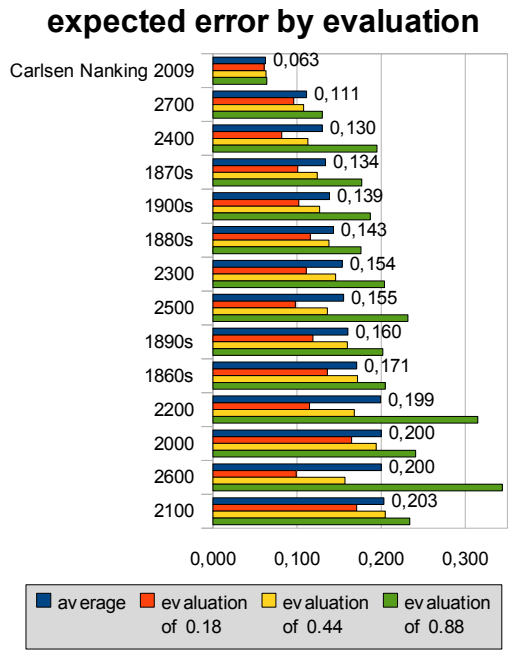


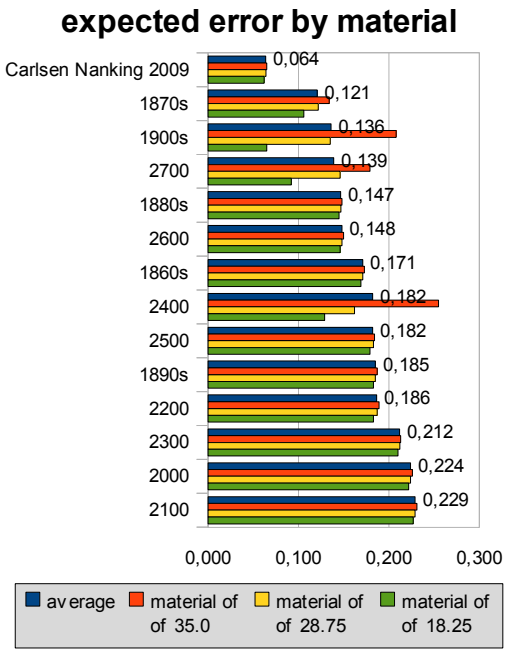


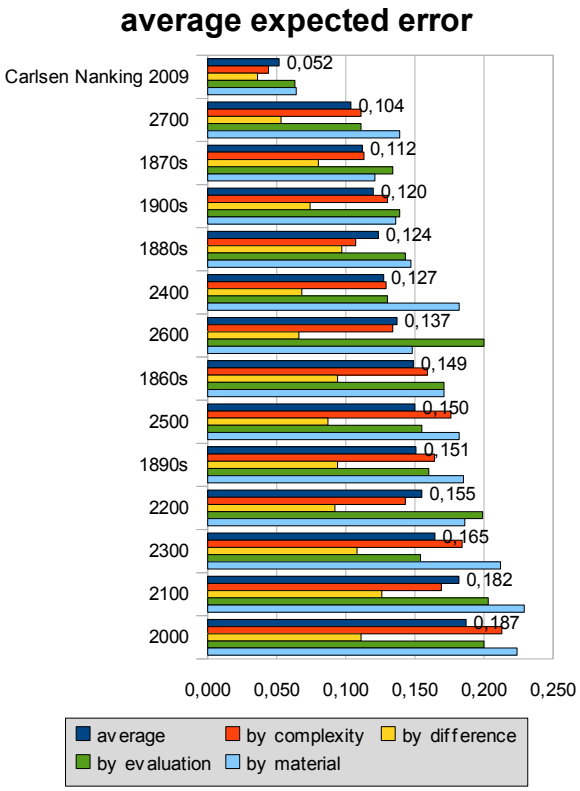
The green curve shows the amount of positions in percentages, the red curve is a statistical regression line, R^2 is the correlation coefficent. The most influential and most trustworthy of them is the difference criterion. In creating all the graphs, the number of moves that included at least 95% of valid positions, served as a basis. To get the expected error values by each difficulty factor, measure ponts must be chosen first. For the purpose of this study, the points were chosen in such a way that they are standing at the transition spots between percentiles of 25%, 50% and 75%: Complexity: 0.00; 0.34; 0.72. Difference: 0.03; 0.13; 0.42. Evaluation: 0.18; 0.44; 0.88. Material: 35.0; 28.75; 18.25.





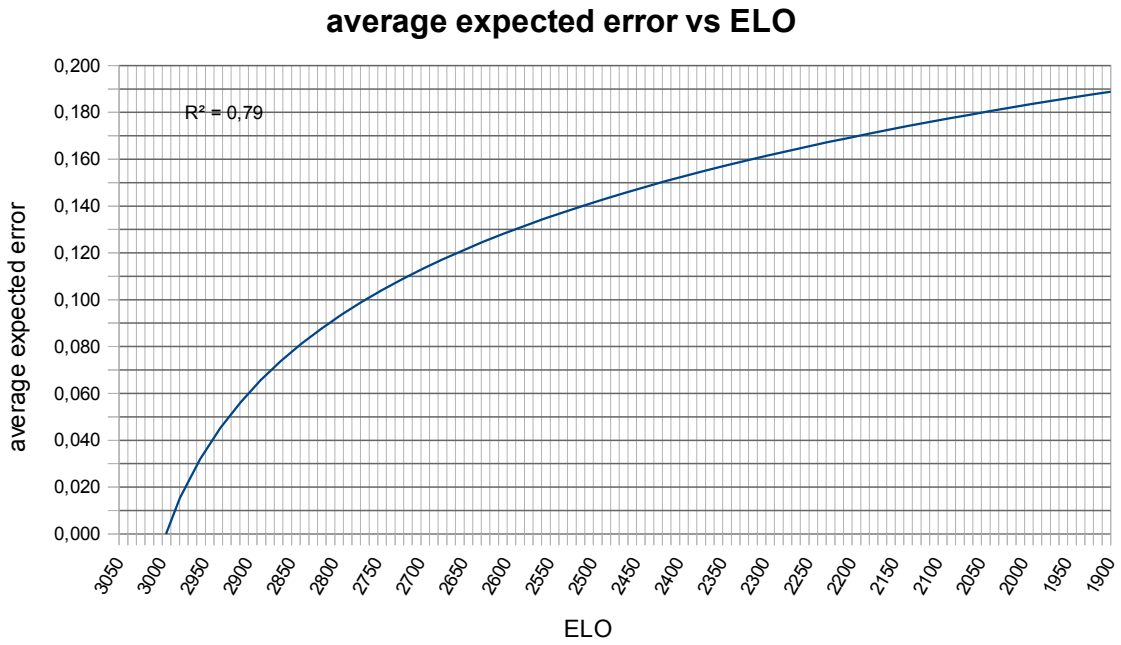




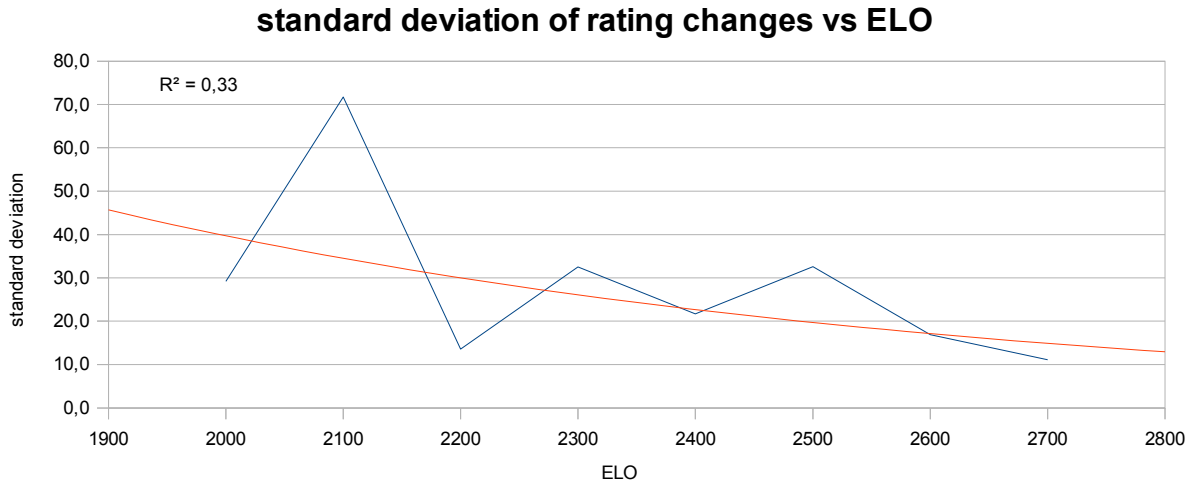


The average expected error is the mean value of all specific expected error values. Having determined its values, it is time to fit it to an identical thinking time; in this study 3 minutes per move (2 hrs per 40 moves) was chosen as a common denominator. Unlike the other difficulty factors, thinking time is taken into account on the basis of ELO rating. Unfortunately it was not feasible to directly apply this effect on the actual average error. The relationship between the accuracy of play and the rating is represented by the graph below. The tme control there is correlated to 180 s per move.

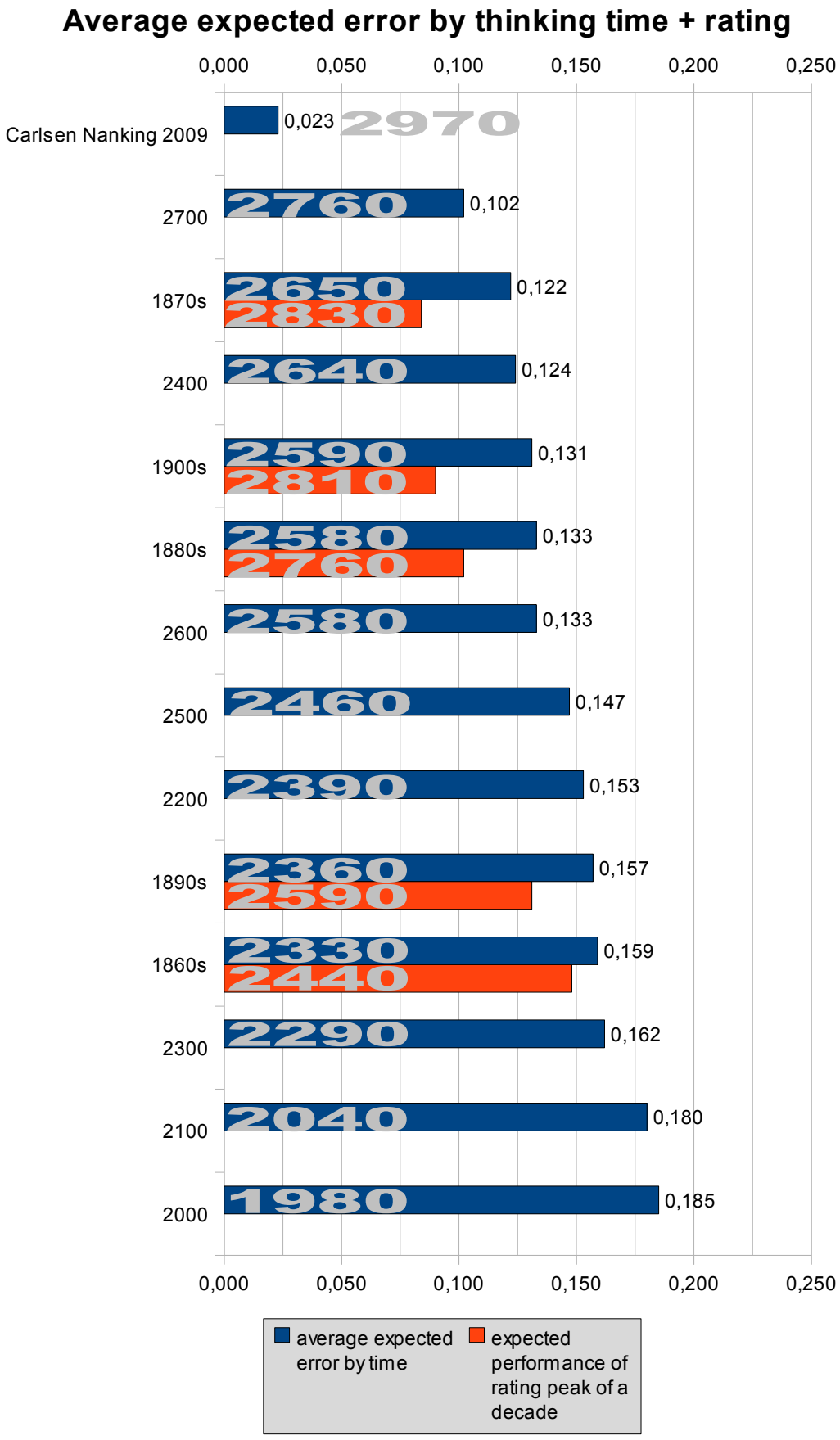
The relation is best described by logarithmic trendline. It hints at well-known fact that the more far back one is as to his playing skill, the easier it is to raise the rating.



Since at a lower level of play similar changes in the accuracy of play should at least theoretically cause bigger relative playing level fluctuations, logically, it must be indicated in FIDE rating statistics. And that's how it actually is! The graph below shows that the standard deviations of rating changes increase in the case of the rating of a player being lower. The highest rating peaks of randomly picked 8 chessplayers by half-years within 2005-2009 were compared.



After applying thinking times on figures represented by the graph 'average expected error', we get a hypothetical playing strength for each player of the past in comparison with today's players. Figures on the bars display the expected rating according to the chart 'average expected error vs ELO'. The red bars represent performances corresponding to rating peaks of the decades.
As can be seen on the graph below, the results are quite unstable, some of them show incredibly unrealistically high playing level. It demonstrates that the study is still in an immature stage, and one is expected to make no far-reaching conclusions yet. First of all, there is a need to increase the amount of games and positions and figure out how to solve the problem of practical, opponen-oriented play and take such aspect of chess into consideration. Computers currently show no understanding whether a mistake is caused by an oversight or insufficient positional knowledge, or a deliberate calculated risk that by raising the difficulty level creates problems for the opponent to solve.


Average expected error by thinking time + rating
0,000
0,050
0,100
0,150
0,200
0,250
Carlsen Nanking 2009
0,023
2970
2700
2760
0,102
1870s
2650
2830
0,122
2400
2640
0,124
1900s
2590
2810
0,131
1880s
2580
2760
0,133
2600
2580
0,133
2500
2460
0,147
2200
2390
0,153
1890s
2360
2590
0,157
1860s
2330
2440
0,159
2300
2290
0,162
2100
2040
0,180
2000
1980
0,185
average expected error by time
expected performance of rating peak of a decade

The following is a list of games included in this analysis. Except Carlsen's games, moves of both participants were taken into account. By 'valid positions', those that satisfy the criteria described above were meant.

1860s

no	White	rating	Black	rating	result	moves	year	opening	number of valid positions
1	Kolisch	2628	Anderssen	2645	0-1	45	1861	B40	
2	Anderssen	2611	Steinitz	2688	0-1	44	1866	C51	299
3	Paulsen	2652	Kolisch	2624	0,5-0,5	29	1861	C00	
4	Anderssen	2660	Paulsen	2663	1-0	29	1862	C65	
5	De Vere	2675	Steinitz	2688	0-1	76	1867	A10	
	average	2645		2662					
				2653					

1870s

no	White	rating	Black	rating	result	moves	year	opening	number of valid positions
1	De Vere	2601	Steinitz	2696	0-1	50	1870	C60	
2	Anderssen	2648	Paulsen	2624	1-0	54	1873	C41	265
3	Zukertort	2663	Blackburne	2617	0-1	61	1878	C11	
4	Anderssen	2638	Zukertort	2663	0-1	49	1878	C65	
	average	2638		2650					
				2644					

1880s

no	White	rating	Black	rating	result	moves	year	opening	number of valid positions
1	Chigorin	2631	Schwarz	2657	1-0	55	1882	C01	
2	Mason	2689	Winawer	2676	0-1	37	1883	C45	276
3	Bardeleben	2651	Tarrasch	2664	1-0	82	1888	D32	
4	Chigorin	2636	Paulsen	2645	1-0	44	1881	B32	
	average	2652		2661					
				2656					

1890s

no	White	rating	Black	rating	result	moves	year	opening	number of valid positions
1	Bird	2603	Blackburne	2665	0-1	30	1892	A02	
2	Schiffers	2662	Walbrodt	2664	1-0	42	1897	C00	308
3	Burn	2669	Marco	2643	0,5-0,5	57	1898	D53	
4	Cohn	2616	Schlechter	2697	1-0	32	1899	C44	
5	Mason	2670	Marco	2653	0-1	45	1894	C27	
6	Teichmann	2638	Charousek	2660	0-1	61	1897	C77	
7	Blackburne	2655	Cohn	2616	0-1	77	1899	C26	
	average	2645		2657					
				2651					

1900s

no	White	rating	Black	rating	result	moves	year	opening	number of valid positions
1	Lasker	2675	Marshall	2658	0,5-0,5	45	1904	B40	
2	Mason	2634	Brody	2602	1-0	55	1900	C45	269
3	Alapin	2658	Blackburne	2656	0,5-0,5	42	1901	B22	
4	Burn	2625	Swiderski	2623	1-0	34	1906	D60	
5	Forgacs	2669	Bernstein	2654	1-0	53	1909	B15	
	average	2652		2639					
				2645					

Carlsen Nanking 2009

no	White	rating	Black	rating	result	moves	year	opening	
									number of valid
1	Carlsen	2772	Leko	2762	1-0	44	2009	C45	positions
2	Carlsen	2777	Topalov	2812	1-0	41	2009	E90	271
3	Wang	2738	Carlsen	2783	0,5-0,5	58	2009	D83	
4	Jakovenko	2743	Carlsen	2782	0-1	63	2009	B92	
5	Carlsen	2787	Radjabov	2757	1-0	25	2009	B30	
6	Leko	2757	Carlsen	2791	0,5-0,5	64	2009	D72	
7	Topalov	2808	Carlsen	2791	0,5-0,5	43	2009	B33	
8	Carlsen	2792	Wang	2741	1-0	69	2009	D17	
9	Radjabov	2747	Carlsen	2796	0,5-0,5	34	2009	D86	
10	Carlsen	2796	Jakovenko	2740	1-0	38	2009	D31	
	average	2772		2776					
				2774					

2000

no	White	rating	Black	rating	result	moves	year	opening	
									number of valid
1	Luodonpää	1991	Pohjala	2022	0,5-0,5	57	2008	A12	positions
2	Buehler	1992	Fiedler	1992	1-0	41	2008	D13	336
3	Falkowski	2002	Stolarczyk	2015	0,5-0,5	116	2008	B26	
4	Boucek	2017	Straka	1978	0,5-0,5	50	2008	B26	
5	Achereiner	2025	Kirk	1979	1-0	40	2008	E97	
6	Schaad	1996	Rickenbach	2023	0,5-0,5	45	2008	A08	
7	Miciak	1987	Stric	2023	0-1	39	2008	E99	
	average	2001		2005					
				2003					

2100

no	White	rating	Black	rating	result	moves	year	opening	
									number of valid
1	Potze	2105	Galje	2121	0,5-0,5	52	2008	B22	positions
2	Kanyadi	2103	Rakaczki	2081	0,5-0,5	45	2008	B01	310
3	Weber	2078	Volkov	2084	1-0	42	2008	D40	
4	Crombleholme	2101	Gustavsson	2096	1-0	50	2008	D63	
5	Li	2108	Sermier	2099	1-0	45	2008	D11	
6	Milonakis	2099	Bras	2082	1-0	50	2008	B14	
	average	2099		2094					
				2096					

2200

no	White	rating	Black	rating	result	moves	year	opening	
									number of valid
1	Kozak	2195	Toma	2212	0,5-0,5	63	2008	D93	positions
2	Tsõganova	2221	Korchagina	2182	0,5-0,5	83	2008	E61	321
3	Shytaj	2212	Zoldan	2208	1-0	42	2008	B07	
4	Masse	2196	Arsenault	2213	1-0	45	2008	E32	
5	Wyss	2181	Kojima	2205	0-1	53	2008	B33	
6	Jagodzinski	2216	Klim	2179	0-1	43	2008	B22	
	average	2204		2200					
				2202					

2300

no	White	rating	Black	rating	result	moves	year	opening
1	Geenen	2304	Ringoir	2321	0,5-0,5	35	2008	B96
2	Barreto Filho	2300	Prates	2298	1-0	89	2008	D11
3	Petersen	2309	Jacobsen	2316	0,5-0,5	48	2008	B06
4	Limontas	2289	Tvarijonas	2277	0-1	41	2008	D02
5	Franciskovic	2275	Kantorik	2288	0-1	44	2008	A21
6	Pachta	2312	Aschenbrenner	2278	1-0	35	2008	E68
	average	2298		2296				

2297

number of valid positions 318

2400

no	White	rating	Black	rating	result	moves	year	opening
1	Marinovic	2408	Nestorovic	2393	0-1	37	2008	C85
2	Löffler	2411	Pinter	2415	0-1	38	2008	D80
3	Housieaux	2411	Debray	2378	1-0	38	2008	D25
4	Brumen	2383	Rogulj	2396	0,5-0,5	41	2008	C26
5	Huss	2378	Gerber	2389	1-0	40	2008	E14
6	Saptarshi	2404	Himanshu	2423	0,5-0,5	68	2008	D01
7	Zvara	2409	Kanovsky	2409	0-1	50	2008	B43
	average	2401		2400				

2401

number of valid positions 317

2500

no	White	rating	Black	rating	result	moves	year	opening
1	Jakubowski	2497	Onischuk	2506	0-1	73	2008	B15
2	Porper	2476	Friedel	2524	0,5-0,5	47	2008	E13
3	Ushenina	2484	Carlsson	2501	0-1	43	2008	A60
4	Vovk	2488	Sulashvili	2495	1-0	43	2008	B19
5	Bhat	2498	Lima	2488	0,5-0,5	46	2008	D53
6	Mainka	2482	Firman	2525	1-0	35	2008	B50
7	Jakubowski	2497	Bachmann Schiavo	2511	0,5-0,5	35	2008	A48
	average	2489		2507				

2498

number of valid positions 330

2600

no	White	rating	Black	rating	result	moves	year	opening
1	Smeets	2604	L'Ami	2610	0,5-0,5	60	2008	B19
2	Malakhatko	2612	Mchedlishvili	2604	0,5-0,5	37	2008	E11
3	Atalik	2585	Pantsulaia	2607	1-0	55	2008	A30
4	Macieja	2606	Rozentalis	2599	0,5-0,5	98	2008	E40
5	Guseinov	2617	Rodshtein	2614	0,5-0,5	35	2008	C12
	average	2605		2607				

2606

number of valid positions 302

2700

no	White	rating	Black	rating	result	moves	year	opening
1	Van Wely	2681	Polgar	2707	0,5-0,5	37	2008	E21
2	Bologan	2682	Naiditsch	2678	0,5-0,5	66	2008	C55
3	Alekseev	2708	Dominguez	2708	0-1	45	2008	B91
4	Alekseev	2711	Jakovenko	2711	0,5-0,5	45	2008	E26
5	Jakovenko	2709	Cheparinov	2687	1-0	50	2008	C67
6	Cheparinov	2687	Gelfand	2720	0,5-0,5	41	2008	C42
7	Wang Hao	2684	Grischuk	2716	1-0	37	2008	C89
	average	2695		2704				

2699

number of valid positions 337