Cognitive Science 34 (2010) 1–9

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01086.x

Does Playing Blindfold Chess Reduce the Quality of Game: Comments on Chabris and Hearst (2003)

Veljko Jeremic, Dragan Vukmirovic, Zoran Radojicic

Faculty of Organizational Sciences, University of Belgrade

Received 27 February 2009; received in revised form 9 September 2009; accepted 5 October 2009

Abstract

Blindfold chess is a special type of chess game where both the board and pieces are not visible to its players. This paper aims to determine whether the quality of the game played blindfolded is lower than when played under normal conditions. The best chess program was used to analyze games played by the world's top Grandmasters under both conditions. We have analyzed the Monaco 1993–1998 data set introduced by Chabris and Hearst (2003). The results showed that although a larger number of mistakes occurred while playing blindfolded, no significant statistical difference between the rapid and blindfold games has been found. Nevertheless, by applying the same methodology to the Monaco 2002–2007 data set a substantial difference between the blindfold and the rapid chess game was noticed. In this paper, we have addressed the possible improvement of the chess game quality and the advances in chess programs that may be responsible for detecting more blunders.

Keywords: Problem solving; Decision making; Expertise; Visualization; Pattern recognition

1. Introduction

For many years, chess has been the focus of researchers interested in memory and decision making (de Groot, 1946, 1978; de Groot & Gobet, 1996). The most important issue has been the unknown reason for Grandmasters' superiority. While one side claimed that recognition of the pattern was more important, the other side persisted in the opinion that the search for and evaluation of possible moves are crucial for chess excellence.

Probably the most prominent advocates of pattern recognition theory, Gobet and Simon (1996) conducted a study to prove their point of view. Their analysis included the performances of Garry Kasparov in simultaneous exhibitions. Their key argument was that

Correspondence should be sent to Veljko Jeremic, Laboratory for Statistics, Faculty of Organizational Sciences, University of Belgrade, Jove Ilica 154, 11 000 Belgrade, Serbia. E-mail: jeremicv@fon.rs

although Kasparov had to play very quickly, rating quality (performance) showed no drastic difference compared to his usual performances. As a proponent of the other point of view, Holding (1985) provided arguments that the search for and the evaluation of moves had much greater importance than pattern recognition. Finally, Chabris and Hearst (2003) presented results showing that the number of mistakes occurring in rapid and classical play differed significantly. Their conclusion was that the time limit reduction significantly affects the quality of game.

In the shadow of pattern recognition versus search dilemma, a specific research area began to develop. Blindfold chess attracted many scientists with its extravagancy. In blindfold chess, a player carries out game(s) having no view of the board and the pieces. For a common observer this seems to demand exquisite memory capabilities. Famous chess players such as Pillsbury, Bourdonnais, Koltanowski, and Najdorf could successfully play 30–45 blindfold games simultaneously. Such a phenomenon easily attracted a crowd, leading to a large number of studies, starting with Binet (1893, 1894), who found that skilled players did not imagine the physical properties of the game, such as the color or pieces. Instead, such players preferred an abstract type of representation. Saariluoma (1991) studied the use of memory in blindfold chess. He found out that chess masters were not able to recall the positions with illegal random moves. The same author has studied decision making using blindfold chess. Saariluoma and Kalakoski (1997, 1998) showed that replacing chess pieces with dots bore no significant effect on the memory performance in both masters and amateurs. One of their conclusions was that the blindfold practice improves the chess skill. Grandmaster Jonathan Tisdall (1997) argued the same. His ability to play chess blindfolded was pointed out as one of the main reasons why he achieved the International Grandmaster title. For further information on the blindfolded chess phenomenon, please refer to Campitelli and Gobet (2005) and Campitelli, Gobet, Williams, and Parker (2007).

In their work, Chabris and Hearst (2003) found no statistically significant difference between the frequency of errors made in a blindfold game and a rapid game. This surprising result was what attracted our attention. The fact that the first author of this paper is an International Grandmaster provided us with the unique opportunity to use his expertise in this research.

2. Method

Similar methods to those of Chabris and Hearst (2003) were applied, and the results of the six Monaco events from 1993 to 1998 were analyzed. In addition to this time span, the Monaco tournaments from 2002 to 2007 were also included in the analysis. In their work, Chabris and Hearst also analyzed the effect of time pressure and they presented a fundamental result that reducing the time limit significantly diminishes the quality of the game. This paper will focus only on the issue of the blindfold versus the rapid game controversy. There are a couple of reasons for this approach. First, we find Chabris and Hearst's conclusion about the classical versus the rapid games issue important and we strongly support their achievement. We also find it very difficult to choose which classical games to incorporate into the analysis. The number of games that the world's best Grandmasters play against each other is so large that it makes it a daunting task to decide which game (and only one may be chosen) to introduce along with rapid and blindfold game played in Monaco. Apart from the above-mentioned reasons, we have only wanted to compare the rapid versus the blindfold games quality as this issue is quite important in our opinion and yet not explored well enough.

During the 1993–1998 and 2002–2007 period, 12 International Grandmasters had competed for the title in the Monaco tournament each year. Organizers invited players according to their chess rating (Elo, 1986). The tournament is organized as round-robin. Grandmasters play two games against each other. One of the games is the ''blindfold'' one in which players have no sight of the actual position and their field of vision is reduced to an empty chessboard projected on the computer monitor in front of them. The other game is the ''rapid'' one, meaning that the players are able to see the actual position. The time limit is almost equal for both types of game, while the only difference is the extra 10 s per move in the blindfold game required for the player to type the move in the computer. Players with white pieces in the first game have black ones in the second game. The draw and the decision on piece colors take place at the beginning of the tournament. This system equals the number of games with white pieces for each player to the greatest possible extent. All of these measures have been made with the aim of maximizing the objective results achieved in this superior tournament.

In order to identify all the blunders, we have analyzed our data sets with the chess program Rybka 3 on AMD Athlon Dual-Core Processor with 2.00 GB RAM using 10 ply-depth search (Vasik Rajlich, Larry Kaufman). Rybka is the world's best chess program and the World Computer Chess Champion for 2007 and 2008. In addition, it is the best-rated program in the world (Karlsson, 2009).

This chess program has analyzed every move in all of our games and reported each ''blunder.'' We have defined blunders as cases in which the actual move played was at least 1.5 pawns worse than the program's choice. Whether the blunder was actually used by the opponents is irrelevant in this analysis. In this research, all the errors that do not affect the outcome of the game were excluded. For instance, if the same side has an advantage of at least 3.0 pawns even after a blunder, the error is not calculated. This criterion of 1.5 pawns has been chosen because the computer chess researchers and the Grandmasters consider this advantage sufficient to win a game (Hartmann, 1989).

3. Analysis of Monaco 2002–2007 games

Results obtained by analyzing our data using the Rybka chess program are summarized in Table 1. As we can see from this table, the rapid game lasted for 46.87 moves. On the other hand, the blindfold game was approximately five moves shorter, with the average of 41.94 moves. The matter of concern in our work is the frequency of blunders committed in these two different conditions. In the rapid game, the average number of blunders per game was 0.53, compared to 0.67 blunders, which occurred in the blindfold games. We have also

Table 1
Exploratory analysis for Monaco 2002–2007

	M ± *SD*	Median ± IQR
Moves		
Rapid	46.87 ± 0.889	44 ± 21
Blind	41.94 ± 0.767	39 ± 21
Blunders per game		
Rapid	0.53 ± 0.04	0 ± 1
Blind	0.67 ± 0.042	0 ± 1
Blunders per move		
Rapid	0.0058 ± 0.0004	0 ± 0.011
Blind	0.0084 ± 0.0005	0 ± 0.014

analyzed the variable number of blunders per move, which is calculated as (number of blunders per game)/(number of moves * 2).[1] As a result, 5.84 blunders per 1,000 moves occurred in rapid games. As expected, the larger number of blunders was committed in blindfold games, 8.42 per 1,000 moves.

According to the results, the Grandmasters made more blunders in the blindfold games. An interesting fact is that the blindfolded games had fewer moves.

As the next step in our work, the Kolmogorov-Smirnov test was used to determine whether the number of moves and the number of blunders for each condition had normal distribution. The Kolmogorov-Smirnov test (Gupta, 1999) indicated that neither of these variables was distributed normally (p's $< .05$). Therefore, we have performed nonparametric tests.

The next step of our analysis was to divide our data in 396 pairs of games (a rapid and a blindfold game make a pair). For instance, a pair is made of rapid and blindfold game played by Aronian and Vallejo Pons in 2007.

We have used the Wilcoxon Signed Ranks Test to compare the number of moves, blunders per game, and blunders per move in each of the conditions (the rapid chess and the blindfold chess). The Wilcoxon Signed Rank test has been very commonly used since it compares distributions when normality cannot be assumed (Gupta, 1999). This is precisely the reason why we have analyzed the two related samples using a nonparametric test. According to this method, the difference in number of moves between the rapid and the blindfold games is considerable, with $p < .001$.

This finding is easy to interpret. Playing a blindfold game is very demanding and players cannot maintain the high level of focus throughout the game. It is the main reason why blindfold games had fewer moves. Within the scope of this paper, a more important issue relates to the frequency of blunders. These results indicate a strong difference ($p = .01$) in the number of blunders per game achieved in different conditions (rapid and blindfold). This is a very important finding because it strongly opposes the Chabris and Hearst conclusion and provides statistical proof for our claim.

We further analyzed the data by comparing the number of blunders per move for both conditions. These results also showed significant difference in the number of blunders per move between the rapid and the blindfold games, with $p < .001$.

One of the important issues in this paper is whether a blindfold game increases the risk of potential blunders. We have calculated Mantel-Haenszel Common Odds Ratio Estimate, OR = 1.545 (95% CI 1.16–2.05; $p = .003$). The important aspect of this part refers to the evaluation of relative risk for blindfold games. This has provided us with an interesting perspective in our work. The relative risk for blindfold game is 1.2391 (95% CI 1.0787–1.422). According to this result, one blunder in a game played under rapid conditions generates 1.2391 blunders if the game is played blindfold.

4. Analysis of Monaco 1993–1998 games

We have applied the same method in this part of our research as in the previous section. The results obtained by examining our data using Rybka chess program are summarized in Table 2. As depicted in Table 2, rapid game was completed in 49 moves. On the other hand, the blindfold game lasted for approximately three moves less, with the average of 45.84 moves. The subject of concern in our work is the frequency of blunders made in these two different conditions. The rapid game yielded the average number of blunders per game of 0.644, compared to 0.75 blunders that occurred in blindfold games.

On the other hand, there were 7.21 blunders per 1,000 moves in the rapid games. As expected, the larger number of blunders was made in the blindfold games, 9.02 per 1,000 moves.

As the next step in our work, we have used the Kolmogorov-Smirnov test to find out if the number of blunders per game and blunders per move for each condition had normal distributions. Neither of these variables had a normal distribution (p's $< .05$). Therefore, nonparametric tests have been performed.

Subsequently, we have divided our data into 396 pairs of games (a rapid and a blindfold game make a pair). We have used Wilcoxon Signed Ranks Test to compare number of moves, blunders per game, and blunders per move for each of the conditions (the rapid and blindfold game). According to these results, the difference in the number of moves between the rapid and the blindfold games is significant, $p < .001$.

Table 2
Exploratory analysis for Monaco 1993–1998

	$M \pm SD$	Median ± IQR
Moves		
Rapid	49.007 ± 0.877	47 ± 24
Blind	45.843 ± 0.847	44 ± 21
Blunders per game		
Rapid	0.644 ± 0.041	0 ± 1
Blind	0.75 ± 0.049	0 ± 1
Blunders per move		
Rapid	0.0071 ± 0.0004	0 ± 0.012
Blind	0.00902 ± 0.0006	0 ± 0.014

Our analysis showed no statistical difference, $p = .108$, in the number of blunders per game achieved under different conditions (rapid and blindfold). This is a very interesting finding because it does support Chabris and Hearst's conclusion. We have further analyzed the data by comparing the number of blunders per move for both conditions. The difference in number of blunders per move between rapid and blindfold games is significant, $p = .035$. Therefore, upon observing the variable number of blunders per game no statistical difference for the two conditions of play (rapid and blindfold) was found. However, scrutinizing the variable number of blunders per move in our analysis led to the result that shows a statistically significant difference.

5. Comparison of Monaco 1993–1998 and 2002–2007 results

Upon conducting individual analysis of Monaco 1993–1998 and 2002–2007 data sets, we focused on comparison of these two periods, to examine quality of play over time. First, we compared blindfold games played in two different periods. The Kolmogorov-Smirnov test indicated that neither number of blunders per game nor number of blunders per move had normal distributions. As neither had normal distribution, both with $p < .05$, we performed the nonparametric Mann-Whitney test. Statistically significant difference has not been found in number of blunders per game in Monaco 1993–1998 and 2002–2007 data sets, with $p = .064$. Same conclusion applies for number of blunders per move, $p = .132$.

Subsequently, we compared rapid games played in two different periods. Again, the Kolmogorov-Smirnov test was utilized to determine whether number of blunders per game and number of blunders per move variables had normal distribution. As neither of these variables had normal distribution, all of them with $p < .05$, we performed the nonparametric Mann-Whitney test. Test results indicated that the number of blunders per game was statistically different for Monaco 1993–1998 and 2002–2007 data sets, with $p = .025$. Same conclusion applies for number of blunders per move, $p = .034$.

It has been determined that number of blunders in rapid conditions is significantly reduced in Monaco 2002–2007 games, compared to the 1993–1998 data set.

6. Discussion

Research papers on blindfold chess have brought interesting insight into the process of thinking and reasoning. The purpose of this research is to determine whether playing blindfolded affects the quality of the game. In their work, Chabris and Hearst (2003) found no statistical difference between the frequency of errors occurring in a blindfold game and a rapid game. They concluded that 6.85 blunders per 1,000 moves were made in rapid games, while in blindfold games the ratio was 7.63 blunders per 1,000 moves. We have examined the same data set as Chabris and Hearst and found no statistically significant difference in rapid and blindfold chess games. Still, we have detected larger number of blunders in both types of the game. In this work, 7.21 blunders per 1,000 moves have been identified for rapid games and 9.02 blunders per 1,000 moves in blindfold chess games. Most probably, this phenomenon occurs as a result of modern chess programs that detect blunders. Significant improvements in chess software and very powerful hardware allowed for reconsideration of data presented in Chabris and Hearst's work. A particularly interesting fact is that newer chess software packages are not only better in fine analysis but also have a larger capacity in determining blunders. Chess players have so far considered that major mistakes could be discovered with different programs yielding equally valid findings.

In addition to analyzing Chabris and Hearst's data set, we have also examined Monaco chess tournaments from 2002 to 2007. Using the same method, we have discovered 5.84 blunders per 1,000 moves in rapid games and 8.42 blunders per 1,000 moves in blindfold games. The most important fact is that we found a statistically significant difference between rapid and blindfold games. In view of these results, we were able to compare the quality of chess games in two periods (1993–1998 and 2002–2007). When observing a blindfold chess game, we can clearly see an improvement, as the 1993–1998 tournaments had 9.02 blunders, while 8.42 blunders per 1,000 moves occurred in 2002–2007. One logical explanation could be that Monaco tournament is unique in the world and players had some difficulties playing (especially blindfold) at the beginning on such a demanding high level. Nevertheless, the real differences in 1993–1998 and 2002–2007 games were noted in the quality level in rapid games. Indeed, 7.21 blunders were committed in 1993–1998 tournaments, while 5.84 blunders per 1,000 moves occurred in 2002–2007. The Mann-Whitney test showed clear statistical difference in number of blunders for these two periods. This is why our statistical approach discovered significant difference between the rapid and the blindfold chess game in the 2002–2007 data set. The number of blunders committed in rapid games was significantly reduced compared to 1993–1998 games. There are several possible explanations for this finding and we will introduce one that the first author of this paper considers the most important. During 1993–1998, chess games lasted for more than 6 h, very often 7 h, compared to Monaco rapid games where each player had a time limit of approximately 30 min. Therefore, considerable difference in time available for thinking was the major obstacle and probably the main reason for that many mistakes. Starting from the year 2000, the World Chess Federation reduced the time limit in most tournaments in the world and, consequently, games have usually lasted for approximately 4 h since then. All chess players, including the best ones, had to adjust to these changes and improve their game when time limit for thinking was reduced so drastically. As a logical consequence of these changes, the rapid play in Monaco 2002–2007 tournaments was noticeably improved.

One can argue that these two periods (1993–1998 and 2002–2007) are not comparable. Undeniably, the average rating of participants at the tournament increased from 2,634 in 1993 to 2,684 in 1996, to 2,728 in 2005, and 2,730 in 2007. Knowing that there is a strong relationship between increasing skill and decreasing blunders (Roring, 2008), the question that could be raised is whether higher rated participants of 2002–2007 tournaments committed less blunders simply because they are ''better'' chess players than their 1993–1998 Monaco tournaments counterparts. A very logical explanation of this phenomenon is a general increase in ratings among chess players. For example, average rating of Top 100 players in July 2000 was 2,644. On the other hand, in September 2009 average rating of Top 100 players was 2,689 (FIDE, 2009). General inflation of rating is constant over the past years, with at least 2–3 points of rating increasing each year. This is why the World Chess Federation (FIDE) officials are exploring new possibilities of improving the rating system (FIDE, 2008).

It is important to mention a potential problem of order of play. Currently at the tournament each player plays two games against a single opponent; the first game is always blindfolded and the second game is the rapid one. This would be true of the 2006 and 2007 tournaments. However, it has not always been so. For at least some of the games mentioned in Chabris and Hearst, players would have played a blindfolded game after a rapid game. We do not know what effect order may have on blunders. It is possible that players just make fewer blunders in the second game. On the other hand, it is also likely that players feel tired after the first game and commit more mistakes in the second game. Either way, the order of play is an interesting issue that should be explored in future research.

We hope that this study will add to the growing body of archival studies in chess (Chabris & Hearst, 2003; van Harreveld, Wagenmakers, & van der Maas, 2007; Roring & Charness, 2007) and that its results could be conducive to a better understanding of the role of mental imagery in expert behavior.

Note

1. Please note that the average of the blunder rates per move will not be exactly the same as if we calculated the blunder rate from average number of blunders per game and average number of moves.

References

Binet, A. (1893). Les grandes memoires: Resume d'une enquete sur les joueurs d'echecs. *Revue des Deux Mondes*, *117*, 826–859. In M. L. Simmel & S. B. Barron (Trans.). (1966). Mnemonic virtuosity: A study of chess players, *Genetic Psychology Monographs*, *74*, 127–162.

Binet, A. (1894). *Psychologie des grands calculateurs et joueurs d'echecs. (Psychology of great calculators and chess players)*. Paris: Hachette.

Campitelli, G., & Gobet, F. (2005). The mind's eye in blindfold chess. *European Journal of Cognitive Psychology*, *17*, 23–45.

Campitelli, G., Gobet, F., Williams, G., & Parker, A. (2007). Integration of perceptual input and visual imagery in chess players: Evidence from eye movements. *Swiss Journal of Psychology*, *66*, 201–213.

Chabris, C. F., & Hearst, E. S. (2003). Visualization, pattern recognition, and forward search: Effects of playing speed and sight of the position on grandmaster chess errors. *Cognitive Science*, *27*, 637–648.

Elo, A. E. (1986). *The rating of chessplayers, past and present* (2nd ed.). New York: Arco.

FIDE (2008). Rating Regulations – The K-Factor. Available at: http://www.fide.com/component/content/article/1-fide-news/3963-rating-regulations-the-k-factor. Accessed September 16, 2009.

FIDE (2009). World top chess players. Available at: http://ratings.fide.com/toplist.phtml?list=men. Accessed September 16, 2009.

Gobet, F., & Simon, H. A. (1996). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grand-master-level chess. *Psychological Science*, *7*(1), 52–55.

de Groot, A. D. (1946). *Het denken van de schaker. (The thought of the chess player)*. Amsterdam: North-Holland.

de Groot, A. D. (1978). *Thought and choice in chess*. The Hague, The Netherlands: Mouton. (Originally published 1965).

de Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess: Studies in the heuristics of the professional eye*. Assen, The Netherlands: Van Gorcum.

Gupta, V. (1999). *SPSS for beginners*. New Delhi: VJBooks Inc.

van Harreveld, F., Wagenmakers, E. J., & van der Maas, H. L. J. (2007). The effects of time pressure on chess skill: An investigation into fast and slow processes underlying expert performance. *Psychological Research*, *71*, 591–597.

Hartmann, D. (1989). Notions of evaluation functions tested against grandmaster games. In D. F. Beal (Ed.), *Advances in computer chess 5* (pp. 91–141). Amsterdam: North-Holland.

Holding, D. H. (1985). *The psychology of chess skill*. Hillsdale, NJ: Erlbaum.

Karlsson, T. (2009). The Swedish rating list. Available at: http://ssdf.bosjo.net/list.htm. Accessed February 2, 2009.

Roring, R. W. (2008). Reviewing expert chess performance: A production-based theory of chess skill. *Dissertations Abstracts International*, *69*(12), 1–85 (UMI No: 3340756).

Roring, R. W., & Charness, N. (2007). A multilevel model analysis of expertise in chess across the life span. *Psychology and Aging*, *22*, 291–299.

Saariluoma, P. (1991). Aspects of skilled imagery in blindfold chess. *Acta Psychologica*, *77*, 65–89.

Saariluoma, P., & Kalakoski, V. (1997). Skilled imagery and long-term working memory. *American Journal of Psychology*, *110*, 177–201.

Saariluoma, P., & Kalakoski, V. (1998). Apperception and imagery in blindfold chess. *Memory*, *6*, 67–90.

Tisdall, J. (1997). *Improve your chess now*. London: Cadogan.