DOI: 10.18290/pepsi-2024-0001

Jan Przewoźnik*
Adam Biela

Thinking Aloud Indicators in Chess Players' Complex Problem-Solving in a Three-Move Experiment

ABSTRACT

The authors present chess players' problem-solving in a three-move experiment. 16 advanced chess players, both men and women, participated in the experiment, which involved solving complex chess problems. The thinking aloud method was used to provide 200 pages of verbal protocols. Nine indicators were abstracted and measured: the total number of all moves considered in verbal protocol (M), the number of successive solving propositions or number of fresh starts in decision tree (branches in decision tree – N), the number of different options (A), the number of successive changes in solving propositions (nn), reinvestigations of the move considered the most (Pmax), the longest immediate reinvestigations of the move (Pser), longest variation (Dmax), the value of the move chosen (V), and thinking time (Time). These indicators can be

---

* We are greatly saddened by the death of Jan Przewoźnik (16 September 1957 – 31 July 2023), who was an outstanding chess player, coach, sports psychologist, a Polish chess champion in 1979, and an international chess champion. He lectured at the West Pomeranian Business School in Szczecin, Poland. He developed a chess training system for young players, used in schools in Poland and abroad. He wrote many publications on chess. Correspondence regarding the paper should be sent to: Adam Biela (ORCID: 0000-0002-3957-8382), John Paul II Catholic University of Lublin, Poland; e-mail: adam@biela.pl.

The authors would like to heartily thank Mr Marek Soszynski from Birmingham for his meticulous proofreading of the manuscript.

interpreted in terms of problem-solving cognitive strategies. This paper demonstrates the methodology of using these indicators to analyze chess players' complex problem-solving in a three-move experiment.

KEYWORDS: psychology of chess; thinking aloud method; three-move-chess-playing experiment; indicators of chess player's complex problem-solving thinking.

INTRODUCTION

What is more important in effective chess thinking: extensive research or pattern recognition (known after de Groot, 1965, and Chase & Simon, 1973, as the *chunking theory*)? In the methodology of chess psychology a leading position is still occupied by the classic work of Dutch psychologist Adriaan D. de Groot (de Groot 1965, 1981; Busato 2006). The experimental research presented in this paper refers to this concept, mainly as it is based on a thinking aloud method. However, de Groot's experiments were based on a one-move chess situation, while our experiment is based on three moves. The data source will be thinking aloud protocols (verbal protocols) of chess players playing three-move games. The thinking aloud method was used to provide 200 pages of verbal protocols.

One way to analyze the thinking aloud protocols used by de Groot is based on solution proposals (de Groot, 1965, p. 104), i.e. the first possible moves to be made by the subject. And, therefore, the whole thinking aloud protocol could be considered as a formula of successive solution propositions. A visual description would appear as shown in Figure 1 (a sample chess position) and Figure 2 (the formula):

Figure 1. The sample chess position. White to move.

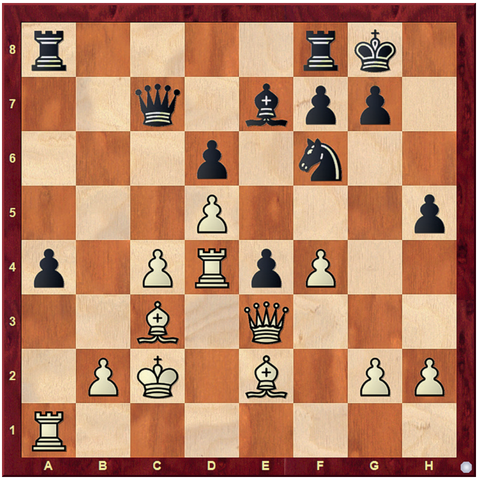


Figure 2. The formula of successive solving propositions. A sample game tree, considered by the subject in a sample position (Figure 1).

A	B	C
1	Rxe4	Rxe4
1	Rxe4	Rxe4
1	Rxe4	Rxe4 Nxe4 Qxe4
1	Rxe4	Rxe4 Nxe4 Qxe4 … Bd3 Re8 Qh7 Kf8
1	Rxe4	Rxe4 Nxe4 Qxe4 Bf6
1	Rxe4	f5 Qe6 Kh7 Bxg7
b	h3	h3 … g4
b	h3	a3
b	h3	Rb8
a	g4	g4 … g5
a	g4	hxg4
a	g4	Nxg4
a	g4	Nxg4 Qxe4
a	g4	hxg4
a	g4	hxg4 h3
a	g4	h3 … hxg4 … Qh3 … Rh1
a	g4	h3 g3

a	g4				g3	Qxg3		
a	g4					Qxg3		… Rxe4
a	g4							Rg1
a	g4						g6	f5
a	g4		a3					
a	g4	g4	Qb6		Ra3			
a	g4	g4	Qb6					
a	g4		a3	b3				
a	g4		Qb6		Ra3			
a	g4	g4						

Note. The choice: a1. g2-g4

EXPLANATION OF FIGURE 2

Column A: The formula of successive solving propositions, Column B: The first move considered, Column C: The decision tree. Possible reasonable moves by White: a = g4, b = h3, c = Ra3, d = Rdd1, e = Rd2, f = Qh3, g = Qg3, h = f5, i = g3, j = Bb4, k = Bd1, l = Rxe4, m = Rg1, n = Kb1, o = Bxh5, p = c5, where a > b > c = d > e = f = g = h = i = j > k > l = m = n.

Therefore, the formula of successive solving propositions of the subject shown in Figure 2 can be presented as follows: l-l-l-l-l-l-b-b-b-a-a-a-a-a-a-a-a-a-a-a-a-a-a-a-a-a-a-a.

The letters in Column A denote the specific chess moves considered by the subject: a = move 1.g4, b = move 1.h3, l = move 1.Rxe4. The closer to the beginning of the alphabet, the more strongly a move is considered. Conversion of chess moves into letters allows for the comparison of formulas from different chess positions. Any single thinking aloud protocol contains a defined number of ongoing moves by the subject. These moves are called "game tree", as considered by the subject.

Even if the researcher can determine the space protocol, he or she cannot be sure that it is compatible, or close to being compliant to the total number of moves considered in the internal space of a subject. For such ephemeral phenomena as a single move, the thinking aloud method is considered inadequate because it does not record all of them. Some of the moves realistically considered by the subject are not verbalized. Individual moves as the basic units of the game tree are too small, so all were included in the protocol. The subject usually – in spite of training – is not able to say “everything”. This makes a single move elude the mind of both the chess player and the researcher from time to time.

However, the reliability of this method can naturally increase if the researcher analyzes only the formula of the successive solving propositions. Each element of this formula is also a single move, but it is usually connected with a further variation (some additional moves). Proposed solutions include not only a single move, but their entire sequence. De Groot represented every single proposed solution by a letter of the alphabet, but this letter constituted a larger entity known as a variation rather than a single move. The letter in the formula meant the current direction of thinking of a chess player. For the researcher, the formula of successive solving proposals acted as a compass, allowing him to determine the direction of a chess player’s current idea.

Of course, oversights can also occur here. They occur when the subject computes a variation in a few seconds and yet does not verbalize it. However, such omissions of solving proposals in the protocol are more noticeable, as they are associated with the onset of a long break in protocol. In this considerable break, the experimenter can draw attention to the earlier training and improve the subject’s verbalization of thought.

It can, therefore, be assumed that the formula of successive solving proposals contains sound data. On the basis of such a formula and a decision tree, several quantitative indicators of individual ways of solving problems on the chessboard can be determined: M, N, A, nn, Pmax, Pser, Dmax, T, W. These indicators are introduced by Adriaan de Groot (1965) to analyze thinking aloud protocols analysis. In this work, only selected names of indicators are modified, and they are also adapted to computerized statistical programs. These indicators give many opportunities for the study of thinking (as was shown in Przewoznik & Soszynski, 2001; Przewoznik, 2019).

M

M denotes the number of all considered moves in the entire decision tree. This can be used to measure moves in chess and analyze expressions in other decision making situations. This indicator allows us to distinguish between those individuals who seek more information for decision making and those who make decisions on the basis of a smaller amount of information.

N

N denotes the total number of subsequent proposals to solve. This number may reflect the subject's mentality type. A large value of N would be consistent with an empirical mentality type, the chess player who prefers a problem-solving approach based on the processing of large amounts of data, and the calculation and verification of a large number of multiple variations. In contrast, a relatively smaller value of N may be characteristic of the theoretical chess player type, whose thinking is less empirical and more deductive without giving specific variations. On the other hand, the value of N may be a function of the position on the chessboard, where a more complex a position implies a greater value of N, **because more variations need to be calculated**. And vice-versa: a simple position may correspond to a smaller value of N. To illustrate, the next indicators will be based on the following example of the formula solving propositions: **a-b-f-g-g-g-h-b-b-c-g-h-a-a-a-a-a-b-a-a,** where letters represent a move; the closer the letter to the beginning of the alphabet, the better the move. In the above example, N = 20, i.e., the total number of letters in the sequence, **excluding** the letter underlined (underlining indicates the decision made).

A

A denotes the set of alternative actions (candidate moves) considered by the subject, here understood as equal to the set of options in a decision-making process. In our example A = 6, since in the formula there appear 6 different solving propositions: a, b, c, f, g, h. Repetitions are omitted. The value of A may have a definitive psychological content, namely, it can be associated with the fluency and semantic versatility of spontaneous thought (Guilford, 1967). It can be assumed that persons characterized by great fluency and versatility would consider many different solving propositions in the process of choosing moves. It may be that fluency of thought will be more closely correlated with the value of N, whereas versatility more closely with the value of A. As in the case of N, the value of A may depend on the situation on the board and not only on the individual traits of the testee.

nn

nn denotes the overall total number of successive changes in the solving propositions. This value applies to all the proposals in the formula, but noted here is each instance of change, each new approach to the problem (including the very first letter as an instance of "change"). In the formula a-a-a-a-b-a nn = 3, because the first four letter a's are treated as a single approach; while in our original example above nn = 11.

Pmax

Pmax denotes the number of re-examined solving propositions. This value marks out all the proposals considered more than once, in other words, all the letters appearing in the formula for the first, second, third time, etc. For instance, in our example a-b-f-g-g-g-h-b-b-c-g-h-a-a-a-a-a-a-b-a-a Pmax equals 8, because that is how often such solving propositions appear.

The psychological significance of Pmax can vary. It may reflect a "functional fixation", when the subject is unable to break through certain barriers during the problem-solving process, and repeatedly directs attention towards the same moves, "going round in circles". When, at the same time, later letters of the alphabet are repeated, the subject analyses qualitatively inferior moves. However, if letters from the start of the alphabet are constantly repeated, then that could be evidence of a good ability to concentrate on the problem. A high value of Pmax could also be evidence of difficulty in making decisions.

Pser

Pser denotes the number of times a solving proposition is reconsidered, but only in the longest single series. From the psychological point of view, this value represents the skill of concentrating on a single chosen solving proposition. In our example a-b-f-g-g-g-h-b-b-c-g-h-a-a-a-a-a-a-b-a-a Pser = 6, for the letter a, between the letters h and b.

Dmax

Dmax denotes the maximum length of calculated variations, the measured number of white and black moves. This value reveals how far ahead the subject is able to calculate variations, to what extent he is able to or wishes to foresee events as they unfold on the chessboard. This value of Dmax can be a measure of an individual's skill in imagining spatial relationships and manipulating them when thinking, or as with other factors it could simply reflect the situation on the chessboard.

T

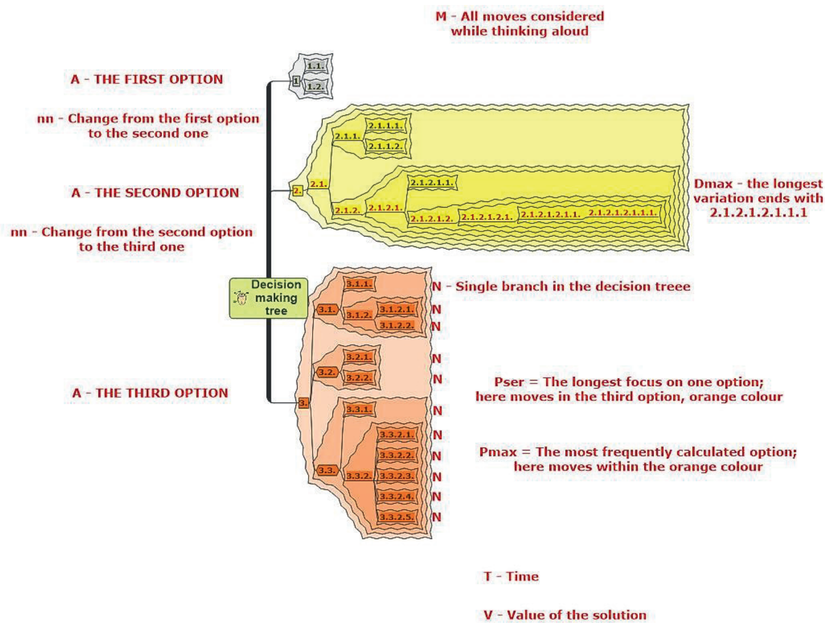
T denotes the time to solve the exercise. On the one hand, this can reflect the tested individual's superior problem-solving speed, and can vary according to cognitive style – reflective versus impulsive. On the other hand, the time to solve the task may simply be a function of the difficulty of the problem.

V

V denotes the value of the solution, 1 or 0.

The indicators described above (M, N, A, nn, Pmax, Pser, Dmax, T, V) are summarised and visualized in Figure 3.

Figure 3. The nine indicators: A visual summary.



The values within the nine indicators mentioned above may quantitatively describe the subjects' specific thinking strategy. Let us assume that the cognitive strategy is an individual's way of data processing and searching for such information that enables the subject to solve the actual problem.

METHODS

One of the problems of research on thinking processes in chess is that previous experiments in the methodological tradition of de Groot analyze only one move on the chessboard. The following question is therefore cognitively intriguing: How to enhance the study of thinking of chess players by analyzing, for example, two or three consecutive moves in an experimental situation, thus exposing all subjects, i.e. the tested chess players, to the same two or three moves?

With the research problem thus formulated, the primary objective of these experiments is to explore the thinking of a chess player during the sequence of moves in complex chess problem-solving situations. The issue is here to analyze the changes in the structure and dynamics of thought processes in a sequence of chess moves. Such an investigation would be a meaningful extension of de Groot's work. Therefore, in our experiment the data source will be thinking aloud protocols (verbal ones) of chess player, who plays three-move game in an experimental situation. Thus, the research problem will be articulated in the following three questions:

> **Q1.** Do chess players' ratings correlate with their thinking strategies – as measured by thinking aloud indicators – in complex chess positions?

The rating of chess players measures the actual chess playing skill of the particular chess players, based on recent performance. In turn, the category of chess player is a lifelong entitlement based on the fulfillment of certain norms in chess tournaments. Thus, both measures of chess-playing skill of the individual chess player really define this skill in differing dimensions. The rating seems to be a more sensitive measuring tool than chess category, more susceptible to change, hence it is reasonable to distinguish between these two skill measures in our analyses of the impact of the chess experience on thinking strategies in chess players.

The first question concerns the impact of chess players' chess experience on their thinking strategies. The next two research questions deal with the relationship between the moves of chess players and the strategies which they use:

Q2. Do chess players who chose good or bad solutions differ in the thinking strategies they use – as measured by thinking aloud indicators in complex chess positions?

Q3. Do the thinking strategies of chess players – as measured by thinking aloud indicators in complex chess positions – differ significantly, when the choice of first moves is compared to the anticipatory thinking strategies for selecting successive moves?

In view of the above-formulated research questions, we propose the following hypotheses:

H1. The rating of the chess players correlates with their thinking strategies – as measured by thinking aloud indicators while solving complex problems in complex chess positions.
H2. Chess players who choose good or bad solutions differ in their thinking strategies.
H3. Chess thinking strategies for choosing first moves differ materially from the anticipatory thinking strategies for choosing successive moves.

The experiment subjects were chosen from the forefront of Polish chess players. The vast majority of people who were invited to the experiment agreed to participate. They were motivated naturally, wishing to do their best in the experimental task and learn about their thinking strategies in chess. The experiment involved 16 adult chess players, 8 women and 8 men. Each subject solved one chess problem, hence the statistical tables values of N are given in relation to the number of participants (8, 16). The subjects represented all of the most important chess titles. Their achievement of a high level in chess guaranteed ease of reporting their own thoughts, handling field names in the experiment, etc. Direct verbalization, where the inner voice speaks, may slightly slow down the execution of the task, but it should not modify the cognitive processes (Ericsson & Simon, 1995a). It can therefore be assumed that the reaction of the chess player is introduced into the short-term memory in a symbolic form and then transferred to the appropriate external behaviour. The rating of the chess players participating in the experiment and their actual chess playing category are collected in Table 1.

Table 1. The subjects' ratings.

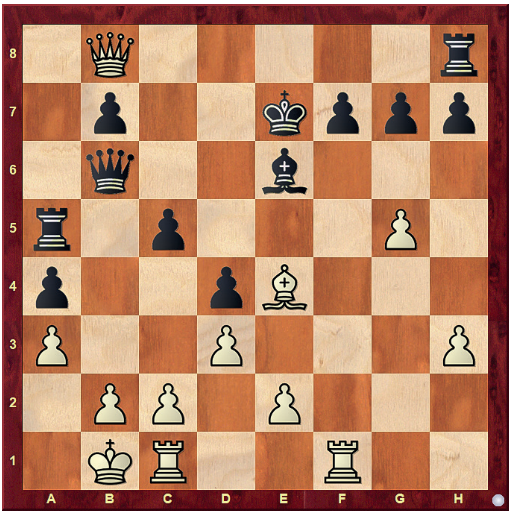
	N	*M*	*SD*
Men and women	16	2229	137.93
Men	8	2326	139.72
Women	8	2133	137.80

The age of the subjects and the length of their chess-playing experience are other factors that can affect the indicators of thinking aloud when making decisions in chess situations. In our experiment the subjects were aged from 20 to 43, their average age was 31. For men the average age was 34 and 28 for women. The difference between the age of male and female subjects is not statistically significant ($t = 1.44[30]$, $p > .10$). This means that both groups of subjects were satisfactorily homogenous in terms of age.

Results obtained by both of Roring (Roring & Charness, 2007) and Charness (1981) suggest that the level of play, as measured by rating, not age, determines the value of chess moves. In designing our experiment, we took into consideration the question of the relationship of age to solving chess problems known from the literature, and especially the method of compensating for the constraints of the elderly that may occur with age (Mireles & Charness, 2002). One of these limitations is the natural proneness to delayed reaction time in aged subjects. Thus in order to better control the above factor we decided to give as much time as 30 minutes to solve the experimental chess task.

On the basis of pilot studies we have found that the most prominent representatives of the Polish national team set the complex chess position for the study of chess strategies by thinking aloud with great accuracy. Figure 4 shows the position that was the basis of tests carried out individually with every chess player as our subject, developed for our experimental purposes.

Figure 4. The chess position presented to the chess players in the experiment.



In the position shown in Figure 4, it is not possible to take the rook on h8 since after 1.Qxh8 Rb5 white soon gets mated, for example, 2.b4 Rxb4+ 3.axb4 Qxb4+ 4.Ka1 Qa3+ 5.Kb1 Qa2#.

White also probably loses after the peaceful 1.Qxb7+ Qxb7 2.Bxb7 Rb8 3.Bc6 Rb6 4.Be4 Rab5 with decisive threats along the b-file. After 1.Qe5 c4 the counter-attack down the f-file gives White some chances: 1.Qe5 c4 2.Rxf7+ Kxf7 3.g6+ hxg6 4.Bxg6+ Ke7 5.Qxg7+ Kd6 6.Qxh8 c3 7.Qb8+ Ke7 8.Qe8 but Black is better. But the immediate 1.Rf7 equalizes! Counterattack down the “f” line! According to the experts from the Polish national team, this position is a good example of a complex situation in chess practice.

Since the thinking aloud method was to provide important data on the strategy of chess thinking, the reactions of the subjects needed to be reliable. The value of this method depends largely on whether the subjects are able to freely and reliably verbalize their thoughts. Therefore, before proceeding with the experiments, the experimenter had to train the chess players to solve problems on the board, so that they would be able to simultaneously verbalize their process of thinking.

Efforts were made, however, in such a training not to teach a specific way of solving problems on the board, which might be subconsciously internalized! One could, for example, unwittingly suggest an empirical approach to solve the problem, one characterized by extensive tree searches, a multitude of variations counted, etc. Here are some of the suggestive questions which should be avoided:

- “Which opportunities are you now checking?” (This question suggests that the subject has given some variations, even if he or she currently is not checking any.)
- “Are you checking any options?” (The subject may suspect that this must be done in order for the study came out well, according to the thinking of the experimenter.)

In the former training an experimenter seeks only to draw the chess player’s attention to accurately reporting their thoughts and avoiding interruptions in speech (saying “Go on”). Each session was discussed, noting existing gaps or inconsistencies in their coverage. Training was terminated when the respondent was no longer able to manage the hassle of verbalizing the thinking process.

The above-described training sought to minimize the influences of various confounding variables – characteristic of the method of thinking aloud – and at the same time increase the compatibility of the subject's interior problem space with his or her protocol space (cf. Lindsay & Norman, 1972). After training, the substantial stage of the experiment started. The subject was familiarized with the test conditions and then exposed to the experimental position. The subject was requested to make a move on the chessboard and at the same time to relate his or her thoughts out loud. Then, the experimenter, who played as Black, responded to the subject's move, and the subject again considered the next move while thinking aloud. After selecting the second move, this situation was repeated again after the third chess player's move on the chessboard. The thinking aloud protocols obtained in such conditions were regarded as an empirical basis for identifying verbal behavior to define each individual chess player's relevant indicators of chess thinking structure.

RESULTS

In this section, outcomes concerning the issues articulated in the hypotheses above will be presented as follows:

1. Thinking strategies of chess players of various playing strengths – measured by their chess rating.
2 & 3. Analysis of thinking indicators when selecting the first, second and third chess move.

In accordance with research outcomes by de Groot (1965) we should not notice statistically significant differences in thinking aloud indicators determined by the chess players' skill level. We can only expect that the stronger chess players will choose stronger moves. Meanwhile, the observation of ratings and indicators of thinking aloud has several important covariants, which corresponds with our Hypothesis 1 (see Table 2).

Table 2. Correlations between ratings and indicators.

		N	A	nn	M	Dmax	Pser	Pmax	s1
1st move	*r*	**.597***	–.133	.136	**.614***	.494	**.532***	.494	**.601***
	p	**.015**	.623	.617	**.011**	.052	**.034**	.052	**.014**
2nd move	*r*	–.015	–.231	–.156	–.026	–.035	.114	.116	–.097
	p	.956	.390	.563	.925	.898	.675	.670	.721
3rd move	*r*	.205	–.156	–.048	.213	–.011	.252	.229	.159
	p	.447	.565	.861	.429	.967	.346	.393	.556

* $p \le .05$, ** $p \le .01$, *** $p \le .001$.

The first remark from analysis of the data collected in Table 2 is that the value of correlation between the chess players' ratings and their thinking aloud indicators in a complex chess situation depends on the order of the move in an experimental situation. Before the first move, half of the indicators (i.e. N, M, Pser, T) reached high correlations with the chess players' ratings (their *r* coefficients ranged from .536 to .614), which are statistically significant. The coefficients of two other indicators (i.e. Dmax and Pmax) also reached a meaningful level $r = .494$ which is $p = .052$. However, similar correlational patterns are not observed at all before the 2nd and the 3rd move in our experiment. Hence, our results are congruent with Campitelli and Gobet's findings, found a strong skill effect in depth of search, rate of search, and number of nodes generated (Campitelli & Gobet, 2004; Gobet, 2019).

If, in turn, we consider the correlation sign in our analysis, it is noticeable that before the first move on a chessboard has occurred, only one negative correlation occurred (for the indicator A), before the second move six to eight indices obtained a negative correlation with a chess players' rating, and before the third move, only three indicators (A, nn, Dmax) received negative correlations.

Now we will explore the Hypothesis 2, inspired by de Groot (1965), that chess players who choose good vs. bad solutions differ in using their thinking strategies. Therefore, we compare the particular chess thinking indicators of the chess players who chose a good solution for the first move on the experimental chessboard – as compared with those who chose a bad one. The outcomes obtained of statistical analysis of the average values of the considered indicators and the appropriate Mann–Whitney *U* tests are collected in Table 3.

Table 3. Indicators of chess thinking strategies in the chess players who chose a valuable or not valuable move on the chessboard (i.e. the first move on the experimental chessboard position).

Outcomes	V0 (n = 3)		V1 (N = 13)		Mann–Whitney *U* test	
Indicators	*M*	*SD*	*M*	*SD*	*z*	*p*
N	53.77	29.791	83.00	22.650	–1.683	.092
A	5.08	1.553	6.00	1.000	–0.968	.333
Nn	13.23	8.136	13.00	6.245	–0.270	.787
M	157.85	95.768	256.33	64.810	–1.683	.092
Dmax	14.31	4.871	17.00	3.606	–1.082	.279
Pser	20.77	13.486	30.67	4.041	–0.810	.418
Pmax	29.77	21.005	39.00	15.588	–0.875	.382
T	794.46	439.212	1001.67	193.143	–0.740	.459

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

The outcomes obtained of statistical analysis of the average values of the considered indicators and the appropriate Mann-Whitney *U* tests. The outcomes evidently show that there are no statistically significant differences in any indicator, where $p \leq .05$. The largest differences are reached in the two indicators N and M, although they only have a level of statistical significance of $p = .092$.

The third of our hypotheses (H3) states that chess thinking strategies for choosing the first move differ materially from the anticipatory thinking strategies for selection of successive moves. In order to explore this hypothesis we have tested the differences between the means of the particular loud thinking indicators with respect to the 1st vs. 2nd, 2nd vs. 3rd, and 1st vs. 3rd chessboard moves. The related outcomes of our analysis are collected in appropriate tables (see Tables 4 and 6).

Table 4. Differences between the means of the indicators with respect to the 1st vs. 2nd chessboard moves, $N = 16$.

Indicators	1st move		2nd move		Mann–Whitney *U* test	
	M	*SD*	*M*	*SD*	*z*	*p*
N	59.25	30.286	13.50	14.642	–4.001	.001
A	5.25	1.483	1.88	1.088	–4.516	.001
Nn	13.19	7.626	2.75	3.992	–4.296	.001
M	176.31	97.332	34.31	38.122	–4.038	.001
Dmax	14.81	4.679	6.75	6.234	–3.126	.002
Pser	22.63	12.790	8.81	10.710	–3.103	.002
Pmax	31.50	19.980	10.31	11.435	–3.303	.001
T	833.31	407.770	155.69	168.355	–4.318	.001

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

From Table 4 we can see that there are very clear statistically significant discrepancies in all of the thinking aloud indicators of chess players between the first and second moves on the experimental chessboard. The level of significance between the particular indicator ranges from $p = .001$ to $p = .002$, which means that all indicators differentiate very highly between the first and the second move in our chess-playing experiment.

Table 5. Differences between the means of the indicators with respect to the 2nd vs. 3rd chessboard moves, $N = 16$.

Indicators	2st move		3rd move		Mann–Whitney *U* Test	
	M	*SD*	*M*	*SD*	*z*	*p*
N	13.50	14.642	9.88	12.355	–0.095	.924
A	1.88	1.088	2.06	0.998	–0.639	.523
nn	2.75	3.992	2.50	2.875	–0.355	.723
M	34.31	38.122	22.81	30.341	–0.019	.985
Dmax	6.75	6.234	5.00	3.670	–0.441	.659
Pser	8.81	10.710	4.31	4.438	–0.430	.667
Pmax	10.31	11.435	5.69	6.140	–0.418	.676
T	155.69	168.355	105.00	122.794	–0.340	.734

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

The Mann–Whitney *U* test outcomes from Table 5, in turn, evidently point out that there are no statistically significant differences between the mean values of any indicators, where the level of *p* is not higher than .523. This means that the discussed indicators do not discriminate between the second and the third move on the experimental chessboard.

Table 6. Differences between the means of the indicators with respect to 1st vs. 2nd moves; $N = 16$.

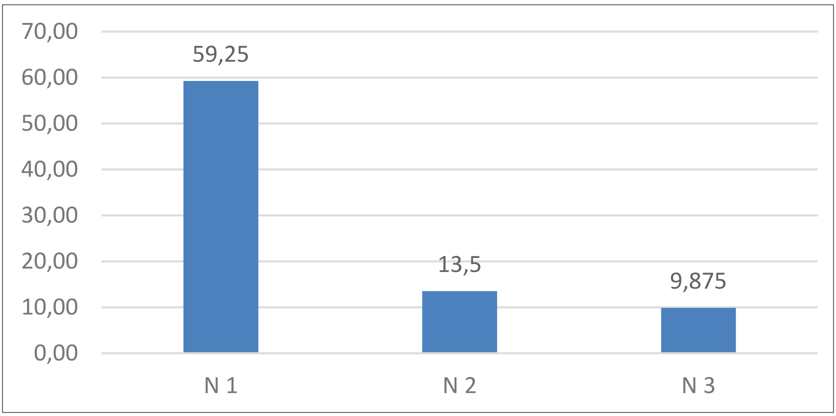
Indicators	1st move		2nd move		Mann–Whitney *U* Test	
	M	*SD*	*M*	*SD*	*z*	*p*
N	59.25	30.29	13.50	14.64	–4.001	.001
A	5.25	1.48	1.88	1.09	–4.516	.001
Nn	13.19	7.63	2.75	3.99	–4.296	.001
M	176.31	97.33	34.31	38.12	–4.038	.001
Dmax	14.81	4.68	6.75	6.23	–3.126	.002
Pser	22.63	12.80	8.81	10.71	–3.103	.002
Pmax	31.50	19.98	10.31	11.43	–3.303	.001
T	833.31	407.77	155.69	168.35	–4.318	.001

*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.

The outcomes of all analyzed indicators collected in Table 6 undoubtedly point to a statistically significant difference between the first move and the third move on the chessboard in our experimental conditions in the sense that the values of the particular indicators are much higher before performing the first move than before the second and the third move as well.

Concluding the analysis of the dynamics of the particular indicators of thinking aloud, we can state that the mean values of these indicators drop monotonically in such a way that the decrease from the 1st to the 2nd and the 3rd move is a kind of rapid falling decrease in all indicator values. As an illustration of such indicator dynamics let us give the distribution of means of the indicator N (see Figure 5).

Figure 5. Dynamics of a typical pattern of thinking aloud indicators (N1 – 1st move, N2 – 2nd move, N3 – 3rd move); N = 16.



After presenting the general regularity in the dynamics of a typical pattern of thinking aloud indicators, we can also ask about any specific differences between the male and female chess players in our experiment. The Mann–Whitney *U* test outcomes show that there are no statistically significant differences between the male and female chess players if the 2nd and 3rd moves are concerned. However, the 1st move differentiated two out of eight indicators in a statistically significant fashion: females have a higher score in the indicator A than males ($p = .040$), while males have a higher indicator Dmax than females ($p = .023$). Three other indicators (M, Pser, Pmax) appeared remarkably higher in the male chess players than in the female ones.

DISCUSSION

Since our research was inspired by de Groot's concept and experiments, regarded today as classic, let's start our discussion of the results of experimental investigations with the source of this inspiration. Our research on chess thinking, on the one hand, confirms some of the conclusions of de Groot's research; on the other hand, however, they indicate many more dimensions which should extend this concept. Let's look first at the results which confirm the discovery of de Groot. First of all, protocol analysis of thinking repeatedly confirmed the phenomenon of progressive deepening, characteristic for thinking in chess. In further studies it would be worthwhile to determine which conditions must be met for a deepening of understanding of the problem in order to be effective, or at least efficient. We illustrate the phenomenon of progressive deepening discovered by de Groot (1965) in Figure 6 in the context of our experimental situation, as a widening and lengthening of the decision tree. This phenomenon is one of the more important ones to be investigated in further research, and particularly, in the context of the threats named "shallow mind" and "digital dementia", to which commentators of research on the digital age draw attention (Carr, 2010; Spitzer, 2012; Desmurget, 2011).

Figure 6. Phenomenon of progressive deepening (bold and italics) discovered by de Groot (1965) and interpreted in our experimental situation.

Rxe4
Rxe4
Rxe4 Nxe4 Qxe4
Rxe4 Nxe4 Qxe4 … Bd3 Re8 Qh7 Kf8
Rxe4 Nxe4 Qxe4 Bf6
           f5 Qe6 Kh7 Bxg7
h3 … g4
   a3
   Rb8
g4 … g5
   hxg4
   Nxg4
   Nxg4 Qxe4
   hxg4
   hxg4 h3
      h3 … hxg4 … Qh3 … Rh1
      h3 g3
         g3 Qxg3
            Qxg3 … ***Rxe4***
                  Rg1
               g6 f5
   a3
g4 Qb6 Ra3
g4 Qb6
   a3 b3
   Qb6 Ra3
g4

The question concerning the conditions for a deeper understanding of the problem in the context of our experimental situation is really addressed by the significance of such indicators as M, N, Dmax, nn. We propose the following interpretation of these indicators:

M – how intensively to explore the reality of a difficult task, so one is not only effective but efficient; this indicator can also inform about the style of data processing (deductive or empirical);

N – how many branches should be added in the decision-making tree to maintain a proper proportion of effort in relation to gain effects;

Dmax – how far to look for insight in order to provide the most important consequences of subsequent actions;

nn – an indicator as to when interference or consideration of other options for action act as a distractor, when they are neutral, and when they can play a positive role, for example, as a kind of review of the entire problem situation.

De Groot (1965) also discussed phase-structure in his analysis of chess thinking, which has a long tradition coming from Dewey (1910), who discovered the structure of problem solving thinking. A direct analysis of thinking aloud of verbal protocols in our three-move chess player's complex problem-solving experimental situation allows us to pick out many examples of such a phase-structure, which de Groot mentions. This is a further confirmation of the fact that the game of chess is an adequate model for the study of problem solving, comparable to the phases and steps discovered in other areas.

Now we can compare our first two hypotheses inspired by the research outcomes by de Groot (1965). As far as Hypothesis 1 is concerned, we found a confirmation of the meaningful correlation between the chess players' ratings and their thinking aloud for six out of eight indicators in the complex chess situation, but only in the first move and not the two subsequent moves. This first move as a reaction to the chessboard experimental position really differentiated chess-playing ratings, but the next two subsequent moves did not. The chess-playing ratings used in our experiment as an interval-type-of-scale (Glickman & Jones, 1999) appeared to differentiate most of the thinking aloud indicators.

Let us now discuss Hypothesis 2, suggested by de Groot (1965). This hypothesis in our research states that chess players who chose a good vs. bad solution differed in their thinking strategies. Unfortunately, our analysis showed that there is no statistically significant difference between the indicators of chess thinking strategies in the chess players who chose a valuable move compared to those who did not choose a valuable move (i.e. the first move on the experimental chessboard position).

So, the hypothesis based on observations of de Groot, which states that the rating of chess players will be positively correlated with the value of the move was not confirmed. Most probably, the chess position selected for the experiment did not have sufficient discrimination power in this respect.

Hypothesis 3 was confirmed in an unquestionable manner. This hypothesis is strictly connected with the structure of our three-move-experiment. It states that chess thinking strategies for choosing the first moves differ materially from the anticipatory thinking strategies for selection of successive moves. The methodology of this experiment assumes that chess thinking strategies are strictly connected with thinking aloud indicators. As expected, these first move indicators differ significantly from the indicators of thinking when choosing successive moves. In the second and third moves all indicators have reduced values. This is the first phase of chess-problem solving thinking, when a chess player structures the main shape of the strategy for the problem solving process. The outcome of this phase in our experiment is the execution of this process in its behavioral form, as the first move on a chessboard. The other two moves seem largely to be a kind of further exploration of the chess-playing situation, information-gathering, in a manner which mainly supports the strategy already formed in the first phase (i.e. before the first move) of thinking.

CONCLUDING REMARKS

The completed experiment and the presented data analysis allow us to make some concluding remarks concerning largely the methodology of experiments on chess-playing thinking. It also offers a perspective on the future development of studies on thinking processes. The main value of our study is that we have shown how it is possible to extend and develop a cognitive perspective to study thinking processes in an experimental situation using a classical approach by de Groot (1965) on chess players' thinking aloud.

Our study pointed out that a three-move experiment is a natural and efficient way to extend the classic methodology of de Groot to explore the cognitive nature of the strategy of human thinking. Thinking aloud protocols with indicators of chess problem solving-thinking when used in a three-move experiment evidently offer a better perspective than the one-move experiment to explore cognitive strategies of human thinking. From the three-move experiment perspective, it is easier to penetrate the dynamics of cognitive processes involved in chess thinking, which are behaviorally revealed in the particular indicators of thinking aloud.

One of the intriguing research problems which might be explored in the three-move experiment is thinking by analogy, which underlies chess-playing mental strategies. The method of thinking aloud, along with subsequent verbal protocol analyses, can develop in the decision-maker a habit of efficient, organized thought consciously applied in various problem solving situations. In thinking by analogy the decision-maker, after training, can initiate a fixed procedure of thought, thereby becoming independent of emotions or other factors that could disturb the thought process (Biela, 1986, 1991). The example might be research on multi-attribute decision support and complexity (Timmermans & Vlek, 1992). Using certain patterns of organized thinking does not at all curb creativity – quite the opposite. Skillfully put to good use, it can foster the discovery of original ideas, as was practically introduced in Przewoznik & Soszynski (2001). We have a lot of possibilities to use the method of indicators as a tool of training in education. For example Gliga and Flesner (2014) demonstrated how chess training has an impact on school performance, memory, sustained attention and creativity. A group of 20 novice primary school students took part in 10 blended learning chess lessons and in a final chess tournament (the chess group). Eighteen control students participated in 10 fun math lessons. Most cognitive skills increased from pre-test to post-test in both groups but the School Performance Test increased significantly more in the chess group. The results obtained by Trinchero and Sala (2016), Sala and Gobet (2016) foster the hypothesis that a specific type of chess training does improve children's mathematical skills, and uphold the idea that teaching general heuristics can be an effective way to promote transfer of learning.

It would also be interesting to experimentally compare the method of indicators and *the chunks theory* (Chase & Simon, 1973; Saariluoma, 2001; Gobet, 2001; Gobet & Clarkson, 2004). According to this theory, the master perceives a position as a sensible, integrated whole and not a sum total of separate black and white pieces. The experience and knowledge of the master enables him or her to integrate and form an opinion about the position very quickly. When forming an opinion, the pieces are seen in more complex combinations of a spatial, functional and dynamic nature. For example, the amateur will notice 6 pieces arranged in rows on the squares: Kg1, Rf1, Bg2, pawns f2, g3, h2, whereas the master will immediately see the integrated whole: "the position with the fianchettoed bishop". Naturally, such an overall glance at positions makes it easier to understand them properly. The comprehension of the strategic essence of certain basic positions is the main task that a chess player must cope with. Therefore, the basic positions with typical solutions are often gathered in theoretical books on chess (Pein & Przewoznik, 1991). By analogy, such *chunking attitude* was observed in the area of medical diagnosis (Reingold & Sheridan, 2011) – eye movement findings indicate that expertise (in chess and radiology) is associated with the ability to process domain related visual information in terms of larger patterns of features rather than isolated features.

At the same time, our experiment verified a practical issue concerning the practice of ratings of chess players in the chess-playing world. We have shown that for six of the nine thinking aloud indicators used, there is a remarkable positive correlation between the values of these indicators and the rating of chess players who took part in our experiment as subjects. It would be probably promising to compare chess strategies (measured by indicators) with cognitive abilities (Burgoyne, Sala, Gobet, Macnamara, Campitelli, & Hambrick, 2016), as results from meta-analysis suggest that cognitive ability contributes meaningfully to individual differences in chess skill, particularly in young chess players and/or at lower levels of skill.

The crucial question is still valid in psychology of chess: What is more important in chess practice: extensive research (nine indicators!) or selective research and pattern recognition (chunking?). Chunking theory has prevailed for years. We hope there are still attractive perspectives for exploring extensive research in nine indicators method.

REFERENCES

Biela, A. (1986). Psychological patterns in predicting disjunction and conjunction of clinical symptoms. *Acta Psychologica*, *61*(3), 183–195.

Biela, A. (1991). *Analogy in Science*. Peter Lang.

Burgoyne, A., Sala, G., Gobet, F., Macnamara, B., Campitelli, G., & Hambrick, D. (2016). The relationship between cognitive ability and chess skill: A comprehensive meta-analyses. *Intelligence*, *59*, 72–83.

Busato, V. (2006a). In Memoriam: Adriaan Dingeman de Groot (1914-2006). *Netherlands Journal of Psychology*, *62*, 2–4. https://doi.org/10.1007/BF03061044

Busato, V. (2006b, November 1). *In Memoriam: Adriaan Dingeman de Groot (1914 – 2006).* Association for Psychological Science. https://www.psychologicalscience.org/observer/in-memoriam-adriaan-dingeman-de-groot-1914-2006

Campitelli, G., & Gobet, F. (2004). Adaptive expert decision making: Skilled Chess Players Search More and Deeper. *ICGA Journal*, *27*(4), 209–216.

Carr, N. (2010). *The shallows: What the internet is doing to our brains*. New York: W.W. Norton & Company.

Charness, N. (1981a). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 467–476. https://doi.org/10.1037/0096-1523.7.2.467

Charness, N. (1981b). Aging and skilled problem solving. *Journal of Experimental Psychology: General*, *110*(1), 21–38.

Charness, N. (1988). The role of theories of cognitive aging: Comment on Salthouse. *Psychology and Aging*, *3*(1), 17–21.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81.

De Groot, A. D. (1965). *Thought and choice in chess*. Mouton & Co.

De Groot, A. D. (1981). Thought and choice in chess: An overview of a study based on Selzean Theory. In N. Frijda & A. D. de Groot (Eds.), *Otto Selz: His contribution to psychology* (pp. 192–255). Mouton.

Desmurget, M. (2011). *TV Lobotomie*. Editions Max Milo.

Dewey, J. (1910). *How we think*. D. C. Heath & Co., Publishers.

Glickman M., Jones, A. (1999). Rating the chess rating system. *Chance*, *12*, 21–28.

Gliga, F., & Flesner, P. I. (2014). Cognitive benefits of chess training in novice children. *Procedia-Social and Behavioral Sciences*, *116*, 962–967. https://doi.org/10.1016/j.sbspro.2014.01.328

Gobet, F. (2001). Chunk hierarchies and retrieval structures: Comments on Saariluoma and Laine. *Scandinavian Journal of Psychology*, *42*(2), 149–155. https://doi.org/10.1111/1467-9450.00225

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four… or is it two? *Memory*, *12*(6), 732–747. 10.1080/09658210344000530

Gobet, F. (2019). *The psychology of chess. Fernand Gobet*. Routledge.

Lindsay, P., & Norman, D. (1972). *Human information processing. An introduction to psychology*. Academic Press.

Mireles, D., & Charness, N. (2002). Computational explorations of the influence of structured knowledge on age-related cognitive decline. *Psychology and Aging*, *17*(2), 245–259. https://doi.org/10.1037/0882-7974.17.2.245

Przewoźnik, J. (2014). *Thinking strategies in complex chess positions.* Wydawnictwo KUL.

Przewoznik, J. (2019). *Nine indicators to improve your chess* (Parts 1 and 2). Chess Base 18. https://en.chessbase.com/post/nine-indicators-method-part-1; https://en.chessbase.com/post/nine-indicators-method-part-2

Przewoznik, J., & Pein, M. (1991). *The Blumenfeld Gambit*. Pergamon Chess.

Przewoznik, J., & Soszynski, M. (2001). *How to think in chess*. Russell Enterprises.

Reingold, E. M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford Handbook on Eye Movements* (pp. 528–550). Oxford University Press.

Roring, R., & Charness, N. (2007). A multilevel model analysis of expertise in chess across the life span. *Psychology and Aging*, *22*(2), 291–299.

Saariluoma, P. (2001). Chess and content-oriented psychology of thinking. *Psichologica*, *22*, 143–164.

Sadler, M., & Regan, N. (2019). *Game changer. AlphaZero's groundbreaking chess strategies and the promise of AI.* New in Chess.

Sala, G., & Gobet, F. (2016). Do the benefits of chess instruction transfer to academic and cognitive skills? A meta-analysis. *Educational Research Review*, *18*, 46–57.

Spitzer, M. (2012). *Digitale Demenz: Wie wir uns und unsere Kinder um den Verstand bringen*. Droemer Verlag.

Timmermans, D., & Vlek, Ch. (1992). Multi-attribute decision support and complexity: An evaluation and process analysis of aided versus unaided decision making. *Acta Psychologica*, *80*(1–3), 49–65.

Trinchero, R., & Sala, G. (2016). Chess training and mathematical problem-solving: The role of teaching heuristics in transfer of learning. *Eurasial Journal of Mathematics, Science & Technology Education*, *12*(3), 655–668.