Gobet, F. (1998). Chess players' thinking revisited. *Swiss Journal of Psychology*, 57, 18-32

Chess Players' Thinking Revisited

Fernand Gobet

ESRC Centre for Research in Development,
Instruction and Training
Department of Psychology
University of Nottingham

Nottingham NG7 2RD
England

Phone: (0115) 951 5402
Fax: (0115) 951 5324
frg@psyc.nott.ac.uk

Running head: Chess Thinking

Author Notes

This research was supported by the Swiss National Science Foundation, Grant No. 8210-30606, and by the National Science Foundation (USA), Grant No. DBS-9121027. I am grateful to Neil Charness, Howard Richman, Frank Ritter, Pertti Saariluoma, Herb Simon, and anonymous reviewers for helpful comments on earlier drafts.

Correspondence concerning this article should be addressed to Fernand Gobet, ESRC Centre for Research in Development, Instruction and Training, Department of Psychology, University of Nottingham, Nottingham NG7 2RD, England.

Abstract

The main result of De Groot's ([1946] 1978) classical study of chessplayers' thinking was that players of various levels of skill do not differ in the macrostructure of their thought process (in particular with respect to the depth of search and to the number of nodes investigated). Recently, Holding (1985, 1992) challenged these results and proposed that there are skill differences in the way players explore the problem space. The present study replicates De Groot's (1978) problem solving experiment. Results show that Masters differ from weak players in more ways than found in the original study. Some of the differences support search models of chess thinking, and others pattern recognition models. The theoretical discussion suggests that the usual distinction between search and pattern recognition models of chess thinking is unwarranted, and proposes a way of reconciling the two approaches.

Keywords

chess, chunking, decision making, expertise, pattern recognition, search, thinking, verbal protocol

Chess Players' Thinking Revisited

What is the key to expertise? Over the years, psychologists have proposed two main explanations: ability to access a rich knowledge database through pattern recognition, and ability to search through the problem space. While no researcher would stress the importance of one of these explanations to the exclusion of the other, the relative importance given to knowledge and search vary in current theories of skilled behavior.

This tension between pattern recognition and search is clearly apparent in research on chess, a domain that has spawned numerous studies, and whose results have been shown to generalize well to other types of expertise. Chess offers several advantages as a domain of research (Gobet, 1993), including rich and ecologically valid environment, quantitative measurement scale of skill, large database of games, and cross-fertilization with research in artificial intelligence.

Basing their inquiry on De Groot's ([1946] 1978) seminal study, Simon and his colleagues (Chase & Simon, 1973; Newell & Simon, 1972) have given the most emphasis to selective search, to knowledge possessed by chessplayers and to perception and memory mechanisms that allow them to rapidly access useful information. They proposed that recognition processes allow search to be cut down typically to less than a hundred nodes and that search does not differ critically among skill levels.

Evidence for this position, which is known as the chunking model or as the pattern recognition theory, converges from several directions. First, De Groot's (1978) data show that most features in the macrostructure of search (including the number of nodes visited and the depth of search) do not differ between top level players and amateurs. Second, data from speed chess (Calderwood, Klein & Crandall, 1988) and simultaneous chess (Gobet & Simon, 1996a), show that strict limitations in thinking time do not impair expert performance much, as should be the case if search **were** the key element of chess skill. Third, chess masters are highly selective and direct their attention rapidly to good moves (De Groot, 1978; Klein & Peio, 1989). De Groot (1978) demonstrated that even chess Grandmasters seldom look at more than 100 possible continuations of the game before choosing a move. Fourth, eye movement studies show that during the five-second exposure of a chess position, Masters and novices differ on several dimensions, such as the mean and standard deviation of fixation durations and the number of squares fixated (De Groot & Gobet, 1996). In particular, Masters fixate more often squares that are important from a chess point of view. As retrospective protocols indicate that very little search is done during these five seconds, these differences suggest that perceptual pattern recognition processes allow Masters to fixate relevant squares more often.

Chase and Simon's (1973) chunking theory, where recognition of known patterns plays a key role, has been shown to apply relatively successfully in several other domains of expertise (Charness, 1992). Its main weakness is the assumption, contrary to empirical evidence (Holding, 1985), that transfer from short-term memory to long-term memory is slow (about 8 s per chunk) even with experts. A revision of the chunking theory (Gobet & Simon, 1996b, in press) has removed this deficiency. In the conclusion of this paper, I will discuss how this theory of memory may apply to problem solving.

Recently, Holding (1985, 1992) argued that the role of pattern recognition was over-emphasized and the role of quantitative search (number of nodes visited) underplayed. Holding proposed three key features of chess expertise: search, evaluation of positions, and knowledge. Note that these elements are not at variance with what the chunking model proposes. For example, both approaches recognize the role of knowledge, and both predict, as was found in empirical research (Holding, 1989; Holding & Pfau, 1985) that strong chess players evaluate positions better, not only when the evaluation applies to a position on the board, but also when it applies to a position anticipated during search. It is the relative importance given to search that differentiates the two approaches. I will refer to Holding's model and similar models giving emphasis to look-ahead search, such as models based on current chess computers, as <u>search models</u>.

Holding's main line of argumentation is that, contrary to what was suggested by De Groot (1978), amount of search is a function of chess expertise—strong players search deeper than weak players. With respect to De Groot's (1978) finding that top-level Grandmasters do not search reliably deeper than amateurs, Holding argues that experimental power may have been too low in this experiment to detect existing differences. Holding also brings forward recent data (Charness, 1981; Holding & Reynolds, 1982), which show that there is some difference in depth of search between weak and expert players. For example, Charness' (1981) data show a small linear relation between Elo points[1] and average depth of search: the search increases by about 0.5 ply (a move for White <u>or</u> Black) for each standard deviation of skill (200 Elo points). Note that in this study, as in Holding and Reynolds' study, the best players were at best Experts, and therefore clearly weaker than De Groot's (1978) Grandmasters, who were world-class level players. To reconcile his results with De Groot's, Charness' (1981) has proposed that depth of search may not be linearly related to skill, but that there is a ceiling at high skill levels, possibly because search algorithms become uniform. Data collected by Saariluoma (1990) suggest that International Masters and Grandmasters sometimes search less than Master players. In (tactical) positions with a 10-minute limit for finding a move, both the total number of nodes searched and the mean depth of search show an inverted U-curve function of skill, with Masters (around 2200 Elo) searching the largest number of nodes (52) and at the largest average depth (5.1 moves). By comparison, Saariluoma's International master and Grandmaster searched, on average, through a space of 23 nodes with an average depth of 3.6 moves.

The relative role of search in chess expertise is theoretically important, well beyond the realm of chess. Do decision-makers rely more on analyzing various alternatives, or on recognizing familiar patterns in the situation? How do these two processes interact? Should the training of future experts—from physicians to computer scientists—lay most emphasis on analytic skills or on building up a huge knowledge database and an automatic access to it? Even though each domain of expertise may have idiosyncratic properties, research on chess may help identify some of the potential conditions under which search, pattern recognition, or some combination of both, may be the best way to cope with the complexities of the environment.

It is therefore important to understand the role of search in chess expertise. Unfortunately, recent empirical data are scarce about chess players' thinking, and no direct replication of De Groot's study is available, in spite of its strong impact in cognitive psychology (Charness, 1992). Newell and Simon (1972) as well as Wagner and Scurrah (1971) used only a handful of subjects. Gruber (1991) had only two skill levels, comparing novices to Experts. Charness (1981), the largest recent source of chess problem solving data, used positions different from the ones used by De Groot (1978), and his experimental procedure differs somewhat, in particular in limiting thinking time to 10 minutes, which may affect variables such as depth of search. Because recent studies have used positions different from the ones used by De Groot, it could be argued that the differences found in depth of search are specific to the type of positions used. Although De Groot (1978, p. 122 ff.) has suggested that most of the statistics he used were relatively stable from one position to another, Charness (1981) has found important differences in some of the variables used in his analyses.

As a consequence of the current theoretical discussion about the role of search, of the importance of De Groot's results and of their lack of replication, I decided to submit data gathered for another purpose to a secondary analysis. This permits replication, with a larger number of subjects, of a subset of De Groot's (1978) seminal study. The goal was to see whether De Groot's results are robust, in particular with respect to the passage of time.

The replication of De Groot's experiment described in this paper was carried out in 1986. The experiment served as a post-test in a study aimed at understanding the role of controllability (Seligman, 1975) in chess players (Gobet, 1992; Gobet & Retschitzki, 1991), where controllability was defined as the degree to which subjects see a correlation between their actions and the outcomes in the environment. Before being confronted with De Groot's task, subjects were assigned to three experimental groups (normal feedback group, manipulated feedback group, control group) according to the type of controllability to which they were exposed. As this manipulation of controllability did not significantly affect any variable that will be discussed later, the data of the three groups will be pooled in this paper.

Method

Subjects

Fifty-one Swiss male chess players participated in this experiment. Three subjects who knew the position "A" of De Groot (see Figure 1), were discarded. The age of the remaining 48 subjects (thereafter, the "Swiss sample") ranged from 18 to 33, with a mean of 25.5 years and a standard deviation of 4.5 years. At the time of the study, four players (all rated above 2400 Elo) had the title of International Masters, and eight belonged to the "extended" Swiss national team. Players were assigned to four skill levels according to their playing strength: level I (Masters; from 2200 to 2450 Elo; mean Elo: 2317), level II (Experts; 2000-2200 Elo; mean Elo: 2101), level III (class A players; 1800-2000 Elo; mean Elo: 1903) and level IV (Class B players; 1600-1800 Elo; mean Elo: 1699). The respective means of age, 27, 26.3, 25.2 and 23.8 years, did not differ statistically across skill levels. Each level consisted of 12 players.

Insert Figure 1 about here

Materials

A competition chess clock informed players about the time elapsed. The position "A" (see Figure 1) of De Groot (1978) was presented to subjects using a standard chess board and chess pieces. A detailed analysis of this position is given by De Groot (1978, pp. 89-90). It was decided to collect the thinking aloud protocols with De Groot's position "A" only, because most of De Groot's results were gathered with this position.

Design and Procedure

As part of the study on the effects of controllability, all subjects received, in order: (a) a short computer-taught instruction on the way to handle positions containing an "isolated Queen's Pawn" (an important strategic feature of chess strategy) and (b) a series of quizzes (presented for 30 seconds each), where subjects had to choose between two proposed moves (see Gobet, 1992, for the detail of these tasks). On the basis of the comments given by subjects after the experiment, it is unlikely that these tasks modified subjects' ways of thinking. Moreover, as noted above, the manipulation on controllability did not yield any effect on the variables measured in this experiment.

Subjects were tested individually. The instruction was to try to find the best move for White, without moving the pieces, as in a competition game. Subjects were asked to think aloud (in their native language, French or German), and were audio-taped. Their thinking time was limited to 30 minutes (none of De Groot's subjects used more than 28 minutes). The experimental instruction was a French or German translation of De Groot's instruction. The experiment ended with the execution of the chosen move on the board.

The verbal protocols were transcribed and Problem Behavior Graphs (Newell and Simon, 1972) were constructed from them. Protocol analysis used the following descriptive variables, chosen both because of their theoretical interest and their availability from De Groot's book: (a) Quality of the chosen move; based on De Groot's and the author's analysis of the position, moves were given a value from 5 (winning move) to 0 (losing move);

(b) Total time to choose a move; (c) Number of different base moves (base moves are the moves immediately playable in the stimulus position (depth 1)); (d) Rate of generating different base moves per minute (this variable is obtained by dividing the number of different base moves by the total time); (e) Number of episodes (an episode is defined as a sequence of moves generated from a base move); (f) Number of positions (nodes) mentioned during the search; (g) Rate of generating nodes per minute (this variable is obtained by dividing the number of nodes by the total time); (h) Maximal and mean depths (both are expressed in plies (i.e. moves for White or Black)); (i) Duration of the first phase (this phase is the orientation period where the player makes a rough evaluation of the position (without search) and notes the possible plans, threats, and base moves);[2] and (j) Number of base moves reinvestigated. Reinvestigations are divided up into two types: Immediate reinvestigations (IR; the same base move is analyzed in the next episode) and non-immediate reinvestigations (NIR; at least one different move is taken up between the analysis of a base move and its reinvestigation). With IR and NIR, the largest number of times a move is (re)investigated was singled out, for each player. The reader is referred to the Appendix for an example of the way these variables are extracted from protocols (see also De Groot, 1978, pp. 119 ff., and Charness, 1981).

Both search and pattern recognition models (in their pure form) predict that strong players choose better moves than weak players, need less time to reach a decision, and generate moves faster during search. Search models predict that strong players search substantially more and deeper, while pattern recognition models do not predict any large difference for these variables. Finally, pattern recognition models predict differences in variables related to selectivity: because strong players identify good moves more rapidly, they should, on average, mention fewer base moves, reinvestigate the same move more often and jump less often between different moves. They also predict that strong players have a shorter first phase. Although Holding's model is not precise enough to make quantitative predictions of these variables, it certainly suggests, given its lack of emphasis on selectivity and pattern recognition, that players do not differ much in these variables.

Results

Comparisons will be made with De Groot's results at two levels: relative difference between groups and absolute values of the variables. First, the different skill levels of this study's sample will be compared with respect to several structural variables in order to see whether there is any difference between them. Next, these skill differences will be compared with those found by De Groot. Then, the absolute values of the variables found in the Swiss sample will be compared with De Groot's. Finally, I will discuss the implications of the results for theoretical approaches based on either pattern recognition or search.

Table 1 gives an overview of the results, with De Groot's data also mentioned for easy comparison.[3] De Groot's Masters (M) and Experts (E) correspond roughly to Masters and Experts of the present study, respectively. De Groot's class players ranged from Class A to Class C players, and may roughly be compared to the Swiss Class A and B players together. Note that both samples show a large variability, a question that will be addressed in the discussion section.

Insert Table 1 about here

Swiss sample

Quality of Chosen Move.

The best move, 1.Ba2xd5, which gives White a winning position, appears 15 times (in about one third of all moves proposed; it also appeared about one third of the time in De Groot's data). The second best move, 1.Ne5xc6, which gives White a solid edge, appears only 3 times (6%), while 21% of De Groot's (1978) subjects chose this move. Two subjects proposed very bad moves, leading to a losing position for White (1.Nc3-a4 and 1.Ne5xf7). As expected, the quality of the chosen moves differs as a function of skill [$F(3,44) = 8.06$, $MS_e = 1.57$, $p < .001$]. Pairwise comparisons with HSD Tukey test show that Masters differ reliably ($p < .001$) from class A and class B players while the other comparisons do not yield significant differences.

Total Time.

Although Masters tend to be faster (11.3 minutes, on average, vs. 16.7 minutes for the others levels pooled), the difference is not significant statistically [$F(3,44)=1.78$, $MS_e = 64.09$, ns].

Number of Nodes.

From Masters down to Class B players, the average number of nodes visited during search is 58.0, 58.3, 56.8 and 33.9. The differences are not statistically significant [$F(3, 44) = 1.11$, $MS_e = 1536.6$, ns]. The maximal number of nodes (177 nodes) was searched by a Master, and the minimal (4 nodes) by a Class A player.

Rate of Generating Nodes.

Although Masters and Experts generate more nodes per minute (respectively 4.8 and 4.1) than Class A and Class B players (respectively 3.2 and 3.4), the differences are not statistically significant [$F(3, 44) = 0.49$, $MS_e = 12.9$, ns]. Only two subjects generated more than eight nodes per minute.

Maximal and Mean Depth.

There is no statistically significant difference between the skill levels for the maximal depth of search [$F(3,44)=1.3$, $MS_e = 19.79$, ns]. In particular, this variable is not reliably larger for Masters than for players from other skill levels: average maximal depth of masters = 9.1 plies ($sd = 3.8$ plies); average maximal depth of the other skill levels pooled = 8 plies ($sd = 4.7$ plies). The deepest line (23 plies) was searched by a Class A player—the statistical results presented in this section are essentially the same when this outlier is removed—and the deepest line for Masters was 14 plies. Note that class B players calculate at the least maximal depth (on average, 6 plies).

There is an effect of Skill for the mean depth of search [$F(3,44)=2.9$, $MS_e = 3.68$, $p<.05$]. The mean depths for Masters, Experts, Class A and Class B players are, respectively, 5.0, 4.6, 3.7 and 2.9 plies. Tukey HSD test indicates a significant difference only between Masters and Class B ($p< .05$). Predicting mean depth from Elo yields the following regression line: MeanDepth = -2.638 + 0.003 * Elo (the coefficient for Elo is significant at the .005 level). This equation indicates that mean depth of search increases linearly about 0.6 ply for each standard deviation (200 Elo points). Interestingly, this linear gain in mean depth of search as a function of Elo is close to the 0.5 ply found by Charness (1981), whose sample ranged from 1284 to 2004 Elo. Note finally that, regressed against choice of move, mean depth of search accounts for about 19% of the variance.

Number of Base Moves.

Masters analyze the least number of different base moves (on average, 3.2 moves) and Class A players analyze the largest number (on average, 6.5 base moves). 58% of Masters analyzed between 1 and 3 base moves, while 41% of Class A players investigated between 9 and 11 base moves. Finally, Experts and class B players analyze on average 4.8 base moves. ANOVA shows the differences to be significant: $F(3,44) = 2.94$, $MS_e = 7.19$, $p<.05$. Tukey HSD test indicates that the significant difference is between Masters and class A players ($p<.025$).

Rate of Generating Base Moves.

The four groups do not differ for this variable [$F(3,44) = 0.08$, $MS_e = 0.06$, ns]. The average rate of generating a base move is 0.38 per minute, with a standard deviation of 0.24 move per minute.

Number of Episodes.

For this variable, there is a steady increase from Masters to Class A players (from 9.2 to 12.4 episodes), then a sharp drop for Class B players (8.6 episodes). The groups do not, however, differ significantly [$F(3,44) = 0.91$, $MS_e = 37.65$, ns].

Duration of the First Phase.

Results show a decrease in the duration of the first phase as a function of skill (Means: Master, 1.1 minutes; Experts, 2.3 m; Class A, 1.6 m; Class B, 4.2 m). ANOVA indicates a significant effect [$F(3,44) = 4.22$, $MS_e = 5.37$, $p=.01$]. Tukey HSD test indicates that Class B players differ significantly from Masters ($p<.01$) and Class A players ($p<.05$).

Number of Reinvestigations.

Although class B players tend to reinvestigate the same base move less often than the other players, Skill does not reliably affect the number of reinvestigations [$F(3,44)=0.67$, ns]. Table 2 summarizes the results for the number of reinvestigations and its sub-categories.

Insert Table 2 about here

ANOVA of the number of Immediate Reinvestigations (IR) indicates a marginal effect of Skill [$F(3,44) = 2.56$, $MS_e = 9.57$, $p < .07$]. Class B players tend to reinvestigate immediately the same base move less often than Masters ($p < .07$). There is no statistically significant difference for the number of Non-immediate Reinvestigations (NIR). Note that Masters produce very few NIR.

Analysis of the maximal number of IR and NIR are consistent with the previous results. The maximal number of IR is proportional to the strength of the players [$F(3,44) = 5.91$, $MS_e = 3.54$, $p < .005$], while the maximal number of NIR is inversely proportional to the strength of the players [$F(3,44) = 3.71$, $MS_e = 0.973$, $p < .02$]. With the maximal number of IR, Tukey test indicates that Masters differ reliably both from Class A and B players, while with the maximal number of NIR, Masters differ reliably only from Class A players.

In summary, it was found that there are (small) skill differences for the quality of the chosen move, the number of base moves, the mean depth of search, the duration of the first phase, the maximal number of immediate and non-immediate reinvestigations. A marginal difference was found for the number of immediate reinvestigations. The other variables did not differ across skill levels. Note that in all these variables showing differences, these differences were between Masters and either Class A or Class B players. In only one case (Time spent for the first phase) did Class A and Class B players differ reliably. In no case did Experts differ significantly from Masters or from either Class A or Class B players.

How much do the variables that have been described predict the quality of the chosen move? A stepwise regression with all the variables discussed in the result section keeps only three variables: Total time, mean depth, and number of maximal reinvestigations. Used in a multiple regression, these variables yield the following equation:

Move = 2.429 - 0.001 * TotalTime + 0.304 MeanDepth + 0.188 * MaximalNumberRI

TotalTime and MeanDepth are statistically significant at the .01 level, while MaximalNumberRI is only marginally significant ($p < .07$). This multiple regression accounts for 35.1% of the variance in predicting the choice of move, which is relatively little, but still more than Elo rating, which surprisingly accounts for only 29.2% ($p < .001$) of the variance. In comparison, Elo rating accounts for much more of the variance in memory tasks where chess positions taken from Master games are presented for 5 seconds. For example, in the sample of 25 subjects ranging from 1680 to 2590 Elo of Gobet and Simon (1996c), Elo rating accounts for 72.3% of the variance in recall percentage. The low power of Elo in predicting the move chosen is due in part to the high variability of results.

Predicting Quality of Move after Partialling Out Search Variables

This experiment does not offer a direct measure of pattern recognition abilities, hence it is not possible to directly study the interaction between search and pattern recognition. As an approximation, one can use the indirect approach of analyzing to what extent Elo rating, which measures chess skill, still predicts Quality of Move after the variables related to search have been partialled out. We saw that Elo rating accounted for 29.2% of the variance for Quality of Move. When one partials out the variables best characterizing search, namely, Mean Depth of Search, Maximal Depth of Search, and Number of Nodes, Elo still accounts for 17.6% ($p < .005$) of the variance. This result indicates that search alone does not account for the quality of the move chosen, and that other factors, probably including pattern recognition, play an important role.

Comparisons with De Groot's Results

A few qualifications are required as to the feasibility of comparing De Groot's results with the Swiss results. First, there are slight differences in the way protocols were recorded (mainly: different languages; tape recorder vs. hand recording). Second, it is difficult to compare directly the skill level groups of the two studies, as De Groot's subjects were not rated according to the (then non-existent) Elo system. As a first approximation, one can use data of Elo (1978, p. 175 ff.), who has retroactively estimated the strength of strong players of the past. According to this source, de Groot's Grandmasters G1, G2, G3, G4, G5, and G6, all world-class players, had an Elo of 2670, 2690, 2620, 2660, 2650, and 2560, respectively, during their best five years. His Masters M1, M2, M3 and M4 had an best 5-year Elo of 2480, 2460, 2480 and 2440, which places them about 100 Elo points above the mean of our Master sample. Third, I did not have access to world class players as De Groot did—nowadays, world class players are simply beyond researchers' financial means. Fourth, De Groot collected very few protocols with Masters and Class players, some of them with variations in the procedure that make them non-usable for comparison. Finally, as the Swiss sample is larger, the statistical tests have more power in this sample than in De Groot's.

The ideal way of comparing the results described in this paper with De Groot's would be to conduct a meta-analysis over the two studies. Unfortunately, this is not feasible, for the reasons mentioned above. In particular, the present study does not have a group comparable to De Groot's world-class Grandmaster group, and De Groot (1978) does not give detailed data for Masters and Class players. For these two groups, I have computed statistics from the protocols given in Appendix II of De Groot (see Table 1). However, this appendix gives only three protocols of Masters, of which one (M1) cannot be used because of differences in the experimental procedure (De Groot, 1978, p. 412), and only two protocols for Class players. With so few observations, it seems unreasonable to apply meta-analytic tools.

Skill differences.

De Groot (1978, p. 319) was mainly interested in high levels of expertise, and focused his attention on comparisons between the Grandmaster and Expert groups (a difference in skill of about 2 standard deviations). His major finding was that the macrostructure of protocols differ little across skill levels—at least with players having the minimal proficiency of Experts. De Groot stated that the only clear differences were that Grandmasters choose better moves, that they reach a decision sooner, and that they orient themselves faster in the position (duration of the first phase). In a re-analysis of De Groot's results, Charness (1981) mentions that there was also a statistical difference in the rate of generating base moves, Grandmasters generating more base moves per minute than Experts.

In the Swiss sample, comparisons were made from Masters to Class B players (a range of 4 standard deviations). There was a clear difference in the quality of the chosen move and in the duration of the first phase, but no statistically significant difference in the time to reach a decision (though the pattern of means indicates that Masters were faster) and in the rate of generating base moves (no indication of skill effect in the pattern of means). There was, however, an effect of skill for several other variables: number of base moves generated, mean depth of search, mean number and maximal number of immediate reinvestigation, and maximal number of non-immediate reinvestigations. Altogether, the present experiment seems to indicate that, in general, strong and weak players differ along more variables than was found by De Groot, even if the absolute differences are small. We turn our attention now to the absolute value of variables.

Absolute values of variables.

Each sample will first be analyzed by pooling the results across skill levels. Table 3 summarizes the results of Table 1 by giving the means of the two samples, pooling across skill levels. De Groot's sample, which includes world-class Grandmasters and relatively few Class players, is stronger than the Swiss sample. One could therefore expect some differences, in particular in the quality of the move chosen. However, in none of the variables[4] presented in Table 3 is there any significant difference (estimated with t-tests) between the two samples. Note also that the values for the total time, the number of episodes and the number of base moves are close to the values given by De Groot (1978, p. 117 and 122) for other positions. The lack of differences could of course be due to the fact that the statistics used are not sensitive enough to distinguish the two samples.

Insert Table 3 about here

Three variables are worth discussing further. First, for all skill levels pooled, the Swiss players tend to produce more immediate reinvestigations (on average 3.2) than De Groot's subjects (on average 2.4). The two samples differ less with respect to the number of non-immediate reinvestigations (1.6 vs. 1.2, on average). Second, De Groot found a mean of 3.1 minutes with the time of the first phase, as compared to 2.3 minutes in the Swiss sample. The fact that the subjects described in the present study produced more immediate reinvestigations and were faster in the first phase could have been due be to an artefact of the experimental procedure (audio-tape vs. note taking).-

Third, as has already been mentioned, the best move (1.Ba2xd5) was chosen with the same frequency (about one third of the protocols) in both samples. This is somewhat surprising, as the Swiss sample did not include players of the strength of De Groot's Grandmasters. Besides the possibility that the position was too simple to differentiate players at high levels of skills, it could also be that today's Masters and Experts are stronger than Masters and Experts of the forties. Testing the latter hypothesis would require investigations going beyond the scope of the present paper, but we can look for some preliminary evidence in the data reported so far. (A more powerful way to settle the question of difference of skill across time would be to study types of position where perfect play is known [for example from computer endgame databases], and to compare the performance in these positions, say, in number of errors, between top-level players of various periods of chess history.)

Perusal of Table 1 seems to corroborate the idea that a progression in strength has taken place as far as the quality of the chosen move is concerned. The Masters from the Swiss sample obtain a rating between De Groot's Grandmasters and Masters; in addition, the Experts from the Swiss sample seem to find better moves than De Groot's experts. However, the differences are not reliable statistically. It is true that the Swiss Masters do not differ from De Groot's Grandmasters [$t(15) = 0.58$, ns], but they do not differ either from De Groot's Masters [$t(12) = -0.89$, ns]. Moreover, the Swiss Experts do not perform significantly better than De Groot's Experts [$t\ (15) = 1.26$, ns]. Finally, the Swiss Class A and B players pooled together do not differ from De Groot's Class players.

Interestingly, other variables show a similar pattern. Except for the rate of generating moves, the mean depth of search, and the number of immediate reinvestigations, the Swiss Masters are closer to De Groot's Grandmasters than to De Groot's Masters. The Masters of the present sample even tend to be more selective than De Groot's Grandmasters (on average, 3.2 base moves vs. 4.2 base moves), though the difference is not significant [$t(15) = 0.46$, ns]. The Swiss Masters are, however, significantly more selective than De Groot's Masters [$t(12) = 2.20$, $p < .05$].

In summary, comparisons between the two samples (all levels pooled) show that they do not differ reliably. Comparisons, skill level by skill level, suggest the Masters of the present sample obtain values closer to De Groot's Grandmasters' than those of his Masters', notably showing a trend towards a better and faster choice (quality of move and time of choosing a move) and a more selective search (number of episodes and number of base moves).

Discussion

The results presented here indicate that (a) chess players from the Swiss sample differ along more variables than did De Groot's (1978) subjects—with the qualification that the differences lay mainly between Masters and class players—and (b) that the average values obtained with the Swiss sample do not diverge significantly from those of De Groot's sample.

The goal of this paper was to replicate a subset of De Groot's ([1946] 1978) results. This obviously had the disadvantage that the conclusions are limited by the particularities of the position used. In addition, this study does not address other interesting aspects of problem solving in chess, such as the role of familiarity with the position, or whether some positions invite players to search more than other positions. These questions are left for further research.

Impact of the Results on Pattern Recognition and Search Models

The replication was in part motivated by the different theoretical accounts given by search and pattern recognition models of chess skill. What is the impact of the empirical results on this theoretical discussion? Both pattern recognition and search models predict that strong players choose better moves, that they select moves faster, and that they generate more nodes in one minute. The first prediction was met, but the second and third were supported only weakly. Search models predict that strong players search more nodes and search deeper. The first prediction was not met, but the second was, with the qualification that the difference lies in the average depth of search, not in the maximal depth of search. Finally, pattern recognition models predict that strong players mention fewer base moves, reinvestigate more often the same move, jump less often between different moves, and have a shorter first phase. All these predictions were met.

What do the models have to say about the large inter-individual variability of the results found in both samples? This variability is compatible with the pattern recognition models, which propose that players acquire patterns for the type of openings and positions they spend time studying and practicing, and therefore build up various "styles" of play (De Groot & Gobet, 1996). Search models are not specific enough about this question, but could account for the variability in the data by assuming that players develop different search algorithms.

Holding (1985, 1992) argued that differences in the depth of search are incompatible with models based on pattern recognition. This is obviously wrong, as pattern recognition should facilitate the generation of moves in the mind's eye, permitting a smooth search. Even from the pattern recognition/chunking model standpoint, it remains somewhat of a surprise that the differences in search are so small between players several standard deviations apart in the Elo scale. First, in comparison with move generation methods relying on processing features of the position from scratch, generation of moves through pattern recognition should allow more nodes to be visited in the search space, as less time and cognitive resources are spent in generating moves. Second, strong players probably associate sequences of moves to patterns of pieces, which should make it easier for them to carry out deep search. Data from Saariluoma (1990, 1992) offer strong evidence for this hypothesis: in positions where one side could mate by playing either of two sequences of moves, Masters usually chose the suboptimal (it necessitated more moves to reach the mate) but familiar sequence of moves.

In the case of the Position "A" used in the experiment, a possible explanation for the rather shallow search shown by subjects is that the position is too easy for Masters—actually, the judgment that White has a decisive advantage in the position could be reached without searching more than 9 plies deep in all variations, assuming enough knowledge to correctly evaluate the final positions (De Groot, 1978). This suggests that pattern recognition, which allows better evaluation of positions, in turn allows cutting down the need for deep search. For an expert player, then, the critical question is not: "How to search as deep as possible?" but: "When to stop searching?" The class A player who performed the deepest search in the Swiss experiment (23 plies) is a case in point: he did not know when to stop, perhaps because he was not able to evaluate properly the positions he was generating. Only when he had reached a very simple endgame could he judge the situation correctly. Pattern recognition, then, not only allows a speedy generation of moves, but also provides position evaluations that enable the search to be terminated at appropriate times. This dual role of pattern recognition may explain why masters do not perform tree searches of different orders of magnitude than weaker players. I will elaborate this dual role below when discussing the template theory.

While Swiss Masters did produce higher mean depths of search than other subjects, they were not conspicuous for the maximal depth of search. It could be that for skilled chess players, it is more important to regularly see slightly more than their opponent than to sometimes search at extraordinary depths. This permits, in the long range, avoiding more errors and seeing more opportunities than the opponent. Simon (1974) has developed a formal model to investigate errors in chess. It could be worthwhile to expand his model by connecting his concept of error to the concept of mean depth search, perhaps by assuming that the probability of making an error in playing a move and the mean depth of search are inversely related.

Integrating Pattern Recognition and Search: The Template Theory

Altogether, the data of this paper vindicate most predictions of the pattern recognition model, but also indicate that there is a difference in the mean depth of search. In addition to the results presented here, other studies, as mentioned earlier, point to search differences between skill levels (Charness, 1981; Holding & Reynolds, 1982; Saariluoma, 1990). This convergent set of evidence calls for a reconciliation of the search and pattern recognition accounts of chess skill. In particular, it is necessary to better connect empirical data supporting the role of search with data showing the importance of pattern recognition in memory and perception tasks, and to develop a computational model accounting simultaneously for both sets of data. Thus, contrary to Holding's (1985, 1992) claims, the correct approach to chess skill—and other skills—is not to focus on a single component, such as search <u>or</u> pattern recognition, but to understand how these two processes interact.[5]

Although the template theory—a modification of the Chase and Simon (1973) theory proposed by Gobet and Simon (1996b, in press)—was mainly developed to account for empirical data from memory research, it also offers a theoretical background for studying pattern recognition and search processes as well. Templates are chunks with slots (variables) that can be rapidly filled in with new information. Slots may store values for the location of individual pieces or groups of pieces (chunks). In addition, templates give access to other types of information, such as potentially good moves or plans, evaluation of the position, etc. Finally, templates may be connected to other templates. For example, a template describing a position reached in a Panov attack of the

Caro-Kann defense after 20 moves may be connected to a template describing some type of endgame that occurs often from this type of position. Such connections may act as macro-operators and allow search through a more abstract space than the move space (for a similar idea in geometry, see Koedinger & Anderson, 1990).

In developing templates through practice and study—and the theory postulates that it takes years to grow several thousands of such templates—Masters acquire knowledge which affects search in two opposing ways. On the one hand, mechanisms are developed that make searching easier. For example, practice may allow the association of not only moves, but also of sequences of moves to patterns of pieces (cf. the computer program described by Gobet & Jansen, 1994). Chunking of moves allows a more selective search (a set of candidates is proposed by some patterns of pieces) as well as a deeper search (time not spent generating moves by other means may be spent searching deeper, and several plies may be readily played in the mind's eye without much conscious search when a chunk of moves is available). In addition, development of templates may also facilitate search, because changing the internal representation of a position when looking forward is made easier by the presence of slots.

On the other hand, templates give access to evaluation knowledge, which, as was shown by Berliner (1981), allows the amount of search to be reduced: if the evaluation of a position is readily at hand, there is no need for searching deeper. The non-monotonic behavior of search across skill levels then follows from the template theory, assuming that players are faster to develop both the piece slots and the association pattern-to-move than to fill in the evaluation slots—perhaps because it is easier to associate concrete bits of information than to learn complex and relatively uncertain evaluations of positions. In a first phase, aspiring Masters learn many pattern-to-move associations and piece slots, allowing them to search faster and deeper; then, they associate more evaluation judgments to templates, and therefore diminish the need for search. This could account for Saariluoma's (1990) results that his Grandmaster and International Master searched less than his Experts, as well as for De Groot's (1978) results when his Grandmasters searched less than his Masters. (The Swiss sample offered a different pattern, where Masters were searching deeper than the subjects of the other groups. This could be due to the fact that the Swiss Masters did not reach a high enough level of expertise—they were clearly weaker than both Saariluoma's and De Groot's strongest players.)

In conclusion, the template theory offers a promising avenue to tie together the concepts of search and pattern recognition, which have not been yet integrated in a single theory of chess skill. As Koedinger and Anderson (1990) correctly note, no chess program has yet been written that both simulates recall experiments and plays chess, let alone plays at master level. The CHREST model (Gobet, 1993; Gobet, Richman & Simon, in preparation), which implements the template theory as a computer program, simulates several critical results from the literature on chess perception and memory, including results that were considered highly detrimental to the original chunking theory. It also offers a framework allowing theories of chess memory and perception as well as theories of problem solving to be integrated in a unified computational model (see Gobet, 1997, for a simplified implementation of this integrative model).

Research in cognitive psychology has shown that many aspects of expertise are specific from domain to domain. However, it has also shown that there exist a few invariants in human cognition, such as the limited size of short-term memory (perhaps four chunks) and the time to encode a new chunk in LTM (about eight seconds). The research reported here converges with previous work to indicate that there also exist strong limits in the time needed to process a state in the problem space (perhaps 8 problem states in one minute). In the future, as in the past, empirical research using chess will help us pinpoint these cognitive invariants.

References

Berliner, H.J. (1981). *Search vs. Knowledge: An Analysis from the Domain of Games.* (CMU-CS-82-104). Pittsburgh: Carnegie-Mellon University, Department of Computer Science.

Calderwood, B., Klein, G.A., & Crandall, B.W. (1988). Time pressure, skill, and move quality in chess. *American Journal of Psychology, 101,* 481-493.

Charness, N. (1981). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 467-476.

Charness, N. (1992). The impact of chess research on cognitive science. *Psychological Research, 54,* 4-9.

Chase, W.G., & Simon, H.A. (1973). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

De Groot, A.D. (1946). *Het denken van den schaker*. Amsterdam, Noord Hollandsche.

De Groot, A.D. (1978). *Thought and choice in chess* (2nd ed.). The Hague: Mouton Publishers. (Revised translation of De Groot, 1946).

De Groot, A.D., & Gobet, F. (1996). *Perception and memory in chess: Heuristics of the professional eye.* Assen: Van Gorcum.

Elo, A. (1978). *The rating of chess players, past and present*. New York: Arco.

Gobet, F. (1992). Learned helplessness in chess players: The importance of task similarity and the role of skill. *Psychological Research, 54,* 38-43.

Gobet, F (1993). *Les mémoires d'un joueur d'échecs*. Fribourg (Switzerland): Editions Universitaires.

Gobet, F. (1997). A pattern-recognition theory of search in expert problem solving. *Thinking and Reasoning, 3,* 291-313.

Gobet F., & Jansen, P. (1994). Towards a chess program based on a model of human memory. In H.J. van den Herik, I.S. Herschberg, & J.W. Uiterwijk (Eds.), *Advances in Computer Chess 7*. Maastricht: University of Limburg Press.

Gobet, F., & Retschitzki, J. (1991). Détresse apprise et jeu d'échecs : Rôle de la familiarité et de la similitude des tâches. *Revue Suisse de Psychologie, 50*, 97-110.

Gobet, F., Richman, H., & Simon, H.A. (in preparation). *Chess players' memory and perception: A unified process model.*

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

Gobet, F., & Simon, H.A. (1996b). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31,* 1-40.

Gobet, F., & Simon, H.A. (1996c). Recall of random and distorted chess positions: Implications for the theory of expertise. *Memory & Cognition, 24,* 493-503.

Gobet, F., & Simon, H.A. (in press). Expert Chess Memory: Revisiting the Chunking Hypothesis. *Memory, 5*.

Gruber, H. (1991). *Qualitative Aspekte von Expertise im Schach.* Aachen: Feenschach.

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

Holding, D.H. (1989). Evaluation factors in human tree search. *American Journal of Psychology, 102*, 103-108.

Holding, D.H. (1992). Theories of chess skill. *Psychological Research, 54*, 10-16.

Holding, D.H., & Reynolds, R.I. (1982). Recall or evaluation of chess positions as determinants of chess skill. *Memory & Cognition, 10,* 237-242.

Holding, D.H., & Pfau, H.D. (1985). Thinking ahead in chess. *American Journal of Psychology, 98*, 271-282, 1985.

Klein, G.A., & Peio, K.J. (1989). Use of a prediction paradigm to evaluate proficient decision making. *American Journal of Psychology, 102*, 321-331.

Koedinger, K.R., & Anderson, J.R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science, 14,* 511-550.

Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Saariluoma, P. (1990). Apperception and restructuring in chess player's problem solving. In K.J. Gilhooly, M.T.G Keane, R.H. Logie & G. Erdos: *Lines of thinking, Volume 2*. New-York: John Wiley.

Saariluoma, P. (1992). Error in chess: The apperception-restructuring view. *Psychological Research, 54,* 17-26.

Seligman, M.E.P. (1975). *Helplessness. On depression, development and death*. New York: Freeman.

Simon, H.A. (1974). The psychological concept of "Losing Move" in a game of perfect information. *Proceedings of the National Academy of Sciences, USA, 71,* 2276-79.

Wagner, D.A. & Scurrah, M.J. (1971). Some characteristics of human problem-solving in chess, *Cognitive Psychology*, *2*, 454-478.

Appendix

Protocol of S21; age 24; level: Expert; ELO 2001

(Translated from French; square brackets indicate information added to the protocol)

First phase. OK. There is an isolated Pawn for White, but it should not be bad, because it's a middlegame position, and it looks rather dynamic, and one can build on it, given that there is a Knight on e5, and one can... It's advantageous. Therefore, one should try not to trade pieces off but to bring an attack on the King's side. Mmm... The Black Bishop is badly placed. Well, the first move that comes to my mind, it's Knight e4. Yes, but it's dangerous because there is the Bishop on c6. I will have to check this later. Take advantage of Black's black diagonal. Maybe try to exchange the Knight on f6 to place the Knight on e4, with gain of tempo, and then, after, to have the outpost on c5. It seems ridiculous to me, because I give up the black Bishop. [2']

Episode 1. Bishop takes f6, Bishop takes f6, Knight e4, Bishop g7 or Bishop e7. After, I cannot progress much. He is holding all the black squares. [2'26'']

Episode 2. What wouldn't be bad either is to overprotect the Knight on e5, with a little move like Rook "f" to e1, and to see what he is doing. [2'59"]

Episode 3. Or Rook "f" to d1. It overprotects my Pawn, which is weak but at the same time dynamic. I'll see. [3'30'']

Episode 4. Bishop h6 doesn't look good. [3'36'']

Episode 5. Knight takes d5. If Bishop takes d5, Bishop takes d5, Knight takes d5, Knight g4... One takes advantage of these squares. Ahhh, but he can take with the Pawn; it isolates the central Pawn for both of us, and then... One does not have much. Ah, maybe the Pawn is on a white square and... [4'24'']

Episode 6. Ah, maybe Pawn b4, eh? It reinforces the advance of the Pawn b5. And then to play on the Queen's side, by trying to bring something on c5... Mmm, Mmm. Especially as it is attacked, moreover, this Pawn, I see now. [6'38'']

Episode 7. I can defend it by Knight c4. No... One takes the Knight away from its good position, which bothers me. [7'12'']

Episode 8. Knight takes d5... [7'25'']

Episode 9. Ahh... it can be dangerous, if he takes it... It can be dangerous if he takes the Pawn b2... [Irrelevant question to experimenter.] No it's not dangerous. [8'08'']

Episode 10. What wouldn't be bad, that's Queen d2. It controls the black squares, and also it allows, maybe, to exchange on d5, followed by [an exchange on] f6, and to be immediately on the black squares of the King's side. Then Queen d2 with the threat Knight takes d5. Either Knight takes d5, Bishop takes e7, Knight takes e7 and Knight g4, with the threat Queen h6 and Knight f6.... It creates holes... Or perhaps? Knight takes d5, Pawn takes d5, Queen f4, I'm attacking. He, he defends. Bishop d7. He is losing the Pawn d5. [10'10"]

Episode 11. Queen d2, again. Queen d2, Rook "f" to d8, Knight takes d5, Bishop takes d5, Bishop takes d5, Knight takes d5, Bishop takes e7, Knight takes e7, Knight g4. Ooooh... it gives play on the Queen's side for Black. [11'20"]

Episode 12. I, I believe that one has to build up, one has to play [Pawn] b4, and after, Rook "f" to e1, and after try to play on the black squares of the Queen's side. I do not see any tactical move. Ahh... Ahh... But on [Pawn] b4, he does Knight takes c3, Rook takes c3. After, he has the outpost on d5, with

Bishop d5, or Knight d5. Let's say Bishop d5. Then, I play Bishop takes d5, Knight takes d5, Bishop takes e7, Knight takes e7, Rook "f" on c1, and afterwards I have the "c" column, but one gets into an endgame, and I have the isolated Pawn. One has to be careful. Ah... that's not an endgame, with two Rooks and one Queen, one shouldn't exagerate. [12'55"]

Well, I play Pawn b4. [13']

Insert Figure A1 about here

Extraction of the descriptive variables

The chosen move (Ab4) gets a value of 1. There were 12 episodes in the protocol, the total time was 13 m and the duration of the First phase was 2 m. The total number of nodes is 52, and the rate of generating nodes per minute is 4 (52/13). The maximal depth is 9 (episode 11 and 12; "no-moves" are not counted). Taking the longest line within an episode, the sum of depths over the 12 episodes is 44, and the mean depth is 3.66 (46/12). The number of (different) base moves is 8 (again, the "no-move" is not counted), and the rate of generating base moves is 0.69 (9/13). For the variables related to the number of reinvestigations, it helps to write down the first move of each episode:

Cxf6 Dfe1 Dfd1 Ch6 Bxd5 Ab4 Bc4 Bxd5 ∅ Ed2 Ed2 Ab4 (Ab4)

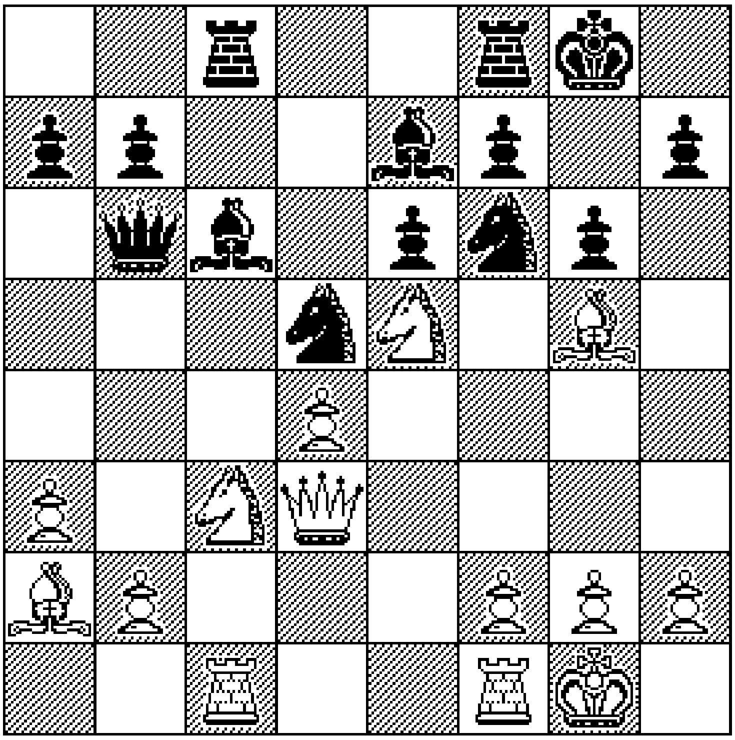
We see that the moves Ab4 and Bxd5 were both reinvestigated once non-immediately, and the move Ed2 was reinvestigated once immediately. We get then a total of 2 non-immediate reinvestigations, 1 immediate reinvestigation, and 3 as total number of reinvestigations. The maximal number of (re)investigations, both immediate and non-immediate, was 2.

Figure Caption

Figure 1. Position “A” of de Groot (1946).

Figure A1. Problem solving behavior graph of S21. Time proceeds from left to right and then down. The following evaluations are used at the end of an episode: + for positive, - for negative, and ? for unknown. ∅ means “no move”. See Figure 1 for an illustration of the chess co-ordinate system.

8
7
6
5
4
3
2
1



a b c d e f g h

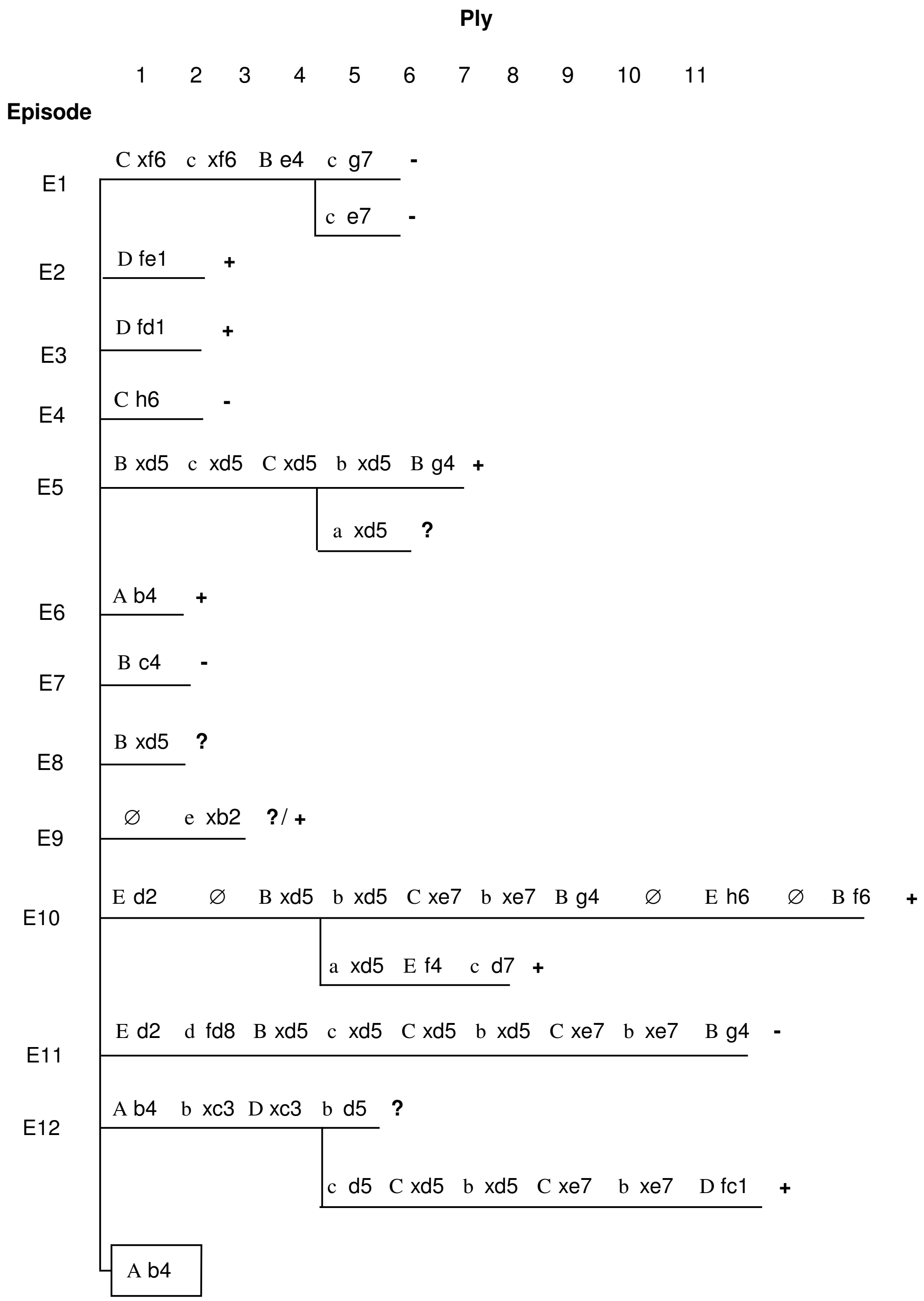

Ply
1 2 3 4 5 6 7 8 9 10 11
Episode
E1
C xf6 c xf6 B e4 c g7 -
c e7 -
E2
D fe1 +
E3
D fd1 +
E4
C h6 -
E5
B xd5 c xd5 C xd5 b xd5 B g4 +
a xd5 ?
E6
A b4 +
E7
B c4 -
E8
B xd5 ?
E9
∅ e xb2 ? / +
E10
E d2 ∅ B xd5 b xd5 C xe7 b xe7 B g4 ∅ E h6 ∅ B f6 +
a xd5 E f4 c d7 +
E11
E d2 d fd8 B xd5 c xd5 C xd5 b xd5 C xe7 b xe7 B g4 -
E12
A b4 b xc3 D xc3 b d5 ?
c d5 C xd5 b xd5 C xe7 b xe7 D fc1 +
A b4

Table 1

Means and Standard Deviations (in Parentheses) of de Groot's (1978) Data and of the Data Presented in this Paper

	de Groot's data				Data presented in this paper			
	Skill level				Skill level			
	Grand-master n=5	Master n=2	Expert n=5	Class n=2	Master n=12	Expert n=12	Class A n=12	Class B n=12
Quality of move	4.8 (0.4)	4.0 (1.4)	2.6 (0.5)	3.0 (1.4)	4.6 (0.8)	3.3 (1.2)	2.3 (1.3)	2.5 (1.5)
Total Time (in min.)	9.6 (3.4)	16.5 (2.1)	12.8 (6.4)	22.0 (8.5)	11.3 (7.4)	14.9 (6.6)	18.5 (9.2)	16.7 (8.6)
# of nodes visited	35.0 (23.8)	93.0 (45.2)	30.8 (18.2)	43.0 (2.1)	58.0 (51.0)	58.3 (35.8)	56.8 (33.4)	33.9 (33.8)
nodes/min.	3.6 (1.3)	5.5 (2.0)	2.4 (1.2)	2.0 (0.7)	4.8 (1.7)	4.1 (2.6)	3.2 (2.0)	3.4 (6.1)

Max. depth (in plies)	6.8 (2.0)	13.0 (2.8)	6.6 (3.0)	5.5 (2.1)	9.1 (3.8)	8.9 (3.6)	9.0 (6.5)	6.1 (3.1)
Mean depth (in plies)	5.4 (1.4)	5.3 (1.3)	4.8 (0.7)	3.5 (0.8)	5.0 (2.3)	4.6 (1.9)	3.7 (2.1)	2.9 (1.1)
# of episodes	6.6 (3.9)	17.0 (4.2)	6.4 (3.5)	12.5 (3.5)	9.2 (7.4)	10.6 (6.2)	12.4 (5.9)	8.6 (4.7)
# of base moves	4.2 (1.3)	8.0 (4.2)	3.4 (1.5)	6.0 (2.8)	3.2 (2.7)	4.8 (2.2)	6.5 (2.8)	4.8 (2.9)
Base moves/min.	.45 (.12)	.47 (.20)	.29 (.09)	.27 (.03)	.38 (.33)	.37 (.19)	.41 (.19)	.37 (.26)
Imm. reinvest.	1.8 (1.9)	4.5 (2.1)	2.0 (1.2)	3.0 (0.0)	4.9 (4.4)	3.7 (3.1)	2.4 (2.7)	1.7 (1.4)
Non imm. reinv.	0.8 (0.8)	4.5 (2.1)	1.0 (1.0)	2.0 (1.4)	1.1 (1.9)	2.1 (3.0)	3.7 (3.1)	2.1 (1.8)

Note. The values for de Groot's Grandmasters and Experts were taken from Tables 8 and 12 of de Groot (1978), except for R (quality of the moves), for which a different classification is used. The values for Masters were estimated from protocols of M2 and M3 and the values for Class players from protocols C2 and C5 in Appendix II of de Groot (1978). De Groot's other protocols of Masters and Class players were either unavailable or were obtained with a different procedure.

Table 2

Means and Standard Deviations (in Parentheses) of the Mean and Maximal Number of Reinvestigations

	Skill level			
Type of variable	Masters	Experts	Class A	Class B
Mean # of Reinvestigations	6.0 (5.9)	5.8 (4.9)	6.1 (3.7)	3.8 (2.7)
Mean # of Immediate Reinvestigations	4.9 (4.4)	3.7 (3.1)	2.4 (2.7)	1.7 (1.4)
Mean # of Non-immediate Reinvestigations	1.1 (1.9)	2.1 (3.0)	3.7 (3.1)	2.1 (1.8)
Max # of Immediate Reinvestigations	5.2 (3.0)	3.4 (1.5)	2.5 (1.4)	2.3 (1.0)
Max # of Non-Immediate Reinvestigations	1.4 (0.7)	2.0 (1.2)	2.7 (1.2)	2.2 (0.7)

Table 3

Means and Standard Deviations (in Parentheses) of Groot's and this Paper's Samples

	de Groot's data	Data presented in this paper
Quality of move	3.6 (1.2)	3.1 (1.5)
Total Time (in min.)	13.5 (6.4)	15.3 (8.2)
# of nodes visited	42.9 (30.0)	51.8 (39.3)
Nodes/min.	3.2 (1.6)	3.9 (3.5)
Max. depth (in plies)	7.4 (3.3)	8.3 (4.5)
Mean depth (in plies)	4.9 (1.2)	4.0 (2.0)
# of episodes	8.9 (5.2)	10.2 (6.1)
# of base moves	4.7 (2.4)	4.8 (2.8)
Base moves/min.	0.37 (0.14)	0.38 (0.24)
Immediate reinvestigation	2.4 (1.7)	3.2 (3.2)
Non immediate reinvestigation	1.6 (1.6)	1.2 (0.8)

Time for the first phase	3.1 (1.3)	2.3 (2.5)

Footnotes

[1] The Elo rating scale is an interval scale ranking competitive chess players, with a standard deviation of 200. Grandmasters are generally rated above 2500 Elo, International Masters above 2400 Elo. Masters are rated in the range 2200-2400, Experts 2000-2200, Class A players 1800-2000, Class B players 1600-1800, and so on.

[2]De Groot (1946) notes that it is a difficult variable to operationalize precisely, because the line between enumerating potential base moves and starting search is not sharply drawn.

[3]For the mean depth of search, de Groot (1946) did not give any data, but Holding (1985) has estimated them by dividing the number of nodes by the number of episodes.

[4]Statistical tests could not be carried out for the time of the first phase, because the individual data in De Groot's sample were not available.

[5] The attempt to combine pattern recognition and search has a long history (e.g., it was one of the main concerns of Chase, Newell, and Simon). As a more recent attempt, one can mention Saariluoma's (1990) theory of apperception and restructuration. As shown elsewhere (Gobet, 1993), however, this theory is not quite successful: it is vaguely stated, and does not go beyond Newell and Simon's (1972) idea of means-end analysis.