#28

APPLIED COGNITIVE PSYCHOLOGY, VOL. 4, 109–128 (1990)

Chess Expertise in Children

DIANNE D. HORGAN and DAVID MORGAN

Memphis State University

SUMMARY

This paper reports several studies of chess expertise in children who play competitive chess. The first study examines (1) the relationship between experience and skill among 113 school-age children (grades 1 through 12); and (2) the relationship between chess skill and scores on various spatial and logical abilities tests among the top 15 players. Improvement in skill is related to experience, and chess players score higher than average on the Raven's Progressive Matrices. Also, scores on a chess-specific test, the Knight's Tour, correlate with scores on the Raven's. The second study reports three experiments with 59 Ss involving chess-specific tasks in memory, perception, and similarity judgements. The first two experiments replicated and extended Chase and Simon (1973). The third experiment, which asked Ss to judge similarities of chess positions, demonstrated that similarity judgements become more global and abstract with increased skill. The final section describes qualitatively how children's chess expertise compares to that of adults. Drawing upon Anderson (1985), we focus on some distinctive features of children's chess play and on some successful techniques in coaching young players.

Although much of recent research on decision-making and problem-solving has stressed the limits of rationality and how far humans deviate from 'good' decisions, chess is a situation in which humans can make unusually sound decisions. The literature on expertise has focused on chess and chess masters partly because of these very high-quality decisions. What is particularly striking is that children—not normally known for their rationality—can compete with adults on an even basis in chess tournaments. Further, chess skill among children is not limited to a very few extremely gifted prodigies; many children who are exposed to good chess coaching become competent players.

The study of childhood expertise is important for several reasons. First, models of expertise in general, and of chess skill in particular, must be able to account for high levels of competency among young children. Also, understanding how they acquire their skill is not only important theoretically, but may have important practical implications for enhancing other types of learning.

Research on expertise (e.g. Anderson, 1985 and Dreyfus and Dreyfus, 1986), however, has been limited to adults. It is not known whether those models of expertise apply to children. There have been only a few studies of any kind with child chess players. Christiaen and Verholfstadt (1978) studied fifth-graders for 2 years, during which time an experimental group studied chess after school, 1 day a week. After the 2 years the experimental group performed better on Piagetian tasks, significantly better on school tests, and better on standardized tests than did the control group. Chi (1978) demonstrated that child players could remember

0888–4080/90/020109–20$10.00

Received 21 August 1988
Revised 10 September 1988

more pieces from a chess scheme than could adult non-players, thus demonstrating that knowledge of the domain can be more important than age when asked to recall a complex array.

While both Christiaen and Chi used children as subjects, none of Christiaen's subjects were officially rated by the United State Chess Federation, and only one of Chi's had an official rating. Virtually no research exists with children who have acquired sufficient expertise to compete with adult tournament players. In contrast, the present research used a sample of nationally rated, competent players and also, in more depth, analysed performance of a subsample of the very top young players.

This research represents a beginning attempt to explore child chess expertise. It is therefore largely descriptive. The first part of the paper reports a correlational study of 113 members of a scholastic chess club. The second part is a series of experiments conducted at a state scholastic tournament. The third part is a discussion of the nature of childhood expertise and chess coaching techniques.

PART 1. CORRELATIONAL STUDY

After an exhaustive review of 'virtually all of the available research on chess skill', Holding (1985) put the relative contribution of practice and abilities at the top of his list of unsolved problems. While Anderson (1985) has claimed that master chess players 'have achieved their expertise through practice' (p. 232), Holding questioned whether practice is sufficient, saying that it is difficult to believe that practice alone would make anyone a grandmaster. Indeed, the existence of 11- and 12-year-old chess masters who have only played chess for several years makes it difficult to hold the view that practice is a sufficient explanation of expertise. (Further, there is no evidence that other kinds of experience can explain expertise. Pfau (1983), for example, reports only a non-significant correlation of .10 between hours of chess playing and rating.) In addition to practice, it seems obvious that spatial ability ought to be related to chess skill. Yet Holding concluded that 'There is almost no evidence that chessplayers excel in spatial ability' (p. 228). In fact, he argued, 'the data that are now available suggest, mainly by default, that chessplayers are almost entirely like other people' (p. 227). Along with Holding and other chess researchers, we believe experience is a necessary, but not sufficient, condition for expertise, and that spatial abilities are related to chess skill. It is difficult to measure these relationships, however, since other factors (such as talent, motivation, intelligence, opportunities) are no doubt of primary importance.

Because our sample is relatively homogeneous on many factors (all bright, middle-class children with the same opportunities for play and coaching), it is a good sample in which to demonstrate a relationship between experience and skill. We hypothesized that in this homogeneous sample there would be a correlation between experience and improvement in chess skill. Experience was measured by the number of non-tournament games played during club hours. This measure presumably reflects general chess activity level and not just game playing, since the number of games played is confounded with other types of experiences such as more feedback from coaches and other players or more study.

Virtually all studies of skill acquisition have found the relationship between practice and performance to be a power function, indicating that practice has diminishing effects (Anderson, 1985). We hypothesized that the same relationship would be found for acquisition of chess skills.

In the first study, several tests of spatial and logical abilities were administered to the top 15 players in the sample. We hypothesized that chess players would score above average on these abilities.

METHOD

Subjects

Our sample came from one chess club at a small school (700 students, grades K through 12) outside of Memphis, Tennessee. At the time of our study, 113 students (16 per cent of the student body) of all ages were members of the chess club and were included in the sample. This high level of interest in chess is very unusual and reflects an unusually supportive environment for chess players. Table 1 shows the students' grade levels, ratings at the beginning of the study, and gender. The children are coached primarily by the high school mathematics teacher (rated about 2000 by the U.S. Chess Federation) with assistance from a science teacher who is a chess master. The students play before and after school and during the lunch hour (some only occasionally). They participate in tournaments locally, regionally and nationally. Some of them attend a 1-week summer chess camp.

Table 1. Characteristics of subjects in study 1

	September rating										
	<1000		*1000–1299*		*1300–1499*		*>1500*		*Subtotals*		Totals
Grade	M	F	M	F	M	F	M	F	M	F	
1–3	17	12	3	0	1	0	0	0	21	12	33
4–6	10	9	8	3	2	0	2	0	22	12	34
7–9	10	4	5	1	4	0	0	0	19	5	24
10–12	6	6	5	1	2	0	2	0	15	7	22
Subtotals	43	31	21	5	9	0	4	0	77	36	113
Totals	74		26		9		4		113		

Skill ratings. The U.S. Chess Federation ranks tournament players (regardless of age) based on their wins and losses against other rated players. The ratings are derived from probability theory and are a good measure of a player's skill. The mean for all U.S. tournament players of all ages is 1500, and the standard deviation is 200. Only fairly serious players participate in tournaments, so a player with a rating of 1500 is very good relative to chess players in general. Above 2200 is a master. In general, a child with a rating of 1100 (two standard deviations below the mean) can beat most recreational, non-tournament-playing adults who consider themselves to be fairly good chess players. An informal survey of parents of our subjects revealed that virtually all the children can beat their parents. (The majority of the parents do not play at all.)

Although the sample came from only one small school, it included a fairly large proportion of the top child chess players in the nation, particularly at the younger end of the scale. For example, six of our primary grade children (grades 1 through 3) have been listed among the top 10 players under 8 in the United States. Most people naively believe that any child who becomes proficient at chess must be an extremely rare prodigy (probably with grandmasters for parents). On the contrary, particular chess coaches consistently produce strong players, year after year—even though the specific children move on. In fact coaches often say that, given a few months of training, any motivated and bright 10-year-old can become a proficient player. In other words, the skills discussed in this paper are not limited to a very select few extremely gifted children; they are skills that presumably many above-average children can learn.

The elite subsample. In order to learn more about the top players, the top 20 club players were invited to participate in a special tournament where they would be given tests between rounds. Fifteen of those children participated. Although this group may be too small to draw unequivocal conclusions, their results provide suggestive results.

Procedure

We ran correlations on both the whole sample and the elite subsample. For the whole sample we examined their chess records for one academic year. These records included national ratings in September and in May, their grade level, the number of club games played and whether they won, lost or drew.[1] The elite group was given the following tasks: (1) the Raven's Progressive Matrices (sets A–E); (2) the Knight's Tour (a chess-specific test that is believed to be related to chess aptitude); (3) a Piagetian task (the plant task designed analogously to the 'colorless-liquids problem' to measure combinatory logic in formal operations—see Kuhn and Brannock, 1977, for details); and (4) a game reconstruction task in which Ss were asked to reconstruct a previous game from memory.

The Raven's Matrices (Raven, Court and Raven, 1985) is a widely used test involving perception of relationships in geometric figures and is considered a measure of logical abilities as well as spatial abilities. Although norms for children are limited, we felt the type of reasoning required was similar to chess reasoning. (We have given adult players the test, and they concur.) We chose this test also because it can be used for a wide range of subjects, is easily administered and untimed. The Knight's Tour, described by Holding (1985), requires subjects to tour the whole board with a knight, landing on all squares except those with a pawn or controlled by a pawn. Figure 1 shows the board set-up for the task. The task is scored according to the number of moves, the number of errors, and the time required to complete the task. The task is repeated; the first tour is believed to measure aptitude while the improvement on the second tour measures trainability. Holding (1985) stressed the need for normative data on the Knight's Tour since existing data are promising, but very scarce. He reported a non-significant correlation of 0.30 among 25 adult club players' ratings and their Knight's Tour scores. He also reported Radojcic's (1971) finding that among a large sample of Czech chess-playing children, the four boys who turned in the fastest times for their age groups are all now prominent players.

[1] Note that the national ratings are not computed using the club games so these numbers are not confounded.

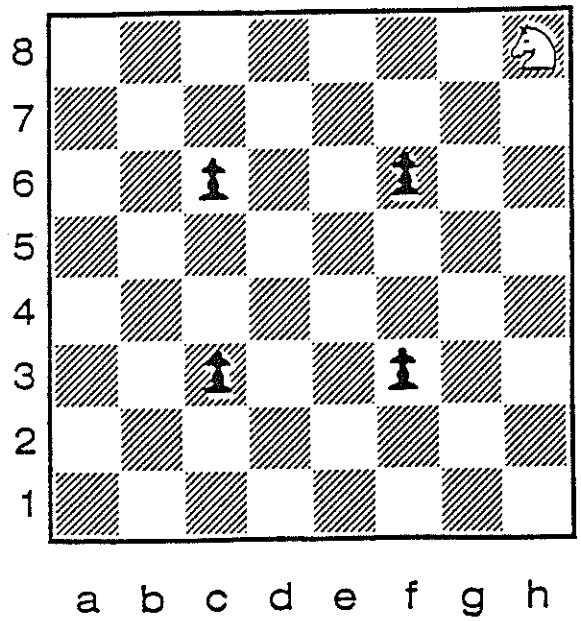


Figure 1. Board set-up for the Knight's tour

Subjects were also asked to reconstruct their most recent game from memory, indicating their best and worst moves, how far ahead they looked before each, and how many alternatives they considered at that point. For comparison purposes, a group of adult players at another tournament were asked to do the same task. A chess master was given both sets of protocols and the corresponding score sheets from the actual games. The master analysed the reconstructions and described the differences he saw between child and adult reconstructions.

To test whether the power law of practice held for our subjects, national ratings were plotted as a function of the number of club games played over a 21-month period, both measured on a logarithmic scale.

Results and discussion

Statistics for the groups are given in Table 2. As can be seen, although the groups are very similar in terms of grade, they differ in terms of their ratings, the number of games played in the previous academic year, and in the number of rating points earned. But the within-group differences are also large: in the elite sample the range for number of rating points gained was from 6 to 922; the range for number of games played was from 20 to 231.

Results from the entire sample

Age. Age, as measured by grade, is significantly related to rating (r=.468, p<.001) but only weakly and non-significantly related to games played (r=.120). Older subjects tend to have higher ratings, but do not play significantly more games. Since the subjects in our sample varied tremendously in terms of their ages

Table 2. Characteristics of sample. Means and standard deviations

Measures	Group			
	Entire sample[a]		Elite subsample[b]	
	M	SD	M	SD
September rating	1004	307	1380	190
Grade	6	3.5	5.9	2.9
Number of games played	40.6	49.3	101.4	66.5
Number of rating points gained	142.0	200.5	283.6	293.3

[a] $N = 113$.
[b] $n = 15$.

and starting points in September, it is useful to consider the partial correlation between number of games played and the initial rating in September, with the effects of grade removed ($pr = .251$, $p < .01$). This significant correlation means that the players rated higher in September played more games as the school year progressed independently of age.

Experience. Since it is possible that these good players improve independently of number of additional games played, partialling out both grade and September's rating yields a clearer picture of the relationship between experience and improvement. Table 3 shows the partial correlations of games played, wins, and ratings in May.

Table 3. Partial correlations between games and rating with grade and September rating partialled

	Games played	May rating
Wins	.918**	.491**
Games played	—	.452**
May rating		—

$p < .01$

Wins and games played are, of course, trivially related since number of wins is a proportion of games played. Holding age and September's ratings constant, there is a significant relationship between experience in terms of games played, and May rating ($pr = .452$, $p < .001$). Again since wins are related to games played, wins are also related to new rating ($pr = .491$, $p < .001$). In sum, the more improved players played more and won more. This does not, of course, demonstrate that more experience leads to more improvement and hence more wins: players may be motivated to play more when they are winning.

The elite subsample

Although our limited test battery and sample size does not allow us to draw firm conclusions, as an exploratory measure, subjects were given a standard Piagetian task, a spatial abilities tests, and the Knight's Tour.

Age results from the elite subsample. Table 4 shows the relationships between the various measures and grade. As in any sample of children of different ages, the

Table 4. Correlations between abilities measures and grade for elite subsample[a]

	Raven's	Knight's Tour	Piaget
Grade	.727*	.404	.563*
Raven's	—	.619*	.578*
Knight's		—	−.285
Piaget			—

[a]($n=15$)
*$p<.05$

Raven's Progressive Matrices scores and the Piagetian scores correlate with grade and with each other. The Knight's Tour correlates with the Raven's scores, suggesting that it measures similar spatial reasoning abilities. The negative correlation between the Knight's Tour and the Piagetian task suggests that the Knight's Tour does not measure logical abilities similar to those necessary for the Piagetian task. The Knight's Tour is not significantly related to grade.

In this more elite group there was no relationship between grade and rating ($r=.059$), but there was a non-significant, negative relationship between grade and number of games played ($r=-.345$). This may result from the fact that the younger children play faster, and hence may play more games in the same amount of time as the older players. Age may act as a suppressor variable. So here, too, to clarify results, the results are presented with grade partialled out (Table 5).

Table 5. Partial correlations between abilities measures and ratings for the elite subsample with grade partialled out[a]

	Raven's	Knight	Piaget	Improve
Rating	.335	.380	−.044	.340
Raven's	—	518*	.298	.221
Knight's		—	−.076	.212
Piaget			—	−.330
Improve				—

[a]$n=15$.
*$p<.05$.

Experience results from the elite subsample. Controlling for age, there was a significant relationship between the number of games played and rating ($pr=.511$, $p<.05$). (For the whole sample, $pr=.251$, $p<.01$). There is a moderate relationship between number of games played and improvement (improvement = the difference between May and September's rating), ($pr=.402$). The correlation between number of games played and rating may show only that better players play more—not that higher ratings result from more practice. Similarly, the correlation between experience and improvement may show only that winning motivates more play. While the results are not unequivocal, they represent, as of now, the first documented relationship between experience and skill in chess.

Figure 2 shows the relationship between number of games played and performance for four subjects. For this figure we choose four subjects from the elite subsample who were top-rated for their age groups and who had been active players for the 21 months for which club records were available. The age of each subject at the beginning of the 21-month period is shown. Although there are wide individual differences in both number of games played and rating, all four subjects show a straight-line function when both number of games and performance are plotted on a logarithmic scale.

Abilities results from the more elite subsample. A stepwise regression, with improvement as the dependent variable, shows that September's rating accounts for the most variance (R^2=.652). R^2 changes to .770 with the addition of the Raven's test, and to .873 with the addition of number of games played. Although the sample size is small, this suggests that abilities of the kind measured by the

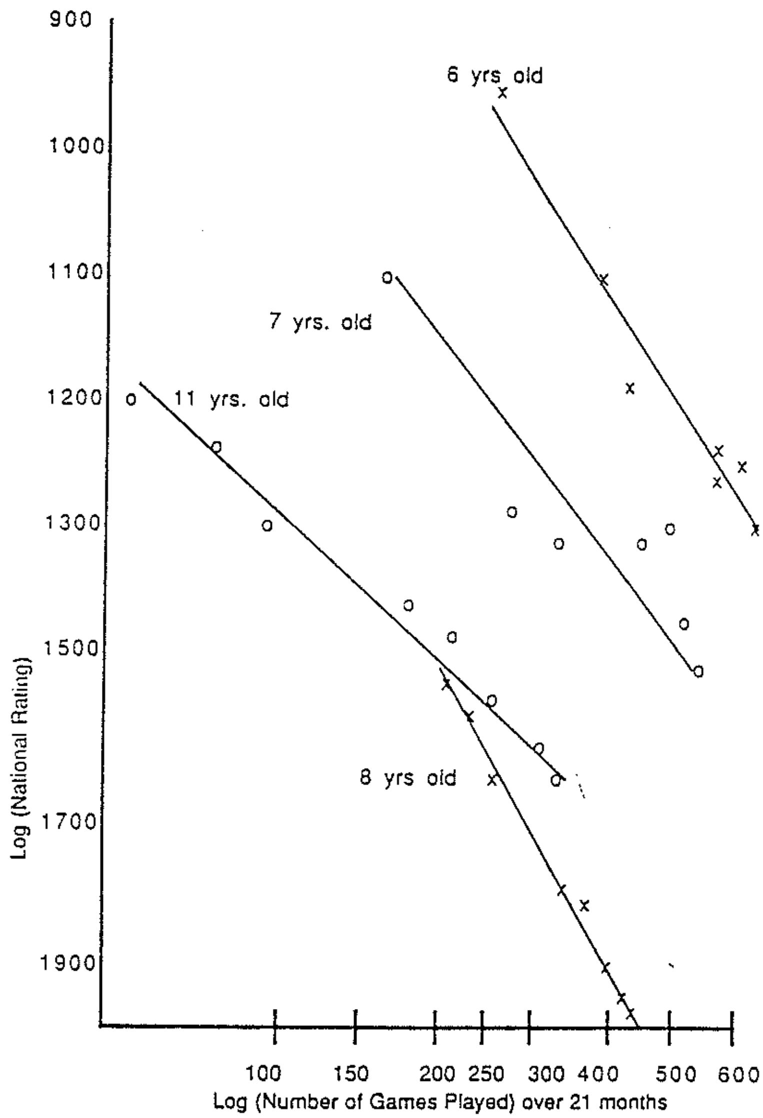


Figure 2. The relationship between number of games played and skill level for four top-performing subjects.

Raven's and number of games played contribute significantly to improvement. Grade does not make a significant contribution.

While spatial and logical ability of the type measured by the Raven's may not differentiate very much among the elite subsample because of the restricted range, it may differentiate good players from non-players. The mean for our elementary players (N=8, mean grade = 4.25) on the Raven's was 37.7; the 75th percentile for fifth-grade children is 39. The older students (mean grade = 8.3) averaged 53.3 on the Raven's; 54 is at the 90th percentile for 20-year-olds. Clearly the elite players in our sample score above average on the Raven's.[2] The Piagetian task showed no significant correlation with chess skill; in fact, among the seventh- through twelfth-graders, the two highest-rated players had the lowest scores on the Piagetian task. Overall, however, the elite players scored above average. Typically formal operational thought does not begin until adolescence. The task is scored from 0 (concrete thought) to 5 (complete mastery of combinatory logic), with scores in between representing levels of transition. Ss in grades 1 through 6 averaged .89 and those in grades 7 through 12 averaged 4.0. Thus, the elite group appears to be above average in terms of attainment of formal operations as well as on the Raven's.

Results from the game reconstruction task for the elite subsample. According to the analysis of the chess master, the children typically did not look ahead more than one move (even those with ratings above 1500). Adults, on the other hand, often looked ahead five moves. Children as well as adults tended to describe their best moves as those that followed their opponent's weakest moves. In general, pauses in reconstruction corresponded to big decision points in the game and places where players indicated having considered the most alternative moves. Winners knew exactly the turning point in their games, while losers were not as specific in defining the turning point. Children were much worse at reconstructing their games than were adults. They very rapidly reconstructed their game through the opening (the book moves), but slowed down and made many reconstruction errors after that point.

In general, the chess master reported that memory for games reflected some of the chess knowledge of players, but that children seemed to remember less than adults, not look ahead, and not consider as many alternatives. These outcomes suggest that child players rely on deliberate analytic abilities less than do adult players. Child players may develop other skills to compensate for their lower level of analyticity.

PART 2. MEMORY, PERCEPTION, AND SIMILARITY

The second study presents results from three experiments conducted at a state scholastic chess championship. Experiments 1 and 2 are replications and extensions

[2] We have tested seven adult masters and experts on the Raven's Advanced Matrices with interesting results. Six of the seven subjects (including one 66-year-old) had scores at the 99th percentile. With age, scores on the Raven's typically drop dramatically. We suspect that chess-playing may help maintain spatial skill. The one subject who did not perform well on the test (he only got four correct) is known for his calculative, analytic style. While the other players describe their play with spatial and fluid metaphors, he typically describes series of moves.

of the classic memory and board reconstruction studies of DeGroot (1946) and Chase and Simon (1973), which showed a relationship between skill level and the number of pieces that can be recalled from briefly presented scenes from real games. In experiment 1, on half the trials, contextual information was added before presentation of the boards to explore whether this would enhance performance. Experiment 3 follows a similar line. In this experiment we were interested in the extent to which similarity judgements become more abstract as skill increases.

Our hypotheses for the second part of the paper are listed below. Each will be discussed in an introductory section for each experiment.

1. For children, as with adults, better players have better memories for chess boards. In particular, there will be a correlation between skill level and the number of pieces recalled from a chess scene.
2. Because better players' reconstructions will reflect better chess knowledge, a chess master will be able to estimate skill level from the reconstructed boards.
3. Recall of chess scenes will be enhanced by providing a contextual framework to help the subject organize the scene. This contextual framework will consist of non-piece-specific information about the strategic/tactical considerations inherent in the scene.
4. In the similarity judgement task, with increasing skill and age, judgements will be based more on abstract similarity and less on superficial, piece-specific similarity.

General method

Subjects and proceduresе

Participants in the Tennessee State Scholastic Chess Tournament with national ratings over 1100 were invited to do experimental tasks between rounds. Their teams were given one dollar for each player who participated, and individuals were given raffle tickets with a chance to win prizes for each task in which they participated. Subjects first performed the memory/recall task, then the similarity judgement task, and finally the board reconstruction task. Because of time constraints, not every subject completed all three tasks. Table 6 shows the ages and ratings of the 59 subjects who participated. Subjects are divided into primary (grades 1–3), elementary (grades 4–6), junior high (grades 7–9) and high school (grades 10–12). These are the divisions used in scholastic chess. Players may play in the division corresponding to their age or in any higher division. In this sample (as in the sample in study 1), age and rating were significantly correlated ($r=.480$, $p<0.01$).

Table 6. Number of subjects at each grade and skill level

Grade	Rating					Total
	11–1299	13–1499	15–1799	18–2099	2100+	
1–3	4	4	0	0	0	6
4–6	12	5	1	0	0	18
7–9	8	6	4	0	0	19
10–12	5	4	4	2	1	16
Totals	28	18	9	3	1	59

Experiment 1: Context and memory

Chase and Simon (1973) claimed that 'chess skill depends in large part upon a vast, organized long-term memory of specific information about chessboard patterns' (p. 279). Highly skilled players supposedly have vast numbers of chunks stored in long-term memory, whereas less skillful players have fewer and smaller chunks. Examples of chunks include specific configurations such as 'fianchettoed bishops'. Chase and Simon's view, sometimes called the 'recognition–association model of chess skill' has been criticized by more recent researchers (e.g. Charness, 1976; Frey and Adesman, 1976; Craik and Lockhart, 1972, Goldin, 1978; Lane and Robertson, 1979; Holding and Reynolds, 1982; see Holding, 1985 for a review of the debate). Holding, for example, argues that the focus of the recognition–association model de-emphasizes the importance of the search for possible moves. Since which move is best depends on the whole configuration of the board, and not simply on familiar configurations of a few specific pieces, the search for the best move is likely to be guided by chess knowledge that is of a more general and abstract nature. Thus, while research conducted within the recognition–association framework has been instrumental in identifying many of the structural components from which whole board configurations are constructed, further research needs to be conducted in which the more global and abstract properties of the resulting chess positions are emphasized. Selecting moves is the key task in chess, and this can only be done well by considering global properties of the game.

Consistent with Holding, we have observed that much of chess coaching consists of giving children a general, non-piece-specific conceptual framework which they can use to understand chess positions. These frameworks often contain dynamic, strategic information. We hypothesized that if superior memory effects result from some kind of global or non-piece-specific chess knowledge (as Holding and other critics of the recognition–association model believe), then memory for chess scenes can be improved by providing a non-piece-specific contextual framework to help the subject organize the scene. This contextual information should facilitate the player's taking a more 'process-oriented' perspective on the board.

Method

We replicated the non-random board part of the Chase and Simon study, but with one task modification. On half the trials, before seeing the board, a 'context' was introduced: the child was given a brief general comment mentioning the strategic/tactical considerations, but not mentioning any specific piece. Examples include: 'It's white's move. Black has just captured one of white's pieces, and white will probably capture back', and 'It's white's move, and white is ready to attack'.

Sixteen diagrams were presented of games in progress: eight middle-games and eight end-games. In half the middle-game positions an exchange was in progress; this tends to make it more difficult to determine the meaning of the configurations. The other half of the middle-game positions (called 'quiet' positions) had no exchanges in progress and are easier to 'understand' even without the benefit of having seen the progression of the game. Half of the end-game positions were from actual games, and half were positions prepared by Fischer, Margulies and Mosennfelder (1966) to illustrate particular principles. Thus the positions represented a wide range of situations. All were realistic, but some were more difficult to understand than others.

Subjects were shown two practice items followed by 16 stimulus items. Forty-eight different orders were presented; each subject saw eight positions preceded by context (two regular middle-games, two middle-changes with an exchange in progress, two regular end-games, and two Fischer end-games) and eight similar positions with no context. For context items, subjects were read the context statement. Subjects then studied each position for 10 seconds. The diagram was removed and subjects were asked to reconstruct the board from memory. A photographer photographed each reconstructed board to allow later scoring.

In addition, a chess master was asked to analyse the photographs of reconstructed boards for information about error patterns, and to determine whether he could estimate skill level from the reconstructed boards.

Results and discussion

Overall, performance was significantly correlated with both age (r=.272, p<.05) and with rating (r=.266, p<.05). Also, there was an overall effect for context; players correctly placed more pieces when given a context statement (t(47) = 2.04, p<.05). When boards were presented *without the context*, performance was significantly correlated with both age (r=.377, p<.01) and with rating (r=.301, p<.05). When boards were presented with the context manipulation, age and rating were less important. The context 'levelled' the performance, resulting in lower and non-significant correlations, (r=.167 for age and r=.230 for rating). This result means that, with the context, there were fewer age differences and skill-level differences. Context apparently allowed lower-rated and younger players to compensate.

Context effects, age, and type of position. Although the sample size is small, and the numbers of players in each age group varied, the results suggest that the effect of the context depended on both age and type of board position. Subjects may have dealt with context in different ways, depending on whether or not they 'needed' the context. If they recognized and understood the position, context was not necessary; if the position was much too complex and subjects were not able to take advantage of the context, then context did not improve performance. It may be that context facilitates performance only if the child is close to understanding the position. In that case the context may cue recognition, or it may provide enough extra information for the child to be able to analyse the position. Overall, younger children showed more of a context effect than did older players; performance on context items was negatively correlated with age: r=−.290, p<0.05. This result does not tell the whole story, however. The age differences can be attributed mostly to the primary children, but junior high students also benefitted from the context. Elementary and high school players showed no difference between context and non-context positions.

Interestingly, primary and junior high players benefitted from the context on *different* kinds of positions. Although the differences are not significant, the pattern is suggestive. The primary children were helped most by the context in the end-game situations (4.16 more pieces correctly placed) and in the 'quiet' middle-game positions (3.40 more pieces correctly placed). In fact, on the end-games with context, primary-age children remembered as many pieces as did high school players! End-game positions and quiet positions are characterized by abstract, underlying principles which were captured in the context statements. They are the easier of the four kinds of positions to understand. On other, more complex positions (the middle-game positions with exchanges and the Fischer positions), context did not help primary children. Presumably they were not sophisticated enough to take advantage of the context.

Junior high players also found the context helpful. They were most improved, however, on the positions *with exchanges* (3.58 more pieces) and on the Fischer problems (2.56 more pieces). These two types of positions are the more difficult. Here junior high players, unlike primary players, may have been sophisticated enough to use the contextual information. Elementary children showed no overall context effect. High-schoolers found the context to be a hindrance or of no help on all types of positions. The overall pattern suggests that providing a global organizing principle in the form of a context statement may or may not be helpful, depending on the cognitive stage of the child.

Chess master's analysis. In reconstructing the boards, subjects paid special attention to the centre of the board, to the positions of the kings, and to prominent or unusually placed pieces. Fewer errors occurred in these positions and more correctly placed pieces were located within these parameters. The boards were set up so that white was on the subject's side and, not surprisingly, more correctly placed pieces and fewer errors occurred with white.

The master examined boards without knowing the rating of the person who reconstructed the board. The master was asked to rate each subject's set of reconstructed boards on a three-point scale representing how well the reconstructions reflected chess knowledge. His ratings agreed closely with actual skill ratings (r=.68, p<.01). His only misjudgements were for those subjects he rated '2'; two of those he rated as average on ability were actually two unusually high-rated third- and fourth-graders and two were lower-rated high-school players. In other words, most of his few misjudgements were for subjects whose skill deviated from what would be expected for their age. However, all of the subjects he rated as highest and lowest on ability (1s and 3s) were at the top and the bottom of the ratings. A master can identify a player's skill from a reconstructed board unless the player's skill level varies from what is expected based on the player's age. Although knowledge was more important than age in the master's ratings, knowledge or skill level alone did not distinguish reconstruction performance. When age and skill level deviated from expectation (high skill, young age or low skill, older age), the master erred in judgements of skill.

Experiment 2: Board reconstruction

The second experiment is a replication of Chase and Simon's (1973) board reconstruction task. This task was designed by Chase and Simon to separate the effects of memory versus perception. According to Chase and Simon, pieces placed on the reconstructed board between glances represent a perceptual 'chunk'. Presumably players segment the board into 'chunks' which increase in size with skill and, in our study, with age.

Subjects and method

Thirty-five subjects who had participated in experiment 1 participated in this experiment. In the time between tasks, subjects had also participated in the third experiment and had played at least one round in the tournament. The tournament ran for 2 days, and most subjects performed experiment 1 on the first day and experiment 2 on the second day.

Following Chase and Simon, we asked subjects to reconstruct a chess position, but they did not need to rely on memory since they were allowed to look back at the target board as needed. The taks in experiment 1 relied on the subject's memory as well as the way the information was encoded. In an attempt to separate encoding (the perceptual process) from memory, Chase and Simon (1973) used this reconstruction task. Chase and Simon found that masters had larger chunks and could reconstruct a position with fewer look-backs.

Results and discussions

Size of chunks and number of look-backs. The youngest and the lowest rated subjects in our study averaged about 2½ pieces per chunk. The highest-rated subject (a master) was able to reconstruct the entire board with only a single glance, an average of 23½ pieces.The average size of the first 'chunk' was significantly related to grade ($r=.38$, $p<.05$) and marginally to rating ($r=.25$). The larger correlation with grade presumably reflects memory constraints that change with age.

Primary and elementary grade subjects needed about 10½ look-backs to complete the board. Junior high and high school players needed about 8. Contrary to the results for the average size of the first chunk, performance, as measured by number of look-backs, showed only a marginal negative correlation with grade ($r=-.25$) and a significant negative correlation with rating ($r=-.36$, $p<.05$), with more advanced subjects needing to look back fewer times. Here the higher correlation with rating suggests that the better players are able to use their chess knowledge to organize information more efficiently, even if they are young and hence have a shorter memory span. Thus the size of the first chunk may be related more to memory constraints while the number of look-backs may relate more to organization of the board based on chess knowledge.

Experiment 3: Similarity judgements

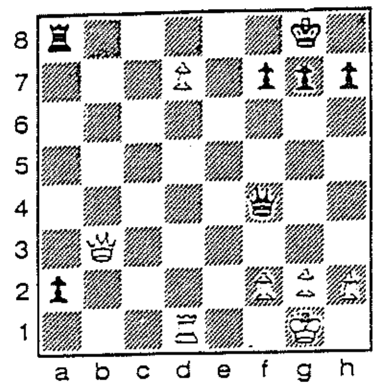
The third experiment examines judgements of similarity between chess positions to determine whether subjects base those judgements on superficial or more abstract features. According to Anderson (1985) and Dreyfus and Dreyfus (1986), a crucial feature of increased expertise is the move from a more concrete to a more abstract representation of features. Thus we predicted that, as skill and age increase, there will be a movement toward judgements based on more abstract features. In this case superficial similarities refer to piece similarities while abstract similarities refer to similarities in strategy.

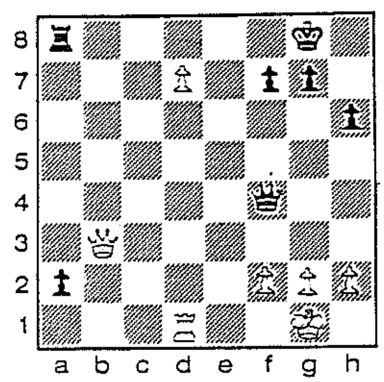
Subjects and method

Forty-seven subjects who also participated in experiment 1 were shown a chess board scene for 45 seconds, the original scene was removed, and the subject was asked to say which of two new boards was more like the original board. Both response and latency were recorded. Each subject saw eight different sets of boards.

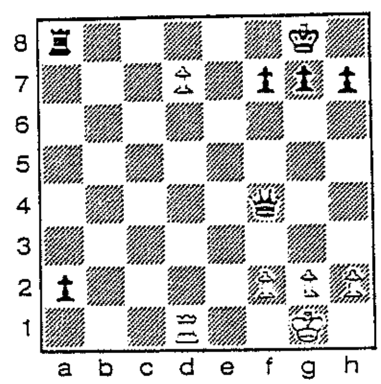
TARGET

"White to move"





Superficially similar



Deeply similar

Forced mate: White pawn moves to d8, becoming a queen. The black rook moves to d8, taking the new queen. The white rook moves all the way to d8, taking the black rook and checkmating black's king.

Figure 3. Examples of stimulus items for the similarity judgement task.

In each set, each of the two subsequently presented boards differed only slightly from the original board (see Figure 3.) One (the superficially similar board) differed from the original board in this respect: one piece from the original board was no longer positioned on its original square, and instead was positioned on an immediately adjacent square. The other subsequently presented board (the abstractly similar one) also differed from the original by only a piece, but was similar in that the same brief sequence of moves led to a forced checkmate. In contrast, the superficially similar board always altered significantly the global strategic features of the board, relative to those of the original board.

Results and discussion

Better players were more likely to choose the deeply similar board (r=.365, p<.05). In fact, all subjects with ratings over 1700 chose the deeply similar boards in all cases. The correlation between grade and deep similarity judgements, however, did not reach significance ($r=.154$). Players who chose the deeply similar boards took more time to reach their decisions ($r=.510$, $p<.01$). These results, along with those from the other experiments, demonstrate that, with increasing skill, children (like adults) move from a more concrete to a more abstract representation of features.

PART 3. CHILDHOOD EXPERTISE AND TRAINING

According to Anderson (1985), experts are able to make better decisions because they have learned through vast experience how to process information in their area of expertise in ways that are more efficient. For example, experts learn to perceive recurring patterns and associate solutions with those patterns, to represent problems in terms of abstract features, and to develop better memories for information involved in the solution of problems. While there is more to expertise than practice, in Anderson's view, extensive practice is essential. If quantity of practice is important for adult expertise, then children's much faster chess skill acquisition must differ in some important way. Watching an adult who has spent most of his leisure time for 20 years playing and studying chess struggle against a 7-year-old who has devoted an hour or so a day to chess for 8 months is a convincing argument against the view that practice *per se* is the key to expertise. Another striking example of high skill without years of practice is with regard to lightning or speed chess, where players are required to make lightning-fast moves. The usual explanation for the masters' and grandmasters' abilities to maintain good-quality games under speed conditions is that their vast experience leads to automatic responses which are based on pattern recognition. While this may be correct for adults, it does not explain how, in 1986, a 10-year-old (non-master) took second place in the Tennessee State Speed Championship against a number of experienced masters. Clearly much more work is necessary to identify the specific experiences that allow children to acquire complex chess skills within a relatively short period of time.

Krogius (1976) suggests that expertise gained as a child differs from expertise acquired later in life. He offers some startling data showing that grandmasters who learned chess as a child played at their peak for more years, and made fewer blunders, than grandmasters who learned chess as adults. He compares early-acquired chess knowledge to a native language; chess was for those players a first language. Holding (1985) suggests that early activity is not only typical of the strongest players but it might be necessary for the highest levels of expertise. He says 'the temptation is to conclude that the early programming of brain function is a necessary prerequisite for chess mastery' (p. 32). One can speculate whether that early programming will affect other kinds of cognitive functioning.

In some superficial ways, children operate like Anderson's experts: they tend to use intuition rather than careful analytic processes, and often ignore many of the details. We have observed children and adults of about the same skill level playing together and it is clear that children play differently. First, and most notably, children play much faster. Consistent with results from other developmental research, child chess players tend to be less reflective and more impulsive than adults. Children typically do not ponder the alternatives in as much detail as do adults. This means that although the adult and child may have similar ratings, the child actually performs as well as the adult in much less time and with much less deliberate analysis. One reason children play faster is because they do not generate long lists of alternative moves—they 'satisfice'. That is, they search until they find a satisfactory move (not necessarily the best move), then cease generating alternatives. Satisficing can be a very useful and efficient heuristic, but it may lead to errors.

Training

Without training or study, few chess players play well. Just learning to move the pieces and playing with other novices results in very slow progress. We visited other schools where enthusiastic teachers who knew little about chess encouraged daily play. Players in those schools had no sense of strategy and very little skill. One of our 7-year-old subjects played a simultaneous match with the entire third grade and remarked that it was difficult for him to 'see' what his opponents were doing because their moves made no sense. The moves, although legal, appeared almost random. What these practised but unskilled players lacked was (1) teaching of principles; (2) process feedback (they only experienced outcome feedback, whether they had won or lost); and (3) specific chess drills. Each of these three topics will be considered.

Teaching of principles

Coaches do not wait for players to discover the principles. They are taught explicitly. Opening systems are memorized and practised. Players are urged to study chess theory. Information is presented as a systematic body of knowledge.[3] When most educators think of gifted and highly motivated students, they assume discovery learning is preferred and memorization is undesirable. What we have found is that young chess players are very adept at and enjoy memorizing openings, learning their names, and classifying them. This pleasure in acquiring a large database is seen, particularly among boys, in collecting information from baseball cards or information about many kinds of dinosaurs. Children's games are usually strongest in the opening, where the moves tend to be more book moves, and principles are rather concrete (e.g. 'move both centre pawns two squares each'). These book moves are not specific piece configurations. They are dynamic systems, with variations and underlying principles and strategies. Children are not typically taught many openings, but each opening consists of several variations and many moves.

Process feedback

Chess offers unusual opportunities for *process* feedback. In tournaments, players write down all their moves. They then replay their games with coaches or other players, trying rejected alternatives and testing what the outcome might have been.

[3] Interestingly, recorded chess games from the previous century are of lower quality than games played today. The reason usually given is that, prior to the existence of a large published body of chess literature, players had to discover principles on their own. Now players have access to a wider, more systematized knowledge base.

Learning to analyse one's own performance objectively provides an excellent lesson in how to maximize skill. In chess, a player has little opportunity to rationalize losses; children learn to be objective about their own performance. In addition, their improvement is readily measured by increased ratings.

Specific chess drills

Chess problems. Chess coaches use a number of interesting training techniques. One is the use of chess problems. Much like case studies constructed for business students, these are problems designed to illustrate a specific principle. Irrelevant details are omitted. Like other kinds of puzzles, they are highly motivating since the learner knows there is a solution.

Speed play and longer analysis time. Paradoxically, players are trained both to play faster and to play slower. Children tend to play fast without much evaluation of alternatives, so coaches have them take more time with moves. In the similarity judgement experiment, longer analysis time was correlated with a deeper level of analysis. But coaches also stress speed training. Children may be especially able to benefit from speed training since it fits their natural tendencies to approach the world in a whirlwind fashion. Playing chess rapidly forces a global perspective, since there is no time to consider details, and hence helps develop intuitions.

Blindfold chess. Another common training technique is to practice playing blindfolded. This forces the player to rely on visualization. When evaluating alternatives several moves ahead, the physical board and pieces can get in the way. The player with good visualization skills can more easily 'see' the board as it might look under different lines of play. Consequently, this practice results in more flexible thinking.

GENERAL DISCUSSION

The following conclusions can be drawn from these studies and observations:

1. Although practice *per se* is not sufficient for expertise, experience is important, especially appropriate training and feedback. Study 1 demonstrated a relationship between level of club activity and improvement in national rating. Some of the characteristics of chess experience that make chess skills learnable by children include: opportunities for process feedback as well as outcome feedback; the opportunity to try out discarded alternatives; unusually accurate feedback about performance; structured, systematic presentation of principles; and special drills that force alternative modes of thinking.
2. For children, spatial and logical abilities such as those measured by the Raven's Progressive Matrices and the Knight's Tour may be useful for identifying chess talent.
3. For children as well as adults, skill level is correlated with the number of pieces correctly remembered from a chess scene.
4. The suggestive age and board type differences in context effects point to the importance of further research on the role of context. Appropriate

COGNITIVE REORGANIZATION AND THE DEVELOPMENT OF CHESS EXPERTISE

DIANNE D. HORGAN, KEITH MILLIS, and ROBERT A. NEIMEYER

Foundations of Education, Memphis State University,
Memphis, Tennessee 38152

One current research strategy in the study of expertise is to compare experts and novices. An important aspect of decision making involves looking for similarities among problem types. Little is known about such processes. We used grid technique to examine similarity judgments associated with different levels of chess expertise. Novice, expert, and master chess players evaluated 4 sets of 12 chess boards. Average FIC scores showed a curvilinear relation to expertise, suggesting increasing differentiation followed by integration in cognitive frameworks for construing board positions. Additional cognitive measures based on move generation tended to support and extend this structural model.

One current research strategy in the study of expertise is to compare experts and novices. Anderson (1985) and Dreyfus and Dreyfus (1986) argue that the stages in the development of expertise are similar across domains. Much of this work, however, has focused on chess players; in fact, Newell and Simon (1972) have referred to chess as the "fruitfly" of cognitive psychology. [See Holding (1985) for a comprehensive review of chess research]. Chess is a complex game that pushes human information processors to the limits of their cognitive abilities. As players gain expertise, they come to represent and organize their knowledge in more efficient ways. One vital question about expertise, then, is how players at different skill levels

A shorter version of this paper (entitled Predecision Processes in Chess: Masters, Experts, and Novices) was presented at the Third International Conference on Thinking in Honolulu, Hawaii, in January 1987 and will appear in the selected proceedings of that conference, published by Lawrence Erlbaum and Associates in 1988.

This research was supported by a Faculty Research Grant to the senior author. We also gratefully acknowledge the support of the Center for Applied Psychological Research, Department of Psychology, Memphis State University. Helpful comments on earlier drafts were provided by Terence Horgan and Stuart Dreyfus. We especially thank our subjects who so patiently participated in our studies and shared with us their chess expertise.

structure and represent their knowledge. How do players perceive and understand new board positions?

We used personal construct theory (Kelly, 1955) and repertory grid technique (Fransella & Bannister, 1977; Neimeyer & Neimeyer, 1981) to explore how construct organization changes across skill levels. We used repertory grid technique because of its usefulness with single subjects, and because we believed that the judgments made by players would be highly idiosyncratic. We wanted to study the idiosyncratic dimensions or constructs that players use to understand elements (in this study, chess board positions), but we also wanted to examine the structure of those dimensions as a function of skill level. In particular, we were interested in assessing construct system differentiation as a function of skill level. A good deal of research suggests that skill is systematically related to level of cognitive development in organizing knowledge about both persons and objects (Adams-Webber, 1979).

SIMILARITY JUDGMENTS

In chess, as well as in other domains of expertise, some of the most important and revealing differences between novice-level and expert-level performance occur when making judgments of similarity (see Dreyfus & Dreyfus, 1986). Although the notion of similarity plays a fundamental role in theories of knowledge and behavior (Kelly, 1955; Tversky, 1977), relatively little is known about how people's similarity judgments change as they acquire expertise. Anderson (1985) and Dreyfus and Dreyfus (1986) argue that experts represent problems in terms of domain-specific features that are more abstract than those features used by novices. In personal construct terms, this may reflect the expert's access to more superordinate dimensions of comparison that identify similarities based on increasingly abstract features of the problems. This process would result in a prototype being representative of a wider range of instantiations as one increased in expertise, or what Mancuso and Eimer (1982) describe as "widening the range of convenience" of a particular construct. Chess offers us an opportunity to examine the kinds of categorization that players use at different skill levels, as well as the degree of hierarchical organization within their systems of categorizations.

To appreciate the role of similarity judgments in chess, consider the process of finding and evaluating moves. Before any evaluation of potential moves takes place, players must *identify* the type of position with which they are faced. They do this by looking for *similarities*

with other positions. This analysis may be at a superficial level or at a more superordinate level, depending on the player's knowledge, goals, and situation. Players then *generate alternatives* for more in-depth evaluation. Presumably, which alternatives are chosen for deeper evaluation results from the process of judging similarities.

Dreyfus and Dreyfus (1986) as well as Anderson (1985) believe that one way in which information processing becomes more efficient as one progresses through the stages of skill acquisition is that one becomes more effective in judging similarities. In their view, when similarity judgments become automatized, problem solving is enhanced in at least two ways: (1) similarities to other known problems are noticed, and problem solving can occur through analogy with those known problems, and (2) capacity is freed up for more strategic processing. That is, experts are able to quickly classify new problems on the basis of similarities to already known problems. This early stage of analysis and problem classification is sometimes referred to as the predecision stage where progressing is more automatic and less analytic. [See Holding (1985) for a thorough discussion of this shift to more automatic processing by chess masters and how it frees up limited information processing capabilities for deeper analysis and strategic thinking.]

This series of studies attempts to uncover some of the ways in which chess players at different stages of expertise judge similarity and generate alternatives. The first study considers players' judgments of similarity for board positions. The second study (which includes two experiments) examines candidate move generation by tournament players as a function of their skill level. We used players' U.S. Chess Federation ratings to determine skill level. These ratings are based on tournament play and reflect win/loss records against players of varying skill ratings. Three levels of players served as subjects: (1) *novices* (players just below the mean for U.S. tournament players), (2) *experts* (players in the top 7% of U.S. tournament players), and (3) *masters* (players in the top ½% of U.S. tournament players).

STUDY 1: COGNITIVE STRUCTURE FOR SIMILARITY JUDGMENTS

This study was designed to allow us to examine the content and structure of individual players' similarity judgments. We used repertory grid technique (1) to determine the dimensions that different players use to understand board positions, (2) to measure the independence of the elicited dimensions, and also (3) to measure the extent to which various boards are seen as similar along the elicited dimensions. We hypothesized that the differences in structure and content at different skill levels would reflect differences analogous to those discussed by Anderson and by Dreyfus and Dreyfus. We also hypothesized that the differences in content and structure at different skill levels would be reflected in differences in move generation in Study 2.

METHOD

Subjects

Given the time-intensive nature of this study (over 5 hours per subject) and the difficulty of recruiting players at the highest skill levels, only 6 subjects participated: 2 novices, 2 experts, and 2 masters. A master has a skill rating about 2200; one of our masters was rated approximately 2250 and the other 2500, where 1500 is the mean for all tournament players and 200 represents one standard deviation. One of the novices was a graduate student in psychology who had played occasionally since childhood and had played seriously about one year. The other novice was a practicing psychologist who had played recreational chess most of his life and had begun playing in tournaments about one year prior to the study. Both experts were in their late twenties to early thirties and had played regularly in tournaments since their teens. The lower master was in his early thirties and had played competitively since his teens. The higher-rated master (actually a Senior Master as well as a FIDE Master) was Stuart Rachels, who was 16 at the time of the study. He started playing at age 9, and became the youngest American master at 11. He is currently one of the top players in the United States. In this sample, then, rating and number of hours spent on chess are not perfectly related; the two 1400-level players had less tournament practice, but among the experts and masters there appears to be no relationship between experience and rating.

Procedure

Materials: Repertory Grid Technique

Forty-eight games were chosen from fairly obscure sources such as state chess association newsletters. Positions were chosen from

these games at fairly random points, and pictorial representations of those positions were then prepared. Sixteen were at fairly early points in games, 16 were at midpoints, and 16 were in fairly late stages of play. The boards were not meant to represent any particular style of play or chess-relevant dimensions; rather, they were meant to be an essentially random sample of positions from real games.

In preparation for the repertory grid administration, the experimenters grouped the boards into 4 sets of 12 boards each. Set A consisted of four early games, four middle games, and four late games. Set B was made up of all early games; Set C, all middle games; and Set D, all late games. The first step in the grid procedure consisted of eliciting subjects' personal dimensions for forming similarity judgments about the boards. In an individual administration, each subject was shown three boards and asked to name one way in which two of them were similar but different from the third. This procedure was repeated 12 times. The dimensions that the participant named were entered on a grid form, with the term describing the way in which they were similar on the left and the term describing the way in which the third was different on the right. The presentation of the sets of three was counterbalanced so that each board was presented three times. After seeing 12 sets of boards to elicit 12 dimensions, the subject was asked to rate each of the 12 boards on each of the 12 dimensions. A 13-point scale, ranging from -6 to $+6$, was used, with a zero indicating the neutral point. Figure 1 gives an example of a completed grid for one of the subjects. This procedure was repeated with the other 3 sets of 12 boards. The completion of all four grids required approximately five hours per subject. Subjects were paid for their participation. Further specifications for "ratings grid" procedures are described by Neimeyer and Neimeyer (1981).

The grids were scored for a number of features using the Reptest Scoring Program devised by Landfield (1977, 1983). Most important for our purposes was the calculation of the "functionally independent constructs" (FIC) score. This score represents the number of functionally independent clusters of dimensions used by each person, with the independence being measured by the degree of dissimilarity of element (board position) ratings. A high FIC score, representing dimensions that are mostly independent of one another, suggests a high degree of cognitive differentiation—all boards are judged to be different. A very low FIC score, on the other hand, indicates that the person is not differentiating boards along many separate dimensions. There are two quite different reasons why this

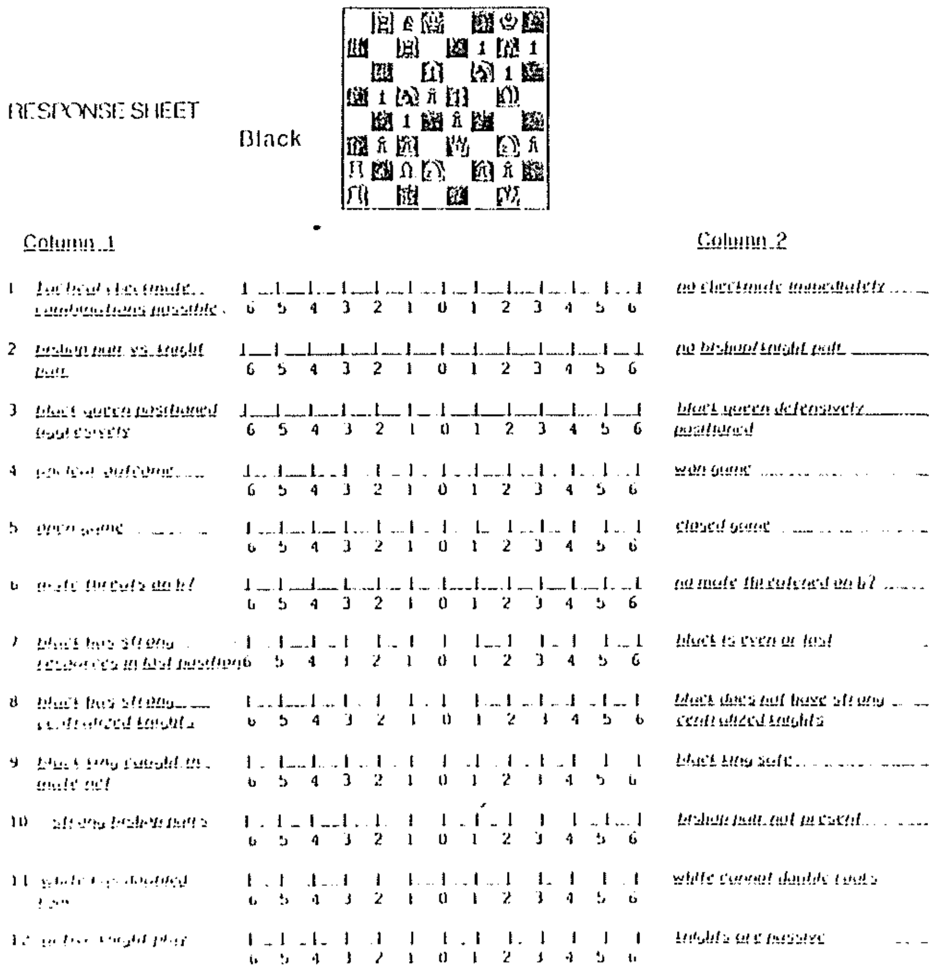


FIGURE 1 Example of a completed grid.

might be so: either (1) because the person's judgments are relatively global and noncomplex, compared with the judgments of those with high FIC scores, or (2) because the judgments are more abstract and integrative than the judgments of those with high FIC scores. Thus, a low FIC score can be interpreted most accurately by considering the content of the constructs involved.

Results

Functionally Independent Constructs

The highest possible FIC score would be 12, since subjects generated 12 dimensions per set. An FIC score of 12 would then show that all dimensions are orthogonal (each construct is separate and statistically independent). Lower scores indicate fewer constructs, and therefore more correlated dimensions.

These results suggest that there is a curvilinear relationship between rating and the mean number of functionally independent constructs; the number of independent constructs increases from novice to expert, but then decreases from expert to master. For each subject, the mean number of FICs was computed by averaging over the four different sets of boards. Novices had a mean of 8.5 independent clusters, experts a mean of 10, and masters a mean of 5.1. The differences in the means were significant according to a one-way analysis of variance ($F(2,3) = 9.05$, $p = .05$). Also, the quadratic component was reliable ($F(1,3) = 8.17$, $p = .06$).

One way to interpret these data is that constructs are moderately interrelated in the novice stage because the novice does not yet know enough to identify all the relevant features and has a simplistic understanding of their relationships. The novice only uses a few constructs of board positions. In comparison, the experts have access to more constructs and features; hence their constructs are more differentiated. The expert "sees" a lot in each board. From the expert's perspective, the boards are extremely complex. Finally, masters show a high degree of interrelatedness in their similarity judgments, presumably because their construct clusters are at a more "abstract," superordinate, hierarchial level. The masters have, in some sense, simplified and integrated their perceptions. Figures 2–5 show examples of sets of constructs from a novice, expert, master, and senior master. The labels given to the constructs were generated by a master who did not participate in the study. He looked at all the dimensions that formed a cluster and tried to identify the basis for their similarity. Subjects confirmed his analysis. Those labels presumably reflect the superordinate category.

The difference between the novice and the master representations is that the master's lowest dimensions are highly meaningful and related to the superordinate category in an important and ab-

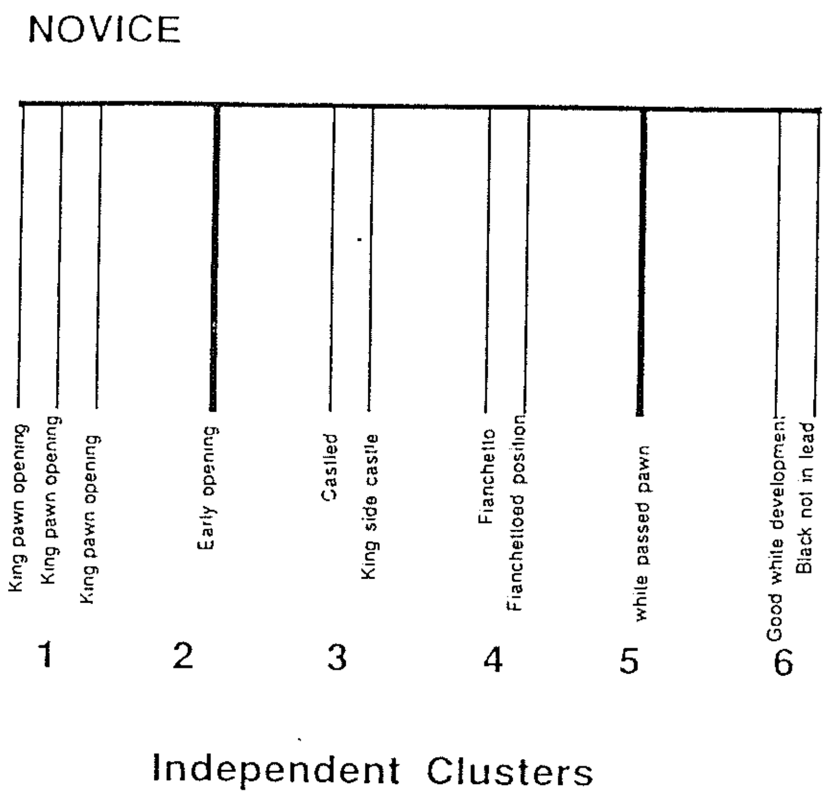


FIGURE 2 Example of dimensions *from* a novice.

stract way. The master's dimensions constitute a hierarchy. In the novice representation, the correlated nodes are redundant (indicated by the thin lines).

An examination of the content of the other dimensions and the constructs used by players at each stage supports this interpretation. In those other dimensions and constructs, as in the examples in Figs. 2-5, we found that the progression from novice through expert to master status begins with constructs of some generality based on insufficient knowledge (novice), followed by detailed and specific constructs (the experts), followed by much more hierarchically organized or abstract constructs (the masters). In general, the master who analyzed the clusters and the players themselves agreed that the constructs generated by novices refer to the very specific piece configurations, and the correlations appear to result from co-occurrences that are specific to these particular boards. They do not represent

sophisticated chess knowledge. Constructs generated by experts were based on general chess features rather than on features specific to these boards. Almost all their construct "clusters" contained only a single dimension, since each dimension carried independent information about the boards.

The two masters produced very different kinds of constructs. The lower master's constructs represented very abstract, superordinate notions such as degree of complexity or strength. In general, for the senior master, each dimension was quite specific, but the relationships within each construct were general and abstract, as in one construct in which the three correlated dimensions were judged by

FIGURE 3 Example of dimensions from an expert.

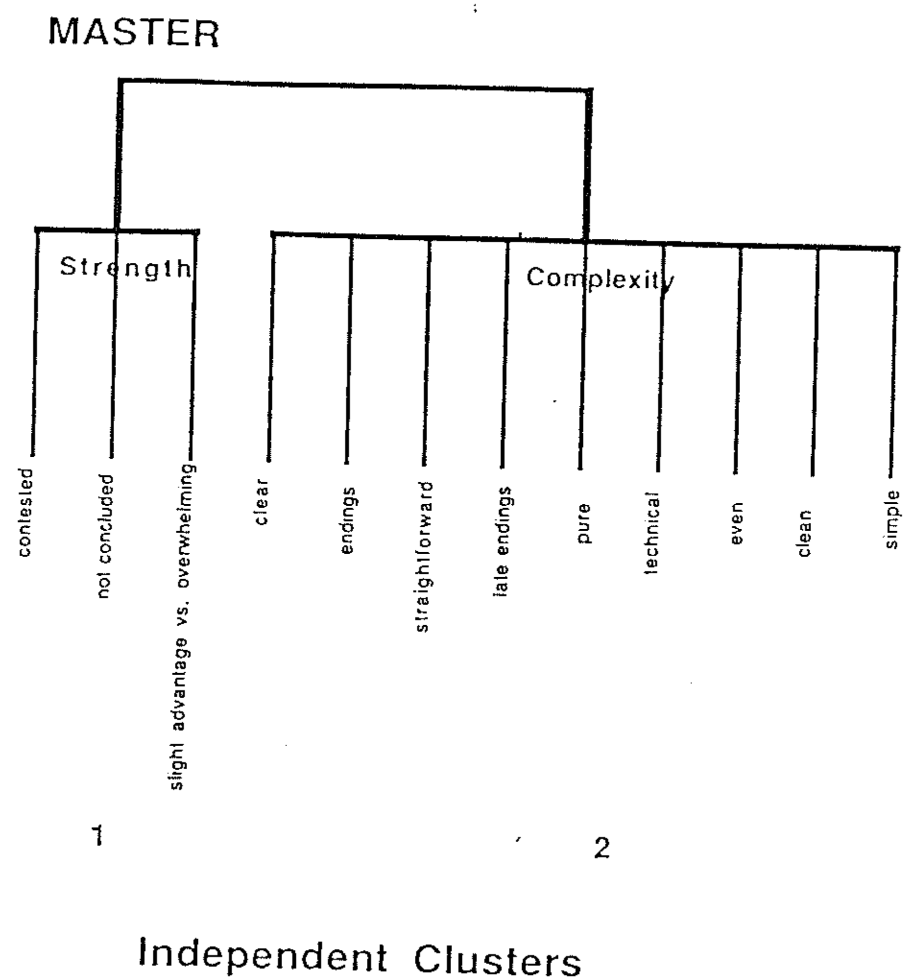


FIGURE 4 Example of dimensions from a master.

the nonparticipating master to be similar in that white had penetrated black's squares unopposed. If this analysis is correct, then the masters will be able to access and manipulate their categories in a more efficient manner while the experts' search processes will take longer. In addition, the masters should also generate fewer candidate moves since they have "pruned" or simplified their perception of the board. The experts' hierarchies, in contrast, are quite wide, with more branches to consider.

This analysis closely parallels Dreyfus and Dreyfus's description

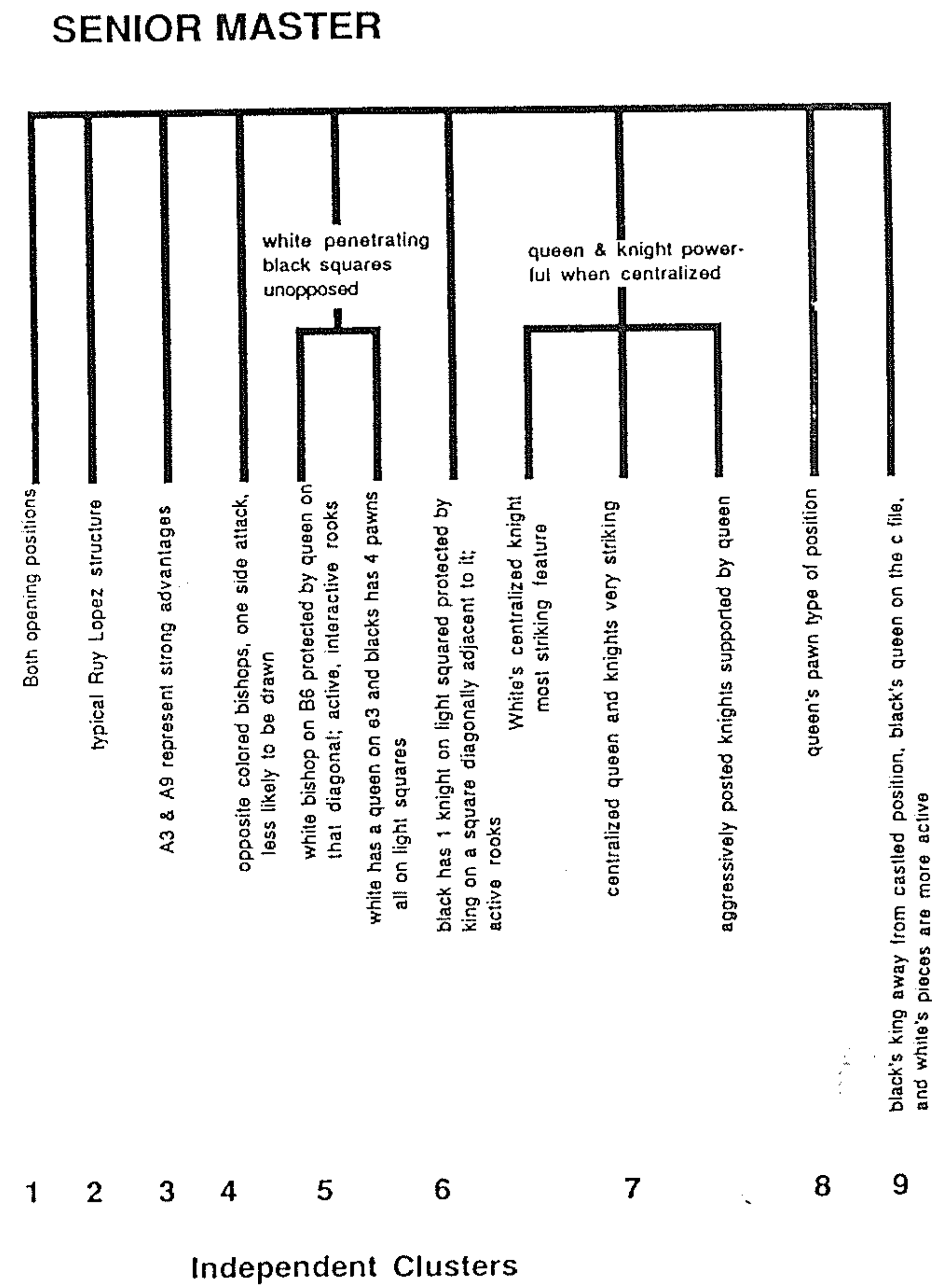


FIGURE 5 Example of dimensions from a senior master.

of stages in the development of expertise. In their view, advanced beginners (our novices), through experience, learn to recognize domain-specific situational features, and learn procedural rules that cite such situational features. Later, learners show more flexibility in their use of rules. The stage of proficiency (our masters), is characterized by the emergence of "holistic similarity recognition." The performer spontaneously sees the current situation as similar in certain salient ways to previous ones, and thereby spontaneously sees an appropriate organizing plan. In Anderson's terms, the players have reached the autonomous stage in which domain-specific procedures are employed more rapidly and automatically.

Stages of the Game

One striking difference between novices' dimensions and masters' dimensions was in their explicit references to the stage of the game. All players spontaneously used stage of the game to describe similarities. Novices' categorizations of early, middle, and endgames were consistent and related to the number of pieces on the boards. Consistent with both Anderson and Dreyfus and Dreyfus, they relied on concrete features that were independent of the context. (For example, the opening ends when either side castles.) The two masters, in contrast, used the terms "early," "middle," and "end" in a much more flexible way. For example, one scored a position with only 6 pieces with an extreme score of (6) on a dimension "middle game" and one with 28 pieces with a (6) on a dimension "endgame." When questioned about this, the master remarked that an endgame is "when you are thinking endgame strategy and is not directly related to number of pieces." The novices, in contrast, defined openings and endgames with rules that were not sensitive to the context.

We also examined the number of functionally independent constructs as a function of stage of the game. In chess, early, middle, and endgames are played with very different strategies. Players typically have differential strengths and weaknesses for each stage.

So far we have considered differences among players at different levels; however, differences in the task demands at each stage of the game should also produce different knowledge representations. Although the differences among means fail to reach significance when submitted to a one-way repeated measures ANOVA, the curvilinear pattern exists: there were 8.2 constructs in the set of boards with the early games, 8.8 in the middle games, and 7.8 in the late games. (The number of functionally independent boards showed the same pattern: 8.3 in the early games, 8.7 in the middle games, and 8.0 in the late games.) Figure 6 shows the number of constructs for each stage by each skill level.

During the opening stage, the moves tend to be standard moves described in books, and principles are rather concrete (e.g., "move both center pawns two squares each"). These standard openings have names and are classified by types. Players at all levels used names of openings for dimensions in this set. The contrast in stages of the game was minimal (but curvilinear) for experts: both early and late games had 10.5 constructs, versus 11 for middle games. This again suggests a great deal of detailed knowledge. Novices also showed a curvilinear pattern. Of the three stages, they had the fewest FICs for openings: 7.5 constructs for early games, versus 10 and 9.5, respectively, for middle and late games. Presumably the novices' overall lack of knowledge was most pronounced in the early stages where skill is most closely tied to memorization of standard openings. They had fewer constructs because they knew fewer openings. Masters showed a decrease in the number of FICs as the game progressed: 6.5 for openings, 5.5 for middle games, and 3.5 for endgames. Because of the more concrete nature of opening knowledge, it may be more difficult to represent constructs in as abstract a way as knowledge of later stages. Higher levels of skill (and hence more integrated and abstract knowledge representation) are needed in the middle and endgame. The dramatic drop in the number of constructs for late games among the masters may reflect the tendency for excellent players to think about their endgame from the earliest stage of play. Masters typically talk in terms of what kind of endgame they want to play and strategies to achieve that goal. Thus, their endgame knowledge and strategy may be the most integrated and abstract. Since it underlies much of their earlier play, it makes sense that it would be represented by fewest constructs. Fewer constructs combined with extremely high skill presumably indicate a more intuitive and automatic processing mode. Thus the master's declining number of constructs as the game progresses may indicate that their strategies become more abstract and integrated as the game progresses. For them, the game may start with standard openings classified hierarchically in a number of ways, then move to a more abstract set of principles in the middle game, and finally, in the endgame it may boil down just to a few abstract principles such as complexity or force.

If we look at the content of the senior master's constructs we see that kind of progression. Table 1 provides an example from each stage. For the highest skilled player, the constructs in the early games focused more on very chess specific knowledge—standard

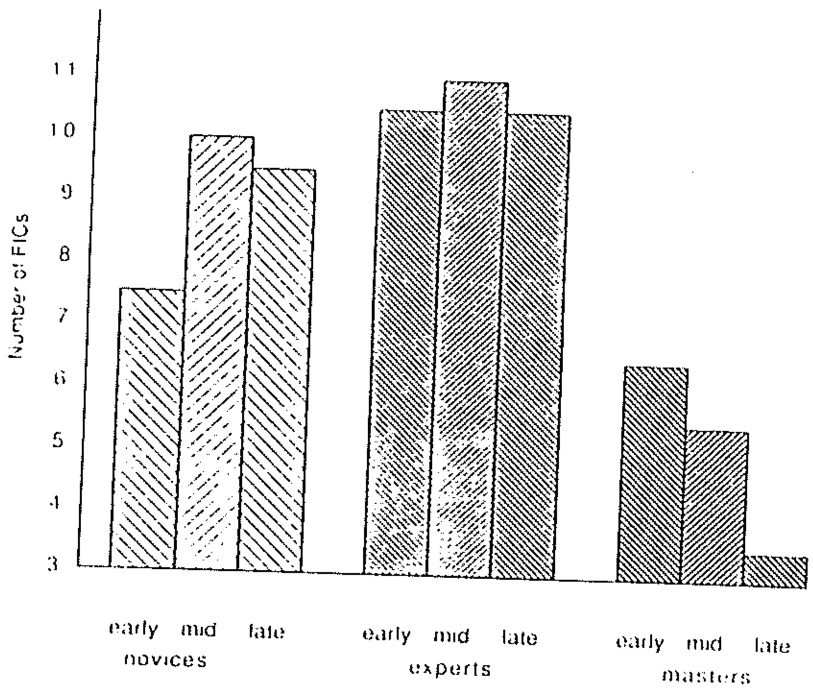


FIGURE 6 Number of FIC's by stage of game and skill level

TABLE 1 Constructs from a Senior Master

Early games[a]	Middle games[b]	Endgames[c]
e4 Opening	White queen and light squared bishop can combine to attack	Pure endgame involving pawn races
Not d4 opening	Very complicated with each side having two knights and a bishop	Even on material
Not benko gambit		Even on material and endgame

[a]This construct can be characterized as "d4 openings" versus "not d4 openings."
[b]This construct relates to an attack on white squares. Several other constructs used by this player related to the color of the squares, which is a much more abstract feature of the game than what is typically attended to by less skilled players.
[c]This construct focuses on the relative strength of the two sides.

openings. In the middle game, the constructs also are chess-specific, but do not refer to specific pieces or specific squares. In that sense, they are more abstract. In the late games, the constructs focus on general concepts such as strength and complexity that are not specific to chess.

Boards

Another way to understand the progress from redundancy to integration is to examine, for each individual player, the extent to which that player judged the various boards to be different from one another. Thus, while the FIC constructs scores indicate the number of independent constructs used to describe the boards, the FIC board scores indicate the number of independent *boards*. The same curvilinear pattern was found; means were 8.7, 9.9, and 5.2 for the number of functionally independent boards for novices, experts, and masters, respectively ($F(2,3) = 7.3$, $p = .07$). Thus masters saw the boards as more similar than did either novices or experts; this accords with the parallel result concerning FIC constructs scores, namely, that the *dimensions* they used to describe the boards are more similar.

Although each player saw some boards as similar to other boards, there was little consistency across players as to which boards were similar. Out of 48 boards, only 8 pairs of boards were highly correlated for two or more players. These were very early positions that were standard openings. Thus, while the structure of the subjects' judgments followed a clear general progression, each player has an idiosyncratic way of analyzing the boards and of judging similarities.

In summary, Study 1 suggests a developmental progression in domain-specific cognitive structure associated with expertise in chess. With increasing expertise, players moved from more specific and unsophisticated similarity judgments regarding board positions (novices) to more general, but differentiated, construals of similarities (experts). At still higher levels of expertise, masters attained considerable consolidation of their dimensions through the use of highly abstract dimension clusters focusing on strategic features of the games. And, at the highest level of expertise that we examined, the senior master appeared to generate dimensions that were specific but abstractly interrelated, while retaining the level of cognitive consolidation characteristic of the lower-rated master's responses. Further research is needed to determine whether the development of expertise is curvilinear as our data suggests, or whether our senior master's trend toward more dimension clusters may represent a new period of differentiation, which might be followed by yet another period of integration, perhaps at the grandmaster level.

Study 1 also demonstrated the usefulness of repertory grid technique for the study of expertise. Although the *content* of the dimensions showed great idiosyncrasy, grid technique allowed us to examine group trends in the *structure* of these judgments.

We hypothesized that move generation would be compatible with judgments of similarity. In particular, players in groups characterized by the use of more independent dimensions ought to generate more moves. Study 2 tests that hypothesis.

STUDY 2: MOVE GENERATION

Experiment 1: Number of Moves

Several questions are of interest. Psychologists (e.g., Parnes & Meadow, 1959; Osburn, 1962; Parnes, 1961) have often assumed that "quantity breeds quality." This notion is supported by numerous reported positive correlations between fluency of ideas and quality of ideas. Researchers have also reported that creative figures such as Thomas Edison had an unusually high and fast rate of idea production (Thurstone, 1952). The prediction, then, according to this hypothesis would be that masters would produce more candidate moves than less skilled players. This quantity of production would be a key to their superior performance. DeGroot (1946), however, reported no differences in the number of candidate moves generated by masters versus other players. (The masters' moves were, of course, superior). We sought to determine whether DeGroot's finding holds for all stages of expertise. We also wanted to examine the content of the candidate moves generated.

Method

In this study, 35 players (15 novices, 14 experts, and 6 masters) were asked to examine 11 board positions for not more than 60 seconds each and identify candidate moves—moves that they would consider making and would evaluate further. The 60-second time limit avoided long periods of in-depth analysis and evaluation. We asked a master (who did not participate in the study) to group all the moves from all the subjects for each board position into categories

reflecting goals or general strategies. For example, one strategy might be to gain control of the queen's side and there might be a number of moves that would further that goal. The master had no difficulty grouping moves into goals. Thus we were able to count not just the number of moves generated by players of different strengths but also the number of goals they considered.

Results

Figure 7 shows these general results for subjects divided into masters, experts, and those below expert.

To assess whether the level of chess expertise has any impact on the number of generated candidate moves and the number of goals used, we performed a 3 (rating levels) by 11 (boards) mixed analysis of variance, with rating as the between-subjects factor and the boards as the within-subjects factor. There were significant differences among the three rating levels for both the number of generated moves and the number of goals, $F(2,32) = 4.44$, $p < .05$, and $F(2,32) = 8.12$, $p = .001$, respectively. For both moves and goals, the means showed the same inverted U pattern as appeared in Study 1.*

Expert-level players "see" everything and hence generate the most candidate moves, the most goals, and the most moves per goal. Apparently the masters have already, in the predecision stage of automatic processing, eliminated some of the moves and goals. The masters have "freed up" some information processing time on the front end. Interestingly, looking only at masters versus all other players as a group (combining experts and novices), we found the same pattern as DeGroot: an average number of moves of 2.12 for masters and 2.11 for lower-rated players. The sharp differences in goals and moves generated by experts and those below experts clearly suggests that they should be analyzed separately. DeGroot's combination of these groups may have obscured the curvilinear relationship.

*Interpretation of results for the number of moves is complicated by a significant rating by board interaction ($F(20,320) = 1.98$, $p = .03$). Inspection of the cell means indicated that this can be attributed to two boards on which experts and masters generated the same number of moves, which was more than the number generated by the novices. On the other nine boards the curvilinear relationship held. Because of the possible effects of different boards, we replicated our findings with another set of board positions and another group of subjects in the second experiment of Study 2.

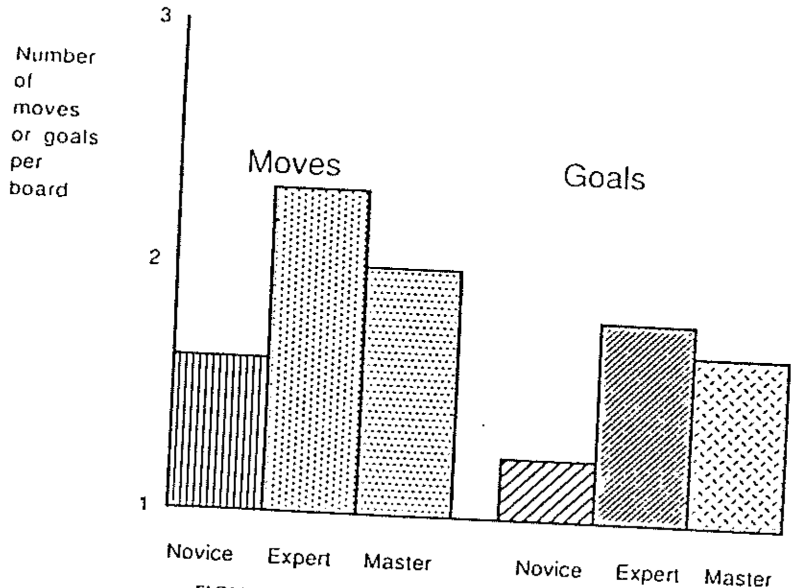


FIGURE 7 Moves and goals by skill level.

Overlap Scores

For each player, we computed how much agreement or "overlap" there was between his candidate moves and those of all the other players. For computing each overlap score (for subject X and subject Y) we used the following equation (taken from Graesser, 1981):

$$\text{Overlap score} = \frac{2 \times \text{number of common moves}}{\text{total number of moves generated}}$$

An overlap score ranges from 0 to 1.0 in magnitude and can be interpreted as the proportion of agreement on the generated moves between the two players. The means presented in Table 2 were computed by averaging over all overlap scores of subject n with every subject in rating level j. The means on the diagonal refer to the degree to which players within each rating level agree with each other. Both masters and experts tend to agree among themselves roughly one-half of the time on the moves that they consider, whereas novices agree among themselves only one-third of the time. Off-diagonal scores indicate agreement between levels. The means suggest that

TABLE 2 Degree of Overlap in Generated Moves by Skill Level

	Means of overlap scores		
	Master	Expert	Novice
Master	.49	.49	.37
Expert		.50	.37
Novice			.32

not only do novices agree among themselves one-third of the time, but they also agree to the same extent with masters and novices.

These results indicate that as a group, novices show less agreement among themselves about the candidate moves than do masters or experts. Masters and experts agree with each other significantly more often than they agree with novices. So despite the highly idiosyncratic dimensions among experts and masters, those dimensions seem to lead to similar moves.

Experiment 2: Response Time

If, in fact, masters generate more "automatic" moves than novices, then one would expect that masters take less time in generating moves than novices. Experiment 2 tests this hypothesis.

Method

The subjects were 12 players who had not participated in the previous study (3 novices, 7 experts, and 2 masters). They were asked to generate moves for 10 positions (positions different from the ones in the previous study), without taking time to evaluate each move. Besides recording the number of candidate moves, we also timed how long they took to generate each move.

Results

Table 3 shows the average number of moves per board and the average time to generate a move.

Although the effect of rating level on the number of generated candidate moves was not statistically significant ($F(2,9) = 2.00$, $p = .19$) when submitted to a rating level by board ANOVA, the means are in the same curvilinear pattern as in the previous experiment. In

TABLE 3 Average Number of Moves and Time to Generate Moves by Skill Level

	Number of moves	Time per move (minutes)
Novice	2.83 (1.24)[a]	.40 (.19)
Expert	4.31 (1.71)	.22 (.09)
Master	3.10 (.99)	.23 (.16)

[a]Standard deviations are shown in parentheses.

fact, for 9 of the 10 boards, the means across the 3 levels show the curvilinear pattern ($p = 0.21$, with a sign test). There was a main effect due to rating on the average time spent per move ($F(2,9) = 7.37$, $p < .05$). As predicted, novices took a longer time to generate moves than both experts and masters, who did not differ from one another.* The fact that masters and experts, in general, generate moves more quickly indicates that their representations of chess knowledge and their processes of making similarity judgments are indeed more efficient than novices.

CONCLUSION

Levels of chess expertise evidently differ from one another in important qualitative ways, rather than as a linear function of the amount of experience and practice. Our results on similarity judgments and on move generation are nonlinear and yield an inverted-U curve. This suggests progression through distinctive cognitive stages, with major reorganizations of knowledge resulting in the qualitative differences in performance at each stage. Progress through these stages appears to involve an increase in the differentiation of context-sensitive, nonredundant dimensions as one moves from novice to expert, followed by an increased ability to integrate and abstractly consolidate those dimensions as one moves from expert to master. This sequence of construct differentiation and subsequent hierarchical integration appears to reflect a general process underlying construct system development (Adams-Webber, 1979). Experts are more adept than are novices in their information processing; hence they perceive more nonredundant dimensions of similarity, more alternative goals, and more candidate moves than do novices. Masters, on the other hand, appear to have relegated even more of their processing to an automatic phase than have the experts. Hence the masters' similarity dimensions are more integrative and more hierarchically organized than those of the experts, and the master's goals and candidate moves are fewer in number than those of the experts, reflecting more thorough pruning in the early stages of analysis.

If these conclusions are correct, then expertise in chess is much more than storing large amounts of domain-specific information obtained through extensive experience. The overlap results demonstrate that as skill increases, there is more agreement on the output—which candidate moves to consider. However, the wide variation in the content of the dimensions and the lack of agreement on board similarity demonstrate that wide differences during processing can converge on similar moves. Because of these wide individual differences, repertory grid technique is especially effective as a research tool. The expertise literature asserts that the process of becoming an expert is similar across different knowledge domains (e.g., Anderson, 1985). If this is true, then we would expect our results to generalize to other domains of expertise.

*There was a rating by board interaction ($p < .05$). The interaction occurred because the masters took a longer time than the novices to generate the moves for two boards. According to the master who served as our consultant, these boards appeared superficially simple, but actually had complex underlying themes. The underlying themes went undetected by the novices. Since 8 of the 10 boards showed that masters and experts took less time to generate moves than novices, we believe that, in general, our results hold. Obviously some boards will follows a different pattern because of the deep complexity of some positions.

REFERENCES

Adams-Webber, J. R. (1979). *Personal construct theory: Concepts and applications.* New York: Wiley.

Anderson, J. R. (1985). *Cognitive psychology and its implications.* New York: W. H. Freeman.

DeGroot, A. D. (1946). *Het Denken van den Schaker.* Amsterdam: North Holland.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine. The power of human intuition and expertise in the era of the computer.* New York: The Free Press.

Fransella, F., & Bannister, D. (1977). *A manual for repertory grid technique.* London: Academic Press.

Graesser, A. C. (1981). *Prose comprehension beyond the word.*New York: Springer-Verlag.

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

Kelly, G. A. (1955). *Psychology of personal constructs.* New York: Horton.

Landfield, A. W. (1977). *Nebraska Symposium on Motivation* (pp. 127-178). Lincoln, NE: University of Nebraska Press.

Landfield, A. W. (1983). *Reptest Scoring Program.* Unpublished manuscript, University of Nebraska.

Mancuso, J. C., & Elmer, B. N. (1982). Fitting things into sorts: The range corollary. In

J. C. Mancuso & J. R. Adams-Webber (Eds.), *The construing person*. New York: Praeger.

Neimeyer, G. J., & Neimeyer, R. A. (1981). Personal construct perspectives on cognitive assessment. In T. Merluzzi, C. Glass, & M. Genest (Eds.), *Cognitive assessment* (pp. 188–232). New York: Guildord.

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

Osburn, A. F. (1962). Developments in creative education. In S. J. Parnes & H. F. Harding (Eds.), *Source book for creative thinking* (pp. 20–29). New York: Charles Scribner's Sons.

Parnes, S. J. (1961). Effects of extended effort in creative problem solving. *Journal of Educational Psychology, 52*, 117–122.

Parnes, S. J., & Meadow, A., (1959). Effects of brainstorming instructions on creative problem-solving by trained and untrained subjects. *Journal of Educational Psychology, 50*, 171–176.

Thurstone, L. L. (1952). Creative talent. In L. L. Thurstone (Ed.), *Applications of psychology* (pp. 18–37). New York: Harper & Brothers.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Requests for reprints should be sent to Dianne Horgan, Foundations of Education, Memphis State University, Memphis, TN 38152.