

Templates in Chess Memory: A Mechanism for Recalling Several Boards

Fernand Gobet and Herbert A. Simon

**Department of Psychology
Carnegie Mellon University**

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

Address of correspondence:

Herbert A. Simon
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

Running head: Templates in Chess Memory

Abstract

This paper addresses empirically and theoretically a question derived from the chunking theory of memory (Chase & Simon, 1973): To what extent is skilled chess memory limited by the size of short-term memory (about 7 chunks)? This question is addressed first with an experiment where subjects, ranking from class A players to grandmasters, are asked to recall up to 5 positions presented during 5 seconds each. Results show a decline of percentage of recall with additional boards, but also show that expert players recall more pieces than is predicted by the chunking theory in its original form. A second experiment shows that longer latencies between the presentation of boards facilitate recall. In a third experiment, a Chessmaster gradually increases the number of boards he can reproduce with higher than 70% average accuracy to nine, replacing as many as 160 pieces correctly. To account for the results of these experiments, a revision of the Chase-Simon theory is proposed. It is suggested that chess players, like experts in other recall tasks, use long-term memory retrieval structures (Chase & Ericsson, 1982) or templates in addition to chunks in STM, to store information rapidly.

Templates in Chess Memory: A Mechanism for Recalling Several Boards

A regrettable finding of cognitive psychology is that the human cognitive system is full of severe information processing limits, in particular: a limit on the short-term memory (STM) capacity (about seven chunks), a limit on the amount of information that can be learned in a given time (about seven chunks in one minute) and a limit on the rate of searching through problem solving states (perhaps 10 states per minute). A felicitous finding of cognitive psychology is that, as experts in various domains demonstrate to us, these limits may be (partly) circumvented: trained subjects recall up to 100 digits dictated at a pace of 2 seconds each, physicists spot rapidly the solution to a difficult problem, and chess masters are able to play simultaneous games when blindfolded.

A chunking model of chess memory

Research in cognitive psychology has pinpointed three features of expertise that are visible across domains: the importance of pattern recognition, the importance of selective search, and rich knowledge in the domain of expertise. A first attempt to tie these three features together in a single theoretical framework was made by Chase and Simon (1973b). These authors applied their theory to the game of chess, a domain that has historically provided a rich source of data and insight for studying expertise (Charness, 1992).

Structure of the Model

Skilled chessplayers (especially Masters and Grandmasters) are able to retain almost complete memory of unfamiliar chess positions shown to them for only a few seconds. To explain this performance, which does not extend to pieces arranged randomly on the board, Chase and Simon (1973a, 1973b), making use of the EPAM model of perception and memory (Simon & Feigenbaum, 1964), proposed that, in the course of acquiring their skill, chessplayers stored chunks in long-term memory (LTM) corresponding to patterns of pieces. Each chunk consists of a small pattern that recurs frequently in the chess positions encountered while playing. To account for their performance, assuming that only about a half dozen chunks can be held simultaneously in short-term memory, skilled players would have to accumulate a store of chunks estimated roughly at 50,000 or more (Simon & Gilmarin, 1973).

The chunking model also provided a theory of the processes underlying chess skill. Skill, according to this theory, has two main components: ability to search the tree of possible moves and their potential consequences highly selectively, and ability to evaluate positions and to discover potentially strong moves. Both abilities are based on recognition of features (familiar chunks) on the chessboard. The search of the skilled player is guided by heuristics, or rules of thumb, that permit it to be restricted to a small tree of possibilities (usually less than 100). The heuristics, in turn, rest upon recognition of familiar patterns or chunks. Recognizing familiar chunks reduces the need for "look-ahead" search by giving access, for example, to information stored in memory about moves that may be advantageous when that chunk is present. (E.g., "If there is an open file, consider moving a rook onto it."). Rapid recognition of patterns or chunks is also essential for evaluating the positions at the termini of searches. This recognition and memory retrieval capability explains how skilled players are able to play rapid-transit or simultaneous games at a relatively high level of competence -- games that do not allow time for much look-ahead search (Gobet & Simon, in press-a). At the same time it

accounts for the role of selective search in playing more deliberately, at a higher level of skill.

More specifically, Chase and Simon's model explains the main qualitative results in the recall task (Masters' superiority with game positions, but minimal skill differences with random positions) in the following way: during the presentation of a position taken from a game, familiar patterns of pieces on the board are recognized as chunks, and a pointer to these chunks is placed in STM, the size of which is limited. Because strong players possess both more and larger chunks in LTM, they recognize more and larger patterns on the board, and therefore recall more pieces from the positions. In the case of random positions, however, few patterns are recognizable; hence Masters' superiority almost vanishes.

Problems with the Model

This model, which Chase and Simon themselves presented with reservations, and as only a first approximation, has been challenged, mainly on two grounds: the assumption that information is stored in a short-term store during the recall task, and the assumption that information is organized in chunks whose number reaches about 50,000 for professional chess players. We will investigate the short-term memory assumption in this paper, while the question of chunking is addressed in detail in another paper (Gobet & Simon, in press-b).

Effects of Interfering Tasks

The first kind of evidence challenging Chase and Simon's use of STM comes from a set of studies employing the Brown and Peterson interference paradigm. Brown (1958) and Peterson and Peterson (1959) had found that interfering tasks, such as

counting backward in threes from an arbitrary number, had a strong effect on the recall of three-consonant trigrams.¹

Charness (1976) obtained quite different results when he inserted a delay of 30 sec between the presentation of a chess position and its recall, with or without instructions to rehearse and with the delay interval occupied or not by an interfering task. Under none of these conditions did the interference impair notably the overall performance. Charness found a decrease of only 6 to 8 percent in number of pieces recalled, small in comparison with the decrease observed with trigrams, but did find a substantial increase in the latency of the first piece reported. Interference using chess tasks (such as finding the best move or naming the pieces in a different position) did not produce stronger impairment in recall.

Charness (1976) interpreted the long latency for recalling the first piece as evidence that much of the information is stored in LTM, proposing that during this delay traces are organized and undergo, during the first seconds of presentation, deep processing that protects them against retroactive interference. In this case, chess experience determines the speed with which more elaborated codes are accessed.

Using a similar experimental approach, Frey and Adesman (1976) presented two positions for 8 seconds each to their subjects, who then had to count backward for 3 or 30 seconds. Finally, they had to reconstruct the 1st or the 2nd position, without prior knowledge of which was going to be chosen. Results indicated only a small loss of performance in comparison with a control condition where only one board was presented. Frey and Adesman (1976) concluded that Chase and Simon's model had to be given up in favor of a model stressing depth of processing (Craik & Lockhart, 1972).

However, the findings of Frey and Adesman, although raising some doubts about the chunking theory, do not seem actually to require this proposed revision. If one takes into consideration learning time per chunk (5 to 10 sec for previously meaningless material, according to Newell and Simon, 1972), and the relatively long presentation time (8 sec per position) used in this experiment, the theory accounts for the Frey and Adesman

results. A little arithmetic establishes this: assuming that the 2 largest chunks (say 3 pieces each) are encoded into LTM during the 16 seconds of the two board presentations, and that the 7 “slots” of the STM contain chunks of 2 pieces each, a total of 20 pieces can be retained for the 2 positions, matching the results obtained in Frey and Adesman's experiment (their subjects replaced correctly about 10 pieces per position). On the other hand, Charness' data are not amenable to such an explanation, because of the more rapid presentation rate he used and the higher recall percentages of his subjects.

Level of Processing

The second kind of evidence used to argue against the completeness of the Chase-Simon theory comes from studies using the level-of-processing concept (Craik & Lockhart, 1972). Several authors have shown that the presence of supplementary information about the position, even of an abstract kind, enhances subjects' performance. Goldin (1978) obtained such results by asking her subjects to study the previous moves of the game. She found that stereotyped, highly typical positions were better recalled by all subjects than atypical positions, and that previous study of the game significantly increased the correctness of the responses as well as the confidence that subjects placed in them. Frey and Adesman (1976, exp. 1) observed similar results when presenting the moves leading to the position to be remembered.

However, in both Goldin's and Frey and Adesman's experiments, the level-of-processing variable is confounded with presentation time, for the subjects in the deep processing condition also received substantially more time to study the position or the moves leading to it. As the position on the chess board changes slowly, much can be learned, and stored in LTM, about a particular position from study of the positions leading up to it.

In another experiment (Exp. 2), Frey and Adesman presented six slides for 2 sec each, containing cumulatively the pieces of one position. The critical variable was the semantic meaningfulness of the sequences (pieces displayed in chunks according to the definition of Chase & Simon [1973a]), or the lack of it (pieces displayed by columns). The meaningful groupings produced better recall of the position. Similarly, Lane and Robertson (1979) observed that recall varied as a function of the level of semantic significance with which subjects could examine the position. Players who had only to count the number of pieces on white and black squares obtained worse results than players who were asked to judge the position and to try to find the best move. This difference disappeared, however, when subjects were notified in advance that they would have to reconstruct the position. Additional evidence on the role of contextual information is provided by Horgan and Morgan (1990), and Cooke, Atlas, Lane and Berger (1993). These results would at least call for adding attentional variables to the Chase-Simon model, for the experimental manipulations altered the amount of attention directed toward the chunks already stored in memory.

Simultaneous and Multiple Games

Tangential evidence that the Chase and Simon model may be too simple comes from the fact that strong chessplayers are able to play blindfold simultaneous games without much loss in playing strength, despite the load imposed on memory. Saariluoma (1989) has shown that Grandmasters can encode up to 4 game positions rapidly when pieces are dictated at a pace of 2 sec per piece (which would mean 6-8 seconds for chunks of size 3 or 4, respectively). Ericsson and Oliver (cited by Ericsson and Staszewski, 1989), asked a subject (Expert level) to memorize two positions, and then probed him with questions like “what piece is on c4?” They found that switching between the two positions is quite rapid (2.4 seconds when the probing is random, 1.9 seconds when the

probing alternates between both positions). Although the studies reported in this paragraph do not directly address the question of STM capacity, because the relatively long encoding time would permit considerable transfer of information to LTM, they suggest that skilled chessplayers probably use a more complex representation than a list of chunks in STM.

Other Interference Effects

Motivated by Baddeley's (1986) theory of working memory, Bradley, Hudson, Robbins and Baddeley (1987) studied the effect of interfering conditions during the presentation of chess positions. They found that a verbal task (repeating "the" every second) had only a minimal effect on performance, while a task aimed at occupying either the visuospatial system (typing a predefined series of keys on a calculator with the index finger of the non-preferred hand) or the central executive (generating letters randomly) caused a significant decline in performance (more than 2/3 in comparison with the control group). These authors observed a similar pattern of results during the resolution of tactical problems, but the performance decrease was not as drastic (only 1/3). Some of these results — effect of visuospatial interference and absence of articulatory interference — have been replicated by Saariluoma (1992), whose subjects, by way of interference, had to count the number of minor pieces (Bishops and Knights) on the board or solve chess quizzes. As chess stimuli are visual, not auditory, these findings are consistent with a chunking theory, but difficult to reconcile with Charness' (1976) findings, which showed little interference even by visual intervening chess tasks.

Summary of Problems

In summary, while the chunking model gives a good qualitative account of expert memory for single positions, a number of questions can be raised about its quantitative fit to the data we have cited. It appears that skilled players sometimes retain more chunks than could be held simultaneously in short-term memory. Other anomalies have also been observed that suggest that the Chase-Simon model may need modification.

In order to define more accurately the limits of the original model and alternative models, we will test their predictions against the findings of two experiments that use more demanding experimental tasks than those used by Charness (1976) and Frey and Adesman (1976). We will also report a longitudinal experiment that is still being extended (Experiment 3). After analyzing the findings of the three experiments, we will propose an elaboration and revision of the EPAM-based theory that incorporates into the model LTM retrieval structures similar to those already identified in other expert memory performances (Ericsson & Staszewski, 1989). With this modification both the old and the new evidence can be accommodated in a single theory that extends beyond chess to expert memory in general.

Independently of our work, Cooke et al. (1993) have recently used a similar multiple-board technique, testing a different theoretical hypothesis. We will discuss the similarities and differences between the two techniques and compare the Cooke et al. results with ours after the analysis of our second experiment.

Experiment 1

A logical extension of Frey and Adesman's (1976) study of memory for either of two positions is to ask subjects to reconstruct both positions. Moreover, this procedure can be extended to more than two positions. Models based solely on STM encoding

predict that recall should increase more or less steadily to a value fixed by STM capacity, and that the number of pieces replaced should remain constant while the percentage of pieces replaced decreases as this limit is exceeded. On the other hand, models hypothesizing a rapid encoding in LTM predict a continuing increase in number of pieces recalled with more boards, and therefore a more or less constant percentage of recall, perhaps with some loss due to interference in LTM.

Methods

Subjects

Thirteen subjects participated voluntarily in this experiment: 3 Masters (mean ELO²: 2453), 4 Experts (mean ELO: 2180) and 6 Class A players (mean ELO: 1883). The sample mean ELO was 2106 (sd = 260), with a maximum of 2510 and a minimum of 1783. Age ranged from 18 to 49, with a mean of 28 and a standard deviation of 8.6. The subjects were recruited from the Fribourg (Switzerland) Chess Club and from players participating in the Nova Park Zürich tournament, and were paid SFr 10.- (SFr 20.- for the players having a FIDE title). These subjects also participated in the copy task experiment presented in Gobet and Simon (1994a).

Materials

In order to check against the possibility that the strong players had superior memory capacities, we also presented as the control task three random positions (mean number of pieces=25), constructed by assigning the pieces from a normal game position to squares on the chessboard according to random numbers. Subjects received these three positions before the multiple board task.

For the multiple board task, we selected 26 positions from Lisitsin (1958), Wilson (1976), Reshevsky (1976), Euwe (1978), Moran (1989) and Smyslov (1972), using the

following criteria: (a) the position was reached after about 20 moves; (b) White is to move; (c) the position is "quiet" (i. e. is not in the middle of a sequence of exchanges); and (d) the game was played by (Grand)masters, but is obscure.

These positions were randomly assigned (differently for each subject) to 5 sets corresponding to the five conditions (sequences of 1, 2, 3, 4 or 5 positions, respectively). We varied the number of trials in order to maintain more or less constant the total number of positions (five, on average) per condition. Thus, subjects received 5 trials with one position to recall, 3 with two positions each, 2 with three positions each, 1 with four positions, and one trial with five positions to recall. Keeping the number of trials the same for all numbers of positions would have lengthened the experiment considerably, creating problems of fatigue and motivation (with 3 trials per condition, a total of 45 positions would have to be presented; with our design, only 26). The mean number of pieces in a position was 25 ± 1 for the 5 groups.

Positions were presented on the screen of a Macintosh SE/30, and subjects had to reconstruct them using the mouse (See appendix in Gobet & Simon, 1994a, for a description of the experimental computer software).

Design and procedure

Subjects were first familiarized with the goal of the experiment and instructed on how to use the program for reconstructing positions. They received then, in order, the control task (recall of random positions) and the multiple board recall task. In the latter task, subjects began with reproducing single positions, then the number of positions was incremented by one, up to 5 positions. Each position was presented for 5 seconds, followed by a dark screen for 5 seconds. At this point, subjects either received a new position or, after the prescribed number had been presented, had to start to reconstruct the position(s). During the reconstruction of multiple positions, only a single board was depicted on the screen at a time; subjects could switch to other boards by clicking in the

appropriate box.³ No indication was given of whether White or Black was to play the next move, and no feedback was given on the number of pieces correctly recalled. After the 26 positions had been presented, the experimenter asked the best subjects whether they would agree to try to remember more than 5 positions. No player showed enthusiasm for further trials.

Dependent variables are the number and percentage of pieces correct; the order of reconstruction; the numbers and kinds of errors; the interpiece latency and the latency before placing the first piece; and the number and maximum size of chunks.

Results

As no correlation was found between age and our dependent variables, we shall proceed without attention to this variable. However, a strong correlation was found between scores and time spent on reconstructing the positions during the experiment. We shall discuss this point at the end of the results section.

Percentage and Number of pieces correct

Random Positions.

Masters were somewhat better than the others in the recall of random boards (mean percentage for Masters = 21, sd=4.6; for Experts=16, sd=5.0; for Class A players=12, sd=5.0). Masters placed correctly about 5 pieces, Experts about 4 and Class A players about 3. The differences are not, however, significant at the .05 level: $F(2,10)=3.46$, ns. This result agrees with those found in the literature (see Gobet & Simon, 1994b for a discussion of results for random positions).

Multiple Positions.

As subjects were not required to reconstruct the positions in any particular order, for scoring, we had to identify which reconstructed position matched which stimulus position. The matching was based on subjects' comments and strategies (some subjects always replaced the positions in the same order) and on salient chess features (e.g. type of opening or presence of a conspicuous strategic or tactical motive). Except for a few rare cases with weak players, the matching was quite obvious.

Figure 1 (upper panel) shows the percentage of pieces correct for each Skill level and Number of positions to recall (percentages for single random positions are also shown for comparison). Analysis of variance indicates a main effect of Skill [$F(2,10)=22.47$, $p<.001$] and of Number of positions [$F(4,40)=9.11$, $p<.001$]; the linear component is significant: $F(1,10)=36.90$, $p<.001$. No interaction is found: $F(8,40)=1.39$, ns. All groups show a decrease in percentage of pieces replaced correctly when more than one position was to be recalled.

Masters' results are biased downward for the five-position condition, as one subject (the Grandmaster) did not pay sufficient attention when the first position of this trial was presented to him on the screen, and suddenly recalled a position of the preceding trial -- a position he had not previously been able to reconstruct and had been very troubled about. If we discard this subject's results, the Masters' percentage for the recall of 5 positions rises to 50%.

 Insert Figure 1 about here

The lower panel of Figure 1 expresses the results in terms of numbers of pieces replaced correctly. We see that the Masters are able to replace up to 60 pieces. Experts show a ceiling of about 40 pieces, and Class A players are not able to recall more than 20

pieces. ANOVA on the number of pieces gives similar results as that on percentage correct.

Order of reconstruction and serial position

Verbal learning research has shown that the recall of nonsense syllables in a sequence may show the so-called primacy and recency effects. From a theoretical standpoint, an approach assuming STM storage would predict the recency effect in such a curve, while with rapid encoding in LTM the recall probability for a given position would depend essentially on the strategy used. In our analyses, we have assessed the order in which subjects recall the various positions by the order of placement of the first pieces of these positions.

Although subjects used several strategies, some of them favoring recall of the last positions, others of the first positions, the order of reconstruction does not seem to be a function of skill level. Therefore, we will combine the data for all three skill levels.

For trials of 4 and 5 positions and with all subjects pooled, the curves of number of subjects recalling (partly) a position and of average percentage of total pieces recalled as a function of serial order of the positions are U-shaped; there is both a primacy and a recency effect (see Table 1). A one-way analysis of variance, using serial position as within-subject variable and percentage correct as dependent variable, indicates that the effect of serial position is not statistically significant with the recall of four positions [$F(3,36) = 1.89, p > .10$], but is significant with the recall of five positions [$F(4,48) = 3.36, p < .05$]. In both cases, trend analysis finds a statistically significant quadratic term at the level of .05. There are essentially no differences by Skill in how often the board best remembered is the first one reconstructed.

Insert Table 1 about here

Analysis of errors

Following Chase and Simon, (1973b), we have divided errors into errors of omission and of commission. The number of errors of omission is defined as the number of pieces in the stimulus position minus the number of pieces placed by the subject. The errors of commission are the pieces placed wrongly by the subject.

Table 2 gives the average number, per position, of errors of omission (upper panel), and the errors of commission (lower panel).

Insert Table 2 about here

For the errors of omission, one finds a significant effect of Skill level [$F(2,10)=18.61$, $p<.001$] and an effect of Number of positions [$F(4,40)=15.80$, $p<.001$; linear component: $F(1,10)=36.33$, $p<.001$]; the interaction approaches significance [$F(8,40)=2.03$, $p=.068$].

For the errors of commission, analysis of variance does not show any main effect [$F(2,10)=2.90$, ns., and $F(4,40)=1.00$, ns.], but indicates an interaction [$F(8,40)=2.31$, $p<.05$]. The pattern of means indicates that Masters tend to make more errors of commission with an increase in the number of positions, whereas Class A players show an opposite trend. Experts do not show a clear pattern.

Inter-piece latencies

The median inter-piece latency is not affected by Skill and/or the number of positions. Analysis of variance, computed after log-transformation of latency times, indicates no main effect [$F(2,10)=.83$, ns.; $F(4,40)=.71$, ns.] nor any interaction [$F(8,40)=1.79$, ns.].

Latency for the first piece

Charness (1976) found that an interfering task affected only marginally the recall of a chess position, but that the time needed for the subjects to place the first piece on the board was then substantially longer than for recall without interference. This result, which is theoretically important in suggesting that subjects are proceeding differently in the two cases, has apparently not been replicated.

One can view recall of multiple boards as an interference paradigm, where each position interferes, proactively or retroactively, with recall of the others. As we permitted any position to be recalled first, our experiment is, however, not wholly similar to Charness'. Keeping in mind this difference, we compared the latency in placing the first piece when one position was presented with the latency when more than one position was to be remembered.

All players show the same pattern: the time to place the first piece is several times longer in the recall of multiple boards than in the recall of a single position. With one position to recall, the median latency to place the first piece averages 3.1 sec for Masters, 3.6 sec for Experts and 2.8 sec for Class A players. With several positions, the respective median latencies are 6.0, 5.0 and 3.9 sec. The standard deviation within groups increases when several positions are to be recalled. Analysis of variance, performed after log-transformation of the medians, indicates a main effect of Number of positions [$F(1,10)=8.77$, $p<.02$], no effect of Skill level [$F(2,10)=1.14$, ns] and no interaction [$F(2,10)=.40$, ns].

Number and size of chunks

STM capacity is defined by number of chunks, and not by the total amount of information stored. Two parameters fix the limit on recall from STM in Chase and Simon's model: the number of chunks (limited to around 7) and the size of chunks (assumed to asymptote at 4 or 5 pieces). As some of our subjects remembered much more than what these parameters predict, an important question is whether they exceeded the limit on the size of chunks or on the number of chunks.

We define a chunk as a sequence of pieces (correct or incorrect) placed within latencies of less than two seconds each. Because moving the mouse slowed the replacement in comparison with standard chess pieces and boards, we computed an adjusted interpiece latency time: the time to move the mouse, once a piece is selected, is subtracted from the interval between the placement of two pieces. We have shown elsewhere (Gobet & Simon, 1994a) that this adjustment yields results close to those found by Chase and Simon (1973a) for recall as well as copy tasks.

Note that our definition does not encompass pieces placed individually: pieces that may occupy one slot in STM. We decided not to count pieces placed individually as separate chunks because many of them are almost certainly the product of inference, of guessing, or of a later return to complete a chunk initially restored incompletely.⁴ If so, they do not indicate the size of STM chunks. By excluding them, we obtain a conservative estimate of the number of chunks in STM.

For the number of chunks, analysis of variance shows main effects of Skill [$F(2,10) = 10.0, p < 0.05$] and Number of boards [$F(4,40) = 21.80, p < .001$], as well as an interaction [$F(8,40) = 3.45, p < .005$]. Class A players produce fewer chunks than Experts and Masters, who do not differ. For Experts and Masters, the number of chunks increases dramatically with additional boards (from 2.3 chunks with one board to 9.9 chunks with five boards), while the class A players' increase is not as strong (from 2.2 chunks with one board to 4.5 chunks with five boards). The number of chunks is the same for the three

skill levels when only one board is presented, while it differs substantially with additional boards. While averages are within the 7 ± 2 range, some individuals are well above it, with some players replacing up to 13 and 15 chunks.

The median size of the largest chunk differs as a function of skill [$F(2,10) = 15.44$, $p < .005$], with the largest chunks of stronger players larger than those of weaker players. There is also a main effect of the number of boards [$F(4,40) = 11.38$, $p < .001$] and an interaction [$F(8,40) = 2.87$, $p < .05$]. The sizes of the largest chunks of Masters and Experts differ substantially between the presentation of one board (on average, 19.7 pieces for Masters and 17 pieces for Experts) and the presentation of several boards (12.4 and 9.8 pieces, respectively), while Class A players do not show such a difference (mean for 1 position = 6.5 pieces; mean for two up to five positions 4.7). The ratios of one board to several boards are however roughly comparable between skill levels (1.6, 1.7 and 1.4 for Masters, Experts and Class A players, respectively).

In summary, we have found that (a) although the average numbers of chunks is within the 7 ± 2 range, some subjects clearly exceed this span (up to 15 chunks) -- even with our conservative estimate of the number of chunks; (b) Masters' and Experts' largest chunks are bigger when only one position is presented than when several are presented, while Class A players' chunk size does not vary much with the number of boards; (c) the largest chunks of stronger players are larger than those of weaker players. The first and second of these findings are inconsistent with Chase and Simon's model.

Total time

One finds a strong correlation between the time spent in reconstructing a position and the amount of recall. The Pearsonian correlation between average time per board and percentage correct is .94 for one position, .93 for 2 positions, .94 for 3 positions, .90 for 4 positions and .74 for 5 positions. For example, regressing average time per position on

score for 2 positions gives an excellent fit: $\text{Time} = 27.42 + .98 * \text{Percentage}$; $r^2 = .87$. Each additional percent correct produces an increase of one second.

Of course, if more chunks and more pieces are held in memory, whether STM or LTM, more time will be required to recover them from memory and replace them on the board. However, the times associated with high percentage of recall are too large to explain on this basis only, and alternative explanations must also be considered, in particular, differences in motivation or fatigue. The cognitive explanation is that the subjects who were more successful in the task had more information stored in memory to retrieve; the motivational (or fatigue) explanation, that those who were more strongly motivated (less fatigued) stuck to the task more tenaciously. Although there is no positive basis in the data for choosing between these explanations, there was also no overt behavior of subjects, other than time spent, suggesting that those who performed less well were less motivated. All the subjects appeared to persist in the task until they believed they could not recover additional pieces from memory, the latencies for the final pieces generally being quite long.

Comments on strategies

The recall of one position, and a fortiori of several positions, is a difficult task for subjects below the Master level. Weaker subjects found the presentation rate too fast, and complained that they had just time to notice some superficial features and not to grasp the meaning of the position. In contrast, Masters get to the heart of the position after an exposure of 5 sec, even being able to propose some good moves at this point. Their comments are indicative of their way of coping with this task and highlight the following features: (a) rapid perception of the pawn structure; (b) rapid recognition of the type of opening that the position may come from;⁵ (c) less often, recognition of the position itself or of a position similar to it. These features allow Masters to draw inferences about the locations of some non-recalled pieces.

Masters' attentions tend to be attracted by the atypical locations of some pieces. Finally, Masters show special difficulty in remembering the positions of the rooks and of the pawns located on the "a" and "h" files. This may be due to the fact that, at this stage of the game, rooks and "a", "h" pawns (a) have about the same likelihood of having or not having moved a square or two, and (b) do not generally play a crucial role in the position. An alternative, perceptual, explanation is provided by the peripheral location of these pieces (normally, first and last ranks for the rooks, and side columns for the pawns "a" and "h").

Subjects comment on the difficulty of organizing their memories as the number of boards increases. A typical remark is that replacing one position leads to the loss of other positions. Weaker subjects explicitly avoid encoding the whole position and focus instead on some precise part of the board (e.g., the White or Black King's-side position, or the pieces of only one color). In general, overt mnemonic strategies become more elaborate as the level of players' skill decreases.

Discussion of Experiment 1

To summarize, when several positions are presented, the percentage of pieces recalled decreases as the number of positions increases, whereas the total number of pieces recalled increases, for the stronger players, beyond the approximately 35 pieces (say, 7 chunks with average size of 4 or 5 pieces) predicted by Simon and Gilmore (1973). Second, subjects' comments underline the difficulty of the task and the necessity, especially for the weakest players, to develop additional memory strategies beyond the ones normally employed.

Third, the latency before placing the first piece is longer when more than one position is presented. From a theoretical point of view, it is unclear, however, whether

this increase is due to a retrieval from LTM (Charness' hypothesis), or to subjects' choice of the order in which they recall the different positions.

Fourth, the number of pieces placed correctly increases with the time spent on the experiment, to an extent that cannot be explained by the time needed actually to place the additional pieces on the board. Fifth, there is a U-shaped recall curve for successive positions in the recall of 4 and 5 positions, recall being best for the first and last positions. Sixth, the number of errors of omission increases with additional boards. Seventh, stronger players replace larger chunks and more chunks (the latter particularly when several boards were presented). Some subjects unquestionably placed chunks exceeding the 7 ± 2 range (up to 15 chunks). Finally, the largest chunks were larger with the presentation of one board than with the presentation of several boards.

The comments of the subjects, as well as the large number of omissions with multiple boards, and the serial position effects, indicate that encoding multiple positions in LTM is not as automatic and easy as was suggested by Frey and Adelman (1976) and Charness (1976), and that subjects spend a non-negligible part of their processing time monitoring information stored in STM. The increased time needed to recall the first piece, the correlation between performance and total time and the numbers of chunks recalled suggest that subjects do access LTM to retrieve some information. Finally, the presence of a U-shaped recall curve, exhibiting a "primacy" effect and a "recency" effect, is usually interpreted as demonstrating a conjugated action of STM and LTM.

Clearly, changes are required in Chase and Simon's (1973b) model, which relies wholly on STM, to account for the fact that Masters have an increased level of recall for more than one position, when STM should already be saturated. Some subjects recalled far more than the 7 ± 2 chunks of Miller (1956). If the assumption of a limited STM is correct, the implication is that some of these chunks are encoded into LTM. The median size of the largest chunks suggests that the estimate of Chase and Simon (4 or 5 pieces) is

too low (in particular when only one position is remembered we found an average largest chunk size of 14.4 pieces).

A similar, preliminary experiment, not reported here (see Gobet, 1993a) produced results generally comparable to those found in Experiment 1. A main divergence was that Masters in the preliminary study performed, on average, less well than Masters in Experiment 1. One difference between the two experiments which may explain this quantitative difference was that the latency between the positions was 2 seconds in the preliminary experiment, and 5 seconds in Experiment 1. This difference may have allowed subjects of the latter experiment to encode more information into LTM. To clarify these matters, we ran Experiment 2, in which we manipulated systematically the latency between the positions, while presenting 2 and 4 positions.

Experiment 2

Methods

Subjects

Five subjects participated in this experiment: one Grandmaster (ELO: 2575), one International Master (2410), one Master (2280) and two Experts (2050 and 2098). They were paid for their participation. Four of these subjects participated in experiment 1 of Gobet and Simon (1994b). Their ages ranged from 24 to 33 years.

Materials

37 positions were taken from the same sources as were used in experiment 1, and five random positions were generated as in experiment 1.

Design and Procedure

Subjects received first, presented individually for 5 seconds each, 5 game positions and 5 random positions. Due to scheduling constraints, the grandmaster received only two game and two random positions. Subjects then received four blocks of 3 trials each, two of the trials in each block having two positions each, and the other trial, four positions. The presentation time was 5 seconds for each position. In the different conditions, the latencies between successive positions within the four blocks were 1 sec, 2 sec, 5 sec and 10 sec, respectively. Three subjects received the latencies in an ascending order, and the two other subjects, in a descending order. The mode of presentation and of reconstruction was the same as in the previous experiment. The mean number of pieces for all Latencies between Boards x Number of Positions cells was 25 ± 2 , and the order of the positions was randomly assigned for each subject.

Results

Percentage and number of pieces correct

With one game position to recall, Masters got an average of 92.3% correct, and Experts an average of 64.5%, results consistent with Experiment 1. With one random position, the Masters' average was 21.1% and the Experts' was 17.7%, also consistent with Experiment 1.

As there was no difference in performance between ascending and descending orders, results will be grouped. Figure 2 shows the number of pieces correctly replaced as a function of the latency between the boards and the number of boards to remember (upper panel: Masters; lower panel: Experts). For both groups, the number of pieces recalled nearly doubled, on average, between two and four boards (but increased less than proportionately from one board to four). The number of pieces recalled also increases

substantially with increased latencies. The gain for Masters between 1 sec and 10 sec latencies is 71% with two boards, and 63% with four boards. The gain is not as big with Experts (40% with two boards and 33% with four boards). With 10 sec latency and 4 boards (a total of 100 pieces), Masters are able to replace 82.9 pieces and Experts 44.9 pieces. Between the 2 sec and 5 sec latencies, there is no substantial difference. The poor performance of Masters with 4 boards and an interlatency time of 5 sec is due to the grandmaster missing one board and mistakenly recalling a board from the previous trial.

Insert Figure 2 about here

Serial position effect

No serial position effect is apparent in the recall of 4 positions, when subjects are pooled across levels of skills and positions across presentation times (first board mean: 50.5 % ; second board mean: 48.3 % ; third board mean: 51.1 % ; fourth board mean: 47.6 %). When Skill is taken into account, Masters show a (statistically not significant) serial position effect (respective means: 73.0 %, 60.0 %, 43.3 % and 72.6 %), but the Experts do not (means: 35.4 %, 40.5 %, 56.3 % and 30.9 %). Finally, there was no difference in recall between positions presented during the first half of the session and positions presented during the second half [$t(78) = -1.12$, ns].

Discussion of Experiment 2

The increase in recall of pieces with increase in latency between successive positions provides an explanation, at least qualitatively, for the superiority in performance of comparably skilled subjects in Experiment 1 over the preliminary Experiment

reported in Gobet (1993a). However, the improvement in Experiment 2 only appeared clearly with inter-position latencies of 10 seconds, while it also appeared in comparing inter-position latencies of 5 seconds in Experiment 1 with 2 second latencies in the preliminary experiment. We have no explanation for this quantitative difference. In other respects, Experiment 2 showed great consistency with the previous experiment. For Masters, but not for Experts, there was again a primacy and a recency effect.

In both Experiments 1 and 2, the number of pieces recalled for multiple boards seems to exceed, especially for Masters, any plausible estimate of the capacity of short-term memory alone, even with chunking. We must look, therefore, for long-term memory mechanisms that would hold some of the information recalled without requiring 8 seconds per chunk for transfer to LTM, a parameter that was derived from verbal learning experiments and which has a strong empirical basis in those settings.

Comparisons with results in Cooke et al. (1993)

In order to compare the Cooke et al. (1993) Experiment 3 with our Experiments 1 and 2, we must describe the differences between the experiments. First, we used 5 seconds exposure in our experiments, while Cooke et al. used 8 seconds. Second, we used 5 second latencies between the presentation of successive positions in Experiment 1 and various times in Experiment 2; Cooke et al. used 1 second. Hence, Cook et al.'s subjects had available a total presentation time (exposure plus between-board latencies) of 9 seconds per board, our subjects total presentation times ranging from 6 to 15 seconds in different conditions.

Finally, we discouraged subjects from guessing “beyond a reasonable level”, while it is not clear from their paper what instructions Cooke et al. used about guessing. That our subjects did little guessing may be seen from the relatively low ratio of errors of commission to errors of omission.

In general, Cooke et al.'s results accord with ours: subjects recall more pieces with additional boards, but the percentages decrease. The difficulty that our subjects have shown with four and five positions suggests that there may be a limit on the number of boards (instead of the number of chunks) that can be kept in memory. As even their strongest subject, rated at 2515 ELO, was able to retain only 7 boards out of 9 presented boards, Cooke et al.'s data are consistent with such a limit. Cooke et al. report a lack of recency effect, while there was some evidence of such an effect in our Experiment 1 but little in Experiment 2. Differences in strategies may explain these different outcomes. Cooke et al. do not provide any data on the interpiece latency, on the latency to place the first piece or on variables related to chunk size and number.

Cooke et al. explain the high performance of their subjects on recall of multiple boards by distinguishing between "perceptual" and "conceptual" memory, within a depth-of-processing framework. Conceptual memory, in their view, makes use of "higher level" conceptual processes to retain information that is not obtained from perceptual processes. Beyond naming these categories, they do not propose specific mechanisms for the "higher level" learning, or a specific theory of memory organization to accommodate it. Below, we propose an alternative, but not inconsistent, explanation that does not require these concepts, that employs specific mechanisms operating on EPAM's STM and LTM, and that is supported by substantial converging evidence on expert memory in other task domains.

The results of our experiments on memory for multiple chess positions create real difficulties for the model of Chase and Simon as a complete model of chess memory. Our experiments on memory for multiple boards, together with the experiments of Cooke et al. show that players at all levels of skill can retain a far larger number of chunks than appears from their recall of single boards. The Masters in Experiment 1 continued to increase the numbers of pieces they placed correctly from about 23 (out of 25) with a single board to 60 (out of 100) with four boards. If, from the single board, we estimate

average chunk size at about 4 pieces (4 pieces x 7 chunks = 28 pieces for Masters), then the Masters appear to be recalling about 15 chunks from the four boards, far more than we would expect them to be able to hold in STM.

As we commented earlier, if only STM capacity were involved, players should be able to remember about as many pieces from a single board (limited only by the total number of pieces in the stimulus position) as from a series of boards, provided that the presentation times were not so long as to allow the transfer of additional chunks to long-term memory. The only other way in which they could increase their total storage, given a fixed number of "slots" in STM, would be to select the largest chunks they encountered for retention in STM. The data do not show an increase in maximum chunk size with number of boards that would support this explanation.

Both in our experiments and in Cooke et al.'s, the average percentage of pieces replaced correctly by Masters decreased with an increasing number of boards, although the average number of pieces replaced correctly increased. This again suggests a possible limit in the number of boards that can be recalled with a specified minimum average percentage correct. Using De Groot's (1966) "average position," Cooke et al. (p. 342) estimated the number of pieces that could be placed correctly by pure guessing at 43.7%. In Cooke's et al.'s experiment, the 2 Masters drop from an average of about 90% with 1 board to about 50% with 5 boards, and to about 40% with 9 boards (estimated from their graph), a performance attainable simply by employing knowledge about average locations. In our Experiment 1, the 3 Masters drop from an average of 95% with 1 board to an average of 39% with 5 boards, again not more than is obtained by using knowledge about typical locations of pieces. Even Cooke's et al.'s best subject recalled more than 43.7% (on average, 11 pieces) on "only" seven boards when nine were presented. If we suppose, in the nine-board experiment, that even as few as 25% of the pieces (on average, 6 pieces) are replaced correctly by guessing "usual" locations, available from LTM, the

remainder from STM, then we have to account for only 35 pieces in STM, which could be achieved by storing 7 chunks averaging 5 pieces each.

In view of the decrease with number of boards in the percentage of pieces replaced correctly, Cooke et al.'s conclusion, that the number of positions that can be recalled at a better-than-chance-level is at least nine (Cooke & al., 1993, p. 344), seems dubious.

Before attempting to construct a revised theory that could deal with the complexities we have noted, we thought it necessary to run a third experiment that would seek to extend the number of boards a chessmaster could reconstruct using mnemonic methods that had been employed successfully in other memory tasks, especially memory for rapidly presented digit strings (Richman, Staszewski & Simon, 1995). We now describe that experiment and its findings, and then return to the task of explaining in a simple and consistent manner all of these complex experimental results.

Experiment 3

The study presented in this section is a first report of a long-term experiment that we intend to continue. In it, we investigate how far a Master can push the limits of chess memory, using various mnemonics. The main question is how easily, and how far, he can go beyond the limit of about 7 boards remembered, assuming each board constitutes one large chunk (more specifically, one large retrieval structure, or template, in STM). We believe that the results obtained to date in this experiment shed important light on the nature of expert memory and allow us to test the generality of the concept of retrieval structures.

Indirectly, this experiment also addresses two criticisms of Experiment 1 mentioned by a reviewer. First, the presentation order in that experiment always went from 1 board to 5 boards. Subjects' poorer percentage-wise performance with many

boards may have been due to fatigue and waning motivation. Experiment 2, where no difference in recall performance was found between the first half and the second half of the trials, suggests that order confounding was not a serious problem. In the new experiment, only one trial is given each day, so that no such confound is present.

Second, the use of mouse and computer instead of an actual board and pieces to reconstruct positions may have slowed down subjects in the other experiments, and therefore impaired their performance in a task where speed of reconstruction could be crucial. Because the subject in the present experiment is highly practiced in reconstructing positions on the screen, it is unlikely that he was impeded by the experimental setting. (The alternative apparatus used by Cooke et al. (1993) -- reconstruction with position and board--also slows down Ss, because they have to move physically when the number of boards becomes large, say more than 6).

Method

Subject

A single S (the first author of the paper⁶) has been participating in this experiment for more than one year. A former chess professional turned psychologist, he holds the title of International Master. At the beginning of the experiment, his international rating was 2380 ELO, and his USCF rating 2396 ELO. This ranks him roughly among the 250 best players in the US. Except during the first two or three months, when he began to replay games from chess magazines (an essential part of Masters' training which he hadn't done for four years), S did not play much chess, apart from a few rapid-transit games.

Materials

The positions⁷ were taken from databases of recent games (from 1975 to 1993). The positions within a database were presented in a random order. They were all taken after Black's 20th move. A position that was selected was taken sometimes in the middle of an exchange or of some other tactical development, situations normally eliminated in most chess research. For this reason, the average typicality of positions was somewhat lower than in most published chess research, and the atypical positions were perhaps more difficult to reconstruct. S was a bit confused by these positions in the first sessions, but then got used to them.

Positions were displayed by the same program that was used in the first two experiments. As in the previous experiments, S could switch during the reconstruction among the various positions, which were accessible through buttons on the screen. He could choose any order during the reconstruction, and come back to any position at any time.

A MacIIci was used for the first 36 sessions, then a Mac PowerBook 160. S did not express any preference between the two modes of presentation. The experimental session took place (almost) daily (from Monday through Friday), at about the same time each day (between 9 AM and 10 AM). The first 35 sessions took place in S's office, the subsequent sessions at his home. Both sites were quiet.

Procedure

Each position is presented for 8 seconds. The time between positions and between the last position and the display of the reconstruction board is 2 seconds. There is no limit on the time allowed to reconstruct the positions.

A session begins with the presentation of two warm-up positions, followed by the multiple position task proper. The number of positions for a given session is determined

in the following way: (a) the minimum number of positions (other than the warm-up positions) is four; (b) if no more than one position in the previous session was below 60% correct, then the number of positions for the next session is increased by one; else, it is decreased by one. The experiment started with four positions.

S could pace his experiment as he wished. He would typically concentrate before the experiment on the names on his cue list (see below) matching the current number of positions in the multiple-board trial, then do the two-position warm-up, then concentrate again on the cue-list, and then perform the multiple-board task.

After the reconstruction, S compared the order of his reconstruction with the order of presentation, and made necessary changes so that the two sets of positions would match. The matching between positions was almost always obvious. In some cases, however, S found the matching process tedious and time consuming--which was a good motivation for him to get the order correct.

After a session, S received from an auxiliary program feedback on his performance: namely, percentage correct for each position and a detailed list of the errors he committed. He then jotted down a few comments about the session, mainly the labels and associations he used with his cue-list. Progress and difficulties were discussed weekly in a research group, of which S was member.

Results

We begin by reporting results when S was not using his retrieval structure (baseline). We then describe the retrieval structure he used, as well as training techniques he employed in order to enhance his memory. The progression of his performance over trials is presented last.

Baseline (results without using retrieval structure)

In order to estimate S's baseline performance, S completed the first three sessions without using his cue list. He was tested twice on four positions (per cent correct: 50% and 55%, respectively; number of pieces correct: 50 and 55, respectively), and once on five positions (per cent correct: 54.3%; number of pieces correct: 68). On 2 positions, his average was 79.9% correct, i.e. about 40 pieces correct. These results are in line with those of the Masters of Experiment 1 and of Cooke et al. (1993).

Recall with a retrieval structure: The cue list

S's mnemonics are based on classical techniques (see Yates, 1966) and on more recent developments in extraordinary digit memory research (Chase & Ericsson, 1982, Staszewski, 1993). S's main technique was to associate each position with the corresponding element in a pre-learned list, which serves as a retrieval structure. During recall, each element of the list serves as a cue to the retrieval of the corresponding position. In order that the elements of the list could lead easily to meaningful associations with the positions, S choose as elements the names of chess world champions, in the order of their reigns (see Table 3, second column). Around the 10th week, S abbreviated the names to their first syllables (see Table 3, third column), in order to pronounce them subvocally more rapidly during the association phase. We have then a variation of the method of loci, which we may call method of magistri. The cue-list was not used for the warm-up trials (recall of 2 positions).

Insert Table 3 about here

We illustrate the use of the cue-list by a few examples, going from rich to poor associations. In the first two examples, S recognizes the type of position, retrieves a verbal label for it, and associates the label with the name on the cue-list. In the third example, a relatively useless label (concerning only 2 pieces) is associated with the cue-list name. In the last example, the position is identified, but no association is made.

(1) Position #5. Name on the list: Euwe.

“A Panov attack. Black has a strong Knight on d5, typical for Euwe’s play.”

(2) Position #6. Name on the list: Botvinnik.

“A Grünfeld defense, as in the match Karpov-Kasparov, Seville. Botvinnik used to play the Grünfeld”.

(3) Position #1. Name on the list: Steinitz.

“White has the Bishop pair. Steinitz liked the Bishop pair.”

(4) Position #2. Name on the list: Lasker.

“A Maroczy without g6.”

Before a trial, S visualized the faces of the champions on his cue list, and recalled -- rapidly -- some information about their playing styles and the types of positions they played. For the first 12 sessions, S scanned the list in a linear fashion. From around session 13 to around session 70, he scanned the list by (mentally) placing each world champion in one of the corners of the room starting from the front wall. In the later sessions, he used a hierarchical organization, grouping the 13 players by group of threes (group of four for the last four players).

S devised, outside the sessions, a few exercises to improve his ability to use the retrieval structure, both to access the names and to make associations. Altogether, he spent about 10 hours, distributed over three months (from around session 36 to around session 70) doing the following exercises: (a) drilling to access a name from a given

number (2 -> Lasker); (b) drilling to access a number, given a name (Lasker -> 2); (c) associating name and positions by groups of three, without time pressure; (d) repeating to himself the cue list as fast as possible; (e) labeling positions taken randomly from books and chess magazines

The following difficulties sometimes occurred with the use of the retrieval structure: (a) the access to a name was too slow, disrupting the recall process; (b) no association could be made between the cue name and the position; (c) the position could not be categorized, making it hard to form an association; (d) although an association was formed, S did not remember it during recall.

S's progression over trials

Figure 3 depicts S's progression for the first 150 trials (grouped by blocks of five trials (days)), spanning 37 weeks. The dependent variable is the total number of pieces correct.

 Insert Figure 3

For the warm-up trials with 2 positions, there is a gradual improvement, from the 40 pieces (80%) of the baseline and 36 (72%) average in the first block to a near-perfect 47 piece (94%) average in the last block. The regression line is: $\text{Number_of_pieces} = 35.75 + 0.31 * \text{block}$ ($r^2 = 0.69$; $p < .001$). Thus, S gains about one third of a piece for every block of 5 sessions.

For the multiple-board task proper, two “jumps” may be singled out. The first one is after the first block (44 pieces correct), where S was even below the baseline performance (53 pieces, on average)⁸, to the fourth block (79 pieces). There is then a long

plateau, with a small peak at blocks 10 and 11 (89 and 94 pieces) up to block 16, which initiates a dramatic rise, with a maximum at block 18 (152 pieces). S's performance then decreases slightly, and plateaus at 123 pieces on average per block. In only four cases did S manage to place more than 160 pieces correctly:

Session 89: 9 positions attempted, 163 pieces correct;

Session 90: 10 positions attempted, 160 pieces correct; (only attempt with 10)

Session 130: 9 positions attempted, 167 pieces correct;

Session 150: 9 positions attempted, 178 pieces correct.

Expressed in terms of the average number of boards attempted per block of 5 sessions, the results are as follows: up to block 15, only two blocks average more than 5 boards (6 boards in block 10, and seven in block 11). Then, after a peak of 9 boards (block 18), S averages 8 positions (from block 19 to 22), and later 7 positions (from block 23 to 30). Note that, up to the present time, although S has almost doubled the number of boards he can recall above the criterion (of no more than one board less than 60% correct), and more than tripled the number of pieces replaced correctly, the number of boards he can replace above the criterion is still within the usual STM span.

As the experiment progressed, S increased linearly the average percentage of pieces correct per position, both for the warm-up trials and for the multiple-board task. The regression equations are:

$$\text{Percent_N_boards} = 54.97 + 0.71 * \text{block} \quad (r^2 = 0.51, p < .001)$$

$$\text{Percent_2_boards} = 71.82 + 0.58 * \text{block} \quad (r^2 = 0.68, p < .001)$$

That the percentage of pieces correct per position increases is a bit surprising, as this was not what S was training for; he was seeking to reconstruct all boards at at least the 60% level. Theoretically, his increase in percentage correct may be explained by S gradually expanding his discrimination net and learning new templates (see general discussion, below). However, this explanation is not entirely convincing, because the time S spent on chess during this experiment is very small compared with the time he spent during his earlier career. It is more plausible to suppose that over the course of the experiment, he may be gradually recovering his earlier knowledge. It may also be that what S has added to his templates are slots useful for recalling positions, while what he had developed earlier were slots useful for playing chess.

A few additional features of the experiment merit comment.

(a) Motivation. S was not strongly motivated at the beginning of the experiment, which he took as a whim of the second author, but he was gradually seduced by the task and became curious about how far he could go. His daily performance became an important part of his weekly routine. A bad performance would, in some cases, vex him for the rest of the day, a good one would exhilarate him for a few hours.

(b) Interruptions. In two cases, S interrupted the experiment for two weeks. Interestingly, he seemed to do better on the multiple-board task after the interruption. The first interruption occurred between the 8th and 9th experimental weeks (between sessions 36 and 37), the second after the 34th and 35th experimental weeks (between sessions 143 and 144). Over the final week before the first interruption, S averaged 41 pieces correct with 2 positions and 62 with n positions. After resuming, his respective averages were 38 and 86. Before the second interruption, his respective averages were 46 and 118. After the interruption, 46 and 124. We speculate that interruption decreased the proactive inhibition in LTM.

(c) Verbal labeling. As shown in the examples illustrating the cue-list, S often uses names -- mainly opening names, such as “King’s Indian”; “Queen’s Gambit with a minority attack” -- in order to label the position. That verbal labeling is important in recalling positions has been noted in chess research on several occasions before (e.g. De Groot & Gobet, in press). While position labeling is generally advantageous for S, two qualifications are worth noting. First, some positions, although labeled correctly, produce a low recall percentage. This occurs typically when the label does not differentiate enough and is compatible with several typical positions. Second, some positions that S could not label, even after the feedback phase of the session, are well recalled. In these cases, S has either identified the position as a whole (“I know this type of position”) even though no name could be associated with it, or noticed several small chunks, or various relations of attack between the pieces.

Discussion of Experiment 3

Cooke et al. seem to have been too optimistic in their conclusion that chess masters could recall (easily) at least 9 boards. The subject of this experiment, only slightly weaker than their best Master (2396 vs. 2515 USCF rating; a one-half standard-deviation difference), has reached a (temporary?) ceiling after 150 sessions with 10 positions attempted, and with something of a plateau around 8 positions reconstructed at or above criterion. (In appendix 1, a mathematical model shows that such a plateau can be expected given S’s current probability of scoring, in a position, above the 60% criterion.) Although this does not prevent future progress, it indicates that his performance in the recall of multiple boards, although based on LTM encoding strategies, probably also requires the retention in STM of the information that identifies the templates for these positions.

In our present understanding of this experiment, S’s performance relies on two processes: (a) recognition of the position -- with or without verbal labeling -- so that the

appropriate template can be retrieved and its slots filled; and (b) association of this template with the current name on the cue list. Position recognition is a direct test of S's chess expertise, and we hypothesize that the only way to improve this skill very much would be to train again to a professional level in order to elaborate his perceptual discrimination net (EPAM net) and to increase the number of templates for recalling positions. Interestingly, as shown by the increase in average percentage per position, both for the warm-up positions and the multiple-board positions, S seems to have gradually improved this part of his skill. The second part, acquiring a more powerful retrieval structure, seems more accessible to "conscious" strategies, like SF's and DD's construction of their retrieval structures (Chase and Ericsson, 1982; Staszewski, 1993). This part has been harder for S to improve than we thought it would be at the beginning of this experiment.

General Discussion

To explain the complex data we have reviewed, we propose a modified theory of chess memory that makes a minimum of special assumptions. We conjoin the chunking mechanism of Chase and Simon with another mechanism that has shown considerable success in explaining other expert memory performances. This is the mechanism of retrieval structures or templates which, operating within the EPAM model of perception and memory, has matched in great detail the performance of experts who have learned to recall long strings of digits presented to them at a rate of one every second or two seconds (Richman, Staszewski & Simon, 1995).

One basic question in the theory of chess memory is what part of the performance employs short-term memory, and what part long-term memory. A second question is what learned structures in long-term memory, whether learned as part of the acquisition of chess skill or learned specifically to improve performance in this recall task, can increase

the number of boards that can be reconstructed, and to what extent. Our answers to these questions draw extensively on the theory, developed initially by Chase, Ericsson and Staszewski, and formalized and tested in detail by Richman, Staszewski and Simon (1995), of memory for rapidly presented digit sequences.

1. The model we propose assumes a short-term memory of "normal" capacity, measured as 7 ± 2 chunks (familiar and recognizable patterns) or as the number of chunks that can be rehearsed in about two seconds (Miller, 1956; Baddeley, 1986; Zhang & Simon, 1985).

2. It assumes a discrimination net (EPAM net) that has learned to recognize and discriminate among a large number of such chunks, and has stored information about them in long-term memory.

3. Among the chunks relevant to chess stored in long-term memory are (a) clusters of pieces (up to five or more), (b) position templates, and (c) a retrieval structure pointing to a sequence of such templates.

(a) The clusters are the chunks of the Chase-Simon theory. As they are the familiar patterns of pieces that are commonly encountered in chess game positions, they can be recognized by expert players whenever they are present on the board, the number that can be recognized varying with experience and skill. The number of such chunks in the long-term memory of a master is probably in the neighborhood of 50,000 or more.

(b) The templates are patterns of the chess board found in familiar openings and lines of play, again of the sorts frequently encountered in games. The templates specify the locations of perhaps a dozen pieces in the position (thus specifying a class of positions), but also contain variables (slots) in which additional information can be placed, thus fixing the positions of additional pieces. Perceived individual pieces or chunks of type (a) can be assigned to slots in a matter of a second or two. Some of the slots may have default values that can be altered. The templates are implicitly acquired

by chess players in the course of their study of games, both those they play and those they examine in the chess literature.

(c) The retrieval structure is a deliberately acquired structure in long-term memory that is used to store identifiers for the list of individual positions encountered in the course of a multiple board trial. We described above the retrieval structure, using the names of world champions, that was learned and used by S in Experiment 3.

The main novelty of this theory is the hypothesis that, in addition to the fixed chunks specified by Chase and Simon, there are chunks (retrieval structures and templates) in long-term memory that contain variables (slots), and that these slots can be filled rapidly with new perceptual information. This hypothesis derives from previous empirical observations on memory. The EPAM model (Simon & Feigenbaum, 1962), using discrimination nets to access memory, postulates that when new items are learned, different processing times are required: (1) to elaborate the discrimination net and (2) to store information about the new chunk at the leaf node. Evidence from other recall tasks (Simon, 1976, esp. pp. 72-73) suggests that most of the time required for learning new chunks (8 seconds per chunk) is spent in elaborating the net, and that information may be stored at existing leaf nodes rather quickly (less than 1 second per item). The idea of rapidly modifiable long-term memory structures (retrieval structures) was then shown capable of explaining the remarkable memory performances of mnemonists, including the subjects who trained themselves to recall rapidly presented sequences of digits (Chase & Ericsson, 1982; Ericsson & Polson, 1988; Ericsson & Staszewski, 1989; Staszewski, 1988, 1993).⁹

The single template hypothesis

If we accept the "standard" theory (Newell & Simon, 1972), the durations of the stimulus presentations are too short (5 seconds per board) to allow the chunks to be transferred to long-term memory; for at 8 seconds per chunk the time to store information in LTM is nearly a minute for seven chunks. However, the eight-second estimate is based on experiments where the subject must first learn to differentiate the response chunk from other chunks, and then store it in LTM in association with the corresponding stimulus. In the present experiments the subject needs simply to recognize familiar chunks, whose content is already stored in LTM and add information to them, without expanding the discrimination net. The empirical evidence suggests that the time required for such an addition to a chunk in the EPAM theory is about one second.

This hypothesis is not an ad hoc explanation for a previously discovered phenomenon, but has strong independent sources of support. The study of phenomenal memory performances, which, on an anecdotal basis goes back to classical times, and which has been revived in systematic studies in recent years by Chase, Ericsson, Staszewski and others, explains the phenomena convincingly on the basis of precisely this mechanism.

Chase and Ericsson (1982) and Staszewski (1988) have shown that subjects possessing normal STM spans and normal abilities to acquire new information in LTM can learn, but only after many months or even several years of training, to recall sequences of 100 or more digits when these are read to them at rates of a digit every second or two. Similar performances recorded in the past have always been associated with some variant of the so-called "method of loci," in which the mnemonist acquires a fixed retrieval structure in LTM by study over some period. The retrieval structure is often a building stored in memory ("Memory Palace") composed of many rooms, with furniture on which objects can be placed (in imagination). When a list is to be memorized, the mnemonist proceeds, in imagination, through the rooms in a fixed order, placing the successive items on successive surfaces ("slots").

Chase, Ericsson and Staszewski showed that their trained mnemonists used retrieval structures that they acquired deliberately during their training, and of which they were fully aware. In the case of sequences of digits, the structures were tree-like, chunks of three or four digits being associated with each successive twig of the tree. In other experiments by Ericsson and Polson (1988), restaurant waiters retained multiple customers' orders, and the templates were standard menus with a slot for each course. The mnemonists and the trained waiters also acquired, or had previously acquired, familiar chunks, which were stored in LTM. In the digit sequence recall, these chunks were running times (it happened that the subjects were experienced runners) and ages; in the case of the waiters, they were standard menu items.

In a first attempt to apply the idea of retrieval structures to chess (Richman, Simon & Gobet, 1991; see also Ericsson & Kintsch, 1995, for a similar approach) we proposed a retrieval structure in the form of a single chess board with slots for storing chunks in association with the squares. When information was stored from two or more successive positions, this could be done on a single retrieval structure, or separate copies of the structure could be used. With the first alternative, some information had to be stored with each chunk to identify the position it came from.

While the single retrieval structure hypothesis will account for data from interference experiments (Charness, 1976; Frey & Adesman, 1976), it suffers from some weaknesses. In Gobet's (1993a,b) simulations, which used such a retrieval structure, some data on the location of chunks did not fit when the structure was used, but did when it was not. Second, such a scheme predicts better recall with random positions than is actually observed. Third, it predicts that subjects can encode, in the recall of several boards, only one piece per square. However, the protocols of subjects in the multiple board task show that they sometimes encode several pieces on the same square for different positions, or several chunks in the same part of the board. Finally, there is relatively little interference between boards, contrary to the prediction of the theory.

The multiple template hypothesis

These shortcomings lead us to propose that previously-stored multiple templates are used to remember chess positions. Chess players have seen thousands of positions; and for expert players, most positions they see readily remind them of positions or types of positions they have seen before. They have information about the positions that arise when the Ruy Lopez opening is played, or the King's Indian Defense. A Grandmaster or Master holds in memory literally thousands of such patterns, each of which specifies the locations of ten or a dozen pieces, with revisable defaults for others. We will take these patterns as the templates in our theory. Templates, under various names, are familiar objects in cognitive psychology and in artificial intelligence programs. They have been called "schemas" (Bartlett, 1932), "frames" (Minsky, 1975) or "prototypes" (for example, for chess, Goldin, 1978; Hartston and Wason, 1983). Without any common agreed upon usage of these terms, we have chosen for the purpose of this paper the term "template."

We hypothesize that these templates have slots which can be filled with various types of information, such as the location of pieces on the board, the opening such a position is likely to come from, the potential plans and moves in the position, and so on. If the similarity of a briefly presented position to one of these templates is recognized, the template can be used, with modifications, to store the information about the current board. Figure 4 shows such a template, as postulated for a relatively weak player.

Insert Figure 4 around here

Because chess is primarily a visual task, templates will be accessed mainly through patterns of pieces on the board. Recognizing the appropriate template requires

no more than an EPAM-like discrimination mechanism, like the one proposed by Chase and Simon (1973), and implemented by Gilmarin and Simon (1973). However, while all nodes in the net were previously assumed to store the same kind of chunk, the multiple-template theory postulates that some chunks evolve into more complex structures (the templates) having slots, which may themselves be filled with chunks. The new proposal supposes that the time needed to fill a slot in an already existing LTM template is brief (less than one second).

According to this hypothesis, chessmasters recall a position better than weaker players both because they can retrieve larger chunks (the hypothesis of Chase and Simon) and because the larger chunks are templates with slots. In the recall of several boards, a pointer to the distinct template for each board is placed in STM.¹⁰ The hypothesis therefore predicts that no more than 7 ± 2 positions can be retained, the exact number depending on the time required to access the labels of the schemas. (Cooke et al.'s data are compatible with this hypothesis, as even their best subject's¹¹ performance is within this range). The idea that experts may maintain a number of different retrieval structures in STM is not completely new. Recently, Staszewski (1988) has shown that expert mental calculators keep several retrieval structures in STM when solving arithmetic tasks.

The hypothesis also produces the phenomena attributed to high-level descriptions by Cooke et al. (1993) and Gruber and Ziegler (1990). These authors did not specify how high-level descriptions are accessed, and a high-level description may point to several positions. The multiple-template model removes these difficulties; for recognition of similarity between the current position and positions for which a template already exists in LTM associates the position with the stored template, and variants of positions associated with a single template can be specified by filling slots.

Finally, the multiple-templates version of the theory is not excessively accurate at remembering random positions, which we have seen to be a problem with the single-template hypothesis. Random positions simply do not evoke a rich template. Even in

random positions, however, search processes may disclose adventitious relations (e.g., attack, defense) among pieces or groups of pieces on the board, relations that are not salient to immediate perception. We would expect high-level players to detect more such relations in a given search time than weaker players. This explains the observed modest superiority of Masters over weaker players in random positions (Gobet & Simon, 1994b).

The multiple template hypothesis also explains the rapidity with which these retrieval structures are accessed (through the mechanisms embodied in the EPAM theory); how they are slowly learned (by the EPAM discrimination and familiarization processes); and how they fit together with the data on chunks (perceptual mechanisms direct attention during the creation of the EPAM net). The application to chess only requires EPAM to learn and add to LTM the templates corresponding to previously experienced chess positions.

Why, then, does the number of pieces recalled not increase linearly with the number of positions presented? Because STM must now be used also to hold the identifying labels for the particular templates that have been retrieved and modified. If these labels are verbal and multisyllabic ("it's the Dragon Variation of the Sicilian Defense"), the Baddeley model of STM would predict that only a small number can be retained (Baddeley, 1986; Zhang and Simon, 1985). If they are themselves encoded as perceptual cues, there is still the familiar limit of 7 ± 2 .

Finally, we can conjecture why the largest chunks recalled are bigger when single positions are presented than when positions are presented in sets. (The size ratio was about 1.6:1 in Experiment 1.) Filling in free slots requires some STM capacity, and this resource may be exhausted when more than one position is presented. In this case, only the core of a template, with perhaps a few default values, may be accessed during recall.

An alternative hypothesis lies in the subjects' strategies when replacing multiple boards. Two strategies frequently seen among our Masters or Experts would cause a reduction in the measured size of the largest chunks. The first strategy consists in placing

a few pieces on each board (thereby providing recognition cues to recover the template for each position, and the full chunks, later), and then going back to complete each position. Replacement of chunks piecemeal, in this way, would lead us to underestimate chunk size and overestimate the number of chunks. The second strategy, having a similar effect, consists in placing first, for each position, pieces the subject is unsure of (e.g., the positions of Rooks, of “a” and “h” pawns, of Queens and Kings) and afraid of forgetting.

Comparison with higher-level descriptions

In what respects do Multiple Templates (MTs) differ from the higher-level descriptions (HLDs) postulated by Craik and others? Perhaps the most fundamental difference is that the MT concept is part of a detailed comprehensive process model, while the HLD concept is not. Second, as opposed to MTs, HLDs do not account for slots and rapid encoding into LTM. Third, MTs may or may not give access to high-level descriptions such as type of position or of the opening to which the position belongs. In MTs, "high-level descriptions" become slot-values in the template. In Experiment 3, S could sometimes reconstruct positions without being able to label them verbally. Fourth, (chess) templates have a strong visuo-spatial character. This is lacking in the HLD concept. Fifth, nothing is said about how HLDs are accessed, while we propose that MTs are accessed through a discrimination process like that incorporated in EPAM.

There is nothing antithetical about the two concepts, but the one, MT, is defined in terms of a precise set of mechanisms and processes already incorporated in an existing model tested on other tasks, while the other, HLD, is characterized in only a very general and imprecise way. For this reason, it is easier to design experiments for MTs than for HLDs that put the theory to a sharp test.

Conclusion

The ability of chessmasters to recall the main features of many, perhaps most, of the serious games they have played, as well as hundreds or thousands of games from the chess literature, indicates that, given sufficient time, many distinct templates can be stored in LTM, with associated information about the positions of pieces. We have no data that permit us to estimate how much time is required for such learning, but as it usually requires several hours to play a single game, and nearly the same time to study one seriously, the times available for learning appear adequate to the task. We think that the weight of evidence now available supports the multiple-template hypothesis of chess memory we have just discussed.

To test quantitatively a model as complex as this one requires it to be specified rigorously as a computer program so that the effects of its many interacting mechanisms can be sorted out and compared consistently with the data from many diverse experiments. Our research group has constructed and successfully tested such a simulation for the expert digit-span performance (Richman, Staszewski & Simon, 1995) and has developed a preliminary model for chess memory (CHREST, Gobet, 1993a,b). At the core of these models is the long-established EPAM program, which is already able to predict behavior over a wide range of perceptual and memory tasks, amplified by a mechanism for creating and employing retrieval structures or templates in LTM in the manners indicated above. We intend to report separately on the application of these models to a wide range of empirical chess data.

References

- Baddeley, A. (1986). *Working memory*. Oxford: Clarendon Press.
- Bartlett, F. C. (1932). *Remembering*. Cambridge: Cambridge University Press.
- Binet, A. (1894). *Psychologie des grands calculateurs et joueurs d'échecs*. Paris: Hachette. [Reedited by Slatkine Ressources, Paris, 1981.]
- Bradley, A., Hudson, S. R., Robbins, T. and Baddeley, A. (1987). *Working memory in chess*. Unpublished report, Cambridge, May.
- Brown, J. (1958). Some test of decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10, 12-21.
- Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 641-653.
- Charness, N. (1992). The impact of chess research on cognitive science. *Psychological Research*, 54, 4-9.
- Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16). New York: Academic Press.
- Chase, W. G., & Simon, H. A. (1973a). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Chase, W. G., & Simon, H. A. (1973b). The mind's eye in chess. In W. G. Chase (Ed.) *Visual information processing*. New York: Academic Press.
- Cooke, N. J., Atlas, R. S., Lane, D. M., & Berger, R. C. (1993). Role of high-level knowledge in memory for chess positions. *American Journal of Psychology*, 106, 321-351.

- Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-681.
- De Groot, A. D. (1966). Perception and memory versus thought. In B. Kleinmuntz (Ed.), *Problem solving: Research, method and theory*. New York, NY: Wiley.
- De Groot, A. D. & Gobet, F. (in press). *Perception and memory in chess: Heuristics of the professional eye*. Assen: Van Gorcum.
- Ericsson, K. A., Chase, W. G. & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181-1182.
- Ericsson, K. A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211-245.
- Ericsson, K. A. & Polson, P. G. (1988). An experimental analysis of the mechanisms of a memory skill. *Journal of Experimental Psychology : Learning, Memory and Cognition*, *14*, 305-316.
- Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: mechanisms of exceptional performance. In D. Klahr. & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon*. Hillsdale, NJ: Erlbaum.
- Euwe, M. (1978). *The development of chess style*. New York: McKay.
- Frey, P. W., & Adesman, P. (1976). Recall memory for visually presented chess positions. *Memory and Cognition*, *4*, 541-547.
- Gobet, F (1993a). *Les mémoires d'un joueur d'échecs*. Fribourg (Switzerland): Editions Universitaires.
- Gobet , F. (1993b). A computer model of chess memory. *Proceedings of 15th Annual Meeting of the Cognitive Science Society*, (pp. 463-468). Hillsdale, NJ: Erlbaum Associates.

- Gobet, F. & Simon, H. A. (1994a). *Expert chess memory: Revisiting the chunking hypothesis*. (Complex Information Processing Paper No. 515). Pittsburgh: Carnegie Mellon University, Department of Psychology.
- Gobet, F. & Simon, H. A. (1994b). *Role of presentation time in recall of game and random chess positions*. (Complex Information Processing Paper No. 524). Pittsburgh: Carnegie Mellon University, Department of Psychology.
- Gobet, F. & Simon, H. A. (in press-a). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grandmaster level chess. *Psychological Science*.
- Gobet, F. & Simon, H. A. (in press-b). *Recall of random and distorted positions. Implications for the theory of expertise*. *Memory & Cognition*.
- Goldin, S. E. (1978). Memory for the ordinary: Typicality effects in chess memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 605-616.
- Gruber, H., & Ziegler, A. (1990). Expertisegrad und Wissensbasis. Eine Untersuchung bei Schachspielern. *Psychologische Beiträge*, 32, 163-185.
- Hartston, W. R. & Wason, P. C. (1983). *The psychology of chess*. London: Batsford.
- Horgan, D. D., & Morgan, D. (1990). Chess expertise in children. *Applied Cognitive Psychology*, 4, 109-128.
- Lane, D. M., & Robertson, L. (1979). The generality of the levels of processing hypothesis: An application to memory for chess positions. *Memory and Cognition*, 7, 253-256.
- Lisitsin, G. M. (1958). *Strategia i taktika shahmat*. Moskva: Fisikultura i sport.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

- Minsky, M. (1977). Frame-system theory. In P.N. Johnson-Laird and P.C. Wason (Eds.), *Thinking. Readings in cognitive science*. Cambridge: Cambridge University Press.
- Moran P. (1989). *A. Alekhine: agony of a chess genius*. Jefferson, N.C.: McFarland & Co.
- Murdock (1961). The retention of individual items. *Journal of Experimental Psychology*, 62, 618-625.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Peterson, L. R. & Peterson, M. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, 58, 193-198.
- Reshevsky, S. (1976). *The art of positional play*. New York: McKay.
- Richman, H. B., Simon, H. A. & Gobet, F. (1991, July). *Applying retrieval structures to chess memory*. Paper presented at the international conference on memory, Lancaster.
- Richman, H., Staszewski, J., & Simon, H.A. (1995). Simulation of expert memory with EPAM. *Psychological Review*, 102, 305-330.
- Saariluoma, P. (1989). Chess players' recall of auditorily presented chess positions. *European Journal of Cognitive Psychology*, 1, 309-320.
- Saariluoma, P. (1992). Visuospatial and articulatory interference in chess players' information intake. *Applied Cognitive Psychology*, 6, 77-89.
- Simon, H. A. (1976). The information storage system called "human memory". In M. R. Rosenzweig and E. L. Bennett (Eds.), *Neural mechanisms of learning and memory*. Cambridge: MA: MIT Press.
- Simon, H. A., & Feigenbaum, E. A. (1964). Effects of similarity, familiarization and meaningfulness in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 3, 385-396.

- Simon, H. A., & Gilmarin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, 5, 29-46.
- Smyslov, V. V. (1972). *My best games of chess*. New York: Dover.
- Staszewski, J. J. (1988). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- Staszewski, J. J. (1993). A theory of skilled memory. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum Associates.
- Wilson, F. (1976). *Lesser-known chess masterpieces: 1906-1915*. New York: Dover.
- Yates, F. A. . (1966). *The art of memory*. Chicago: The University of Chicago Press.
- Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypothesis. *Memory and Cognition*, 13, 193-201.

Appendix 1

Probability of Success in Replacing Pieces on Multiple Boards

Suppose that S attempts to replace the pieces from a sequence of n boards and that the probability of achieving a replacement of at least 60% on any board, selected at random, is q per cent. Then the probability of a successful trial (defined as a trial in which not more than one of the n boards has a replacement of less than 60%) is shown in the table below. The equation is:

$$\text{prob. success} = q^{(n-1)} (q + n(1-q))$$

It can be seen in Table A-2 that for the player to succeed more often than fail with 6 boards requires $q=.75$; with 7 boards; $q=.775$; with 8 boards, $q = .8$; and with 9 boards, $q = .825$. To break even with 11 boards calls for a q just over .85.

In the multiple board task (without the warm-up positions), S obtains 73% of the boards above criterion in the 150 first sessions. The percentage increases to 77% if only the last 75 sessions are considered, which gives S an equilibrium just below 8 boards, according to this mathematical model.

 Insert Table A-1 around here

Appendix 2

Comparison of chess templates with DD's retrieval structure

As the data on two subjects, SF and DD, retrieving lengthy number strings have been a major basis for the skilled expert memory theory (Chase and Ericsson, 1982) and the concept of retrieval structure, it is informative to compare the chess templates we have put forward in this paper with the structure DD learned in order to perform the memory task. Table A.1 summarizes the main points of similarity and dissimilarity. To determine whether the same mechanisms can explain both chess memory and digit retrieval, we are now collecting data (Experiment 3) on one subject, trained in a manner as similar as possible to the way SF or DD were trained, to remember as many briefly presented chess boards as possible.

Insert Table A-2 about here

One of the main differences patent in the two columns of Table 3 is that DD both constructed and used his retrieval structure deliberately, while chess players' templates are acquired as a by-product of their chess practice and training. In both cases the creation and refinement of such structures takes years; but, once acquired, the slots in them may be filled rapidly. We do not have direct estimates of the time required to fill slots in chess templates, but we propose an upper bound. Saariluoma's experiments (1989), where positions are verbally dictated at a rapid rate, show that Masters are able to retain a random position dictated at a pace of one piece every 2 sec. The recall is high, but not perfect (60%). As templates are not available for random positions, we cannot derive a direct estimate from these data; but if Masters are no slower in encoding game positions,

where templates are at hand, it follows that template slots can be filled at a rate of 2 seconds or less per slot, with an accuracy of at least 60%.

Comments of players playing blindfold games suggest that they focus their attention only on parts of the board, and that they do not maintain at one time in the ir mind's eye an entire representation of the chess board (Binet, 1894; Gobet, 1993a). Similar remarks are valid for DD: both during encoding and recall, DD only directs his attention, at any given time, to small parts of his retrieval structure.

While DD's retrieval structure was specifically constructed for purposes of memorization, the raison d'être for chessplayers' templates is to allow them to search efficiently while choosing a move. In particular, the templates give access to semantic information likely to be useful in this choice. Another useful characteristic of templates, saving time during both problem solving and memory tasks, is the presence of default values.

Finally, a chess template may point to other templates, describing earlier or later positions in the game (From which opening does this position come? To which endgame is it likely to lead?) These features do not exist for DD's retrieval structure. It seems that chess players templates are more complex structures than DD's retrieval structure: the latter is a tree, the former are networks. DD' retrieval structure is an arbitrary representation of a list of digits, while the chess players' template contains an isomorphic representation of the chess board. Such a template provides many redundancies; for example, a given square is related not only to adjacent squares, but to squares on the same diagonal, or even to plans in which this square plays an important role. DD also uses redundant encoding of information, but of a different kind (Richman, Staszewski & Simon, 1993). There is one respect in which DD's template is more powerful than those used by chess players': it can encode the order in which digits were dictated, whereas there is no evidence that chess players are good at remembering the exact order in which pieces were dictated in auditory presentation.

In sum, while the retrieval structures and chess templates use the same fundamental processing mechanisms (rapid encoding in slots, long construction time, information stored in LTM...), they differ in features that are dictated by the peculiarities of their respective task environments. The characteristics of retrieval structures have been determined for only a few environments. We have already mentioned, in addition to digit lists and chess, waiters' memory for orders from menus, and the classical mnemonic devices like the "Memory Palace." Retrieval structures will be better understood as research is extended to other domains: for example, sports, scientific disciplines, engineering drawing, music and natural language.

Authors' note

The first experiment of this paper was presented at the XXV International Congress of Psychology, Brussels (July 1992). Preparation of this article was supported by grant No. 8210-30606 from the Swiss National Funds of Scientific Research to the first author and grant No. DBS-912-1027 from the National Science Foundation to the second author. Correspondence concerning this article should be addressed to Herbert A. Simon, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213.

The authors extend their thanks to Jean Retschitzki, Howard Richman, James Staszewski and Shmuel Ur for valuable comments on parts of this research as well as to Pertti Saariluoma, James Greeno and two anonymous reviewers for helpful comments on a first draft of this manuscript.

Table 1

Frequency of Recall of Positions as a Function of Presentation Order

	Order of presentation				
	#1	#2	#3	#4	#5
4 positions to memorize					
Number of subjects recalling (part or all of) the position ^a	12	11	8	11	-
Average percent of pieces recalled	40.8	29.5	20.7	38.2	-
5 positions to memorize					
Number of subjects recalling (part or all of) the position ^a	12	10	10	8	10
Average percent of pieces recalled	41	20.8	22.6	14.3	27

Note. The three skill levels are pooled.

^a Maximum=13 subjects.

Table 2

Mean Number (per position) of Errors of Omission and Errors of Commission as a Function of Skill and Number of Positions Presented

Skill level	Number of positions presented				
	1	2	3	4	5
Errors of omission					
Masters	0.4 (0.5)	0.6 (0.7)	7.4 (1.9)	7.2 (0.7)	11.5 (7.3)
Experts	5.4 (4.7)	8.2 (5.3)	8.5 (5.4)	11.0 (4.7)	10.8 (8.8)
Class A	14.1 (5.5)	19.5 (3.1)	20.6 (1.8)	20.7 (1.7)	20.7 (2.5)
Errors of commission					
Masters	0.8 (0.4)	3.1 (2.2)	2.6 (2.8)	2.7 (0.9)	3.8 (2.4)
Experts	2.9 (1.7)	5.3 (3.5)	5.0 (2.5)	4.0 (3.3)	5.4 (4.8)
Class A	3.3 (3.6)	1.6 (1.1)	1.1 (0.8)	0.9 (0.7)	1.1 (1.0)

Note. Standard deviations are given in parentheses.

Table 3

List of the Chess World Champions Used by S as a Cue List (Retrieval Structure)

1.	Steinitz	Stein
2.	Lasker	Las
3.	Capablanca	Cap
4.	Alekhine	Al
5.	Euwe	Euw
6.	Botvinnik	Bot
7.	Smyslov	Smys
8.	Tal	Tal
9.	Petrosian	Pet
10.	Spasski	Spass
11.	Fischer	Fish
12.	Karpov	Kar
13.	Kasparov	Kas

Note. The third column indicates the abbreviations used by S when rehearsing the names subvocally.

Table A-1

Probability of Meeting the Criterion for a Trial of n Boards when the Probability of Scoring at least 60% on a Randomly Chosen Board is q .

q	Number of boards					
	5	6	7	8	9	10
.750	0.6328	0.5339	0.4449	0.3671	0.3003	0.2440
.775	0.6854	0.5941	0.5092	0.4324	0.3644	0.3051
.800	0.7373	0.6554	0.5767	0.5033	0.4362	0.3758
.825	0.7875	0.7166	0.6464	0.5788	0.5150	0.4559
.850	0.8352	0.7765	0.7166	0.6572	0.5995	0.5443

Note. Criterion = not more than one board less than 60% correct.

Example: if q is .775, then there is a probability of .5092 of meeting the criterion in a trial of 7 boards.

Table A-2
Comparison between DD's Retrieval Structures and Chess Templates

Features	Chess Template	DD's retrieval structure
type of information stored	chess information	digits
speed of information encoding into slots	rapid (1-2 sec)	rapid (~ 1 sec)
accessibility of the entire structure at the same time	no	no
duration of construction	years	years
contains indications for actions	yes	?
consciousness during construction	low	high
consciousness for accessing structure	low	high
duration of information in slots	?	lasts hours
presence of default values	yes	no
type of structure	network structure	tree structure
structure is semantically laden	yes ^a	no
presence of redundancy	yes	yes
points to possible future templates	yes	no
points to past templates	yes	no
multiplicity of templates	yes	yes
dimensionality	bidimensional	unidimensional
order matters	no	yes

^a Provided by the topology of the chess board.

Figure captions

Figure 1: Mean percentage (upper panel) and mean number (lower panel) of correct pieces as a function of chess skill and of number of positions to memorize. Mean percentage and mean number of random positions are shown for comparison.

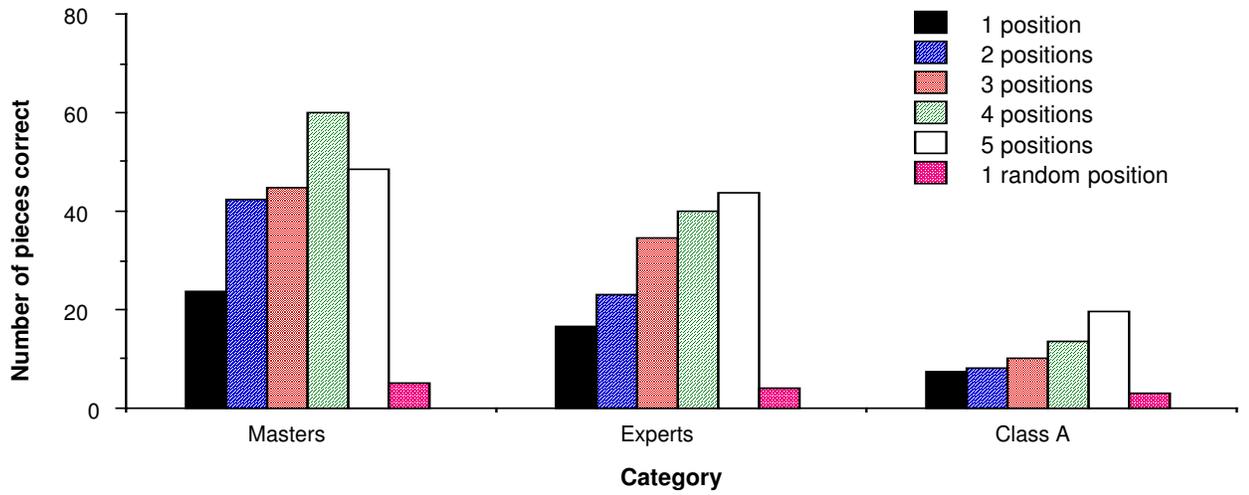
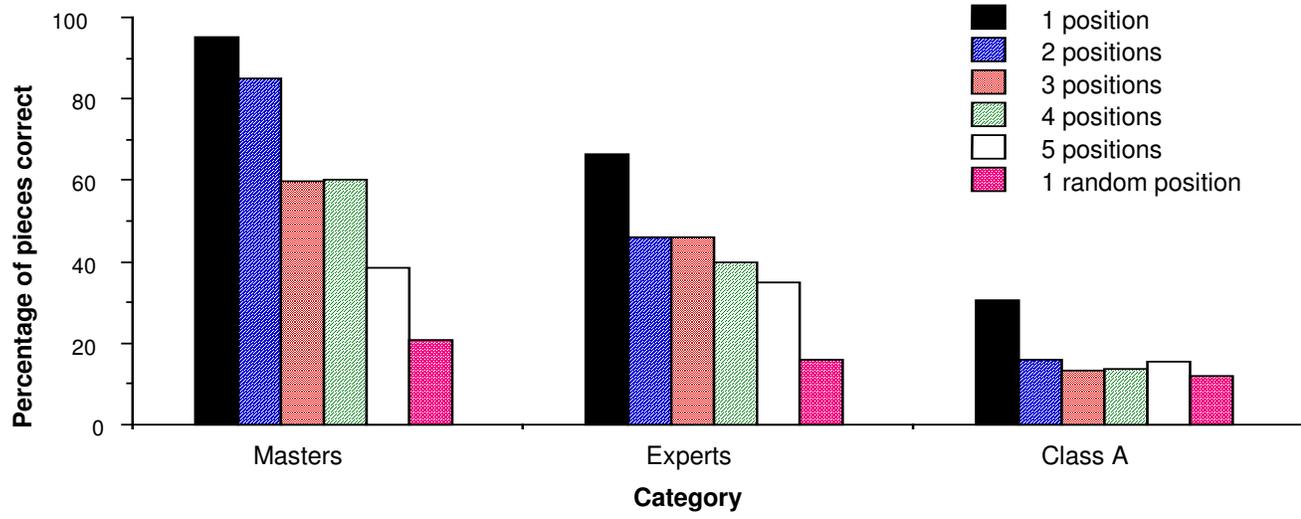
Figure 2: Mean number of correct pieces as a function of number of positions to memorize and inter-position latency for Masters (upper panel) and Experts (lower panel).

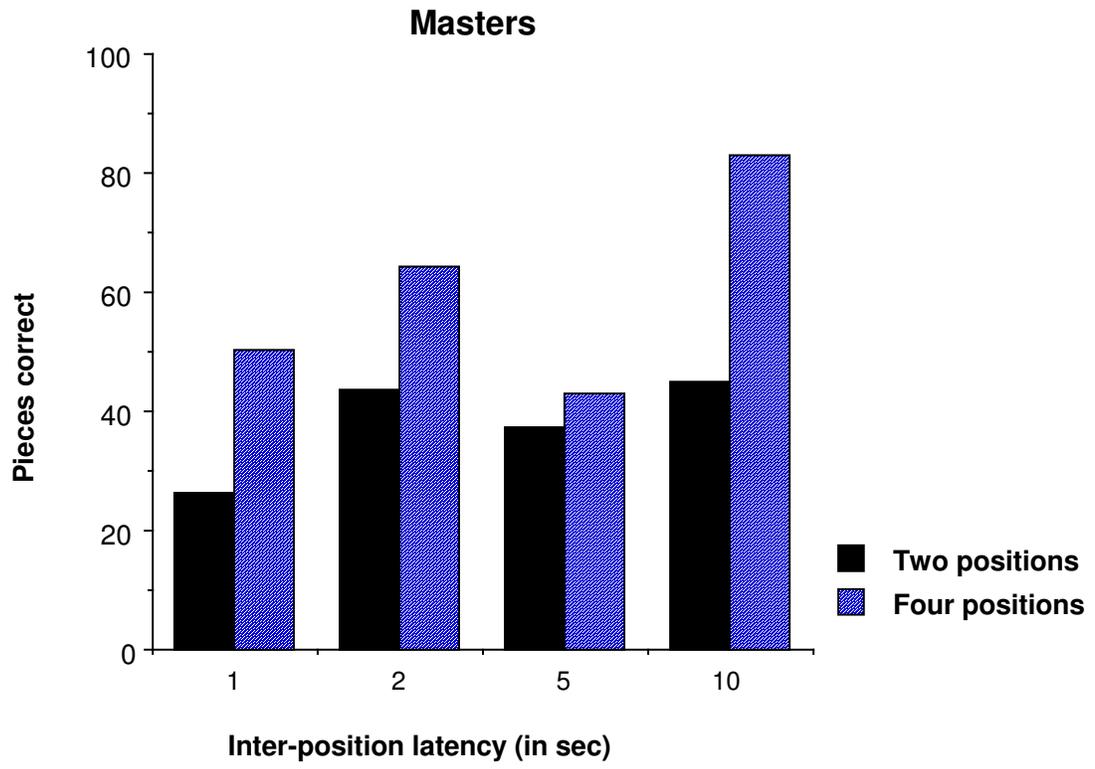
Figure 3: Number of pieces placed correctly by S as a function of the number of blocks of five practice sessions.

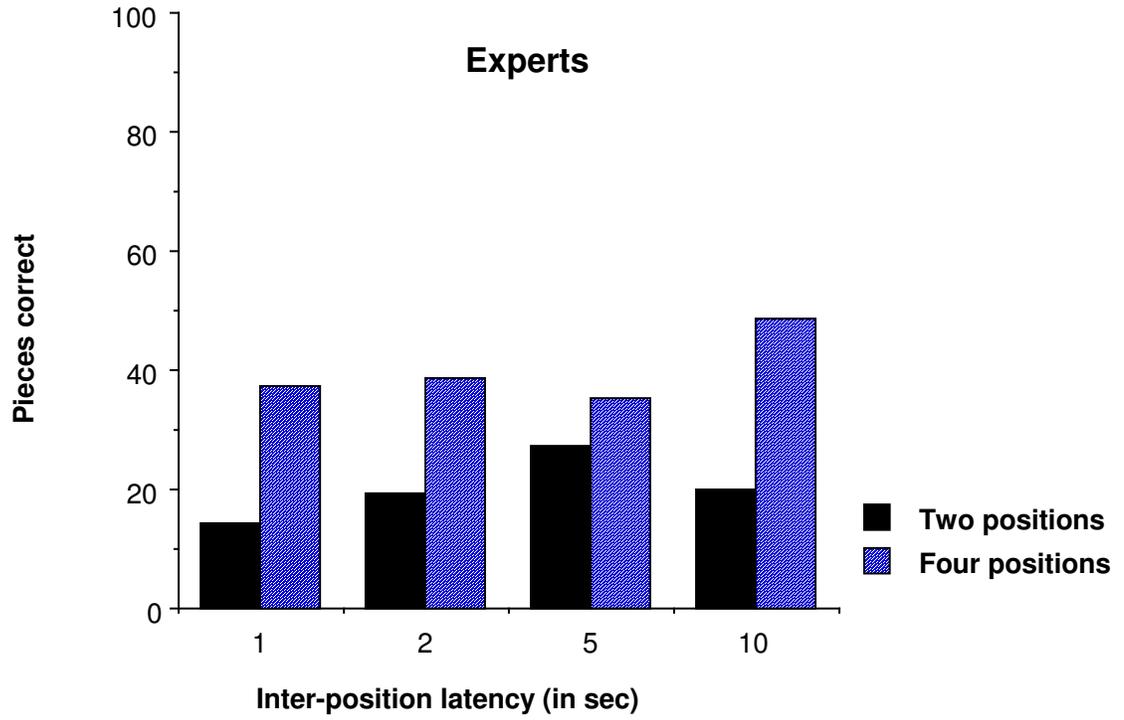
Figure 4: Upper panel: Illustration of the concept of template. The template indicates the stable pieces for a class of positions and contains among others slots for pieces squares openings plans and moves. It also contains links to other templates (e.g. a possible type of position after 10 additional moves). Lower panel: Diagrammatic representation of the same template. Pieces on the board indicates the core pieces in the template and crosses indicates values contained in piece or square slots.

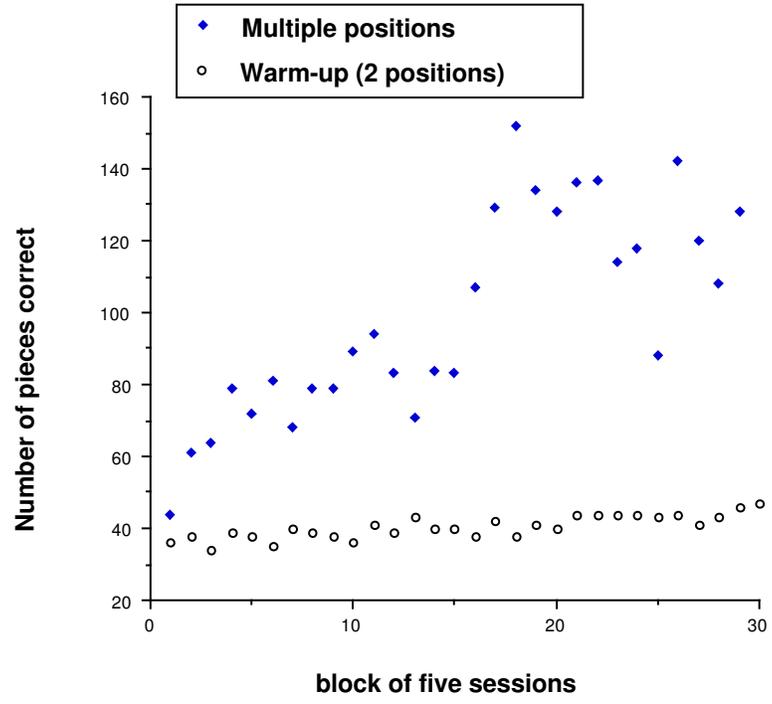
Templates in Chess Memory

65









a)

Template-core:

White ♖c4, ♗d5, ♖e4, ♗f2, ♗g2, ♔g1, ♘c3, ♙e2

Black ♜c7, ♜d6, ♜e5, ♜f7, ♜g6, ♜h7, ♚g8, ♛c8, ♞f6, ♞g7

Slot for pieces:

- ♗ : h2, h3
- ♜ : a7, a5
- ♞ : b8, d7, e5
- ♚ : f8, e8
- ♔ : c2, d1
- ♙ : c1, d2, e3

Slot for squares:

- d2 : ♘, ♙, empty
- e8 : ♚, ♞, empty
- e1 : ♚, ♘, empty

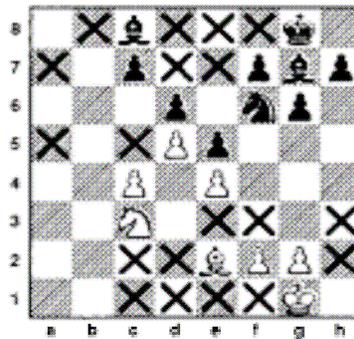
Slot for opening: King's Indian Defense

Slot for plans: Break in the center with f7-f5

Slot for moves: 1... Nf6-e8
1... Nf6-h5

Links to other templates: chunk #231

b)



Footnotes

¹Interestingly, this lack of impairment was also found by Murdock (1961) when a three-letter word was used as stimulus. However, the interfering task had a substantial effect when three such words were to be recalled.

²The ELO rating is an interval scale that ranks competition chess players. Its standard deviation (200 points) is often interpreted as delimiting skill classes. Grandmasters are normally rated above 2500, Masters above 2200, and Experts above 2000. The American rating system (USCF rating) uses the same mode of computation as the international system (ELO). However, because of differences in the games selected for computation, USCF rating is in general about 50 points above the international rating.

³The switch between positions was easy enough to allow the Ss to sometimes reconstruct several boards in parallel or to scan the reconstructed boards rapidly to inform them on their progress.

⁴As chunks are two-dimensional structures, subjects are likely, when replacing pieces serially on the board, to proceed down one path at a bifurcation point in a chunk and neglect pieces on the other path. Later comparison of the chunk in memory with the pieces replaced could notice the omission and repair it. This would afford one explanation of why most trials end in replacements of a sequence of one-piece "chunks." The other explanations (noted in the text) are inference and guessing.

⁵Obviously, this type of recognition is more likely to happen when the opening leading to the stimulus position belongs to the subject's repertoire.

⁶Given the long duration of the experiment, we chose this procedure over the expensive choice of employing another professional player. In incorporating in the experimental design the collaboration of subject with experimenter to find ways of enhancing

performance, we follow the examples of Ebbinghaus, and of the earlier subjects on expert memory for digit strings, SF and DD (Chase & Ericsson, 1982; Ericsson, Chase & Faloon, 1980; Staszewski, 1993). In a test of cognitive abilities, with no deception in the experiment's design, and no possibility for subject deception in an upward direction, an expert member of the research team is an appropriate subject. The research question is how far, using any available knowledge of the process, expert memory can be stretched.

⁷We thank Peter Jansen and Murray Campbell warmly for gathering these positions and coding them into an appropriate format.

⁸We hypothesize that this loss in performance is due to the attention S had to spend in monitoring the use of his retrieval structure.

⁹ Appendix 2 offers a direct comparison between chess templates and the retrieval structures used in the digit-span task studies.

¹⁰There is some evidence from the multiple board experiments that players may access a position after a search of several minutes in LTM without using a pointer held in STM. We propose to account for this (rather rare) phenomenon with an LTM activation mechanism similar to that employed by Richman, Staszewski and Simon (1995).

¹¹ This subject seems to use a mnemonic scheme to access the addresses of these templates rapidly in LTM. If our interpretation of the Cooke and al. data is correct, this subject would then use, as did S in experiment 3, two types of retrieval structures: one to encode the addresses of the positions, and the others (a template for each position accessed) to encode the pieces in the position.