August 18 1996

Gobet, F. & Simon, H. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, 6, 225-255.

Expert Chess Memory: Revisiting the Chunking Hypothesis

Fernand Gobet and Herbert A. Simon
Department of Psychology
Carnegie Mellon University

Send correspondence to:
Prof. Herbert A. Simon
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

Running head: Chunks in Expert Chess Memory

Abstract

After reviewing the relevant theory on chess expertise, this paper re-examines experimentally the finding of Chase and Simon (1973a) that the differences in ability of chess players at different skill levels to copy and to recall positions are attributable to the experts' storage of thousands of chunks (patterned clusters of pieces) in long-term memory. Despite important differences in the experimental apparatus, the data of the present experiments regarding latencies and chess relations between successively placed pieces are highly correlated with those of Chase and Simon. We conclude that the 2-second inter-chunk interval used to define chunk boundaries is robust, and that chunks have psychological reality. We discuss the possible reasons why Masters in our new study used substantially larger chunks than the Master of the 1973 study, and extend the chunking theory to take account of the evidence for large retrieval structures (templates) in long-term memory.

Expert Chess Memory:
Revisiting the Chunking Hypothesis

How can chess masters play high quality games when they are allowed only five minutes for the entire game? How can they recall almost perfectly a position presented for a few seconds? Chase and Simon (1973b) proposed that Masters access information in long-term memory (LTM) rapidly by recognizing familiar constellations of pieces on the board, the patterns acting as cues that trigger access to the chunks. Because these chunks are associated with possible moves, chess masters can generally choose good moves with only moderate look-ahead search. Because storing one chunk in STM gives access to a number of pieces, masters perform remarkably well in recall tasks. As this theory and the consequences that flow from it have had considerable impact on the study of expertise in numerous domains (Charness, 1992), its validity is of interest to cognitive psychology generally.

Chase and Simon carried out little more than an exploratory experiment. They studied only a single Master, a single Expert and a single Class A player. Moreover, the Master was rather inactive in chess at the time of the experiments and performed substantially less well than other Masters who have been tested in the same or similar tasks. In addition, as the subjects used actual chess boards and pieces, the maximum number of pieces they could grasp in one hand could have limited apparent chunk sizes. For these reasons, and because of the amount of attention the experiment has attracted, it seemed important to carry out a new study, not simply as a replication, but in such a way as to overcome the limitations of the original study (especially the two just mentioned) and to re-examine and illuminate some of the issues that have been raised in the literature about that study and its interpretation.

After summarizing Chase and Simon's (1973a) definition of chunk, we answer the major criticisms that have been aimed at the chunking theory, and propose a modest reformulation of the theory that makes different predictions about the size of chess masters' chunks, and especially the largest chunk, than the original version. Comparing a copy and a recall task, we show that the 2-second chunk boundary proposed by Chase and Simon is robust. Comparisons between latencies and frequencies of various chess relations indicate that, in both tasks, different processes are used to place successive pieces within a chunk than to place the first piece in a new chunk.

What is a Chunk?

From the standpoint of the theory, a chunk is a LTM symbol, having arbitrary subparts and properties, that can be used as a processing unit. Each chunk can be retrieved by a single act of recognition. Chunking has been pinpointed as a basic phenomenon in chess expertise at least since De Groot (1946/1978), who noted that chess positions were perceived as "large complexes" by masters. The concept was made more precise by Chase and Simon's (1973a) proposed operational definition of chunks in chess. Comparing the distributions of latencies in a memory task (the De Groot recall task) and a perceptual task (copying a position on a different board), they defined a chunk as a sequence of pieces placed with between-piece intervals of less than 2 seconds.

According to the theory, pairs of pieces that have numerous relations are more likely to be noticed together, hence chunked. Chase and Simon then analyzed the chess relations (attack, defense, proximity, same color and same type) between successively placed pieces in the two tasks and in different types of positions, thereby demonstrating that the probabilities of these relations between successive pieces belonging to a chunk (less than 2 seconds' interval) are much greater than the probabilities between successive pieces not belonging to a chunk (an interval of more than 2 seconds). The large average differences observed add considerable credence to the reality of chunks.

Chase and Simon (1973b) proposed that, during the brief presentation of a chess position, players recognize already familiar chunks on the board and place pointers to these chunks in a short-term memory of limited size. A computer program, MAPP (Simon and Gilmartin, 1973), simulated several experimental findings, including the percentage of pieces recalled by a class A player,[1] the types of pieces replaced and the chess relations between successive pieces in the reconstruction. Simon and Gilmartin estimated that expertise in chess would require between 10,000 and 100,000 chunks in memory (in the literature, this range is often reported simply as 50,000 chunks). Finally, Chase and Simon's theory of memory implies that chunks, upon recognition, would suggest good moves to the masters.

Other Experimental Evidence for the Chunking Hypothesis

The evidence of Chase and Simon (1973a,b) was obtained from a single experimental paradigm. Chunk structures have been identified experimentally in other paradigms as well. Charness (1974) presented pieces verbally, at a rate of 2.3 s per piece. Pieces were either grouped by the experimenter according to the chunking relations proposed by Chase and Simon (1973a), or ordered by columns or dictated in random order. Charness found better recall in the chunking condition than in the column condition, and poorest recall in the random condition. The same results were found when pieces were presented visually, one at a time (Charness, 1974).

Similarly, Frey and Adesman (1976) presented slides, each containing a group of (usually) four pieces, but retaining the pieces from previous slides. Each of the six slides for a position was presented for 2 s. Chunk presentation produced better recall than column presentation; and, in fact, better recall than presentation of the entire position for the same length of time (12 s).

Two important results have been found by Chi (1978), who applied to chess the partitioning technique devised by Reitman (1976) for studying Go memory. Given the diagram of a position, subjects draw boundaries around the groups of pieces they perceived. First, chunks sometimes overlapped. Second, in the recall task, Chi found that subjects took longer, on average, to place pieces crossing a chunk boundary (about 3 s) than to place pieces within a chunk (around 1.5 s). Chi observed that this finding supports Chase and Simon’s (1973a) estimate that it takes at least two seconds to retrieve a new chunk and less than two seconds for within-chunk retrieval.

Freyhoff, Gruber and Ziegler (1992) used a similar partitioning procedure, with the addition that subjects had both to divide the groups obtained in a first partition into subgroups and to combine the original groups into supergroups. Masters created larger clusters at all levels of partitioning than did class B players. In addition, the chunks they detected at the basic level corresponded to the chunks identified by Chase and Simon (1973a). First, their size was, on average, 3.6 pieces for masters, and 2.7 pieces for class B players—reasonably close, given differences in the types of positions used, to Chase and Simon’s 2.5 pieces for the Master and 2.1 pieces for the Class A player. Second, the pattern of relations between pieces was very similar to that found by Chase and Simon. In particular, 74.6% of the pieces within partitions shared three or more relations, as compared with 67.6% in Chase and Simon's data for the recall task.

Gold and Opwis (1992) applied hierarchical cluster analysis to chess players' chunk structures. The variables were the locations of pieces on the board, and their values were their correct or incorrect recall. Clustering was determined by aggregating over subjects the frequency with which pieces of each pair were both placed correctly or incorrectly. The clusters identified with this technique constituted stable and easily interpretable partitions similar to those identified by latencies (e.g. castled positions, chains of pawns, common back-rank piece positions).

Gruber and Ziegler (1990) found that chess players, ranging from average club players to Grandmasters, used knowledge units when sorting a position similar to the chunks identified by Chase and Simon (1973a,b). However, the number of units decreased and their size increased with expertise, stronger players using overlapping sorting criteria that grouped chunks together. Gruber (1991) also found that, when allowed to ask questions about an as yet unseen position, chess experts asked about the past and future path of the game, about plans and evaluations, while novices asked about the locations of single pieces (see also De Groot, 1946, for early investigations on the role of complex knowledge in chess). The template theory, which we will present later, proposes that experts encode knowledge as relations between chunks and store other information besides the location of pieces.

Retrospective verbal protocols do not seem to reveal much about chunking. De Groot and Gobet (in press) found that Masters often give high-level descriptions of a position, such as type of opening or main strategical plans, but almost never mention clusters of pieces sharing relations of defense, attack, and proximity. They do mention what De Groot and Gobet call "visual images," occurring about once per protocol, where such perceptual properties as similarity or contrast of color, and geometrical shapes, dominate over semantic features. De Groot and Gobet propose that chunks are missing from these verbal protocols, first, because units are so self-evident for Masters that they are not conscious of them, and second, because masters may not have verbal labels for many of these perceptual units.

In summary, these experiments support the psychological reality of chunks as defined either by numbers of (chess-)meaningful relations or latency in placement. The two criteria are bound closely together, theoretically and empirically, in the chess recall tasks, as well as in verbal and pictorial recall tasks that involve semantic clustering (Wixted & Rohrer, 1994).

Criticism of the Chunking Hypothesis

The chunking model has spawned considerable empirical work (see Holding, 1985 or Gobet, 1993, for reviews), but has also been challenged on several grounds. We now review the most important of these criticisms, going from general aspects of the chunking theory to the specific way the chunks were identified in Chase and Simon's analysis (1973a).

The Recognition-Association Assumption

One central thesis of the chunking model is that chunks act as cues that, when recognized, evoke access to heuristic suggestions for good moves. Holding (1985, 1992) has challenged this assumption on the grounds that (a) most chess patterns consist of pawns,[2] and pawn structures do not generate many moves; (b) that most chess patterns found by Chase and Simon (1973a,b) are too small to provide useful information; and (c) that pattern recognition is not sufficient to explain chess skill, because it applies only to the initial moves from the stimulus position and does not allow for look-ahead analysis.

The claim (a) that few moves are evoked by pawn structures, is refuted by the significance that chess players attach to pawns. Their importance was recognized already in the eighteenth century by Philidor (1749), who stated that "Pawns are the soul of chess." Whole books (for example Euwe, 1972; Kmoch, 1980) analyze the proper way to handle pawns and describe typical pawn structures. Pawn structures provide information about the squares on which pieces should be placed (e.g. a Knight in front of an isolated pawn) and also about typical pawn moves for given structures. Subjects, while thinking aloud, frequently comment on pawn structures and on moves relevant to them in problem solving tasks (see De Groot, 1949/1978) and even in memory tasks (De Groot & Gobet, in press; Gobet, 1993).

The claim (b) that chunks are too small to generate useful information (Chase and Simon hypothesized chunks of at most 5-6 pieces) may have some truth, although even small chunks can suggest good moves in tactical situations, and chunks or small constellations of them allow recognition of positions of particular types. Moreover, as the experimental part of this paper shows, Chase and Simon probably underestimated chunk size, especially for masters.

The claim (c) rests on a misunderstanding of the theory. Holding states that "...the basic assumption of the pattern-move theory [is] that the better players derive their advantage simply from considering the better base moves suggested by familiar patterns" (Holding, 1985, p. 248), where "base moves" are moves playable in the stimulus position. On the contrary, Chase and

Simon (1973b, pp. 268-272) stated explicitly that recognition of patterns is used not only to generate base moves but also subsequent moves triggered by patterns in the "mind's eye" at deeper levels during search:

> "When the move is made in the mind's eye—that is, when the internal representation of the position is updated—the result is then passed back through the pattern perception system and new patterns are perceived. The patterns in turn will suggest new moves, and the search continues." (Chase & Simon, 1973b, p. 270).

A study by Holding and Reynolds (1982) is often cited as evidence against the recognition-association theory. In their study, the skill of subjects (from 1000 to 2200 ELO) did not correlate with the recall of random positions[3] shown for a few seconds, but effectiveness of the search for the best move in these positions did correlate with skill. However, because pattern recognition is applied recursively during look-ahead, a memory test only on the initial problem position does not really address the recognition-association theory.

Although Chase and Simon only mention chunks as eliciting (initial or subsequent) moves, chunks, especially the large chunks we call templates, can also provide information about the class to which the position belongs, about heuristics, plans, partial evaluation of the position and so on. Pattern recognition then facilitates the generation of moves and plans during search and allows a rapid and precise evaluation of positions met during search. Indeed, pattern recognition can provide the basic mechanisms that are needed for, but are now lacking in SEEK (Search, Evaluation, Knowledge), the model of chess expertise that Holding (1985) proposed in place of Chase and Simon's model.

The Number of Chunks Needed to Reach Expertise

The estimate that 50,000 chunks must be stored in LTM to reach expertise is an extrapolation from the simulations performed by MAPP, the program described by Simon and Gilmartin (1973). Holding (1985, 1992) argued that this number is much too large, and that as few as 2,500 chunks may account for the results obtained in recall experiments, by assuming that the chunks encode relations between pieces but not the location of these pieces. In that case, the same chunk could encode a pattern of pieces on the White and Black side of the board, or a pattern of pieces that had been shifted by several squares. Gobet and Simon (in press-b) and Saariluoma (1994) addressed this hypothesis by asking subjects to recall positions that were modified either by taking a mirror image or by translation. They found that, in comparison with unmodified game positions, the manipulations degraded recall performance. These results undermine the hypothesis that a pattern of pieces can be recognized independently of its position on the chessboard and add support to the estimate of 50,000 chunks.

The Emphasis on STM Storage

According to the chunking model, pieces are encoded in a STM of limited size during the recall task and no new information is then added to LTM. However, studies using interfering material in intervening tasks (Charness, 1974; Cooke, Atlas, Lane, & Berger, 1993; Frey & Adesman, 1976, Gobet & Simon, in press-a) have shown that this material does not interfere much with chess memory, thus implying that, as the interfering tasks reduce retention in STM, some information has to be transferred rapidly to LTM.

Gobet and Simon (in press-a) propose a modified model that is in accord with other recent models of expert memory (e.g., Richman, Staszewski & Simon, 1995), and accounts for the rapid encoding shown by chess masters. The modified theory asserts that—as in the chunking theory—chunks are accessed through a discrimination net. In addition, chunks that recur often in masters' practice and study evolve into more complex data structures, called templates. Templates, besides containing information about a pattern of pieces, as chunks do, possess <u>slots</u> (variables that can be instantiated) in which some new information can be stored in a matter of seconds. In particular, information about piece location or about chunks can be (recursively) encoded into template slots. The basic mechanism allowing this rapid LTM storage is the same as the one proposed by Chase and Ericsson (1982) to account for expert digit memory. (For a similar, but less specific, proposal for rapid storage in existing LTM structures, see Simon, 1976.)

Notice that, although slots in templates can be filled rapidly, hence serve essentially as augments to STM in the domain of expertise, the templates themselves are built up slowly, at normal LTM learning rates. Finally, templates contain pointers to symbols representing plans, moves, strategical and tactical concepts, as well as other templates. These pointers are also acquired at normal learning rates (i.e., 5 to 10 seconds per chunk).

The template idea is compatible with the findings of De Groot (1946/1978), who emphasized that his Grandmaster and Master were able to integrate rapidly the different parts of the positions (Chase & Simon's chunks) into a whole, something weaker players could not do. The integrated representation can depict a typical opening or middle game position. We have mentioned earlier empirical evidence that strong players are able to access rapidly descriptions of the position that are larger than 4 or 5 pieces.

The template hypothesis and the evidence supporting it predict that strong players should replace positions in chunks (templates) larger than the ones identified by Chase and Simon (1973a). This is of course at variance with their findings (which, we have noted, were based on the performance of only a single Master). In order to evaluate this discrepancy between the template theory and the earlier estimates of chunk size we must consider the last set of criticisms aimed at the chunking theory, which relate to using inter-piece response latencies to identify chunks.

<u>The Operationalization of Chunks</u>

Several authors (Freyhoff, Gruber & Ziegler, 1992; Gold & Opwis, 1992; Holding, 1985; Reitman, 1976) have seen difficulties in Chase and Simon's method for defining chunks, among the most important of them: (a) difficulty in identifying chunks by reaction times, (b) impossibility of capturing overlapping or nested chunks, (c) difficulty in assigning pieces erroneously replaced and (d) the assumption that each chunk is recalled in a single burst of activity during board reconstruction. These objections raise serious difficulties if the goal is to cut a chess position into precise chunks but are not fundamental for analyses that relate chunks to the distributions of relations between pieces, as is the case for Chase and Simon's (1973a) study and the present one. Moreover, as we have seen above, various alternative techniques (partitioning, sorting) provide converging evidence that supports the original results of Chase and Simon.

Two other methodological concerns may be mentioned. First, specific latency criteria may not provide unambiguous chunk boundaries because, as

Wixted and Rohrer (1994) showed, in recall both from STM and LTM, latencies generally become longer as successive items are recalled. Successive pieces placed early in recall would be assigned to the same chunk while those placed later in recall, with longer latencies, would be assigned to separate chunks. This could account for the observed larger average size of the early than of the late chunks. We will take up this question in the experimental part of our paper.

Second, subjects in the original study replaced pieces by picking up several of them simultaneously. Hand capacity will limit the number of pieces that can be grasped, hence the estimated size of chunks. In the same line, subjects might grasp pieces more or less randomly, and then look for appropriate locations for them. Our new experimental procedure eliminates these two potential problems.

Overview of the Experiment

Most of the criticisms we have reviewed were either due to misinterpretation of the chunking theory or pointed toward the necessity for postulating some kind of rapid encoding into LTM, a requirement that is now met in the template theory (Gobet & Simon, in press-a). Still, there is warrant for testing further the validity of Chase and Simon's method for identifying chunks: infelicities in the original study; criticisms of the technique used; evidence that Masters perceive a position at a higher level than 4-5 piece chunks; a different prediction of the template theory about the size of chunks; the small number of subjects. In addition, if the close relation between the number of relations joining a pair of pieces and the likelihood of the pair being perceived in rapid succession were confirmed, then numbers of relations, on the one hand, and latencies, on the other, provide converging evidence about the numbers, sizes and character of the chunks that experts perceive.

Chase and Simon (1973a) studied the clusters and timing relations in the output (for earlier uses of such techniques, see Tulving, 1962, and Bower & Springston, 1970), using two experimental paradigms in order to isolate and define chunks. In the copy task, subjects reconstructed a chess position while keeping the stimulus position in plain view. Successive glances at the stimulus position were used as the index of chunking, on the assumption that one chunk is encoded per glance. In the recall task, subjects reconstructed a position presented for 5 s. The time between the replacement of successive pieces was used to segment the output into chunks. Chase and Simon found that pairs of pieces within chunks identified by the copy and recall methods showed the same pattern of relations, but a different pattern from that shown by pairs of pieces belonging to different chunks.

Replications are rare in chess psychology research (Charness, 1988; Gobet, 1993), and the data supporting the 2-s boundary for delimiting chunks have never been replicated experimentally. For reasons already discussed we are more interested in an extension and clarification of the earlier results than an exact replication. The most important difference between our experiment here and the earlier study is that we use a computer display instead of physical chess pieces and board. The new apparatus removes the possible artifact in Chase and Simon's experiments, that chunks may have been limited by the hand's capacity to grasp pieces. We will show that the change in apparatus provides converging evidence supporting the standard method of identifying chunks.

We first analyze the latencies in replacing pieces in the copy task and discuss strategies employed by the subjects. We then compare these results with those obtained in the recall task, focusing on the latencies and the chess relations between successive pieces. Data on the size of chunks will be examined next. Finally, we consider the implications of our results for the chunking theory.

Methods

The copying and recall tasks were given as part of a larger design to the subjects of Experiment 1 of Gobet and Simon (in press-a) and to half of the subjects of Experiment 2 of Gobet and Simon (in press-b). All subjects carried out the copying task (with the same material and instructions) at the beginning of the experimental session, after they were introduced to the computer program used to run the experiments and before the main experimental manipulation of the session. The random positions of the recall task were presented immediately after the copying task. The game positions in the recall task were then given as the initial stage of an experiment on the recall of multiple boards (Experiment 1) and as the control condition of an experiment on the effect of mirror-image modification of positions (Experiment 2). We decided to pool the results from the two experiments, as there was no difference between the two experimental groups nor any interaction of experimental group with the variables discussed below.

Subjects

Twenty-six male subjects participated in the experiment, recruited from players participating in the Nova Park Zürich tournament and from the Fribourg (Switzerland) Chess Club, and were paid SFr 10.- (SFr 20.- for the players having a FIDE title). Their Swiss ELO ratings ranged from 1680 to

2510, with a mean of 2078 and a standard deviation of 233. Subjects were grouped in three skill levels: Masters (n = 5; mean ELO = 2453), Experts (n = 9; mean ELO = 2148) and Class A players (n = 12; mean ELO = 1869). The mean age was 29.7 years (sd = 8.5). The youngest subject was 18 years, the oldest 49 years.

Materials and Procedure

Copy task.

Experiments were run with a Macintosh SE, having a high resolution 9 inch diagonal screen (512 by 342 pixels). The positions were presented on the screen with a 9 x 9 cm chessboard. Individual squares were 11.25 x 11.25 mm. Pieces of standard shape were used. The background was black during the presentation of the board. Between the presentation of one stimulus board and the presentation of the reconstruction display, the screen was black.

The reconstruction display had the following appearance: an empty 9.5 x 9.5 cm empty board (lower left corner of the board 1.35 cm from the lower left corner of the screen), a rectangular box (2.4 x 7.1 cm, 2.2 cm from the right side of the screen) displaying the 6 different kinds of pieces of White and Black, a 11.9 x 11.9 mm box below the previous box where the selected piece was displayed, an “OK” box near the upper left corner of the screen, permitting the subject to choose when to receive the next stimulus. To place a piece, the subject first selected the desired kind in the “pieces box” by clicking the mouse, and then clicked it on the appropriate square, producing an icon of the piece on this square. Each successive piece had to be selected independently with the mouse from the rectangular box displaying the kinds of pieces. Only the mouse was used by the subjects (not the keyboard).

Two numbered boxes were displayed near the top of the screen for switching the display between the position to be copied and the reconstruction board. The two positions (the model and the position being reconstructed) were slightly shifted and of a different size, in order to avoid subjects' using iconic memory to superimpose one on the other.

Log files recorded the following data: time between the selection of a piece and its placement; time between the placement of two pieces (interpiece latency); type of piece placed and its location; removals of pieces and placements outside the board.

Five positions (see Appendix A) were used, 3 taken from master games (with 24, 30 and 26 pieces) and 2 random positions (with 25 and 28 pieces). Random positions were created by randomly reassigning to new squares the pieces of a game position. The five positions and their order of presentation (game - random) was the same for all subjects. The concern in this experiment was not in demonstrating the superior memory for the game as compared with random conditions—a very large superiority, already established beyond reasonable doubt—but in exploring the relation between the two definitions of chunking, the one based on latencies, the other on chess relations between successive pieces. Hence, the confounding caused by presenting the game positions before the random positions was of minor importance for the purposes of the experiment. The first game position was used for practice and is not included in our analyses.

After subjects were introduced to the computer program and, if necessary, to the use of the mouse, they were given the copy task. A position was presented on the screen, and subjects had to reconstruct (copy) it on another board, which they could access by clicking a particular box on the screen. Only one board was visible at a time. Subjects could switch from the stimulus position to the copy as often as they wished. They were encouraged to do the task as fast as possible.

Recall task.

The recall experiments were carried out in the same way as the copy experiments, except that after the stimulus position was presented for 5 s, it was no longer available to the subjects, who had to replace the pieces on the board from memory.

The game positions used in the recall task were taken from master games after about 20 moves with White to move, from various chess sources. The positions were "quiet" (i.e. were not in the middle of a sequence of exchanges). A computer program generated random positions by randomly reassigning to new squares pieces from game positions. For the recall of game positions, subjects' results are based on 4 positions for the subjects who participated in Experiment 2 of Gobet and Simon (in press-b) and on 5 positions for those who participated in Experiment 1 of Gobet and Simon (in press-a). The game positions were randomly selected from a pool of 16 positions for the former subjects and of 26 positions for the latter (see Appendix A). For all subjects, data on random positions are based on three positions. The mean number of pieces per position (random or game) was 25.

The random positions were presented before the game positions (the latter being used also as the initial task of another experiment). As in the case of the copy task, we judged the confounding due to the non-random order of presentation to be acceptable, because we were not primarily interested in comparing the levels of reconstruction of the game and random conditions.

Results and Discussion

We present our results in four parts. First, we analyze the copy task, to establish the relation of within-glance to between-glance latencies. Second, we examine the percentage of correct recall in the recall task. Third, we compare the copy and recall task with respect to the latencies between pairs of pieces and the number of relations between pairs of pieces. We use these findings to establish converging definitions of chunks by (1) a latency criterion and (2) a criterion of number of relations between successive pieces. Fourth, we examine the size distribution of chunks and numbers of chunks.

We will show that our data generally agree well with the data from the earlier experiments of Chase and Simon (1973a and b), with some differences in sizes and numbers of chunks that are more compatible with the revised template theory than with the chunking theory in its original form. In the third section, we will add credibility to the modified chunking theory by showing that there is converging evidence, from latencies and from relations between pieces, that provide alternative, independent but quite consistent ways of defining chunks.

Copy Task

All subjects but one (an expert) were proficient in handling the mouse. The subject who experienced difficulties dictated (using algebraic chess notation) the location of the pieces to the experimenter, who placed the pieces on the board with the mouse. In general, the time to move the mouse once a piece is selected is independent of players' skill ($\underline{r}$ = .05 for game positions and $\underline{r}$ = .01 for random positions). To remove learning effects, the first position is omitted from the following analysis.

In the remainder of this section, we investigate the following three variables: inter-piece latency, total time to study a position and number of times subjects accessed the stimulus position. We then comment on the strategies used and on the role of age.

An important difference in the behavior of our subjects from Chase and Simon's (1973a) in the copy task should be mentioned first. Their subjects studied the stimulus position for a short time (a few seconds), then replaced a few pieces on the copy board, repeating this cycle until all pieces had been replaced. Our subjects (especially the Masters) studied the stimulus position for some dozens of seconds before placing the first piece; later, they rarely revisited the stimulus. This difference in behavior may be related to the differences in the ways in which stimuli were presented and responses made in the two sets of experiments. We will see that, in spite of this difference in strategy, most of our results accord closely with Chase and Simon's.

Latencies between successive pieces.

Like Chase and Simon, we were interested in two modes of placement: (a) within-glance placement (WGP): piece placed without switching back to the stimulus position; and (b) between-glance placement (BGP): piece placed after switching back to the stimulus position;

The latencies between successive pieces will be analyzed using a 3 x 2 x 2 (Skill level x Type of position x Placement Mode) factorial design, with repeated measurements on the two last variables. Because of the skewness of the distributions, medians are used as the measures of central tendency (the means were close to the medians). The first piece placed in each position was omitted from the analysis. Figure 1 shows, for each skill level, type of position and type of placement, the mean of the medians.

Insert Figure 1 about here

One master subject did not produce any BGP when copying game positions (he viewed the board only once before copying it), hence his data were not used when computing the following ANOVAs. There is an important difference between WGP and BGP: WGP latencies are much shorter than BGP. ANOVA indicates this main effect of Mode of placement [$F(1, 22) = 90.74$, $MS_e = 10.3$, $p<.001$]. No main effect of Skill [$F(2, 22) = 0.32$, $MS_e = 13.5$] or of Type of position [$F(1, 22) = 1.63$, $MS_e = 2.15$] are found. A marginal interaction is signaled for Skill x Type of position [$F(2, 22) =3.17$, $MS_e = 2.15$, $p=.062$]. There are no other two-way or three-way interactions: Type of position x Mode of placement [$F(1, 22) = 1.25$, $MS_e = 2.0$]; Skill x Mode of placement [$F(2, 22) = 0.37$, $MS_e = 10.3$]; Skill level x Type of position x Mode of placement [$F(2, 22) = 2.58$, $MS_e = 2.02$].

Mode of placement is thus significant at the .001 level. Besides, Masters show an interesting pattern: in contrast with the other players, their BGPs are much slower with random positions than with game positions. This difference accounts for almost the whole of the (marginal) interaction effect of Type of position x Skill : Experts and Class A players keep almost the same rhythm for the BGPs in both game and random positions.

The WGP latencies are longer than those found by Chase and Simon (1973a). In their data, 80% of the WGP latencies were less than 2 s, with a median around 1 s and a mode around 0.5 s (estimated from their graph). For our subjects, the median is 2.63 s and the mode is about 2.37 s. This difference can be explained by the time needed to move the mouse, which is greater than the time needed to pick up a piece from one's hand or from the side of the board. We have therefore computed a corrected latency, by subtracting from our times the time needed to move the mouse to the destination square once a piece has been selected (this time was recorded in our log files for each placement). Figure 2 reproduces, for all of our subjects, the corrected BGP latencies (180 observations) and the corrected WGP latencies (1283 observations) in game positions. About 79.5% of the corrected WGP latencies are now below 2 s, with a median of 1.37 s and a mode of 1.13 s, in reasonable agreement with the Chase and Simon's data.

As the correction we used may appear a bit ad hoc, we also examined latencies after subtracting the mouse move time estimated from Fitts' Law corrected for errors (Welford, 1968), employing the parameters proposed by Card, Moran and Newell (1983, p. 241-242): $T_{pos} = K_O + I_M \log_2 (D/S + .5)$, where T_{pos} is the positioning time, I_M = 120 msec/bit, D = distance of the target, and S = size of the target. For K_O, the intercept, we used 400 msec, obtained by computing the time to click and unclick the button of the mouse (4 x 100 msec). The corrected distribution of WGP latencies has now a median of 1.49 and a mode of 1.25. We obtained similar results when we fitted these parameters individually for each subject. None of the results we report in this paper are changed if we use the correction based on Fitts' Law instead of the correction based on the time to move the mouse once a piece has been selected.

Insert Figure 2 about here

As in Chase and Simon (1973a), the WGP and BGP distributions are quite different. In the present experiment, 79.5% of the WGP latencies (against only 1.11% of the BGP latencies) are less than 2 s, and 89.3% (against 4.4%) are below 2.5 s. The times, consistent for our three skill levels, are close to Chase and Simon's, although a little slower even after the correction for the mouse. The close agreement adds considerable credibility to the 2-s boundary as a basis for defining chunks in replacement experiments.

The between-glance latency distribution has small peaks at 3.75 and 5.75 s and a median at 7.3 s. BGP latencies are longer than those found by Chase and Simon (means around 3 s), which may reflect differences in strategies used by our subjects. Note also that, because of the design of the program, it was difficult to access the stimulus position and come back to the reconstruction board in less than one second, which may have provided one motive for fewer and longer references to the stimulus.

Total study time.

Total time studying the stimulus position is not identical with the sum of between-glance latencies, for (a) subjects, once they choose a piece, need some time to move it; (b) some subjects did examine the stimulus position after the reconstruction to check for correctness, without placing any new piece. The ANOVA shows a main effect of Skill [$F(2,23)=5.85$, $MS_e = 1265.5$, $p<.01$] and Type of position [$F(1,23)=109.72$, $MS_e = 332.8$, $p<.001$]. In the game positions, time to study the stimulus position seems to be a linear function of chess skill (28.6, 48.5, and 76.9 s from higher to lower skill). In the random positions, Masters are faster than the others (97.0, 98.8, 128.4 s), but slower than would be predicted from their times in game positions.

However, the interaction is not statistically significant [F (2,23)=0.89, MS_e = 332.8, ns].

Number of references to stimulus.

For the number of times, on average, subjects referred back to the stimulus position, the ANOVA shows a main effect of Skill [$F(2,23)=8.31$, MS_e =10.29, p=.002] and Type of position [$F(1,23) = 176.36$, MS_e = 1.62, p<.001]. No interaction is found [$F(2,23)=1.65$, MS_e = 1.62, ns]. The mean number of references to the stimulus decreases with chess skill, and game positions require fewer references (2.5, 4.9, 7.2) than random positions (6.8, 10.8, 12.0).

Subjects' comments and strategies.

When copying game positions, subjects' comments are similar to those uttered during a recall task; in particular, subjects often conjecture from what opening the position comes, and make many corrections in placing lateral pawns (pawns on columns "a" and "h") and rooks.

Although no subject actually refused to copy (or to recall)[4] a random position, this task elicited deep negative feelings. Stronger players tended initially to try to replace pieces by semantic groups; then, encountering difficulty, switched to reconstructing the position in a more systematic way, roughly line by line or column by column. Whatever strategy was used, subjects had to correct numerous color errors.

Copying a random position is more like a problem solving task than a memory task. With random positions, Masters tend to study the stimulus in longer glances than those of weaker players, but to return less often to look at the stimulus position. They try to memorize the position as if it were a game position. The weaker players use a strategy less expensive in memory requirements: they cut the position in small chunks, generally by columns and rows, and copy these chunks. One might say that Masters over-estimate their memory capacity for (randomly placed) chess material through failure to recognize their memory's limits in the absence of meaningful chunks.

Subjects' comments on their strategies are corroborated by a quantitative analysis of the reconstructions. We looked at the number of subjects using a "strict" line-by-line or column-by-column procedure, where "strict" means that only deviations of one square in any direction from the order predicted by the procedure are allowed. Reconstructions that follow the systematic strategy only in part (e.g. the subject starts with the line-by-line procedure, and then switches to a semantic strategy), are not counted as strict. The strict systematic strategy was used by only one Master, on his second random board; but Experts and Class A players used this strategy in copying 38% of the random boards. The strategy was used by 22% of Experts and 33% of Class A players on the first board, 56% of Experts and 42% of Class A players on the second.

Role of age.

Because Charness (1981) has shown that age affects chess position memory, and because age is known to slow motor tasks—in our case, moving the mouse—we checked to make sure that age did not introduce spurious relations into our data. As the age of our oldest subject was only 49, we should not expect large effects. Moreover, age and skill (measured by ELO points) were orthogonal in our sample (r = .09, ns.).

Having used latencies in our analyses to infer the cut-off values between within-chunk and between-chunk placements, we analyzed the correlations between age, on the one hand, and (a) the median interpiece latency, and (b) the median time to move the mouse once a piece was selected, on the other hand, for copying both game and random positions. We obtained the following correlations with age: (a) for game positions: interpiece latency (.22), time to move the mouse (.10); (b) for random positions: interpiece latency (.26); time to move the mouse (.01). As none of the correlations were significant at $p = .10$, we may conclude that age did not much affect time to move the mouse or compromise our estimate of chunk cut-offs.

Age also did not correlate significantly with recall of game positions ($r = .06$) or of random positions ($r = -.36$), though the observed correlation was stronger for the latter. We will omit age as a variable in the rest of our analyses.

Percentage Correct in the Recall Task

With game positions, the percentages of pieces correctly recalled are 92.0, 57.1 and 32.2 for Masters, Experts and Class players, respectively. The corresponding percentages for random positions are 19.0, 13.8 and 12.4. The main effects of Skill [$F(2,23) = 44.41$, $MS_e = 89.17$], of Type of positions [$F(1,23) = 309.20$, $MS_e = 75.92$] and the interaction term [$F(2,23) = 34.17$, $MS_e = 75.92$] are all significant at the 10^{-6} level. In particular, Masters recalled nearly three times as many pieces as Class players in game positions, but only 1 1/2 times as many in random positions. However, contrary to what was found in Chase and Simon (1973a), there was some tendency for the recall of random chess positions to vary with skill in our experiment, although the effect is not statistically significant [$F(2,23) = 2.27$, $MS_e = 34.10$, ns]. We show elsewhere (Gobet & Simon, in press-c) that this result (a small effect of skill on recall in random positions) is observed consistently in other studies.

The levels of recall for both game and random positions, and for all levels of skill, are similar to those that have been observed in the previous studies of these phenomena. The confounding of condition (game-random) with order of presentation did not have any discernible effects on recall levels when the findings of this study are compared with previous studies.

Comparison between Copy and Recall Tasks

In this section, we compare the results from the copy task and the recall task, comparing latencies and inter-piece relations as criteria for defining chunks. The theory predicts the same pattern of relations for pieces within chunks in both tasks. We first compare the latencies between consecutive pieces with the pattern of relations between these same pieces. We next show that the chunks could be defined by numbers of relations between pieces instead of by latencies, and estimate how well the numbers of relations predict the latencies. We then compare the actual pattern of relations in the data with a random pattern of relations.

Correlation between latencies and chess relations of successive pieces.

To demonstrate the psychological reality of the chunks defined by latencies, Chase and Simon (1973a) measured the meaningful relations between pairs of pieces that were placed on the board successively in copying or replacing positions. The chunking hypothesis predicts that there would be many more relations of attack, defense, proximity, shared color and shared type of piece between successive pieces within the same chunk than between successive pieces on opposite sides of a chunk boundary. The hypothesis was strongly supported by their data. We now check whether the findings are supported by the new experiments. We will use the interpiece latencies corrected for the time to move the mouse once a piece has been selected. (We obtained essentially the same results when we adjusted the latencies using Fitts' correction described earlier.)

If chunk boundaries are indicated by latencies > 2 s, then the relations between successive pieces should be different with short than with long latencies between them. In addition, if the same processes determine the latencies in both the copy and recall tasks, then there should be a high correlation between the relations in the two tasks. Specifically, the relations for within-glance placements in the copying experiments should correlate with those for rapid placements (≤ 2 s) in the recall experiments and the relations for between-glance placements in the former should correlate with those for slow placements (> 2 s), in the latter.

We use, as Chase and Simon (1973a) did, the following primitive relations : attack (A), defense (D), same color (C), same piece (S) and proximity (P). Pairs of successively placed pieces are assigned to exclusive categories according to the relations each pair shares. All pieces placed by the subjects are used in our analysis, whether or not they are placed correctly.

Table 1, columns 2-3 and 5-6 show, averaged over all subjects (there was little difference in latency statistics between skill levels), the median latencies between the placement of two successive pieces for each combination of relations, for the Copy task in Random and Game positions, within and between glances. We do not show the statistics for the recall task, as the separation of within-chunk from between-chunk placements in that task on the basis of latency would confound independent with dependent variables. We will later discuss how latencies relate to number of chess relations between successive pieces in the recall task with data for all latencies combined.

Insert Table 1 about here

Chase and Simon had found that small average latencies in the copy task correlate with a large number of relations between successive pieces, while large average latencies correlate with few relations. We find the same pattern in our data: all four correlations of latencies with numbers of relations are negative, although some are not statistically significant. For the within-glance copy placements in game positions, Spearman's rho correlation is -.77, and in the random positions the correlation is -.84. The shortest times are obtained with the PCS and DPCS relations, which mainly appear with pawn formations. In contrast, correlations for between-glance conditions in game and random positions are insignificant (-.26 and -.02). Finally, all but one of the latencies of the within-glance condition of the copying task are below 2 s, the exception being the case where there is only a relation of attack with game positions (this case occurs only in 0.5 % of the observations).

Table 1, columns 1 and 4 show, for both game and random positions in the copy task, that the numbers of within-glance sequences increase rapidly relative to the numbers of between-glance sequences as the numbers of relations between the successive pieces increase. For example, in game positions, there are 19.5 cases of DPCS relations for within-glance sequences for every between-glance sequence, while the ratio is only 1.5 for no relations. That is, sequences with four relations are 13 times as likely to be within-glance rather than between-glance as sequences with no relations. (The same pattern exhibits itself in the recall task, where the corresponding figures are 18.6 and 1.3 for pairs with latencies of less than 2 s and more than 2 s, respectively.)

Figure 3, which plots latency as a function of the number of relations in the recall task, for both short and long latencies combined (skill levels and types of positions are pooled), shows a clear negative correlation between number of relations and latencies, giving results similar to Figure 5 of Chase and Simon (1973a). That the slope is not as steep in our Figure as in theirs may be due to the fact that we have used medians while Chase and Simon used means of the latencies, and the fact that our sample of players is in general stronger than theirs: Figure 10 of Chase and Simon (1973b) shows clearly that, when plotting latency as a function of the number of relations, the slope is inversely proportional to the skill level.

Insert Figure 3 about here

Predicting latency from types of relations.

Which of the five relations largely account for the differences in latencies? For the within-chunk data (again pooled over tasks and types of positions) stepwise regression removes two of the relations (Defense and Attack) from the equation as insignificant. The multiple regression with the remaining relations yields the following equation:

$$\text{Latency} = 1.754 - 0.266 * \text{Same-Type} - 0.287 * \text{Color} - 0.180 * \text{Proximity}$$

The equation accounts for 63.2% ($p < .01$) of the variance. For the between-chunk data, the stepwise regression removes all relations but Same-Type as insignificant. The regression obtained with Same-Type as predictor is not statistically significant. These results indicate that the glue between successive pieces is weak for pieces belonging to different chunks. In summary, the relations of Same Type, Color and Proximity play a major role in predicting the latency when successive placements belong to the same chunk, but not when they belong to two different chunks.

The lack of importance of Attack and Defense relations might be thought somewhat surprising, but is more easily understood when we note, in the game positions, that while 48% of all pairs of pieces, selected at random, have the same color, 28% are the same kind of piece, 11% are in proximity, just 10% have a defense relation between them, and only 2.3% an attack relation. The importance of relations for sequence of placements is closely related to the frequency of their occurrence, although proximity has a larger role than its frequency would predict. (The results are almost the same when computed separately for each skill level.) This state of affairs does not necessarily imply, however, that chess chunks are shaped only by basic Gestalt organizational principles, for proximity, color and kind play an important role in the semantics of chess.

<u>Observed and expected probabilities of sets of relations.</u>

Table 6 of Chase and Simon (1973a) gives the probabilities of the presence of different combinations of relations in the various experimental conditions. We have computed the comparable data for our experiments, but as the two sets of data are very similar, we will summarize the comparison with Chase and Simon instead of reproducing our table in full. (We will be glad to provide the full table on request.) We also compare the observed probabilities with <u>a priori</u> probabilities (for game and for random positions) based on 100 positions and 26,801 pairs of pieces. For example, in 27 cases in game positions, two opposing pieces of the same kind attacked each other (and had no other relation), giving a probability of .001 for the AS relation; and in 8,978 cases a pair of pieces had none of the five chess relations, giving a probability of .335 for the null relation.

We show in our Table 2 the correlations of probabilities among the conditions in our experiment, which can be compared with the corresponding correlations shown in Chase and Simon (1973a), Table 7. Both sets of correlations suggest strongly that the short and long latencies in the recall task have the same meanings, respectively, as the within- and between-glance placements in the copy task. One can see five distinct clusters of correlations: (a) the short-latency probabilities in the recall task (variables 3 and 4) are very strongly correlated ($\underline{r}$ > .89) with the within-glance probabilities in the copy <u>game</u> task (variable 2) and with each other; (b) the between-glance probabilities in the copy task (5 and 6) are very strongly correlated ($\underline{r}$ > .78) with long-latency probabilities in the recall task (7 and 8); (c) the between-glance and long-latency probabilities (5-8) are very strongly correlated ($\underline{r}$ >.78) with the <u>a priori</u> (game and random) probabilities (9-10); (d) the within-glance <u>random</u> probabilities are correlated moderately (.5 < $\underline{r}$ <.75) with all the other conditions; and (e) all the correlations between "within chunk" variables (2)-(4) and other variables (5)-(10) are small to moderate (.15 < $\underline{r}$ <.54). All of these correlations show that within-chunk patterns of relations, whether in the copy or recall task, are quite different from between-chunk patterns, the latter resembling more closely the relations between pairs of pieces selected randomly. A closely similar structure is seen in Chase and Simon's Table 7.

Leaving aside the data for recall of random positions, which were not computed by Chase and Simon, and the a priori probabilities, the correlation between the remaining items in our Table 2 and the corresponding items in Chase and Simon's (1973a) Table 7 is .78, accounting for 61% of the variance in the correlation coefficients. There is good consistency between the two studies in the patterns of chess relations between pairs of successively placed pieces within and between chunks as defined.

Insert Table 2 about here

We also analyzed, for our data, the deviations of the number of observed chess relations from the a priori probabilities, subtracting the a priori probabilities from the observed relative frequencies of a given condition. In agreement with the theory, the within-chunk deviations from a priori probabilities are highly correlated with the number of relations, while this correlation is weaker for the between-chunk deviations. The correlations with number of relations for the within-chunk conditions are: copy game, within-glance: .81; copy random, within-glance: .68; recall game, short latencies: .86; recall random, short latencies: .79. The correlations with the between-chunk conditions are: copy game, between-glance: .61; copy random, between-glance: .56; recall game, long latencies: .69; recall random, long latencies: .31. These results are illustrated graphically in Figure 4, where we have pooled all within-chunk conditions and all between-chunk conditions. From the Figure, we see that, for within-chunk conditions, the placements having few relations are below chance, while the placements having several relations are well above chance. Between-chunk placements, with a flatter trend, are overall much closer to chance.

Insert Figure 4 about here

Relations by time interval.

How robust is the 2-s boundary in the recall task? One obvious test is to tabulate the number of relations for each interval of latencies, as is done in Table 3, with results pooled over types of positions and skill levels. There is a clear pattern: from the short intervals to the long intervals. Below 1.8 s, placements having 3 or 4 relations dominate over placements with 0 or 1 relation. Above 2.2 s, the pattern is shifted. In the interval 1.8-2.2 s, which includes the value of 2 s we have selected as a cut-off, the numbers with few and many relations are almost equal.

Insert Table 3 about here

Convergence of definitions of chunks by latencies or number of relations.

As the relations we have described show up prominently in chunking, we should be able to use them to define whether two successive pieces belong to the same chunk or not, with results similar to those when we use latencies to define chunks. For each pair of pieces, we have computed whether they belong or not to the same chunk in two ways: (a) by using the corrected latency, as before; and (b) by using the number of relations shared by the two pieces. In the former case, two successive pieces belong to the same chunk if the latency between them was less or equal to 2 s. In the latter case, two pieces belong to the same chunk if they had two or more relations. Table 4 presents the results, with all skill levels pooled. In all four conditions, the agreement between the two methods is high for the less-than-two-second cases, and a little less for the more-than-two-second cases. The percentage of placements classified consistently by the two methods is 72% for the task of copying game positions, 64% for copying random positions, 74% for recalling game positions, and 70% for recalling random positions. All four tables have chi-squares with probabilities below .0001.

Insert Table 4 about here

Thus, the two methods of defining chunks produce quite similar segmentations of the output, and the findings reported in Chase and Simon's paper and in this paper would hold about equally well if we defined a chunk as a set of consecutively placed pieces each of which has 2 or more chess relations with the piece previously placed. This provides strong convergent evidence that chunks have psychological reality as structures in LTM.

<u>Latencies as a function of cumulative placements.</u>

We mentioned earlier that Wixted and Rohrer (1994) have shown that latencies generally increase with the number of items previously recalled. Are chunks an artifact of these increasing latencies, early placements being classified as within-chunk and later placements as between-chunk? The chunking theory predicts that the inter-piece latencies will stay more or less constant when pairs of pieces belong to a chunk, but that the inter-piece latencies between chunks may increase as a function of the number of pieces previously placed. The former follows from the fact that chunks are postulated to be stored in LTM, and speed of recovery of their successive elements should be independent of the time when they are copied or replaced. The slowdown between chunks would follow if the players first replace salient chunks, then have to search a little longer to find the less salient, and therefore less easily recognized chunks.

Insert Figure 5 about here

We have computed the average inter-piece latencies for within and between-chunk placements, using as criterion for chunking that two or more relations are shared by two successive pieces. Figure 5 illustrates the results for the recall of game positions. Clearly, the evolution over time of the two variables follows different curves. The within-chunk latencies do show a modest increase of about 50% over 30 pieces, but the between-chunk latencies increase by a factor of 2 over the same interval. Wixted and Rohrer (1994) report inconclusive results on the latencies within clusters: in some studies, latencies increased with position, but did not in other studies.

In summary, the statistics of frequency of chess relations between successive pieces within the same chunk, as compared with successively placed pieces in different chunks, provide strong support for the chunking hypothesis. In spite of the difference in apparatus, the statistics derived from the current experiments agree closely with those reported by Chase and Simon (1973a). Finally, in two analyses extending Chase and Simon's, we showed that there was a considerable agreement in predicting whether a piece belongs to a chunk using either the number of relations or the latencies, and that the between-chunk latencies lengthen significantly over time, but the within-chunk latencies only slightly.

Size of the Largest Chunk and Number of Chunks

The template theory predicts that experts develop larger chunks with practice than are predicted by the original chunking theory. In the recall task,[5] Chase and Simon (1973a) did find a difference in chunk size between their subjects, their Master obtaining, for the first chunk replaced with middle game positions, a mean of 3.8 pieces, while the Class A player had a mean of 2.6 pieces and the beginner a mean of 2.4 pieces. The median largest chunk per position was 5 pieces for the Master with game positions. In the following analyses, using the 2 s cutoff to define a chunk, we discuss mainly the size of the largest chunk in a position, rather than the average size of chunks, because the template theory makes direct predictions about the size of the largest chunk, and also because skewness argues against using the arithmetic mean to measure chunk size.

Insert Figure 6 about here

Our data differ strikingly from Chase and Simon's in the sizes of chunks at all skill levels. Figure 6 shows the mean (for subjects at each skill level) of the median (over positions) of the largest chunk[6] as a function of the experimental condition. (These data were obtained by taking the median, for each subject in each experimental condition, of the largest chunk in each position. The means were taken across subjects at the same skill level). For game positions, the size of chunks varies with skill levels [$F(2,23) = 11.81$, $p<.001$], but size does not differ significantly between the copy and recall tasks [F(1,23) = 1.56, ns.] The interaction term is not statistically significant. For the Masters, the mean of the median largest chunks is 16.8 for the recall task, and 14 for the copying task. In a few cases, the entire recall consisted of a single chunk. In the Chase and Simon experiment, the largest chunk recalled by the Master was 7 pieces. Experts and Class A players also produce relatively large chunks in our experiment. Note also that the mean of the median largest chunks for random positions is constant across skill levels (6.3 pieces [sd = 2.1] in the copy task and 4.6 pieces [sd = 1.7] in the recall task, for all subjects pooled) and is well over what would be predicted by a theory postulating the visual encoding of individual pieces in STM.

For the number of chunks, Chase and Simon (1973a) found that their Master recalled more chunks in the recall task than the other players. This was one of the most troublesome of their findings, for the original model postulated that the difference in recall between players of different skills was to be explained by chunk size differences not by differences in the chunk capacity of short-term memory. By contrast, Figure 7 illustrates the number of chunks—defined as groups of at least two pieces placed with an interpiece latency of less than two seconds—found in our results, both for game and random positions and both for the copy and recall task. For the game positions, there is a main effect of Skill [$F(2,23) = 4.35$, $MS_e = 1.72$, $p < .05$], and of type of presentation [$F(1,23) = 7.11$, $MS_e = 1.58$, $p < .05$] as well as an interaction [$F(2,23) = 12.92$, $MS_e = 1.58$, $p < .001$]. The number of chunks replaced is inversely related to skill level for the copy task, while it shows an inverted U-curve for the recall task, with Class A players recalling the least chunks. Note that the difference between Masters and Class A players is small in the recall task: less than one chunk.

In summary, the current experiments fit better than the original Chase and Simon experiments did the hypothesis that size and not number of chunks accounts for the superiority of players of higher skill in recalling game positions (cf. Figures 6 and 7). The size of the largest chunks for Masters and Experts also supports the hypothesis that they frequently retrieve templates (large chunks with slots) that characterize the position as a whole.

Insert Figure 7 about here

For the random positions, we find again an inverted U-curve, both with the copy and the recall tasks. The main effects of skill [$\underline{F}(2,23) = 4.84$, $\underline{MS}_e = 1.50$, $\underline{p} < .05$] and type of presentation [$\underline{F}(1,23) = 240.1$, $\underline{MS}_e = 1.68$, $\underline{p} < .001$] are significant, while the interaction is only marginally significant [$\underline{F}(2,23) = 3.23$, $\underline{MS}_e = 1.68$, $\underline{p} < .06$]. Finally, for the two types of positions, Subjects produce fewer chunks in the recall task than in the copy task, which simply reflects the fact that multiple chunks do not have to be held in STM in the copy task.

For the recall task, few pieces are placed individually (on average 0.9 for random positions and 1.5 for game positions). Thus, even if we assume that none of these pieces are guessed, which is probably not the case, the total of chunks plus pieces placed individually (2.0 and 4.0 for random and game positions, respectively), agrees reasonably well with Zhang and Simon's (1985) estimate that the capacity of visual STM is about 3 chunks.

General Discussion

Five main reasons led us to test further the validity of the Chase and Simon method for identifying chunks. First, empirical data reviewed in the introduction suggested that Masters perceive a position at a higher level of organization than the chunks described by Chase and Simon. Second, the template theory, a refinement of the chunking theory, predicts maximum chunk sizes larger than those predicted by the original theory. Third, we wanted to see whether the number of relations could predict chunks that are consistent with the chunks predicted by the inter-response latencies, not only at an aggregated level, as in Chase and Simon's (1973a) study, but also at a detailed level. Fourth, we wanted to address the concern that chunking might be an artifact of the total time spent to replace a position, as might be concluded from the data presented by Wixted and Rohrer (1994). Fifth, we wanted to check whether a different apparatus (a computer display as contrasted with actual chess pieces and board), could account for differences with Chase and Simon's study in the size and relational richness of chunks, where the handling of pieces by subjects and size of grasp may have affected the way subjects chunked the position.

The first part of the results section was devoted to the copying task. We started by noting a difference between the strategic behavior of our Masters and Chase and Simon's Master: the former tended to spend more time than the latter in studying the position before reconstructing parts of it. Despite this difference in strategy, we were able to replicate the main results of Chase and Simon's paper. First, the distributions of latencies between successive pieces are different for within- and between-glance placements. Second, the latency distributions, corrected for the time required to move the mouse, are close to those of Chase and Simon (1973a), despite the differences in the experimental apparatus and in Masters' strategies. Specifically, more than three quarters of the within-glance latencies (corrected for the time needed to move the mouse) are below 2 s. Third, strong players tend make faster between-glance placements in game positions than amateurs, but not within-glance placements.

Although the 2-s boundary is only an approximation, it seems to be a reasonable approximation, for in our replication of the copy task, 79.5 % of the latencies for placements within a glance (as against only 1.11% of the latencies for placements between glances) were less than 2 s. This provides additional reason for thinking that the definition of chunks employed by Chase and Simon and used here is not arbitrary but reflects subjects' perceptions of the board.

In the second part of the results section, we analyzed the recall performance of our subjects. In agreement with other studies, stronger players were clearly superior with game positions. There was some indication that they were slightly superior also with random positions, a result that did not appear in Chase and Simon's data.

The third set of analyses compared chunking between copy and recall tasks. Analyzing the relation of latencies to the numbers and probabilities of chess relations present, we found, as Chase and Simon did, three main phenomena. First, latencies are shorter when the number of relations within a chunk is larger. Second, numbers of chess relations in within-glance placements in the copy task and in placements within 2 s in the recall task were strongly correlated, as were numbers of relations in between-glance placements in the copy task and placements over 2 s in the recall task. Third, the size of chunks increased with skill, accounting for skilled players' superiority in the recall task. These results, based on a larger sample (26 subjects) than that of the earlier experiment, support the major findings of Chase and Simon (1973a) and corroborate their hypothesis that the same information processing mechanisms, operating on chunks stored in long-term memory, determine the time intervals in both the copy and recall tasks.

Another result that strengthens the concept of chunk was provided by the behavior of interpiece latencies over successive placements. Wixted and Rohrer (1994) showed that latencies generally become longer for successive items recalled, which may suggest that chunks are defined artifactually: successive pieces placed early in recall would be classified as within-chunk, those placed later in recall, with longer latencies, would be classified as between-chunk. Our data showed that, when we plot the latencies of pieces that have 2 or more chess relations with their predecessors, we find a small trend in size over time; when we plot the latencies of pieces that have 0-1 chess relations with their predecessors, we find a strong positive trend, with the latter slope more than twice the former. This result is consistent with the numerous findings, reported by Wixted and Rohrer, of clustering of semantically related items in recall from semantic memory, where clusters were also defined by inter-item latencies, and where the within-cluster times do not increase, or increase only slightly, over successive clusters.

Except for experiments on semantic clustering by Tulving (1962) and others, all of the experiments cited by Wixted and Rohrer involve retrieving already familiar items (chunks) from memory; hence the correct interpretation of their finding is that when successive whole <u>chunks</u> are retrieved, the latencies between chunks will grow with time. This does not imply that there will be any systematic increase over the period of recall in within-chunk latencies for successive chunks, nor do our data show substantial differences of this kind. Hence, the Wixted and Rohrer (1994) summary of findings in the literature on free recall is wholly consistent with the findings on chunking in the chess literature.

In general, there was little difference among skill levels in the nature of the chunks: all show the same pattern of relations, which led us to pool the skill levels in the presentation of results. One reviewer has suggested that this lack of difference implies that chess information is represented in the same way by players of different skills, differing in quantity, rather than in qualitative organization. If so, chess expertise would differ from physics expertise, where it has been shown that the representation of information differs qualitatively with skill. However, as the largest chunks of highly skilled players are much more complex than those of weaker players, important qualitative differences may be present in these large templates.

Maximum chunk size was substantially larger in the current experiment than in the experiments of Chase and Simon. As the number of chunks in the recall task differ little between the strongest skill level and the weakest, this supports the hypothesis that strong players use templates, supplemented by smaller chunks, to encode information rapidly. The superiority of strong players for game positions is then due not to the number of chunks (Masters replaced only one chunk more than Class A players, in the recall task), but to the presence of a few large chunks. Although we did find, as Chase and Simon did, some differences in the number of chunks between skill levels, the ordering of numbers by skill was wholly different in the two studies. This difference in chunk size may have been produced by the difference in experimental procedure (grasping pieces by hand in the earlier study) that split large chunks into small ones.

In all other respects, our distributions of chess relations are much the same as those found by Chase and Simon. The large size of chunks we have found adds support to the template theory, inasmuch as the difference in the number of chunks by skill was small and as this number was within the supposed capacity of visual STM (Zhang & Simon, 1985).

Are the chunks of our subjects, especially the Masters, real or are they the product of some artifactual feature of our material? A first possible confound is that the positions we used might come from very typical opening variations that are likely to have been overlearned by skilled players. While this is somewhat possible for the positions of the copy task, the positions we have used in the recall task do not, as far we can judge, belong to typical openings situations (see Appendix A), and it is unlikely that the typicality of positions has inflated our estimate of the size of chunks.

A second potential confound is the effect of "serialization" (only one piece can be replaced at a time) on the structure of chunks. Whereas Chase and Simon's apparatus forced subjects almost physically to chunk pieces, it could be argued that our experimental procedure allows subjects to search in memory for a new piece/chunk while still busy replacing the previous piece. Such a time-sharing strategy should however level the inter-piece latencies and destroy chunks. (It cannot be assumed that Masters have superior motor skills in placing pieces, for experience with the use of computer and mouse was evenly distributed among the subjects in our experiments—skill level accounts for less than 0.3 % of the variance when used to predict the time to move the mouse.)

A reviewer has proposed that the serialization could "artificially concatenate small chunks into large ones." It is unclear why this should be the case more with Masters than with Class A players. The latters' larger chunks (at most 6-7 pieces) are not unusually large and can easily be explained, for example, by a few common patterns, like castling positions. Finally, because our subjects were somewhat slower than Chase and Simon's even after the correction for mouse movements, it cannot be argued that the 2 s boundary favored the former in comparison with the latter. Altogether, our results concerning the size of chunks seems to survive critical analysis, and we may state confidently that Chase and Simon (1973a) underestimated the size of chess chunks.

Why then did Chase and Simon's subjects, and particularly their Master, not find such large chunks? When a physical chessboard and pieces are used, the maximum chunk size is limited by the number of pieces the hand can grasp; with our present apparatus, it is not limited in this way. When this limit is absent, large chunks may be recalled that represent core information of familiar or typical positions. In both cases, the player has spent enough time studying this position or type of position to have acquired large, well differentiated templates for it. (Chase & Simon, 1973a, also conjectured that such chunks might be present, but their data did not provide much evidence for them.)

Another possible factor in the differences in chunk size at Master level between the Chase and Simon experiments and ours is that their Master was somewhat out of practice (a score of only 71.5% for recall of middle game positions, as compared with 92% for our Masters), and, being in his midforties, may have been slower in replacing pieces, causing some chunks to be divided in scoring them.

The size of chunks in the random positions also calls for brief comment. Most of them seem to be built up either from dynamic chess relations (in particular, pieces close to or attacking a king) or geometric patterns (pieces and pawns forming a square or located on the same diagonal). In these cases, subjects may hold in STM descriptions of the pattern (e.g.: [Slot #1: black pawns], [Slot #2: on the same diagonal], [Slot #3: starting from the square a1], [Slot #4: number of pawns is 4]) rather than simple chunks enumerating the pieces (see Gobet and Simon, in press-b, for more data on recall of random positions).

Conclusion

In the introduction of this paper, we presented four major criticisms of the chunking theory of expert memory in chess. The results of a new experiment, together with evidence already in the literature, clearly establish that an augmented chunking theory—the template theory described here—meets all these criticisms. Summarized in one sentence, the message of this paper is that chunks are larger than estimated by Chase and Simon, but that, as they showed, the pattern of relations between two pieces placed successively are radically different when the pieces do or do not belong to the same chunk. Our explanation is that the large chunks are built around templates that encode information, acquired by strong players over years of practice, about typical and familiar positions, and that provide rapidly fillable slots for additional chunks of information about the current position. There is considerable evidence for the use of such templates in expert memory performances of other tasks.

The chunking hypothesis has sometimes been misinterpreted as a claim that recognition of familiar patterns and retrieval of moves associated with them is almost the sole basis of expertise in chess (e.g., Holding, 1985). The correct claim, and the one actually made by Chase and Simon and in this paper, is that skill in playing chess depends both on (a) recognizing familiar chunks in chess positions while playing games, and (b) exploring possible moves and evaluating their consequences. Hence, expertise depends both on the availability in memory of information about a large number of frequently recurring patterns of pieces, and upon the availability of strategies for highly selective search of the move tree. Expert memory, in turn includes slowly acquired structures in long-term memory (retrieval structures, templates) that augment short-term memory with slots (variable places) that can be filled rapidly with information about the current position.

References

Bower, G. H., & Springston, F. (1970). Pauses as recoding points in letter series. *Journal of Experimental Psychology, 83,* 421-30.

Card, S., Moran, T. P., & Newell, A. (1983) *The psychology of human-computer interaction.* Hillsdale, NJ: Erlbaum.

Charness, N. (1974). *Memory for chess positions: The effects of interference and input modality*. Unpublished doctoral dissertation, Carnegie Mellon University.

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

Charness, N. (1988). Expertise in chess, music and physics : A cognitive perspective. In Obler, L. K. & Fein D. A. (Eds.) *The exceptional brain : The neuropsychology of talent and special abilities.* New York, Guilford Press.

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.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?*, Hillsdale, NJ: Erlbaum.

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.

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

De Groot, A. D. (1978). *Thought and choice in chess*. The Hague: Mouton Publishers.

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

Euwe, M. (1972). *Jugement et plan*. [Judgment and plan]. Paris: Payot.

Frey, P. W., & Adesman, P. (1976). Recall memory for visually presented chess positions. *Memory & Cognition*, *4*, 541-547.

Freyhoff, H., Gruber, H., & Ziegler, A. (1992). Expertise and hierarchical knowledge representation in chess. *Psychological Research, 54,* 32-37.

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

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

Gobet, F., & Simon, H. A. (in press-b). Chunks in chess memory: Recall of random and distorted Positions. *Memory & Cognition.*

Gobet, F., & Simon, H. A. (in press-c). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review.*

Gold, A. & Opwis, K. (1992). Methoden zur empirischen Analyse von Chunks beim Reproduzieren von Schachstellungen. *Sprache & Cognition, 11*, 1-13.

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

Gruber, H., & Ziegler, A. (1990). Expertisegrad und Wissensbasis. Eine Untersuchung bei Schachspielern. *Psychologische Beiträge, 32*, 163-185.

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

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

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

Kmoch, H. (1980). *La force des pions aux échecs.* [Pawn power in chess]. Paris: Payot.

Lories, G. (1987). Recall of random and non random chess positions in strong and weak chess players. *Psychologica Belgica, 27,* 153-159.

Philidor, F. A. D. (1749). *L'analyse du Jeu des Echecs*. [The analysis of the game of chess]. London.

Reitman, J. S. (1976). Skilled perception in GO: deducing memory structures from inter-response times. *Cognitive Psychology, 8*, 336-356.

Richman, H. B., Staszewski, J., & Simon, H. A. (1995). Simulation of expert memory with EPAM IV. *Psychological Review, 102*, 305-330.

Saariluoma, P. (1994). Location coding in chess. *The Quarterly Journal of Experimental Psychology, 47A,* 607-630.

Simon, H. A. (1976). The information storage system called “Human memory”. In M. R. Rosenzweig & E. L. Bennett (Eds.), *Neural mechanisms of learning and memory*. Cambridge: MA: MIT Press.

Simon, H. A., & Gilmartin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, 5, 29-46.

Tulving, E. (1962). Subjective organization in free recall of “unrelated” words. *Psychological Review, 69*, 344-54.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition, 13,* 193-201.

Welford, A. T. (1968). *Fundamentals of skill.* London: Methuen.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical evidence. *Psychonomic Bulletin & Review, 1*, 89-106.

Appendix A: Positions used in the copy and recall tasks

The positions are given in Forsyth notation. In this notation, the board is scanned from rank by rank from a8 to h8, then from a7 to h7, and so on to h1. Letters indicate pieces (uppercase for White, lowercase for Black). Digits refer to the number of empty squares between pieces or the board's boundaries. Ranks are separated by slashes.

<u>Copy task</u>

<u>Game positions.</u>

2kr3r/pp3ppp/2p1p3/q4n2/3P4/3BQ2P/PPP2PP1/1K1R3R/

r4rk1/4b2p/pnqpbpp1/np2p3/4P3/2P1N2P/PPB2PPN/R1BQR1K1/

rq1r2k1/pb3pp1/1p1bpn1p/8/3P4/1B1Q1N2/PP1B1PPP/2R1R1K1/

<u>Random positions.</u>

1Q2r3/2RP2N1/2p5/3q2R1/p1n2pPk/pP3PP1/1Kp4b/pBPr1p2/

PPR3P1/3P3b/1nr1pKBq/pPP5/PnpRpp2/6N1/N2k1Q2/p2r1p2/

<u>Recall task</u>

<u>Random positions.</u>

1pP2RKb/3P2pP/RQ2k1nP/p1q4P/4N1p1/3r4/1p5B/1B4rb/

1B1PQp2/1pb2P2/qp3bNp/p1PnP1PK/1k2p2B/1Pp5/r7/3R3P/

NP1pp3/1P3Br1/P7/3bR2P/4p1p1/r4p2/1b5K/1Pk1R3/

<u>Game positions.</u>

The positions used in the recall task were randomly sampled, for each subject, from either one of the following pools of positions:

(a)

r2r1nk1/pp2qpbp/5np1/3p1b2/1P1Pp3/PQN1P3/1B2BPPP/R1R2NK1/

r3rbk1/1bq2pp1/p1n4p/1ppnpN2/4Q3/2PP1N2/PPBB1PPP/3RR1K1/

r3r3/2nq1pbk/p2p1npp/1p1P4/PPp1P3/2N1BP2/3QB1PP/1R3RK1/

2knr2r/p1p1nb2/1p1q1p2/3p2p1/3P3p/1NPB1N1P/PPQ2PP1/R3R1K1/

2n1r1k1/ppq3pp/2b1prn1/P2p4/2pP1P2/B1P2NP1/2PQ2BP/RR4K1/

r3kr2/1bq1bp1p/4p3/p2pNp1Q/1p1P1P2/8/PPP1N1PP/1K1R3R/

2b1r1k1/1p3pb1/1qpp1np1/2n4p/2PNPP2/1PN1R1PP/3Q2BK/B7/

r2r2k1/p1q1bppp/1p2p3/3pP3/2pP4/2P2N1P/PP2QPP1/3RR1K1/

2rr2k1/1b2qppp/2p1pn2/ppN5/3P4/3BP3/PPQ2PPP/1KR4R/

1rb2rk1/4qppp/p7/3pPp2/np3Q2/1N1B2P1/PPP4P/2KR3R/

3rr1k1/pp3ppp/1qp1p3/8/PPQPn3/3NP2P/5PP1/2RR2K1/

2r4k/ppr3pp/2b1pn2/3p1p2/3P1P2/P1N1P3/1P3PBP/2R2RK1/

2rbr2k/p4ppp/3p4/1pqNpP2/4P3/1PP2R2/P5PP/R3Q2K/

r4qk1/2p2r1p/1p1p1pp1/nP1N4/4PP2/2Q3P1/2P3KP/3RR3/

r1b1k3/p4p2/2p5/1p2p1q1/4PpPr/1BP2Q1P/P1P2PK1/1R5R/

3q1rk1/pb3p2/1p2p1p1/6N1/2rP4/2P5/P2Q1PP1/R3R1K1/

(b)

2rqk2r/pp1b2pp/3npn2/3pN1B1/3P4/2PB4/P3QPPP/1R3RK1/

2k3r1/pp1bq3/1np1p3/3p2N1/3b1P2/2NB4/PPP1Q3/1K1R4/

5rk1/r1q1b1p1/pnbp1pPp/2p1pP1Q/Pp2P3/1P1PN3/2PB2BP/R4R1K/

r4r2/1p1q1pbk/2bnp1p1/p2p2Pp/3P1B1P/1P3BN1/P1PQ1P2/2KRR3/

1k5r/pp2qpp1/r6p/3pN3/2pPnP2/2P1P3/P3R1PP/K1QR4/

3nrrk1/pB1b2bp/1p2p1p1/5p2/3P3B/1N2P3/P4PPP/1RR3K1/

2rr2k1/1b2qppp/pb2p3/1p2P3/1P3Pn1/P1NB4/1B2Q1PP/R4R1K/

r2q1rk1/pp6/3p2bb/3Pp2p/6pP/2N3P1/PP2QRPN/5R1K/

2b1nrk1/5r1q/p2p1pp1/1p1Np3/4P3/P2P1R2/1P1B1QPP/5R1K/

r1bk3r/pp3ppp/1n2p3/2b5/2P5/1P4P1/P3NPBP/R1B2RK1/

2b2rk1/p3q1np/2p1p1p1/2Pp1pP1/3P1P2/2NBP3/P1Q4P/1R5K/

r2qk2r/1p1bbp2/1P2p3/p2pPp2/n2N1N1p/3PB3/5QPP/R4RK1/

r3r3/pp2q1kp/2b2pp1/7n/2BQp3/7P/PP2RPPB/R5K1/

2krr3/1ppn2p1/p1n4p/3qN3/3P1B2/2P4P/P1Q3P1/4RR1K/

r3r1k1/pp3ppp/2pR1n2/2n1p3/2P5/2N2BP1/PP2PP1P/3R2K1/

2rq1rk1/4np1p/p5p1/1p1pNn2/3P1P2/8/PP1QN1PP/2R2RK1/

r1b1n2r/1p2q3/1Qp1npk1/4p1p1/P1B1P3/2P1BNP1/1P3P2/R3R1K1/

1q5k/5rbp/1n4p1/QNpb4/1p6/4nN2/PP2BRPP/4R1K1/

2r2rk1/1q2b1pp/p1b2p2/1p1pp3/2n1PP2/2P1B1P1/PPQ3BP/R2N1R1K/

r3r1k1/p2q1ppp/np3n2/3p4/P1pP4/B1P1P3/2Q1NPPP/R4RK1/

1r5r/p2q1k1p/4pppb/1pRp4/3P4/1Q2PN2/PP3PPP/4R1K1/

r2q1rk1/4bp1p/bn1p2p1/p1pP4/Pp3B2/1P1B2P1/2Q1PP1P/R2N1RK1/

2r3k1/pb2qppp/1p1r4/8/4p3/PN1nP3/1P2BPPP/1Q1R1RK1/

r3r1k1/1n3ppp/p2p1q2/1ppP2b1/4P3/1P3Q1P/P1BB1PP1/3RR1K1/

r4rk1/pp3p2/1qnN3p/6pn/2P1p3/P3P3/1PQ1K1PP/3RB2R/

r4rn1/1p3pkp/p2q2p1/Qb1pN3/3P4/5N2/PP3PPP/2R1R1K1/

Table 1

Average Latencies (in Seconds) for the Copy Experiments, for Combinations of the Five Chess Relations: Attack (A), Defense (D), Spatial Proximity (P), Same Color (C), and Same Piece (S).

	GAMES			**RANDOM**		
	(1)	(2)	(3)	(4)	(5)	(6)
Relations	ratio[a]	WITHIN	BETWEEN	ratio[a]	WITHIN	BETWEEN
–	1.533	1.783	8.717	1.909	1.817	7.058
A	5.940	3.617	6.050	1.644	1.592	5.167
P	0.000	.	4.783	1.796	1.383	7.250
C	3.795	1.608	7.142	2.144	1.617	7.283
S	2.037	1.767	10.675	2.338	1.667	10.217
AP	.	.	.	2.324	1.442	6.867
AS	.	.	.	1.718	1.333	9.292
DC	5.574	1.517	6.683	2.678	1.483	6.417
PC	8.936	1.575	5.942	3.673	1.233	6.275

PS	2.376	1.508	5.733	1.969	1.567	5.433
CS	8.720	1.200	6.733	3.682	1.250	8.017
APS	.	.	.	2.864	1.317	9.783
DPC	9.961	1.567	6.583	5.504	1.433	6.733
DCS	20.125	1.317	5.633	.	.	.
PCS	15.822	1.200	11.250	9.326	1.133	8.450
DPCS	19.530	1.192	7.500	2.864	1.133	10.483

Note. [a] The ratio of the number of within-glance latencies to number of between-glance latencies.

Table 2

Intercorrelation Matrix for the Copy (in Bold), Recall and *A Priori* (in Italics) Chess Relation Probabilities

	1	2	3	4	5	6	7	8	9	10
1. **Within-glance (random)**	1.000	0.701	0.648	0.550	0.752	0.696	0.531	0.656	0.495	0.525
2. **Within-glance (games)**	.	1.000	0.890	0.924	0.308	0.531	0.171	0.470	0.159	0.162
3. ≤ 2 seconds (random)	.	.	1.000	0.907	0.236	0.392	0.184	0.447	0.179	0.173
4. ≤ 2 seconds (games)	.	.	.	1.000	0.205	0.411	0.102	0.413	0.170	0.166
5. **Between-glance (random)**	.	.	.	.	1.000	0.867	0.778	0.829	0.775	0.837
6. **Between-glance (games)**	.	.	.	.	.	1.000	0.795	0.954	0.813	0.834
7. >2 seconds (random)	.	.	.	.	.	.	1.000	0.862	0.907	0.873
8. >2 seconds (games)	.	.	.	.	.	.	.	1.000	0.901	0.903
9. *A priori (games)*	.	.	.	.	.	.	.	.	1.000	0.984
10. *A priori (random)*	.	.	.	.	.	.	.	.	.	1.000

Table 3

Probabilities as a Function of Time Interval for Numbers of Chess Relations

	Time interval (in seconds)										
# of relations	0.2-0.6	0.6-1.0	1.0-1.4	1.4-1.8	1.8-2.2	2.2-2.6	2.6-3.0	3.0-3.4	3.4-3.8	3.8-4.2	4.2-4.6
0	0.015	0.024	0.039	0.068	0.086	0.088	0.141	0.163	0.055	0.147	0.205
1	0.031	0.131	0.170	0.230	0.273	0.339	0.341	0.373	0.495	0.386	0.346
2	0.169	0.243	0.287	0.275	0.308	0.293	0.219	0.310	0.341	0.334	0.321
3	0.447	0.380	0.357	0.343	0.274	0.238	0.229	0.148	0.099	0.093	0.116
4	0.338	0.224	0.147	0.082	0.059	0.044	0.071	0.008	0.011	0.040	0.013
Expected # of relations	3.062	2.653	2.403	2.137	1.947	1.815	1.750	1.469	1.518	1.493	1.388

Table 4

Defining a Chunk (a) Using Corrected Latencies Versus (b) Using the Number of Relations Shared by the Pieces

Copy task

Game positions

	# relations	
Latency	• 2	< 2
• 2 sec	861	167
> 2 sec	245	190

Random positions

	# relations	
Latency	• 2	< 2
• 2 sec	640	248
> 2 sec	286	329

Recall task

Game positions

	# relations	
Latency	• 2	< 2
• 2 sec	1206	357
> 2 sec	143	189

Random positions

	# relations	
Latency	• 2	< 2
• 2 sec	196	76
> 2 sec	28	46

Note. Numbers of pairs of successive pieces with: [short (• 2 sec) and long (> 2 sec) latencies] x [•2 and < 2 relations].

Figure captions

Figure 1: Median latency between the placement of successive pieces, as a function of skill level, type of position, and type of placement (within-glance [WGP] or between-glance [BGP] placements).

Figure 2: Frequency histogram of within-glance latencies and between-glance latencies (corrected latencies for game positions).

Figure 3: Relation between interpiece latencies and the number of relations shared by two pieces successively placed (data from the recall task pooled over type of position and skill levels).

Figure 4: Relation between chess relation probabilities and the number of relations shared by two pieces successively placed (data pooled over tasks and skill levels).

Figure 5: Interpiece latency as a function of the number of pieces previously replaced, for between-chunk and within-chunk placements.

Figure 6: Mean of median largest chunk as a function of skill level, mode of replacement and type of position.

Figure 7: Number of chunks as a function of skill level, mode of replacement and type of position.

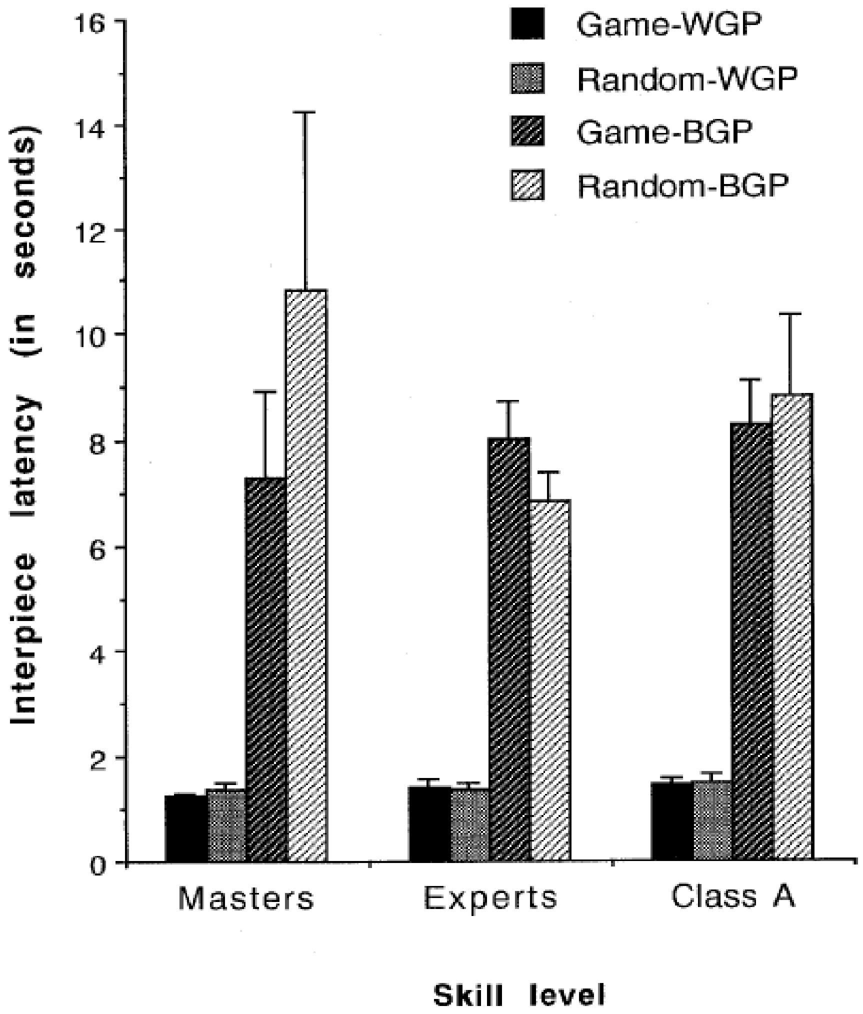

Game-WGP
Random-WGP
Game-BGP
Random-BGP
Interpiece latency (in seconds)
0
2
4
6
8
10
12
14
16
Masters
Experts
Class A
Skill level

Figure 1

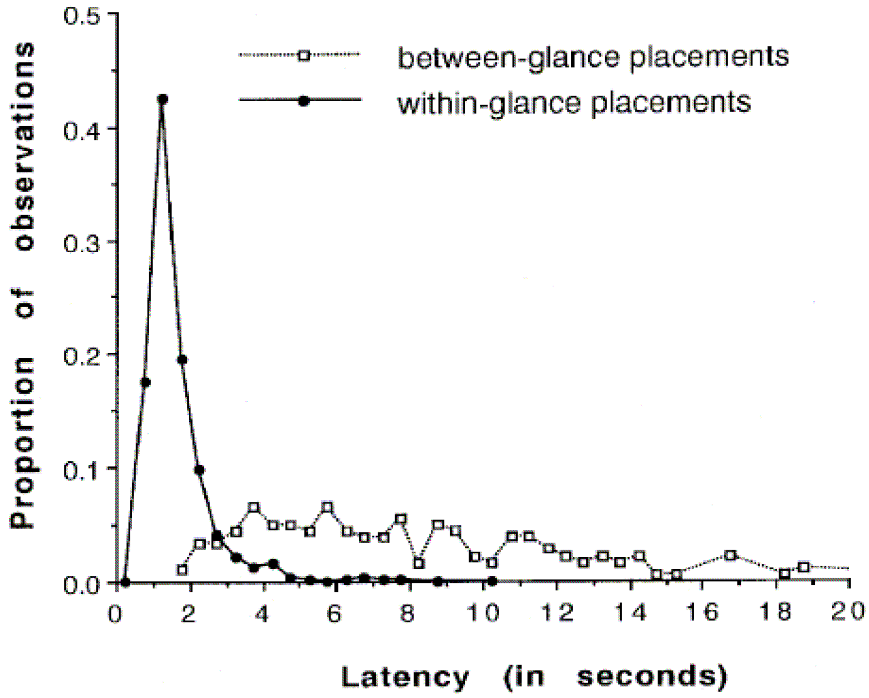

between-glance placements
within-glance placements
Proportion of observations
Latency (in seconds)
0.5
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
12
14
16
18
20

Figure 2

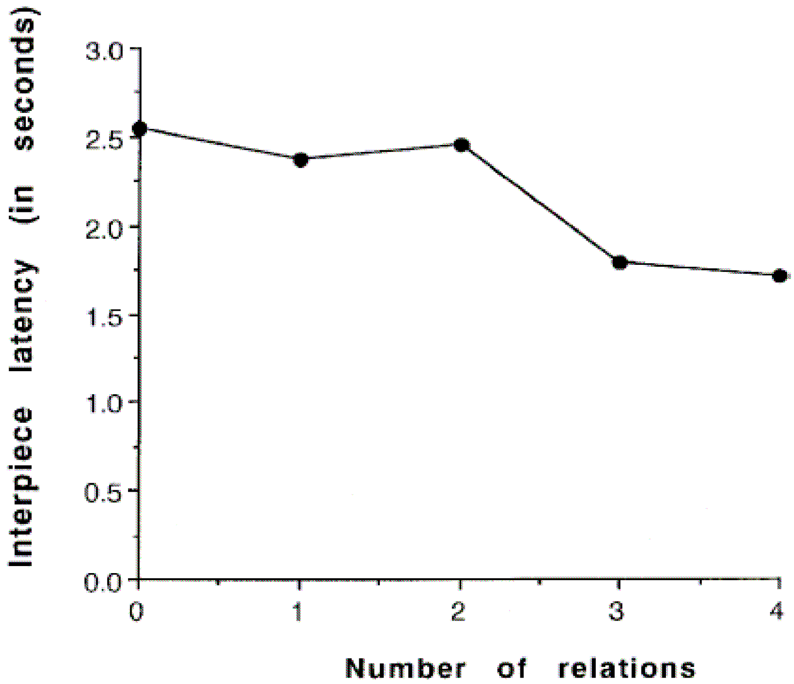

Interpiece latency (in seconds)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
Number of relations

Figure 3

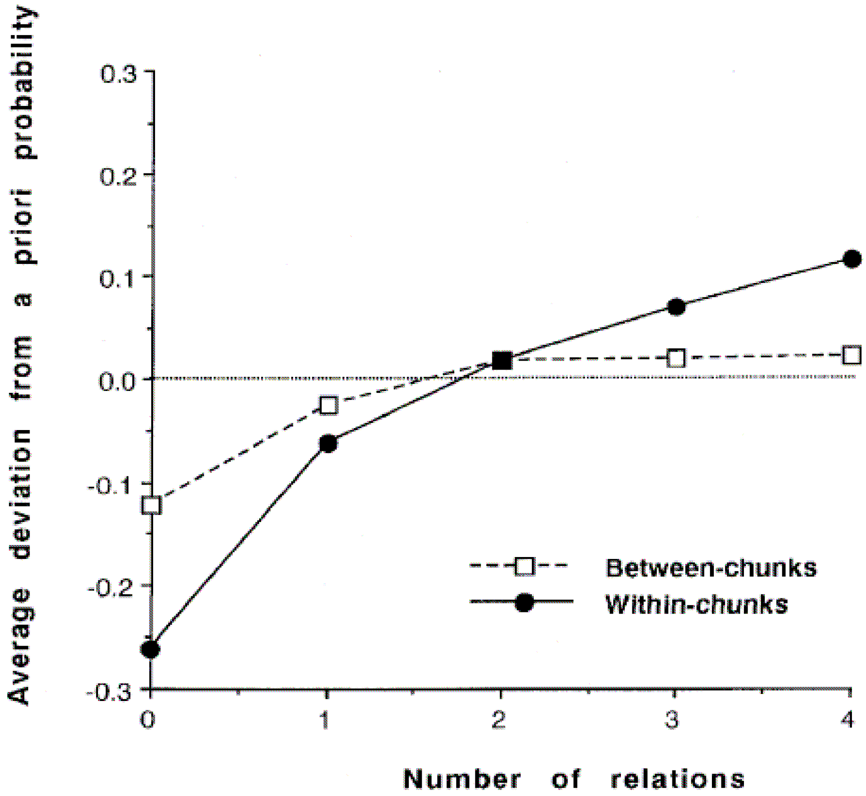

Average deviation from a priori probability
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
0
1
2
3
4
Number of relations
Between-chunks
Within-chunks

Figure 4

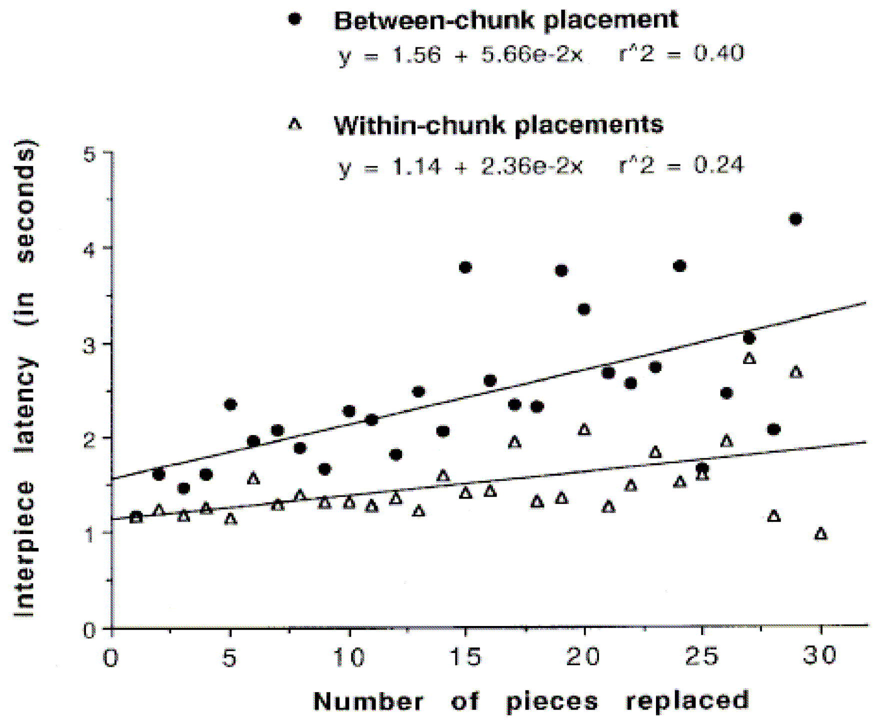

Between-chunk placement
y = 1.56 + 5.66e-2x r^2 = 0.40
Within-chunk placements
y = 1.14 + 2.36e-2x r^2 = 0.24
Interpiece latency (in seconds)
Number of pieces replaced
0
1
2
3
4
5
0
5
10
15
20
25
30

Figure 5

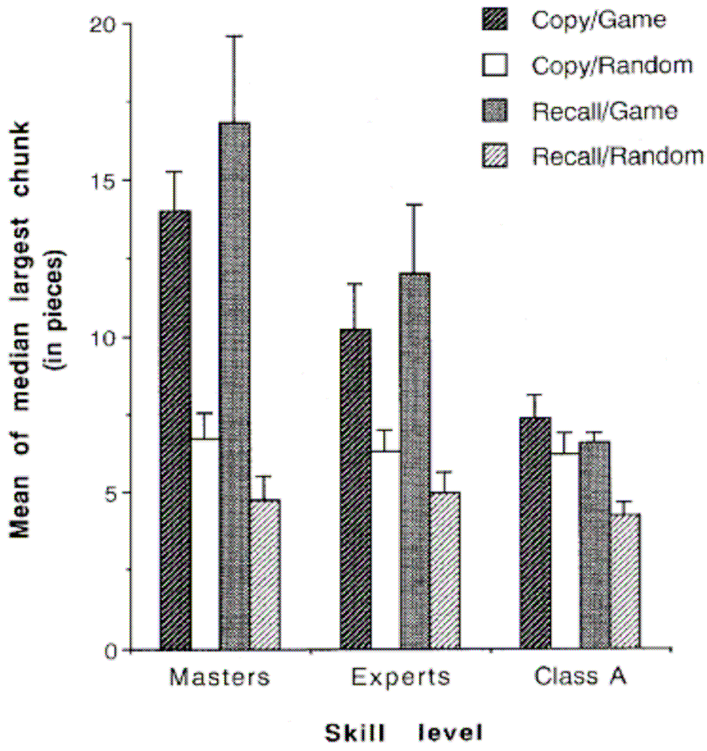
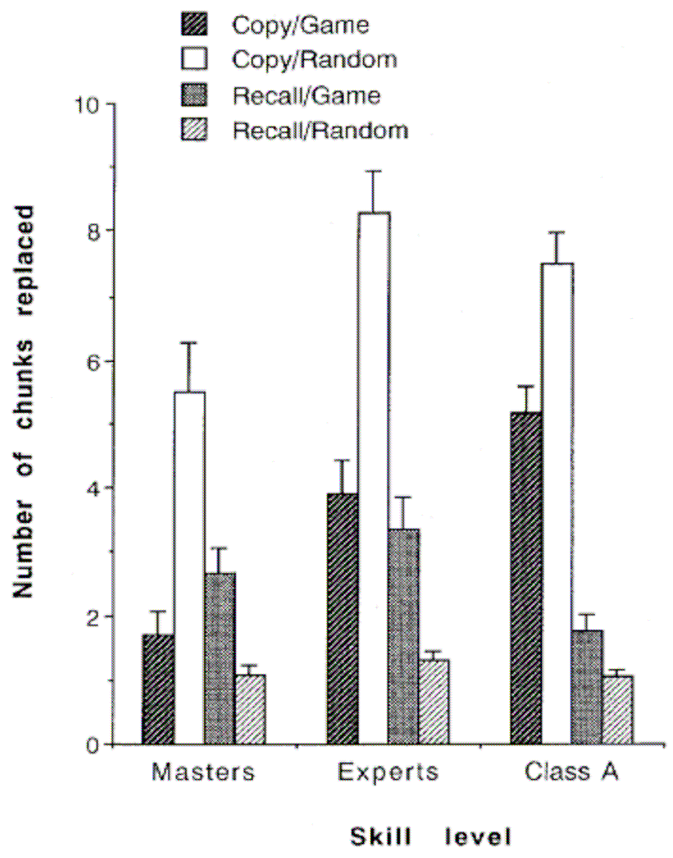

Copy/Game
Copy/Random
Recall/Game
Recall/Random
Mean of median largest chunk (in pieces)
20
15
10
5
0
Masters
Experts
Class A
Skill level

Figure 6



Copy/Game
Copy/Random
Recall/Game
Recall/Random
Number of chunks replaced
10
8
6
4
2
0
Masters
Experts
Class A
Skill level

Figure 7

Authors note

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. Fernand Gobet is now at the Department of Psychology, University of Nottingham (England). 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 Howard Richman, Jim Staszewski and Shmuel Ur for valuable comments on parts of this research. They are grateful to Neil Charness, Pertti Saariluoma, and three anonymous reviewers for extensive and especially helpful comments on earlier drafts.

Footnotes

[1]Competition chess players are ranked by the ELO rating (an interval scale). Its standard deviation (200 points) is often interpreted as delimiting skill classes. Grandmasters are normally rated above 2500, International Masters above 2400, Masters between 2200 and 2400, Experts between 2000 and 2200, Class A players between 1800 and 2000, and Class B players between 1600 and 1800.

[2]Chess players differentiate between *Pawns*, the weakest men and the remaining *Pieces* (King, Queen, Rook, Bishop and Knight).

[3]A problem with this study is that the positions are not completely random. First, some (semantic) constraints were applied in generating the positions used by Holding and Reynolds (1992). Second, a statistical analysis shows that equiprobalitity of White and Black pieces' distribution on the board may be rejected at $p<.001$ (Gobet, 1993). Therefore, the findings of this study are hard to interpret.

[4]Such refusals have been reported by Lories (1987) and Gobet and Simon (1995).

[5]Chase and Simon give no data on the size of chunks for the copying task.

[6]In this entire section, chunks include both correctly placed and incorrectly placed pieces. From a psychological standpoint, incorrect pieces have the same meaning as correct pieces, as the subject may have drawn on erroneous information in memory.