

# Expert Chess Memory: Revisiting the Chunking Hypothesis

Fernand Gobet and Herbert A. Simon

*Carnegie Mellon University, USA*

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 two-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.

## INTRODUCTION

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 recognising 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

---

Requests for reprints should be sent to Dr. Fernand Gobet, Department of Psychology, University of Nottingham, Nottingham NG7 2RD, UK.

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.

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, Nancy Cooke, Robert Crowder, Pertti Saariluoma, and three anonymous reviewers for extensive and especially helpful comments on earlier drafts.

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.<sup>1</sup> 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 experiment and its interpretation.

After summarising 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 two-second 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?

A chunk is an 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 two seconds.

---

<sup>1</sup> 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.

According to the theory, pairs of pieces that have numerous relations are more likely to be noticed together, hence chunked. Chase and Simon then analysed the chess relations (attack, defence, proximity, same colour, and same type) between successively placed pieces in the two tasks, thereby demonstrating that the probabilities of these relations between successive pieces belonging to a chunk (less than two seconds' interval) are much greater than the probabilities between successive pieces not belonging to a chunk (an interval of more than two 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 recognise 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 & Gilmartin, 1973), simulated several experimental findings, including the percentage of pieces recalled by a Class A player, 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, 50,000 is often mentioned). Finally, Chase and Simon's theory of memory implies that chunks, upon recognition, would suggest good moves to the Masters, which could then be further evaluated by limited look-ahead search.

## Experimental Evidence for the Chunking Hypothesis

Chunk structures have been identified experimentally in paradigms other than that employed by Chase and Simon (1973a,b). It has been shown that presenting pieces (visually or verbally) as chunks allows better recall of a position than presenting pieces by columns or in random order (Charness, 1974; Frey & Adelman 1976). Chi (1978), using a partitioning technique, found that chunks sometimes overlapped and 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). Freyhoff, Gruber & Ziegler (1992) using a similar partitioning and sub-partitioning procedure, also found that Masters created larger clusters at all levels of partitioning than did Class B players. In addition, the chunks detected at the basic level corresponded in size and interpiece relations to the chunks identified by Chase and Simon (1973a). Gold and Opwis (1992), applying hierarchical cluster analysis, found partitions similar to those identified by latencies. Finally, the early suggestions of De Groot (1946) on the role of complex knowledge in chess have been corroborated in sorting tasks (Gruber & Ziegler, 1990), in guessing experiments (Gruber, 1991) and verbal retrospective protocols from a recall task (De Groot & Gobet, 1996). In general, the higher the skill level, the more often the verbalisations refer to abstract knowledge, the less often to chunks similar to those identified by Chase and Simon (1973a,b).

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).

## The Template Theory

The chunking model has spawned considerable empirical work (see Holding, 1985, and Gobet, 1993, for reviews), but has also been challenged on several grounds. For example, Holding (1985) criticised the recognition-association assumption (chunks act as cues that, when recognised, evoke access to heuristic suggestions for good moves)—but Gobet and Simon (1996d) obtained results supporting this assumption. Holding (1985) also proposed that Simon and Gilmartin's (1973) estimate of roughly 50,000 chunks needed for expertise is much too large. However, recent data (Gobet & Simon, 1996b; Saariluoma, 1994) support this estimate.

The main weakness of the chunking model is its assumption of encoding only into STM during a typical five-second recall task, with no new information being added to LTM. However, studies using interfering material after one or several positions have been presented (Charness, 1974; Cooke, Atlas, Lane, & Berger, 1993; Frey & Adesman, 1976, Gobet & Simon, 1996a) have shown that this material does not interfere much with chess memory, thus implying that most of the retained information has been transferred rapidly to LTM.

Gobet and Simon (1996a) propose a modified model that accounts for these findings and is in accord with other recent models of expert memory (e.g. Richman, Staszewski & Simon, 1995). (For a similar, but less specific, proposal for rapid storage in existing LTM structures, see Simon, 1976.) The modified theory continues to assert that chunks are accessed through a discrimination net. In addition, chunks that recur often in Masters' practice and study evolve into larger and more complex data structures, called *templates*, typically representing a dozen or so pieces as they are placed in a particular chess "opening".

Templates, besides containing information about a pattern of pieces, as chunks do, possess *slots* (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. Slots are created at board locations where there is substantial variation in pieces or groups of pieces among positions in the class represented by the template. This rapid LTM storage uses the same basic mechanism as the retrieval structure mechanism proposed by Chase and Ericsson (1982) to account for expert digit memory. (Ericsson and Kintsch's, 1995, alternative proposal of a single hierarchical retrieval structure to store any type of chess position is not precisely enough specified to be tested against empirical data.)

Although slots in templates can be filled rapidly, and hence augment STM in the domain of expertise, the templates themselves are built up slowly, from chess experience, 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 emphasised that his Grandmasters and Masters 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. It is by this means that strong players are able to access rapidly descriptions of the position that are larger than four or five pieces.

The template hypothesis predicts that strong players should replace positions in chunks (template) 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 in predicted chunk sizes, we must consider the criticisms of the use of inter-piece response latencies to identify chunks.

## The Operationalisation of Chunks

Several authors (Freyhoff et al. 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 would be serious if the goal were to cut a chess position into precise chunks, but are not fundamental for relating chunks to the distributions of relations between pieces, as in Chase and Simon's (1973a) study and the present one. Moreover, as we have seen, 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, simple 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 manually. 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 either misinterpreted the chunking theory, or suggested the necessity for postulating some kind of rapid encoding into LTM, a requirement that is now met in the template theory (Gobet & Simon, 1996a). 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; see also Tulving, 1962, and Bower & Springston, 1970) used 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 to detect chunking, on the assumption that one chunk is encoded per glance. In the *recall task*, subjects reconstructed a position presented for five seconds. 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.

The data supporting the two-second boundary for delimiting chunks have never been replicated. 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, removing 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 analyse 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.

## METHOD

The copying and recall tasks were given as part of a larger design to the subjects of Experiment 1 of Gobet and Simon (1996a) and to half of the subjects of Experiment 2 of Gobet and Simon (1996b). 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). As there was no difference between the two experimental groups nor any interaction of experimental group with the variables discussed here, we pooled the data from the two experiments.

### 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 (see Footnote 1). 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 nine-inch diagonal screen ( $512 \times 342$  pixels). The positions were presented on the screen with a  $9 \times 9$ cm chessboard. Individual squares were  $11.25 \times 11.25$ 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 contained: an empty  $9.5 \times 9.5$ cm board (lower left corner of the board 1.35cm from the lower left corner of the screen); a rectangular box ( $2.4 \times 7.1$ cm, 2.2cm from the right side of the screen)

displaying the six different kinds of pieces of White and Black; a  $11.9 \times 11.9$  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 were used, three taken from master games (with 24, 30, and 26 pieces) and two 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) were the same for all subjects. The concern in this experiment was not in demonstrating the, already established, superior memory for the game as compared with random conditions, 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 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 five seconds, it was no longer available to the subjects, who had to replace the pieces on the board from memory.

The 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 the pieces from game positions. For the recall of game positions, subjects’ results are based on four positions for the subjects who participated in Experiment 2 of Gobet and Simon (1996b) and on five positions for those who participated in Experiment 1 of Gobet and Simon (1996a). The game positions were randomly selected from a pool of 16 positions for the former subjects and of 26 positions for the latter. 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 also used as the initial task of another experiment). For the same reasons as in the case of the copy task, we judged the confounding due to the non-random order of presentation to be acceptable. (All positions used in this experiment can be obtained in a computer format from the first author.)

## RESULTS AND DISCUSSION

First, we analyse 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 tasks 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.

Our data generally agree well with the data from Chase and Simon (1973a, 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 exhibit converging evidence that latencies and relations between pieces provide alternative, independent but quite consistent ways of defining chunks. (As age did not correlate significantly with skill, interpiece latency, time to move the mouse, or recall percentage, we omit it from our analyses.)

### Copy task

All subjects but one (an Expert) were proficient in handling the mouse. The remaining subject dictated (in algebraic chess notation) the piece locations 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 ( $r = .05$  for game positions and  $r = .01$  for random positions).

There was an important difference in the behaviour of our subjects from those of Chase and Simon (1973a) in the copy task. Their subjects studied the stimulus

position for 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 may be due to the differences in the ways in which stimuli were presented and responses made in the two sets of experiments. Nonetheless, 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 analysed using a  $3 \times 2 \times 2$  (Skill level  $\times$  Type of position  $\times$  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 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.

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. WGP latencies are much shorter than BGP [ $F(1, 22) = 90.74$ ,  $MS_e = 10.3$ ,  $P < .001$ ]. 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 marginally significant interaction is signaled for Skill  $\times$  Type of position [ $F(2, 22) = 3.17$ ,  $MS_e = 2.15$ ,  $p = .062$ ]. There are no other significant two-way or three-way interactions: Type of position  $\times$  Placement mode [ $F(1, 22) = 1.25$ ,  $MS_e = 2.0$ ]; Skill  $\times$  Placement mode [ $F(2, 22) = 0.37$ ,  $MS_e = 10.3$ ]; Skill level  $\times$  Type of position  $\times$  Placement mode [ $F(2, 22) = 2.58$ ,  $MS_e = 2.02$ ].

Placement mode 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  $\times$  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 in 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 therefore used a *corrected latency*, subtracting from our times the time needed to move the

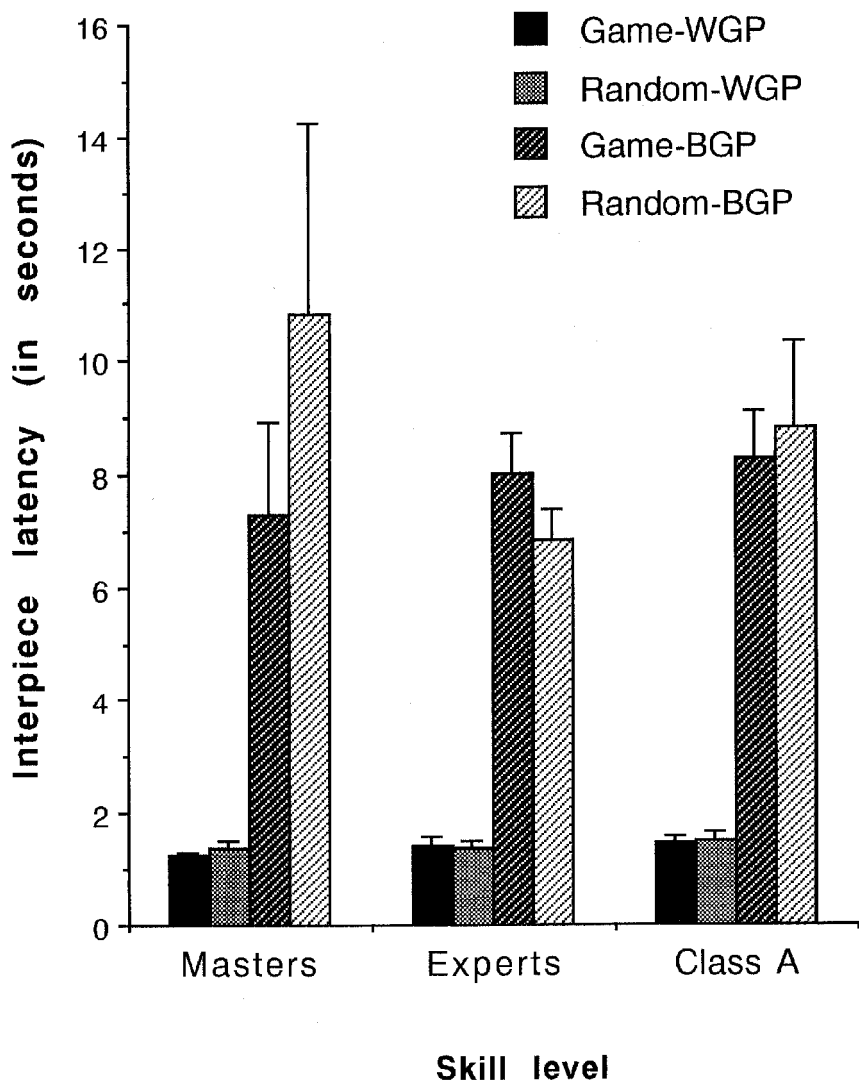


FIG. 1. Median latency between the placement of successive pieces, as a function of skill level, type of position, and type of placement (within-glance [WGP] or between-glance [BGP] placements).

mouse to the destination square once a piece was selected. Figure 2 reproduces, for all of our subjects, the corrected BGP latencies (180 observations) and 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 Chase and Simon's data.

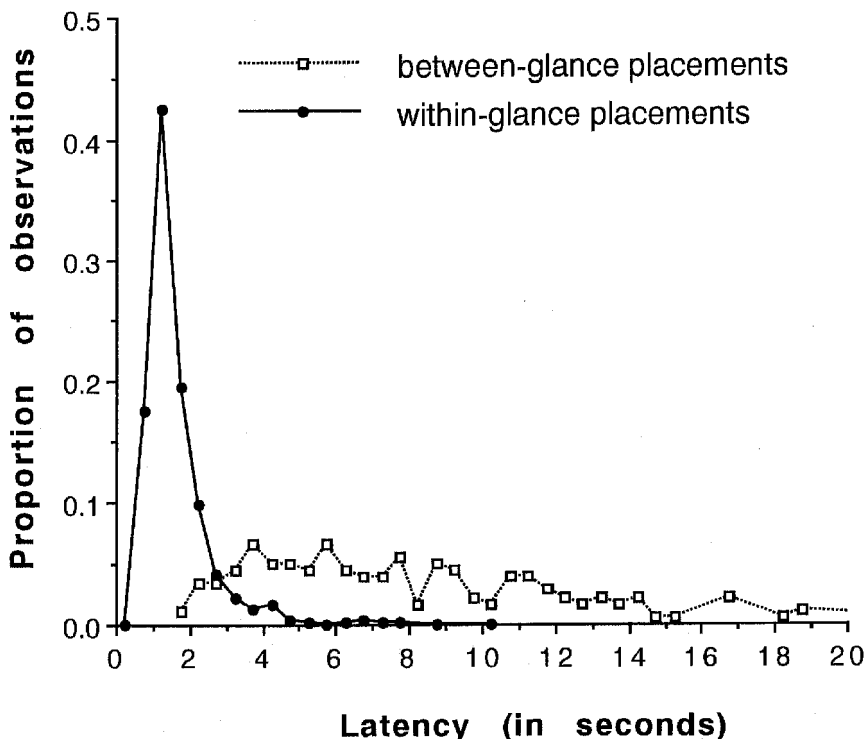


FIG. 2. Frequency histogram of within-glance latencies and between-glance latencies (corrected latencies for game positions).

As this correction 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).<sup>2</sup> The corrected distribution of WGP latencies now has 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 is 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.

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

<sup>2</sup> We employed the parameters proposed by Card, Moran, and Newell (1983, pp.241-242):  $T_{pos} = K_0 + I_M \log_2 (D/S + .5)$ , where  $T_{pos}$  is the positioning time,  $I_M = 120$  msec/bit,  $D$  = distance of the target, and  $S$  = size of the target. For  $K_0$ , the intercept, we used 400 msec, obtained by computing the time to click and unclick the button of the mouse ( $4 \times 100$  msec).

and Simon's, although a little slower even after the correction for the mouse. The close agreement adds considerable credibility to the two-second 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. Also, because of the program design, it was difficult to access the stimulus position and return to the reconstruction board in less than one second, which may have provided a motive for fewer and longer references to the stimulus.

*Total Study Time, References to Stimulus, and Subjects' Strategies.* Total time studying the stimulus position is not identical with the sum of between-glance latencies, for (a) subjects, having chosen a piece, need time to move it; (b) some subjects examined the stimulus position after the reconstruction to check for correctness. 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].

For the mean number of times 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).

Verbal comments indicate that Masters tend to study random positions in longer glances than those of weaker players, but to return less often to look at the stimulus position. They try to memorise the position as if it were a game position. The weaker players use a less expensive memory strategy: they cut up the position, generally by columns and rows, and copy these chunks.

## 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 A 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 A players in game positions, but only  $1\frac{1}{2}$  times as many in random positions. However, contrary to Chase and Simon (1973a), the recall of random chess positions varied somewhat with skill in our experiment, although the effect is not statistically significant [ $F(2,23)=2.27$ ,  $MSe=34.10$ , ns]. We show elsewhere (Gobet & Simon, 1996c) that 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 this study is compared with previous studies.

## Comparison between Copy and Recall Tasks

The theory predicts the same pattern of relations for pieces within chunks in both the copy and recall tasks. Comparing the latencies between consecutive pieces with the relations between these same pieces, we show that the chunks could be defined by numbers of relations instead of by latencies, and estimate how well 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, which predicts many more relations of attack, defence, proximity, shared colour, and shared type of piece between successive pieces within the same chunk than between pieces on opposite sides of a chunk boundary, was strongly supported by their data. We now check the findings, using 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.)

If chunk boundaries are indicated by latencies  $>2$  s, then the relations between successive pieces should be different with short than with long latencies. In addition, if the same processes determine the latencies in both the copy and recall tasks, then the relations for within-glance placements in the copying experiments should correlate with those for rapid placements ( $\leq 2$  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 did Chase and Simon (1973a), the following primitive relations: attack (A), defence (D), same colour (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 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.

Chase and Simon had found that small latencies in the copy task correlate with numerous relations between successive pieces, while large latencies correlate with few relations. In our data, also, 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

TABLE 1  
Copy Experiments

<i>Relations</i>	<i>Games</i>			<i>Random</i>		
	(1) <i>ratio<sup>a</sup></i>	(2) <i>Within</i>	(3) <i>Between</i>	(4) <i>ratio<sup>a</sup></i>	(5) <i>Within</i>	(6) <i>Between</i>
–	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

Average latencies (in seconds) for the copy experiments for combinations of the five chess relations: Attack (A), Defence (D), Spatial Proximity (P), Same Colour (C), and Same Piece (S).

<sup>a</sup> The ratio of the number of within-glance latencies to number of between-glance latencies.

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 in the within-glance condition of the copying task are below two seconds, 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 ratio of within-glance sequences to between-glance sequences increases rapidly as the numbers of relations between the 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. The same pattern appears in the recall task, where the corresponding figures are 18.6 and 1.3 for pairs with latencies of less than two seconds and more than two seconds, 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

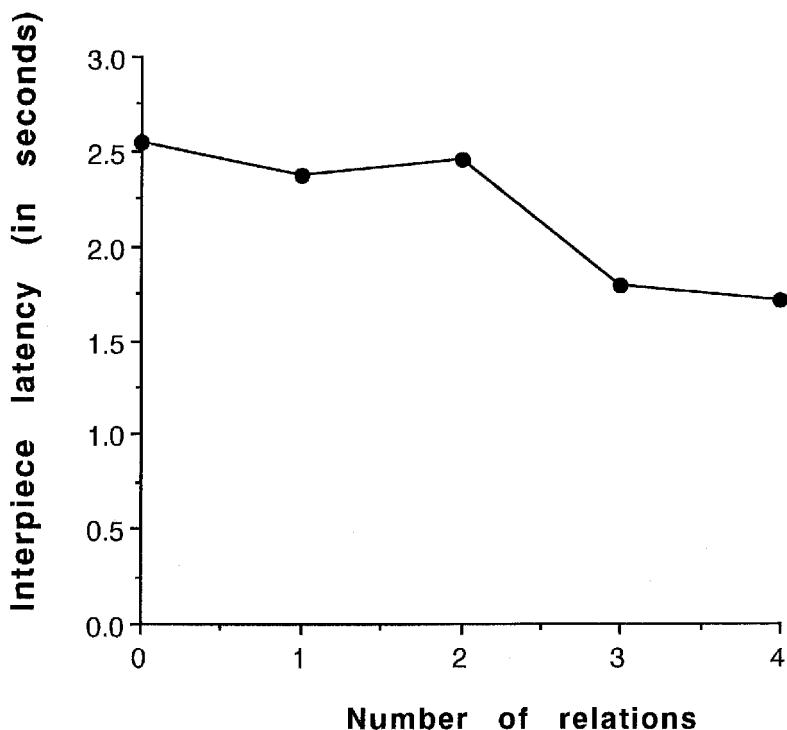


FIG. 3. Relation between interpiece latencies and the number of relations shared by two pieces successively placed (data from the recall task pooled over type of position and skill levels).



positions are pooled), shows a clear negative correlation between number of relations and latencies, similarly to Fig. 5 of Chase and Simon (1973a). Our gentler slope 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: Fig. 10 of Chase and Simon (1973b) shows clearly that the increase in latency as a function of the number of relations is inversely proportional to the skill level.

*Predicting Latency From Types of Relations.* Which relations largely account for the differences in latencies? For the within-chunk data (again pooled over tasks and types of positions) stepwise regression removes Defence 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, 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, Colour 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 Defence relations 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 colour, 28% are the same kind of piece, 11% are in proximity, just 10% have a defence 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. This does not necessarily imply, however, that chess chunks are shaped only by basic Gestalt organisational principles, for proximity, colour, and kind play an important role in the semantics of chess.

*Observed and Expected Probabilities of Sets of Relations.* Table 6 of Chase and Simon (1973a) gives the probabilities of occurrence of different combinations of relations in the various conditions. As the comparable data for our experiments are very similar, we will summarise instead of reproducing our table in full. (We will provide the full table on request.) We also compare the observed probabilities with *a priori* 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.

Our Table 2 gives the correlations of probabilities among the conditions in our experiment, corresponding to the correlations in Chase and Simon's (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 ( $r > .89$ ) with the within-glance probabilities in the copy *game* task (variable 2) and with each other; (b) the between-glance probabilities in the copy task (5 and 6) are very strongly correlated ( $r > .78$ ) 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 ( $r > .78$ ) with the *a priori* (game and random) probabilities (9–10); (d) the within-glance *random* probabilities are correlated moderately ( $.5 < r < .75$ ) 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 ( $.15 < r < .54$ ). Hence, within-chunk patterns of relations, in either 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. The structure in Chase and Simon's Table 7 is closely similar.

Omitting 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 correlations, and demonstrating good consistency in the patterns of chess relations between pairs of successively placed pieces within and between chunks as defined.

We also analysed 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 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 for 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 in Fig. 4, where we have pooled all within-chunk conditions and all between-chunk conditions. 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 are overall much closer to chance.

TABLE 2  
Intercorrelation Matrix for the Copy (in Bold), Recall and *A Priori* (in Italics) Chess Relation Probabilities

	<i>I</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
1. <b>Within-glance (random)</b>	1.000	0.701	0.648	0.550	0.752	0.696	0.531	0.656	0.495	0.525
2. <b>Within-glance (games)</b>		1.00	0.890	0.924	0.308	0.531	0.171	0.470	0.159	0.162
3. <b>≤ 2 seconds (random)</b>			1.000	0.907	0.236	0.392	0.184	0.447	0.179	0.173
4. <b>≤ 2 seconds (games)</b>				1.00	0.205	0.411	0.102	0.413	0.170	0.166
5. <b>Between-glance (random)</b>					1.000	0.867	0.778	0.829	0.775	0.837
6. <b>Between-glance (games)</b>						1.000	0.795	0.954	0.813	0.834
7. <b>&gt; 2 seconds (random)</b>							1.000	0.862	0.907	0.873
8. <b>&gt; 2 seconds (games)</b>								1.000	0.901	0.903
9. <i>A priori (games)</i>									1.000	0.984
10. <i>A priori (random)</i>										1.000

TABLE 3  
Probabilities as a Function of Time Interval for Numbers of Chess Relations

# of relations	Time interval (in seconds)										
	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

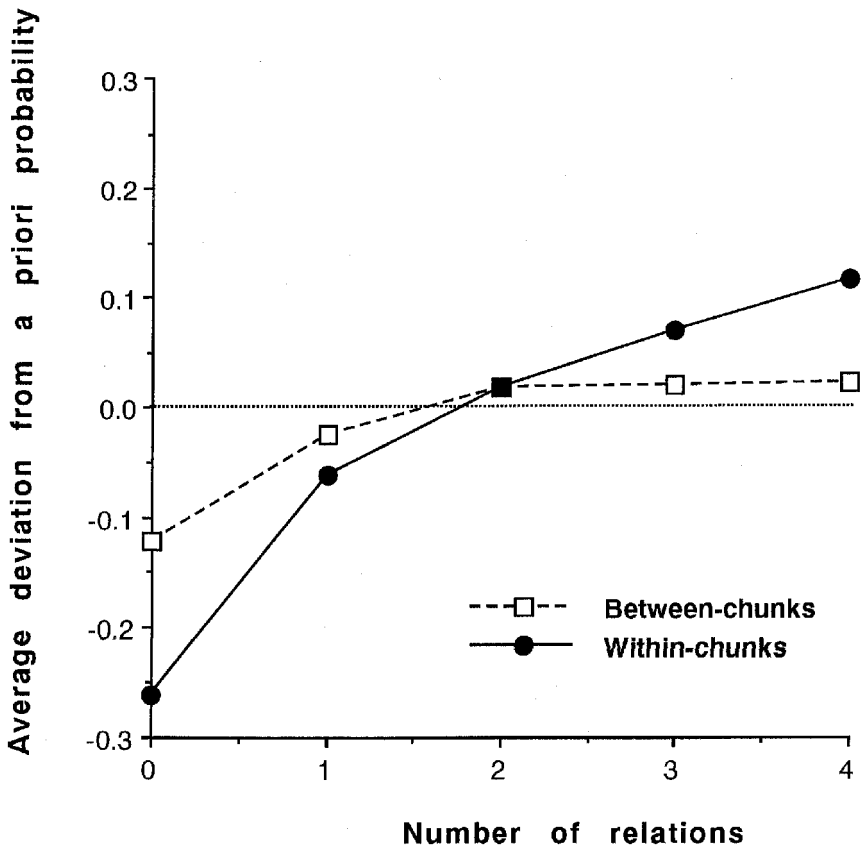


FIG. 4. Relation between chess relation probabilities and the number of relations shared by two pieces successively placed (data pooled over tasks and skill levels).

*Relations by Time Interval.* How robust is the two-second boundary in the recall task? We tabulate number of relations for each interval of latencies in Table 3, with results pooled over types of positions and skill levels. There is a clear pattern. 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.

*Convergence of Definitions of Chunks by Latencies or Number of Relations.* As the relations we have described show up prominently in chunking, using them to define whether two successive pieces belong to the same chunk or not should give results similar to those from using latencies to

define chunks. We have computed whether each pair of pieces belongs to the same chunk or not 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  $\leq 2$  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.

Thus, the two methods of defining chunks produce quite similar segmentations of the output: the findings 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 two or more chess relations with the piece previously placed. This provides strong convergent evidence that chunks have psychological reality as structures in LTM.

*Latencies as a Function of Cumulative Placements.* 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? The chunking theory predicts that the interpiece latencies will stay more or less constant when pairs of pieces belong to a chunk, but that the interpiece latencies between chunks may increase as a function of the number of pieces previously

TABLE 4  
Defining a Chunk

<i>Game positions</i>			<i>Random positions</i>		
# relations			# relations		
<i>Latency</i>	$\geq 2$	$< 2$	<i>Latency</i>	$\geq 2$	$< 2$
<i>Copy task</i>					
$\leq 2$ sec	861	167	$\leq 2$ sec	640	248
$> 2$ sec	245	190	$> 2$ sec	286	329
<i>Recall task</i>					
$\leq 2$ sec	1206	357	$\leq 2$ sec	196	76
$> 2$ sec	143	189	$> 2$ sec	28	46

Defining a chunk (a) using corrected latencies versus (b) using the number of relations shared by the pieces.

Numbers of pairs of successive pieces with: [short ( $\leq 2$  sec) and long ( $> 2$  sec) latencies]  $\times$  [ $\geq 2$  and  $< 2$  relations].

placed. The constancy 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 recognised chunks.

We have computed the average interpiece 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 shows, for the recall of game positions, that the evolution over time of the two variables follows different curves. 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 not in other studies.

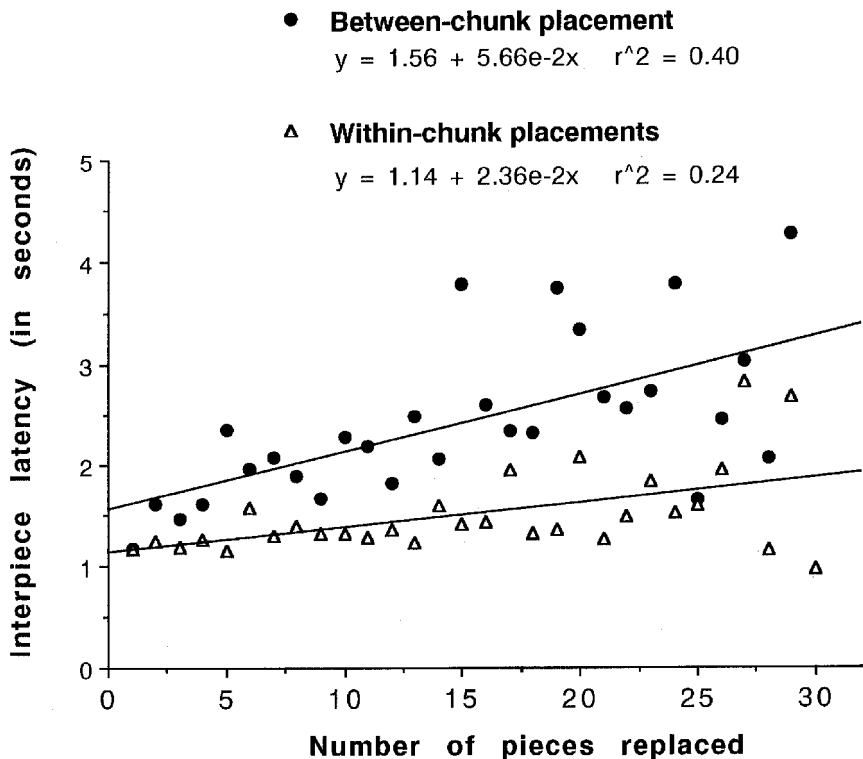


FIG. 5. Interpiece latency as a function of the number of pieces previously replaced, for between-chunk and within-chunk placements.

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. Despite the difference in apparatus, the current experiments agree closely with Chase and Simon (1973a). Finally, in two analyses extending Chase and Simon's, there was a considerable agreement in predicting whether a piece belongs to a chunk using either the number of relations or the latencies, and 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 than are predicted by the original chunking theory. In the recall task,<sup>3</sup> Chase and Simon (1973a) found 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, the Class A player, 2.6 pieces, and the beginner 2.4 pieces. The median largest chunk per position was five pieces for the Master with game positions. In the following analyses, using the two-second cutoff to define a chunk, we discuss mainly the size of the largest chunk in a position, because the template theory makes direct predictions about the size of the largest chunk, and also because skewness argues against using the arithmetic mean of chunk sizes.

Our data differ strikingly from Chase and Simon's in the sizes of chunks at all skill levels. Figure 6 shows the mean (at each skill level) of the median (over positions) of the largest chunk<sup>4</sup> 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 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 seven 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.

<sup>3</sup> Chase and Simon give no data on the size of chunks for the copying task.

<sup>4</sup> 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.

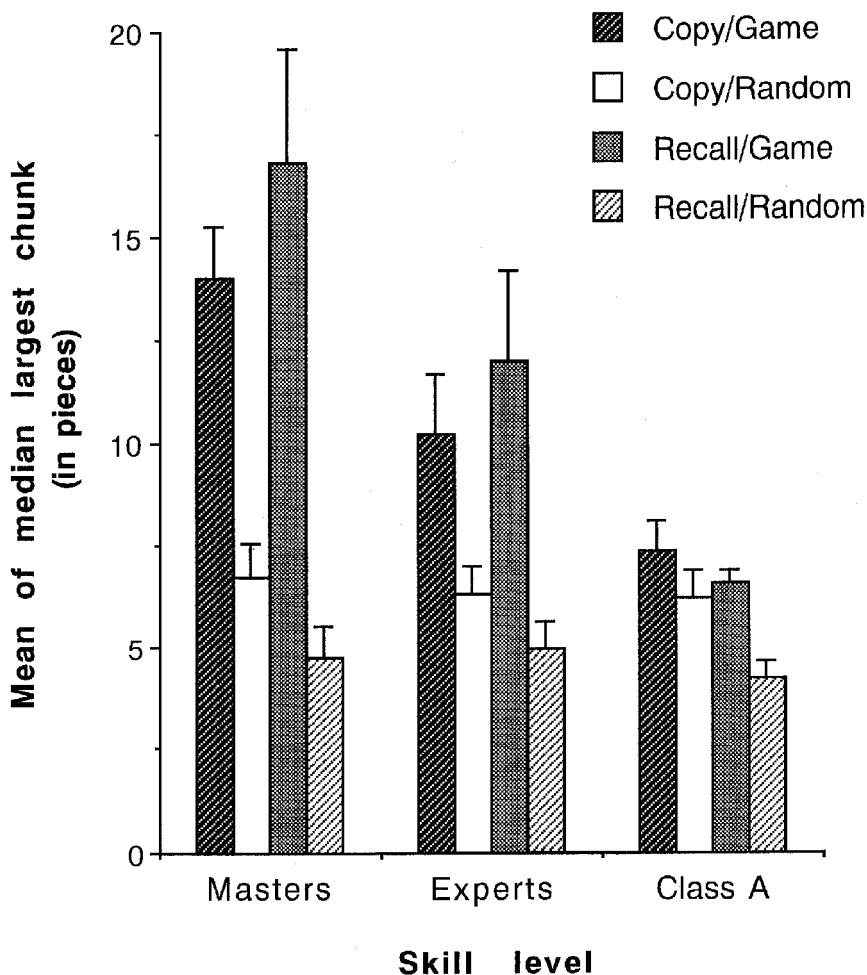


FIG. 6. Mean of median largest chunk as a function of skill level, mode of replacement, and type of position.

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, Fig. 7 illustrates the number of chunks—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. 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. Fig. 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 characterise the position as a whole.

For the random positions, we find again an inverted U-curve, both with the copy and the recall tasks. The main effects of skill [ $F(2,23)=4.84$ ,  $MS_e=1.50$ ,  $P<.05$ ] and type of presentation [ $F(1,23)=240.1$ ,  $MS_e=1.68$ ,  $P<.001$ ] are significant, while the interaction is only marginally significant [ $F(2,23)=3.23$ ,  $MS_e=1.68$ ,  $P<.06$ ]. 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 is 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 three chunks.

## GENERAL DISCUSSION

Five main reasons led us to test further Chase and Simon's method for identifying chunks. First, empirical data suggested that Masters perceive a position at a higher level of organisation 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 would 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 (1973a), but also at a detailed level. Fourth, we wanted to test whether chunking might be an artifact of the total time spent to replace a position (Wixted & Rohrer, 1994). Fifth, we wanted to check whether a different apparatus (a computer display versus actual chess pieces and board), could account for differences with Chase and Simon in the size and relational richness of chunks, where size of grasp may have affected the way subjects chunked the position.

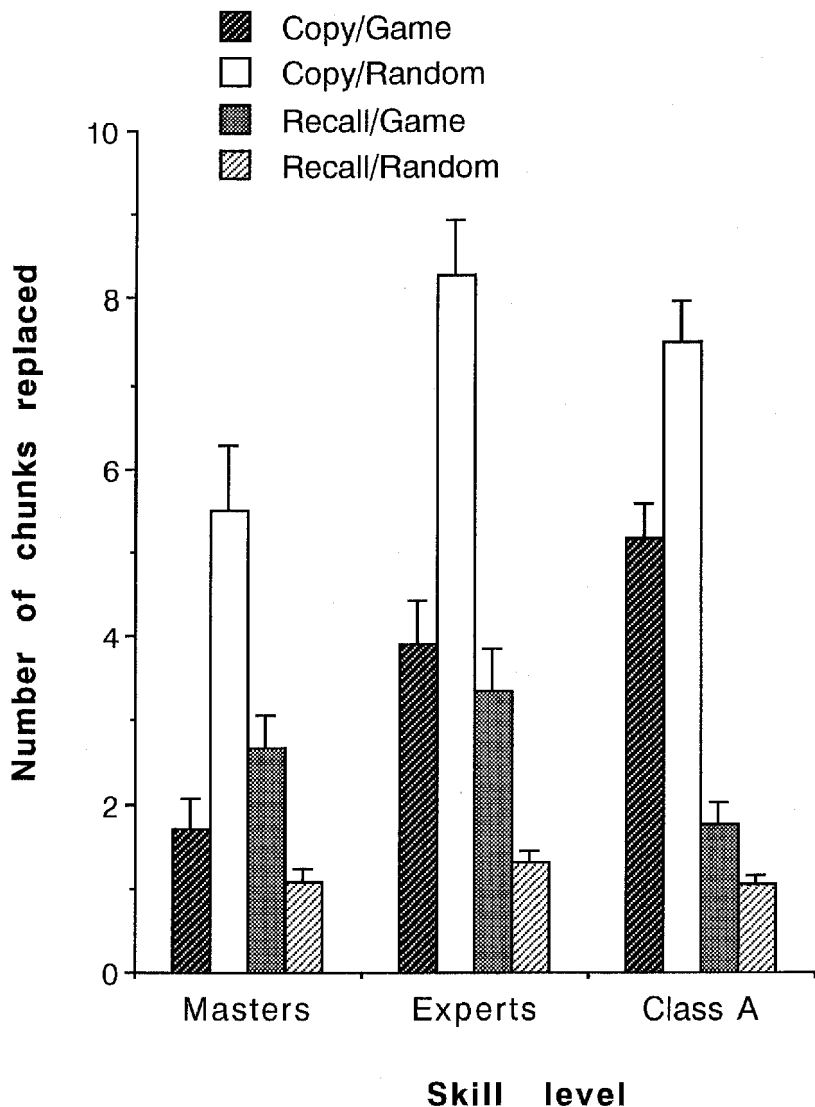


FIG. 7. Number of chunks as a function of skill level, mode of replacement, and type of position.

In the results section we noted a difference, for the copying task, between the strategic behaviour of our Masters and Chase and Simon's Master: the former spent more time than the latter in studying the position before reconstructing it. Despite this difference, we could 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 apparatus and in Masters' strategies. Specifically, more than three quarters of the corrected within-glance latencies are below two seconds. Third, in game positions, strong players make faster placements between glances but not within glances than amateurs.

Although the two-second boundary is only approximate, it seems to be reasonable, for in our copy task, 79.5% of the latencies for within-glance placements were less than two seconds (versus 1.11% for between-glance placements). This provides additional evidence that the definition of chunks employed by Chase and Simon and used here actually reflects subjects' perceptions of the board. In agreement with other studies, the recall performance of stronger players was clearly superior with game positions. They were also slightly superior with random positions, a result that did not appear in Chase and Simon's data.

In copy and recall tasks, in 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 there are more relations within a chunk. Second, numbers of chess relations in within-glance placements in the copy task and in placements within two seconds in the recall task were strongly correlated, as were numbers of relations in between-glance placements in the copy task and placements over two seconds 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), 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.

The behaviour of interpiece latencies over successive placements also strengthens the concept of chunk. Wixted and Rohrer (1994) showed that latencies generally become longer for successive items recalled. Our data showed that the latencies of pieces that have two or more chess relations with their predecessors exhibit a small increase in size over time; the latencies of pieces that have 0-1 chess relations with their predecessors increase more than twice as rapidly. This is consistent with the 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. The correct interpretation of Wixted and Rohrer's finding is that in retrieval, the latencies between chunks will grow with time, but that this does not imply an increase in within-chunk latencies for successive chunks, nor do our data show substantial differences of this kind. Hence, Wixted and Rohrer's survey of the literature on free recall is wholly consistent with the findings on chunking in the chess literature.

In general, chunks at different skill levels all show the same pattern of relations. 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 organisation. If so, chess expertise would differ from physics expertise, where it has been shown that the representation of information differs qualitatively with skill (Chi, Feltovich, & Glaser, 1981). However, as the largest chunks of highly skilled players are much more complex than those of weaker players, important qualitative differences, not reflected in the pattern of elementary relations, may be present in these large templates.

Consistent with the evidence reviewed in the introduction, maximum chunk size was substantially larger in the current experiment than in the experiments of Chase and Simon. The small difference in number of chunks in the recall task between the strongest skill level and the weakest 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 mainly due to the presence of a few large chunks. Although we did find, as did Chase and Simon, some difference in the number of chunks between skill levels, the differences were much smaller in the present study than in the earlier one. This change in chunk size may have been produced by the difference in experimental procedure (grasping pieces by hand in the earlier study). 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 found adds support to the template theory, in as much as the number of chunks at all skill levels was within the supposed capacity of visual STM (Zhang & Simon, 1985).

Are the chunks the product of some artifactual feature of our material? First, the positions we used might come from very typical opening variations that are likely to have been overlearned by skilled players. Although this is somewhat possible for the positions of the copy task, the positions used in the recall task do not, as far we can judge, belong to typical openings situations, and it is unlikely that the typicality of positions has inflated our estimate of the size of chunks.

The second feature is the effect of "serialisation" (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 procedure allows subjects to search memory for a new piece/chunk while still busy replacing the previous piece. Such a time-sharing strategy should, however, level the interpiece latencies and destroy chunks. (It cannot be assumed that Masters have superior motor skills in placing pieces, for chess skill level accounted for less than 0.3% of the variance in speed of moving the mouse.)

A reviewer has proposed that the serialisation 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 latter's larger chunks (at most 6–7 pieces) 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 two-second boundary favoured the former in comparison with the latter. Altogether, our results on the size of chunks 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? With a physical chessboard, the maximum chunk size is limited by the number of pieces the hand can grasp. Without this limit, large chunks may be recalled that represent core information of typical positions. The Master has spent enough time studying these types of position to have acquired large, well differentiated templates for them. We remark again that Chase and Simon's Master was somewhat out of practice, and, being in his mid-forties, may have been slower in replacing pieces, causing some chunks to be divided in scoring them.

In the random positions, most of the chunks 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 a2], [Slot #4: number of pawns is 4]) rather than simple chunks enumerating the pieces (see Goldin, 1979, or Gobet and Simon, 1996b, for more data on recognition and recall of random positions).

## CONCLUSION

In the introduction of this paper, we presented major criticisms of the chunking theory of expert memory in chess. New experiments, together with evidence from the literature, clearly establish that an augmented chunking theory—the template theory described here—meets all these criticisms. Summarised 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 build around templates that encode information, acquired by strong players over years of practice and study, about typical and familiar positions, and that provide rapidly fillable slots for additional chunks of information about the current position.

Schemas and scripts have long been known to play a key role in expert performance and memory, as has been documented, for example, in physics (Chi et al. 1981), medical expertise (Patel & Groen, 1991), reading (Ericsson & Kintsch, 1995), and scientific expertise (Schraagen, 1993). Our theory builds on these concepts by (a) adding mechanisms for creating and accessing templates,

(b) adding the assumption that slots (variable positions in templates) may be filled in rapidly, and (c) linking these to a computational model of cognition, based on the EPAM theory (Feigenbaum & Simon, 1984). A task for future research will be to establish the precise nature of the structures that are required in specific task domains to allow rapid storage of new information.

Manuscript received 23 August 1996

Manuscript accepted 14 February 1997

## REFERENCES

- Bower, G. H., & Springston, F. (1970). Pauses as recoding points in letter series. *Journal of Experimental Psychology*, 83, 421-430.
- Card, S., Moran, T.P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Charness, N. (1974). *Memory for chess positions: The effects of interference and input modality*. Unpublished doctoral dissertation, Carnegie Mellon University.
- 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: Lawrence Erlbaum Associates Inc.
- Chi, M.T. H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Cooke, N.J., Atlas, R.S., Lane, D.M., & Berger, R.C. (1993). Role of high-level knowledge in memory for chess position. *American Journal of Psychology*, 106, 321-351.
- De Groot, A.D. (1978). *Thought and choice in chess*. The Hague: Mouton Publishers. [First publication: 1946.]
- De Groot, A.D., & Gobet, F. (1996). *Perception and memory in chess. Heuristics of the professional eye*. Assen: Van Gorcum.
- Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.
- Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science*, 8, 305-336.
- Frey, P. W., & Adelman, P. (1976). Recall memory for visually presented chess positions. *Memory & Cognition*, 4, 541-547.
- Freyhoff, H., Gruber, H., & Zeigler, 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. (1996a). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, 31, 1-40.
- Gobet, F., & Simon, H. A. (1996b). Chunks in chess memory: Recall of random and distorted positions. *Memory & Cognition*, 24, 493-503.
- Gobet, F., & Simon, H. A. (1996c). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review*, 3, 159-163.

- Gobet, F. & Simon, H.A. (1996d). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grandmaster level chess. *Psychological Science*, 7, 52–55.
- Gold, A., & Opwis, K. (1992). Methoden zur empirischen Analyse von Chunks beim Reproduzieren von Schachstellungen. *Sprache & Cognition*, 11, 1–13.
- Goldin, S. E. (1979). Recognition memory for chess positions: Some preliminary research. *American Journal of Psychology*, 92, 19–31.
- 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: Lawrence Erlbaum Associates Inc.
- Patel, V.L. & Groen, G.J. (1991). The general and specific nature of medical expertise: A critical look. In K.A. Ericsson, & J. Smith, (Eds.), *Studies of expertise: Prospects and limits*. Cambridge: Cambridge University Press.
- 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.
- Schraagen, J.M. (1993). How experts solve a novel problem in experimental design. *Cognitive Science*, 17, 285–309.
- 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., & Gilmarin, 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, 344–354.
- 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.
- Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition*, 13, 193–201.

Copyright of Memory is the property of Psychology Press (T&F) and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.