Chunks in Chess Memory:
Recall of Random and Distorted Positions

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

Complex Information Processing
Working Paper #518

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

has@cs.cmu.edu
412-268-2787

Running head: Chunks in Chess Memory

Abstract

This paper explores various aspects of chunking in chess. The two first experiments investigate whether it is possible to reduce drastically the usual estimate that Masters store some 50,000 chunks by assuming that the same chunk can encode (a) the same chess pattern for White as well as Black pieces and (b) the same patterns at different locations on the board. When positions are modified by mirror image reflection about various axes, recall is impaired; hence it is dubious that many patterns, even those that can be obtained from each other by simple transformations, can be represented by one chunk.

In Experiments 3 and 4, we analyze the recall of random positions and the structural properties of chunks. Finally, Experiment 5 investigates the effect of presentation time on the recall of both game and random positions. Strong players are better than weak players both with game and random positions. Their superiority with random positions is especially clear with long presentation times, but remains after brief presentation times, although smaller in absolute value. The role of interpretation and compilation of chunks is discussed, and the results of the five experiments are confronted with the template theory proposed by Gobet and Simon (1994a).

Chunks in Chess Memory:
Recall of Random and Distorted Positions

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

Chunking has been shown to be a basic phenomenon in memory, perception and problem solving. Since 1956, when Miller published his "magical number seven" paper (Miller, 1956), evidence has accumulated that memory capacities are measured not by bits, but by numbers of familiar items (for example, common words are familiar items). Many studies of expertise, in which chess expertise has played a prominent role, have focused on discovering the size of expert memory, the way it is organized and the role it plays in various kinds of expert performance. The often-cited figure of 50,000 chunks -- familiar patterns of pieces -- in the memories of chess Masters and Grandmasters was proposed by Barenfeld, Gilmartin, Chase and Simon (Chase and Simon, 1973a,b) as an order-of-magnitude estimate, *roughly* comparable to natural language vocabularies of college-educated people. Subsequent studies have raised questions about memory for chunks and the role it plays in expertise.

The chunking hypothesis, and the consequences that flow from it, have great significance both for memory theory and for the application of the theory to teaching and learning. For this reason, we have thought it important to review the evidence on chunking in chess, to repeat and elaborate some kinds of experiments that have produced conflicting or ambiguous findings, and to construct a revised theory of chunking and chess memory generally.

In the domain of chess expertise, chunking has been pinpointed as a basic phenomenon 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 when Chase and Simon (1973a) proposed an operational definition of chunks in chess. Comparing the distributions of latency times in a memory task (the de Groot recall task) and a perceptual task (where the contents of one position had to be copied on a different board) and analyzing the chess relations between pairs of pieces during the reconstruction of the board, they proposed that a chunk may be defined as a group of pieces placed successively with between-piece intervals of less than about 2 seconds[1].

Chase and Simon (1973b) proposed that, during the brief presentation of a chess position, chess players recognize chunks on the board and that pointers to these chunks are placed in a short-term memory of limited size. A computer program, MAPP (Gilmartin and Simon, 1973), simulated several findings from this kind of experiment, including the percentage of pieces recalled by a class A player, the type of pieces replaced and the relations between successive pieces in the reconstruction. Extrapolating from their results, Gilmartin and Simon proposed that expertise in chess would require between 10,000 and 100,000 chunks (in the literature, this range is often replaced by the approximation of 50,000 chunks). Finally, an important implication of Chase and Simon's theory of chunking was that chunks, upon recognition, would suggest good moves to the masters.

Additional experimental evidence supporting chunking in chess

The empirical evidence described in Chase and Simon (1973a, b) was derived from a single paradigm. Other authors have tested the plausibility of chunk structures experimentally in other paradigms as well. Charness (1974) presented pieces verbally, at a rate of 2.3 sec per piece. Pieces were either grouped according to the chunk 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, which in turn permitted better recall than the random condition. The same results were found when pieces were presented visually, one at a time (Charness, 1974).

Frey and Adesman (1976) used a similar technique, presenting slides, each containing a group of (usually) four pieces, but cumulating by retaining the pieces from previous slides on the board. Each of the six slides was presented for 2 sec. Frey and Adesman found chunk presentation superior to column presentation. Moreover, the chunk presentation yielded better recall than presentation of the entire position for the same length of time (12 sec).

Criticisms of the Chunking Hypothesis

The chunking model has spawned considerable empirical work (see Holding, 1985 or Gobet, 1993, for reviews). The model has been challenged, however, on several

[1]We have replicated elsewhere (Gobet & Simon, 1993b) this experiment, with a material slightly different and with more subjects than in the original study.

grounds. Several authors (for example Reitman, 1976; Gold & Opwis, 1992; Holding, 1985) have suggested weaknesses in the way chunks are operationalized in Chase and Simon's experiments. These include the difficulties of determining clustering from reaction times, the failure to capture overlapping or nested chunks, and the assumption that chunks are recalled completely and in a single block. We address these questions of identifying chunks in detail in Gobet & Simon (1994b). In another paper (Gobet & Simon, 1994a), we review studies that address the lack of interference in chess memory (Charness, 1974; Frey & Adesman, 1976), discuss the criticisms of the hypothesis that the pieces are encoded in a STM of limited size during the recall task and propose a modified (template) model that accounts for the rapid encoding shown by chess masters.

In this introduction, we will review two other sets of criticisms. The first questions whether chunks play an important role in chess problem solving (the recognition-association hypothesis). The second questions the estimated number of chunks necessary to reach expertise, usually estimated at about 50,000.

The recognition-association assumption

A central thesis of the chunking model is that chunks act as conditions of a production system, and that their actions contain heuristic suggestions for good moves. Holding (1985, 1992) has challenged this assumption on the grounds that (a) most chess patterns are made 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 base moves (first moves from the stimulus position) and does not allow for look-ahead analysis.

The first criticism, the unimportance of moves proposed by pawn structures, is refuted by the importance that chess players give to pawns in chess. Their importance was recognized already in the eighteenth century by Philidor (1749), who stated that "Pawns are the soul of chess". More recently, several books have been written (for example Euwe, 1972; Kmoch, 1980) on the proper way to handle pawns and on typical pawn structures. In brief, pawns structures provide information about the squares on which pieces should be placed (e.g. a Knight in front of an isolated pawn) and also give indication of typical pawn moves for given structures. Subjects comment frequently on pawn structures and on moves

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

relevant to them in problem solving tasks (see de Groot, 1949/1978) and even in memory tasks (Gobet, 1993).

The second criticism is that chunks are too small, according to Holding, to generate useful information (Chase and Simon hypothesized chunks of at most 5-6 pieces). This may have some truth, although even small chunks may suggest good moves in tactical situations and chunks or small constellations of them may allow recognition of positions of particular types. As we shall see in the experimental part of this paper, Chase and Simon have probably underestimated the chunk size, especially with masters.

The third criticism involves a misunderstanding of the pattern-association 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. However, Chase and Simon made it clear 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 this study, the skill of subjects (from 1000 to 2200 ELO[3]) did not correlate with the recall of random positions[4] 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 on the initial problem position does not really address the recognition-association theory.

Although Chase and Simon mention only moves as the information elicited by chunks, chunks can also provide information about the class to which the position belongs,

---

[3]The ELO rating is an interval scale that ranks competition chess players. Its standard deviation (200 points) is often interpreted as a measure of skill class in chess. Grandmasters are normally rated above 2400 and Experts above 2000.

[4]A problem with this study is that the positions are not completely random, and that Saariluoma's (1990a) qualification "pseudo-random" would suit them better. 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.

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. We will suggest later that pattern recognition can provide the basic mechanisms that are lacking in SEEK (SEarch, Evaluation, Knowledge), the model of chess expertise that Holding (1985) proposes in place of Chase and Simon's model.

The number of chunks

The estimate that 50,000 chunks must be stored in LTM to reach expertise comes from extrapolations from the simulations performed by MAPP, the program described by Simon and Gilmartin (1973). Holding (1985, 1991) has argued that this number is vastly exaggerated, and that as few as 2,500 chunks may account for the results obtained in recall experiments. We will address this question directly in the first and second experiments of this paper.

In this paper, we will start with estimating the number of chunks (Experiments 1 and 2). The control task of these experiments (recalling random positions) will allow us to study in more detail chunking in unstructured positions (Experiments 3 and 4). Finally, we will analyze the role of presentation time for both game and random positions (Experiment 5). The aim of this last experiment is mainly to see how the amount of time available affects players' abilities to chunk positions, and to analyze in detail chunking process in random positions.

Effect of mirror image reflection on the recall of chess positions

The sensitivity of perception to transformations of stimuli has long been a topic of research in psychology. For example, subjects experience considerable difficulty in reading upside-down printed texts, or text that has been flipped so that it reads from right to left with reversed letters. After a substantial number of hours of practice, however, their speed increases approximately to the level for normal text (Kolers & Perkins, 1975). We can learn something of the nature of chunking in chess perception by subjecting the board positions to transformations that alter chunks to varying degrees and in different ways.

As stated above, Holding (1985, p. 109) has suggested that the estimate of 50,000 chunks is too large, and could be reduced by half by assuming that the same chunk represents constellations of either White or Black pieces[5] and further reduced by assuming that constellations shifted from one part of the board to another are encoded by the same symbol. As we interpret Holding's view, chunks could be seen as schemas encoding abstract information like: "Bishop attacking opponent's Knight from direction *x*, which is protected by a pawn from direction *y*", where the exact location in the chess coordinate system is not encoded. The alternative hypothesis is that chunks do encode precise information on piece locations, and therefore that *different* chunks would be activated upon recognition of the same White and Black patterns, or of a pattern that has been shifted by one or more squares. A weaker version of this hypothesis is that both ways of encoding operate simultaneously, the specific one being faster than the non-specific, which requires additional time to instantiate variables (see Saariluoma, 1991, for a similar view).

Information about chunk locations seems to be necessary as a part of the chunk definition because shifting the location of a chunk changes the relations of that chunk with the rest of the board. Suppose, for example, there is a two-piece pattern characterized by the relation pawn-defends-bishop. When the pattern involves a white Pawn at d2 and a White Bishop at e3 and no other piece is on the board, the Bishop controls 3 empty diagonals (9 squares). However, when the pattern is shifted 3 columns to the right and 4 ranks to the bottom of the board (i.e. a white Pawn at g6 and a white Bishop at h7), the Bishop controls only one empty diagonal (one square). Needless to say that two such patterns have totally different roles in the semantics of chess.

At a more general level, to what extent is expertise based on perceptual mechanisms, and to what extent on knowledge of a more conceptual kind? The former alternative would explain expertise as a product of very specific recognizable perceptual chunks and associated productions. The latter hypothesis would explain expertise as based upon general-purpose schemas whose variables can have different values in different situations. In the former case, a necessary, but not sufficient, condition for expertise would be possession of a large number of productions conditioned on specific patterns (e.g., chess patterns noticed on the board). In the latter case, fewer schemas would be needed for expertise, for schemas could be instantiated differently from case to case, but instantiation would increase the time required to acquire a schema (Richman, Staszewski and Simon, 1994).

Saariluoma (1984, 1991) addressed this question by manipulating the location of chunks. In one experiment, he constructed positions by first dividing the original position

[5]Incidently, we note that Holding is wrong in assuming that the lack of distinction between White and Black will divide the estimate by half, because Simon and Gilmartin's program, which the extrapolations stem from, already encode White and Black patterns as a single chunk (see note 2 in Simon and Gilmartin, 1973).

in 4 quadrants, and then swapping two of these quadrants (see example given in Figure 1). (Notice that this type of modification sometimes produces illegal positions.) These positions were then presented for five seconds to subjects ranking from Class C to Expert level. Results of the recall task show that subjects remember well the non-transposed quadrants (not as well, however, as the game positions) but remember badly the transposed quadrants (even less well than the random positions). In addition, a condition where the four quadrants are swapped gives results close to random positions.

Insert Figure 1 about here

A possible criticism of this experiment, however, is that subjects may choose a strategy that avoids the non-familiar portions of the board (the transposed quadrants are easily recognized because they do not fit the color distribution normally found in chess positions). In a second set of experiments, Saariluoma (1991) tried to remove this objection by hybridizing different positions instead of transforming a single one.

He constructed positions by assembling 4 different quadrants from 4 different positions, but retaining the locations of the quadrants on the boards. Although such hybrid positions respect the color partition found in games, some of them may be illegal[6]. In a recall task, Saariluoma found that subjects recall these positions at about the same level as game positions, from which he concludes that encoding maintains location information. These results show moreover that subjects may recall a position very well even when a high-level description of the position (a general characterization of the type of position) is not available.

Insert Table 1 about here

Table 1 summarizes the results obtained in the recall of normal, hybrid and diagonally swapped positions. It can be seen that positions keeping pieces in the same locations produce good recall even if the overall structure of the position has been changed by hybridization. One cell is however missing in this table: how good is recall when

[6]Note that this transformation keeps the pawn structure essentially plausible. Two possible experiments to see whether location matters more for pawns or for pieces suggest themselves: (a) randomizing pawns and leaving pieces intact and (b) randomizing pieces and leaving pawns intact.

location is different but the overall structure is kept intact? This question is important, as it addresses Holding's hypothesis directly: in this case, the chess relations (mainly attack, defense and proximity) are the same between the two position, but the locations of chunks have changed. Independently of Saariluoma, we were also trying to investigate this question of location of chunks. Luckily, our experiments address the question posed by the missing cell, thus supplementing Saariluoma's findings.

In the two following experiments, we will propose a new way to investigate whether two instances of the "same" pattern are represented by a single chunk or by distinct chunks when they are located at different parts of the chess board. Following Holding, such constellations as [King on g1 + pawns on f2-g2-h2] and [King on g8 + pawns on f7-g7-h7], which are very common in chess games, could, ignoring color, be encoded by a single chunk in LTM. The same chunk could encode constellations like [King on b1 + pawns on a2-b2-c2] and [King on b8 + pawns on a7-b7-c7] (ignoring the color distinction and shifting the constellations to the left). This hypothesis of invariance is unobvious, since players may feel at ease in certain positions, but not in the corresponding positions reversing Black and White and thereby shifting the location of chunks on the board (for an informal example, see Krogius, 1976, p. 10). The psychological reality of such generalized chunks must be settled empirically. We will shed some light on the question by using normal game positions and game positions that have been modified from by taking mirror images[7] around horizontal, vertical and central axes of symmetry.

Three points should be mentioned. First we used a transformation by *reflection,* and not by *translation* as in Saariluoma's swapping experiment. Second, our transformations keep the relations between pieces intact, but may change the up-down and/or left-right orientation of these relations. This is a bit unfortunate, but there is no transformation that manipulates location while keeping at the same time the overall chess relations intact and their orientation unchanged. Third, and most important, our mirror image transformations keep the game-theoretic value of the position invariant (correcting, of course, for colors). The only exceptions are positions where one side still has the right to castle before or after vertical or central transformations (this situation occurs rarely in our stimuli).

Holding's assumption would predict no difference in the recall of the various conditions. Our alternative hypothesis, based on analysis of the chess environment, which we will discuss shortly, leads us to predict a continuous decrease in performance in the following order: (a) normal positions; (b) positions modified by reflection about a horizontal axis (horizontal symmetry); (c) positions modified by reflection about a vertical

[7]These modifications will be illustrated in the section Materials of the Methods of Experiment 1.

axis (vertical symmetry) (d) positions modified by both reflections (central symmetry). As we suppose that color is encoded in the chunks, reflecting the board around the horizontal axis through the middle should affect recall performance, however slightly; although most configurations can appear both on the White and the Black sides, some patterns occur almost always on the one rather than the other. (For example, the central pawn structure made of White pawns on c4, e4, and f4 and Black pawns on d6, e6, f7, typical for many variations of the Sicilian defense, is quite uncommon with the reverse colors).

Vertical symmetry will alter recall performance more than horizontal symmetry because the former will produce positions much less likely to appear in normal games than those produced by the latter. In particular, the King's position, which is rich in information in chess, is not basically altered by reflection about a horizontal axis, whereas it is by reflection about a vertical one[8]. Following the same line of reasoning, we predict that recall of positions modified by central symmetry (reflection about both axes) will be harder to recall after modification by horizontal and vertical symmetries.

In summary, after modification, it is harder to find familiar chunks in LTM, and, in consequence, recall is impaired. Impairment of recall will be a function of the kind of modification. Because these modifications leave many configurations recognizable (for example, after modification by central symmetry, White pawns on f4, g3 and h2 become Black pawns on c5, b6, a7, which is a common pattern), and possibly because chess players, if they do not recognize patterns, may find a few chunks based on functional relations present in these positions, recall of modified positions should be greatly superior to recall of random positions.

Experiment 1

Methods

Subjects.

One female and eleven male chess players volunteered for this experiment. Their ratings ranged from 1950 to 2590 USCF ELO[9]. Subjects were classified in three groups: Masters (n = 3, mean ELO = 2417), Experts (n = 6, mean ELO = 2182) and Class A

[8] In most chess games, both sides' Kings remain on the King's side. Thus, for most positions, vertical and central modifications will send the Kings to the Queen's side, a location they occupy in only about 10% of games (Krabbé, 1985).

[9] The American rating system (USCF rating) uses the same mode of computations as the international system (ELO). However, because of differences in the games selected for computation, a player's USCF rating is in general about 50 points above the international rating.

Players (n = 3, mean ELO = 1976). Their ages varied from 17 to 45, with mean = 26 and standard deviation = 8. Subjects were recruited in New York's Manhattan Chess Club, and were paid $10 for their participation. These subjects also participated in experiment 1 of Gobet and Simon (1994, a).

Control Task.

In order to check against the possibility that the strong players had superior memory capacities, we presented five random positions (mean number of pieces=25). These positions were constructed by assigning the pieces from a normal game position to squares on the chessboard according to random numbers provided by a computer. The five positions were inserted randomly among the experimental positions.

Material.

Twenty positions were selected from Wilson (1976), Reshevsky (1976) and Euwe (1978), using the following criteria: (a) the position was reached after about 20 moves; (b) White is to move; (c) the position is "quiet" (i. e. is not in the middle of a sequence of exchanges); (d) the game was played by (Grand)masters, but is obscure. The mean number of pieces is 25. The positions were assigned to 4 groups, comparable as to numbers of pieces and position typicality (as judged by the first author, whose rating is about 2400 ELO). The positions of the first group (normal group) were kept unchanged; those of the second group (horizontal symmetry group) were modified by taking the mirror image with respect to the horizontal axis of the board; the positions of the third group (vertical symmetry group) were modified in the same way, but this time with reflection about the vertical axis. The positions of the fourth group (central symmetry group) were subjected to both modifications simultaneously. Figure 2 illustrates these modifications for a position appearing in each of the 4 permutations.

Insert Figure 2 about here

Positions were presented on the screen of a Macintosh SE/30, and subjects had to reconstruct them using the mouse (For a description of the experimental software, See appendix in Gobet & Simon, 1994b). Each position appeared for 5 seconds; the screen was then black during 2 seconds preceding display of the blank chessboard on which the subject was to reconstruct the position. No indication was given of who was playing the next move, and no feedback was given on the correctness of placements.

Procedure and design.

Subjects were instructed on how the program functioned and how to use the mouse, and received two training positions (one game- and one random position). The 5 positions of the 4 groups as well as the positions of the control task were then presented, in random order (the randomization being on positions, not on type of position). The set of positions and their order was the same for all subjects, in order to facilitate comparisons between subjects.

A factorial design 3x4 (Skill x Type of modification), with repeated measurements on the Type of modification, was used. Dependent variables were the percentage of pieces correctly replaced, the latencies between successive pieces, the mean number and mean largest size of chunks, and the number and type of errors.

Results

Percentage of pieces correct.

No significant correlation was found between recall and age, or between recall and time to perform the task; we shall therefore ignore these variables in the following analyses.

The control task (random boards) shows no differences in the mean percentage of correct pieces among the three skill levels: Master=13.1 (sd=4.9), Expert=15.0 (sd=5.3) and Class A=11.8 (sd=2.6); $F(2,9)=0.48$, ns. The mean percentage across the three skill levels is 13.7, that is, about 3.4 pieces. This result agrees with those found in the literature.

Figure 3 shows the results for the experimental positions. (Random positions are also shown, for comparison). An analysis of variance, computed after log-transformations of the scores, indicates a main effect of Skill [$F(2,9)=4.38$, $p<.05$] and of Type of modification [$F(3,27)=12.59$, $p<.001$]. No interaction is found [$F(6,27)=1.96$, ns]. Trend analysis shows a linear component [$F(1,9)=62$, $p<.001$], and a weaker cubic component [$F(1,9)=5.52$, $p<.05$], the latter being due to the Masters' results with vertical and central symmetry.

Insert Figure 3 about here

In accord with our hypothesis, unmodified positions are recalled best; positions with horizontal symmetry are recalled better than positions with vertical symmetry. However, two points are to be noted. First, the difference between positions with normal and horizontal symmetry is small. The difference seems to be related inversely to chess skill. As the effect is small, more data need to be collected before any strong statement can be made. Second, positions with central symmetry are well recalled by the Masters.

Interpiece latencies.

No statistically significant differences were found for the medians of interpiece latencies. Medians are located around 3 sec, while Chase and Simon (1973 a) found medians between 1 and 1.5 sec. When the time to move the mouse is subtracted from the time between the placement of two pieces, the agreement with the earlier study is fairly good. (For more detail on inter-latency times in experiments using a computer display, see Gobet and Simon, 1994b)

Chunk analysis.

As the chunking hypothesis plays an important role in Chase and Simon's (1973b) model, we analyze in some detail the potential effects of our modifications on the number and size of chunks. Our hypothesis is that the modifications tend to decrease the likelihood of evoking chunks in LTM; this should affect the number of chunks as well as their size. Before investigating this matter, however, we will present some results on the overall distribution of chunks. Throughout this discussion, we define a chunk as a sequence of pieces whose consecutive placement times are separated by less than 2 seconds. As our experimental apparatus (especially the need to move the mouse) has increased the interpiece latencies in comparison with Chase and Simon (1973a), we will use a corrected latency, where the time needed to move the mouse once a piece has been selected is subtracted from the interpiece time.

Let us illustrate the general distribution of chunk sizes with the results of four subjects, M1 (USCF ELO 2590), M3 (2325), E6 (2100) and A2 (1989), who are representative of our sample. The upper panel of Figure 4 shows, for these subjects, the distribution of chunks by size (pieces placed in isolation are size 1), without regard to the correctness of the placements. The distribution shows a clear positive skewness.

Insert Figure 4 about here

Subjects place a substantial number of single pieces. However, most pieces are placed within a chunk. The ratio of number of pieces belonging to a chunk to the number of pieces placed separately is 4.6 for M1, 5.8 for M3, 3.7 for E6 and 2 for A2. The reader should notice the presence of a few large chunks (the 17 piece chunk of A2 is particularly conspicuous[10]).

What is the distribution of chunks if we consider only the pieces placed correctly? The lower panel of Figure 4 illustrates the results for the same subjects as before. Now, only M1 places chunks of more than 10 pieces. In general, only 50% of all placements are correct (more about this in the section dealing with errors). Finally, the ratio, for the correct placements of pieces belonging to a chunk to isolated pieces is 5.9 for M1, 3.4 for M3, 3.8 for E6 and 1.9 for A2.

We return to the main concern of this section: the effect of the mirror image modifications on the size and number of chunks[11]. In the following analyses, chunks are defined as containing correct as well as incorrect pieces.

For the size of the *largest chunk*, no difference is to be noted between skill levels [$F(2,9)=1.44$, ns], nor any interaction [$F(6,27)=1.77$, ns]. There is, however, a significant effect of type of position: $F(3.27)=4.13$, $p< .02$. The largest chunks are bigger in the normal and horizontal conditions (means=7.4, 7.5, respectively) than in the vertical and central conditions (means=5.9 and 6.8, respectively). The largest chunks are particularly small for positions reflected about the vertical axis.

An ANOVA, performed on the log-transformation of the *number* of chunks per position, finds no main effect of Skill [$F(2,9)=1.72$, ns], but a main effect of the Type of modification [$F(3,27)=5.09$, $p<.01$]. For all skill levels together, the mean number of chunks per position is 4.3 for the normal positions, 4.0, 3.8 and 3.5 for the horizontal, vertical and central conditions respectively. Trend analysis indicates a linear component [$F(1,9)=11.85$, $p<.01$]. No interaction is found [$F(6,27)=1.22$, ns].

---

[10] This chunk is robust to changes in the cut-off criterion. Setting this criterion to 1.6 sec instead of 2 sec. still produces a chunk of 15 pieces.
[11] By definition, the minimal size of a chunk is 2 pieces.

Error analysis.

Few studies have analyzed in detail the errors occurring during the recall task. As far as we know, only Chase and Simon (1973b), Charness (1974) and Saariluoma (1991) have made such analyses. We will first classify the different kinds of errors, and then see if the mirror image reflections had any effect on the number and kinds of errors.

Following Chase and Simon, (1973b), we have divided errors in two main groups: errors of omission and errors of commission. The number of errors of omission is defined as the number of pieces in the stimulus position minus the number of pieces placed by the subject. The errors of commission, the pieces placed wrongly by the subject, are subdivided into 7 categories (the first 3 were used by Chase and Simon, 1973b): (a) Wrong piece: the piece is of correct color but of wrong kind; (b) Wrong color: the piece is of correct kind but of wrong color; (c) Close Translation: the piece is located within a distance of ±2 squares of a "correct piece"[12]; (d) Diagonal translation: the piece is located on one of the diagonals drawn from the square occupied by a correct piece; (e) Vertical or horizontal translation: the piece is located on one of the columns or rows drawn from the square occupied by a correct piece; (f) Symmetrical exchange: the location of two pieces has been swapped in comparison with the stimulus position; (g) Other: errors not explained by the above categories.

In the following analyses, we will not compare these categories with each other, because (a) they are not exclusive (e. g., a piece can be explained at the same time by a close translation and a diagonal translation) and (b) a piece can be explained several times within the same category[13] (e. g., a white pawn wrongly placed on d4 may be explained by horizontal translation of a white pawn on a4 and by horizontal translation of a white pawn on h4. Finally we will consider surplus pieces, i. e. cases where subjects place more pieces than were contained in the stimulus position.

Chase and Simon (1973b) found that most errors were omissions. The upper panel of Table 2 shows the mean number of omission errors, and the lower panel shows the mean number of commission errors in our data. Chase and Simon's results are replicated only for Class A players. Masters and Experts make more errors of commission than of omission (with the exception of Masters in vertical symmetry positions).

---

[12]In this definitions, "correct piece" is construed as a piece of the same kind and color in the stimulus position.
[13]In our statistics, all the possible explanations have been considered.

Insert Table 2 about here

As for errors of *omission*, skill levels do not differ statistically [$F(2,9)=1.84$, ns.] and no interaction is present [$F(6,27)=1.22$, ns]. A main effect of Type of modification is found [$F(3,27)=4.96$, $p<.01$]. Trend analysis shows a strong linear trend [$F(1,9)=46.32$, $p<.001$]. Although the patterns of means show that Masters perpetrate more errors of *commission* with positions modified by a reflection around the vertical axis, no main effect nor interaction is found for this variable [$F(2,9)=1.90$, ns; $F(3,27)=0.67$, ns and [$F(6,27)=0.21$, ns]. It is therefore reasonable to conclude that mirror image reflections affect mainly the number of pieces recalled, and not the number of errors of commission.

Most errors of commission are explained by Close Translation (a mean of 9.4 per position) and Vertical and Horizontal translation (mean=7.6). Other errors occur less often (the mean numbers of errors per position explained by Incorrect Type, Incorrect Color, Diagonal Translation and Symmetrical Exchange are, respectively, 1.4, 0.06, 2.2, and 0.1. The number of errors not explained by our definitions is very low (about 0.5 errors per position -- less than 2%). Hence, most placement errors conserve some information on color, type or location of the pieces. This result is partially accounted for by the limited space of the chess board, which implies a high redundancy.

No main effect of Skill or Type of modification is to be noted for these various types of errors. Finally, the number of boards reconstructed with an excess of pieces is small; among 60 positions for each experimental condition, an excess is present in 4, 7, 8 and 6 positions for the Normal, Horizontal, Vertical and Central conditions respectively. No difference is observable between the skill levels.

Note on typicality of positions.

As noted in the methods section, we have chosen to present the positions in the same order for all subjects, in order to allow fine comparative analyses. One consequence of this choice is that each experimental condition has only 5 different positions, making it possible that intrinsic characteristics of the positions, rather than the type of modification itself, play the crucial role. To control the typicality of the positions, we have asked external judges to rate them. The judges were 2 Grandmasters (ELO 2525 and 2465), 2 Masters (ELO 2250 and 2220) and one Class A player (ELO 1900). Typicality had to be rated on a 10-level scale, from 1 (atypical) to 10 (very typical). Table 4 presents the correlations obtained[14].

Insert Table 3 about here

The correlations between the best players (GM1, GM2 and M1) are very weak. This result is interesting in itself, because it suggests that the concept of typicality varies from one player to another, even when they are of nearly equal strength. It also makes it impossible to control more precisely the comparability of the positions belonging to the experimental conditions. Such a lack of agreement among our judges probably can be explained by the fact that, in our choice of positions, we have tended to select *atypical* positions; for example, we have as a rule discarded positions too close to standard opening positions. In experiment 2, we control typicality more directly.

Discussion

In this experiment, we have found that, for all skill levels, our subjects have somewhat more difficulty in recalling positions modified by vertical or central symmetry than positions modified by horizontal symmetry or unmodified positions. None of the modifications decreases the recall percentage to the level of random positions. No subject recognized the types of modification to which the positions had been subjected.

Chunk size analysis showed that the size of the largest chunk per position varies among the four experimental conditions: the largest chunks contain more pieces in the unmodified and horizontally modified conditions than in the others. The number of chunks (for all pieces as well as for correct pieces only) is reduced by the mirror iamage reflections. Finally, the number of omission errors is sensitive to the experimental manipulation, whereas number of errors of commission is not. We will discuss the theoretical implications of these results at the end of experiment 2.

Experiment 2

Experiment 1 has shed some new light on the organization of information in LTM and the effect on recall of location encoding. As the number of our subjects was small, it seemed to us necessary to replicate these findings. This experiment fulfills this goal.

[14]The correlations were computed using the positions of this experiment and the positions used in Experiment 1 of Gobet and Simon (1994a).

Methods

Subjects.

13 volunteers participated in this experiment. They were recruited from the Fribourg (Switzerland) Chess Club and from players participating in the Nova Park Zürich tournament, and were paid SFr 10.- (SFr 20.- for the players having a FIDE title). Ages ranged from 18 to 45, with a mean of 31.3 and a standard deviation of 8.6. ELO ratings ranged from 1680 to 2475. Subjects were assigned to 3 skill levels: Masters (n=2, mean ELO=2452), Experts (n=5, mean ELO=2122) and Class A players (n=6, mean ELO=1855). These subjects also participated in the copy task experiment reported in Gobet and Simon (1994b).

Materials.

We constructed three random positions following the method described in Experiment 1. We then selected sixteen positions from Lisitsin (1958), Wilson (1976), Reshevsky (1976), Euwe (1978), Moran (1989) and Smyslov (1972), with the same criteria as were used in Experiment 1. The mean number of pieces per position was 25. Four of these positions were presented without any modification, 4 each with a horizontal, vertical and central symmetry modification. Positions were randomly assigned to the four groups, in a different way for each subject, with the constraint that the mean number of pieces be 25±1. Positions were presented in a random order. Each subject thus received the positions in random order and with random assignment to type of modification.

Design and procedure.

Subjects received instruction on the goal of the experiment, and could familiarize themselves with the functioning of the program (just as described in experiment 1) and (if necessary) were instructed on how to use the mouse to reconstruct the positions[15]. They received then, in order, the control task (recall of random positions) and the mirror image reflection recall task. Each position was presented for 5 seconds, after which the screen was black for 5 sec. The subject had then to reconstruct the position. No feedback was given on the number of pieces correctly replaced.

[15]One Expert, who had difficulties in manipulating the mouse, used algebraic notation to dictate the positions to the E, who handled the mouse.

A factorial design 3x4 (Skill x Type of modification) was used, with repeated measurements on the Type of modification.

Results

Correlations between age and scores, or between time to perform the task and scores, were not significant. Hence we will not include these variables in the following analyses.

Percentage of pieces correct.

For random positions, Masters obtained a mean of 16% pieces placed correctly (sd=2.8), Experts a mean of 12% (sd=9.8) and Class A players a mean of 12.8% (sd=4.2). Expressed in number of pieces, we get 4 for the Masters and about 3 for the Experts and the Class A players. These differences are not significant [$F(2,10)=.24$, ns].

The effects of mirror image transformations are in the direction of our hypothesis (see Figure 5), but only approach significance. Analysis of variance, computed on the log-transformation of the percentages, shows an effect of Skill [$F(2,10)=18.21$, $p<.001$], and an effect of the Type of modification approaching the critical level [$F(3,30)=2.83$, $p=.055$]. Trend analysis indicates a linear component [$F(1,10)=5.61$, $p<.05$]. No interaction was found [$F(6,30)=0.76$, ns].

(Insert Figure 5 about here)

As the effect of Type of modification is not strong, we grouped the results by the presence or absence of a vertical symmetrical modification. Our reasoning is that this kind of modification has deeper effects on the "normalcy" of a position than a horizontal one: it is not uncommon (with possibly a tempo difference) to meet the same position for White and Black (a horizontal symmetry modification), but one rarely finds, in the chess literature, a position where the king and queen sides were inverted (a vertical symmetry modification). It is thus reasonable to expect a difference between these two groups.

For normal and horizontal modifications together, the mean percentages of correct answers are 90.5%, 47% and 31.8%, respectively, for Masters, Experts and Class A players. For vertical and central modifications together, the respective means are 79.2%, 35.4% and 24.9%, respectively. Computed on the grouped scores, the ANOVA indicates a main effect of Skill [$F(2,10)=17.61$, $p<.001$] and of Type of modification [$F(1,10)=9.85$, $p<.02$], as well as a lack of interaction [$F(2,10)=0.63$, ns]

Inter-piece latencies.

Analysis of variance, computed after log-transformation of the times, shows the absence of a main effect of Skill, [$F(2,10)=2.45$, ns.], a significant effect of Type of modification [$F(3,30)=6.62$, $p<.005$; linear trend: $F(1,10)=12.44$, $p<.01$], and no interaction [$F(6,30)=1.35$, ns.]. The means of the medians of the distributions as a function of Type of modification are 2.44 sec, 2.47 sec, 2.63 sec and 2.60 sec for Normal, Horizontal, Vertical and Central conditions, respectively. Transformations seem to have lengthened the latencies as compared with normal conditions. This effect is particularly clear if one groups Normal with Horizontal conditions, and Vertical with Central conditions.

Chunking.

In this section, we define a chunk as in the first experiment, i. e. a sequence of at least two pieces whose mean inter-piece (adjusted) latency is less than or equal to 2 sec. All analyses of this section have been done after log-transformation of the data. Looking at the *largest chunk* (pieces correct as well as pieces incorrect) per position, one finds an effect of Skill [$F(2,10)=5.75$, $p<.05$], no effect of Type of position and no interaction. The average of the largest chunk per position is 13 for Masters, 7.2 for Experts and 6.4 for Class A players. As for the mean *number* of chunks by experimental condition, one finds a main effect of Skill [$F(2,10)=8.41$, $p<.01$], but no main effect of Type of modification [$F(3,30)=0.01$, ns.] or interaction [$F(6,30)=0.39$, ns.]. The mean number of chunks per position is, pooling the 4 conditions, 3.5 (sd = 1.1) for Masters, 4.0 (sd = 1.6) for Experts and 1.9 (sd = 0.7) for Class A players.

Analysis of errors.

In this section, we will analyze errors in the same way as in the first experiment. The upper panel of Table 4 gives the mean number of errors of omission, and the lower panel the mean number of errors of commission. Comparison of the two panels indicates that Masters make more errors of commission than of omission, the numbers are about equal for Experts, while Class A players make more errors of omission than of commission.

For *errors of omission*, the ANOVA indicates a main effect of Skill [$F(2,10)=16.01$, $p<.001$], but no significant main effect of Type of modification [$F(3,30)=2.44$, $p=.084$]. No interaction is found [$F(6,30)=0.61$, ns]. As the effect of Type of modification is marginally significant, we have analyzed the data by pooling normal with horizontal positions and vertical with central positions, then comparing the two groups. The main effect of Skill remains [$F(2,10)=16.01$, $p<.001$], but this time a main effect of Type of Modification is present [$F(1,10)=12.15$, $p<.01$]. There is no interaction [$F(2,10)=0.97$, ns].

Insert Table 4 about here

As for *errors of commission*, one finds a main effect of Skill [$F(2,10)=5.74$, $p<.05$], no effect of Type of modification [$F(3,30)=0.89$, ns.] and no interaction [$F(6,30)=0.45$, ns.]. Results are similar when experimental conditions are pooled (Normal & Horizontal vs. Vertical & Central). Experts commit the most errors of commission. Note that the tendency for Masters to commit more errors of commission with Vertical and Central positions is not statistically reliable.

In a detailed analysis of errors of commission, the effect of Skill that we have just noted is found again for Incorrect Type, Local Translation, Diagonal Translation and Vertical or Horizontal Translation. No effect of Type of modification is found for these variables. Finally, no effect is found for the remaining errors, that is for incorrect Color, symmetric Exchange and Number of unexplained pieces. On average, the allocation of errors by types resembles that in Experiment 1 (Incorrect Type=1.0, Incorrect Color =.06, Local Translation=7.0, Diagonal Translation=1.6, Vertical or Horizontal Translation=5.6, Symmetric Exchange=.04). Again, the unexplained errors are few in number (.13 per position).

Discussion

This experiment replicates quite well the results found in Experiment 1: positions modified by mirror image reflection, particularly around a vertical axis, are harder to memorize than non-modified positions, but not as hard as random positions; inter-piece latencies are affected by manipulation; when the types of positions are pooled conditionally to the presence of reflection around vertical axis, we found that the modifications affect the number of errors of omission but not of commission; finally, contrary to results from Experiment 1, results from the present experiment did not show any effect of type of position on the size and number of chunks. These results suggest that at least a part of chess knowledge is encoded by keeping information about the precise location of the pieces. Conceptual knowledge of characteristic relations between pieces does not explain the ability of players to recall positions, an ability that also depends on perceptual knowledge of specific chunks that describe pieces at specific locations and is sensitive to small changes in location. Chase and Simon's (1973b) theory offers, at least on this point, a plausible explanation of the processes involved.

The deterioration of the subjects' performance with mirror image reflections of the positions, taken together with Saariluoma's (1991) results, throws doubt on Holding's (1985, p.109) hypothesis that chunks may be recognized independently of the colors of the pieces or their locations on the board. It appears that chess information, or at least much of it, encodes both the color and the precise location of the pieces. That the importance of color encoding decreases as expertise increases can be explained by masters having a larger repertory of chunks, learning rarer chunks as well as more common ones.

As transformation of the positions did affect the number of errors of omission, it appears, having in mind the Chase-Simon model, that, contrary to Holding's supposition, the modifications reduce the number of chunks on the modified boards that are recognizable as familiar patterns already stored in LTM. Since Holding's calculation of the number of chunks required in LTM to attain nearly perfect recall is based on a contrafactual supposition, we must reject his conclusion that the number is much smaller than the 50,000 estimated by Chase and Simon.

Random Positions: a second analysis

Although the random positions in Experiment 1 and 2 were used primarily as a control task, it is instructive to investigate in some detail the behavior of our subjects with this material. First, the literature does not offer very much information on this topic; second, recall of positions that are theoretically devoid of any structure may offer useful information on the way skilled players cope with material that does not allow them to apply their chess knowledge. Implicitly, Chase and Simon's model predicts that, in the case of random positions, each slot in STM is occupied by a symbol pointing to a single piece, perhaps occasionally to two pieces for Masters, who may be able to recognize some uncommon patterns.

As shown above, no significant difference was found among the three skill levels in the recall of random positions. Almost all published results show the same pattern: the best players recall slightly more pieces than weaker players, but this difference is not statistically significant (for example Saariluoma, 1984, Reynolds, 1982; Chase and Simon's (1973a) Master, who obtained even worse results than the Expert and the Class A player, is a conspicuous exception).

Experiment 3

In this experiment, we analyze in further detail the data on random positions from Experiment 1, as compared with game positions. For details on the Subjects, the Material and the Methods, see the corresponding sections of Experiment 1.

Results

Chunks.

Two different patterns can be observed for this variable. Most of the subjects (8 out of 12, including all Masters), are very circumspect and rarely place more than 4 chunks or single pieces in a given position. The remaining 4 subjects are more audacious and replace more pieces, although their results in correct pieces are not better. Table 5 shows the behavior of 4 subjects, the same ones as in our previous discussions, in the recall of random positions. M1, M3 and A2 belong to the group of cautious subjects, E6 to the more audacious.

Insert Table 5 about here

As numerous pieces are replaced erroneously, it is difficult to separate the roles of memorization and guessing. We will therefore focus on correct pieces. Isolated correct pieces are more the exception than the rule (25% for the 4 subjects); with the exception of M3, these subjects place some large chunks (6 and 4 pieces for M1, 5 and 4 pieces for E6 and A2).

One can classify the contents of chunks grossly in three categories: 1) constellations that are plausible in a game; 2) a King surrounded by one or more pieces; 3) a geometrical pattern. Reconstructions of the positions #2 and #13, given in Figure 6, by M1, M3, E6 and A6, illustrate this.

Insert Figure 6 about here

Insert Table 6 about here

The first chunk of M1 (Bc4-Nd3-Re7-Qe3) is an example of the first category. Also for M1, the first chunk of the position #13 is an interesting example of the second category: all pieces of this chunk (Qf7-Rd6-Bd5) converge on the square e6, where the black King is located in the stimulus position. The black King itself is absent in the reconstruction. Finally, the pawns b2-a1-b1-c1 in position #13 constitute a striking geometrical figure, and appear in various fragments in the four subjects' reconstructions.

When no attention is paid to the correctness of pieces building up a chunk, the pattern of means of the largest chunk shows that the larger chunks are bigger for game positions than for random positions (means with game positions: 10.3, 7 and 5.3 pieces for Master, Experts and class A players, respectively; means with random positions: 3.3, 4.3 and 3.7 pieces; $F(1,9)=49.54$, $p<.001$). There is also an interaction between skill and type of position [$F(2,9)=8.20$, $p<.01$]: while the size of the largest chunk correlates with Skill level in game positions, it stays roughly constant in random positions. The main effect of Skill is not significant [$F(2,9)=1.31$, ns]. For the number of chunks, there is a clear effect of Type of position [$F(1,9)= 90.33$, $p < .001$], but no effect of Skill and no interaction. The mean number of chunks, for all subjects pooled, is 4.3 for game positions and 1.7 for random positions.

Errors.

As was to be expected, the number of errors of omission is high for random positions (mean for normal positions: 4.4; mean for random positions: 16.3). The difference is highly significant: $F(1,9)=112.47$, $p<.001$. No difference is found between Skill levels[$F(2,9)=1.20$, ns], nor is there an interaction [$F(2,9)=3.00$, ns]

We have seen that the type of modification (in the mirror image reflection experiment) had no effect on the errors of commission. What about the random positions, which can be considered as extreme modifications of normal positions? Although Experts tend to commit more error by commission than Masters and Class A players both in game and in random positions, the effect is not significant [$F(2,9)=2.53$, ns]. No statistically significant difference is to be noted in the mean number of errors by commission between game (n=6.8) and random positions (n= 5.3), nor any interaction is present. Table 7 offers an overview of the types of errors of commission encountered.

(Insert Table 7 about here)

There is a significant effect of the Type of position for three kinds of errors: Incorrect Type [$F(1,9)=7.84$, $p<.03$], Vertical or Horizontal Translation [$F(1,9)=12.79$, $p<.01$] and Local Translation [$F(1,9)=10.31$, $p<.02$]. No other effect is significant at the .05 level.

Latencies.

As compared with normal game positions, interpiece latency time is not affected by the randomness of the positions [$F(2,9)=0.15$, ns; $F(1,9)=4.50$, ns; $F(2,9)=2.31$, ns]. The median latency averages 3.0 sec for game positions, and 3.1 sec for random positions.

Discussion.

We have seen that there is an important difference in the size of the largest chunk recalled between the recall of random and game positions, respectively. However, a striking feature of our analysis of the recall of random positions is the presence of numerous chunks, occasionally as large as 6 or 7 pieces, with stimuli supposed to be devoid of any structure. Some of the chunks in random positions may occur in normal games, and their recall may therefore be explained by an access to LTM. However, such an explanation does not hold for all chunks. It is plausible that subjects prefer to construct a small number of new chunks rather than to memorize a list of pieces without any relation among them. It is unclear whether this strategy is the most efficacious (scores are less than what Chase and Simon's (1973b) model would predict), but the approach is probably close enough to the normal activity of chess players.

Three error types differ in the recall of normal as opposed to random positions: Incorrect Type, Local Translation and Vertical or Horizontal Translation. It is an interesting question whether these errors are the product of inference mechanisms or of memory processes.

Experiment 4

In the following analyses, results on the recall of random positions and game positions have been pooled from Experiment 2 of this paper and Experiment 2 of Gobet and Simon (1994a). This pooling seems reasonable for the random positions, as all subjects received them at the same point in their experimental sessions (after the copying task). The temporal order for the normal positions was, however, slightly different in the two experiments: in experiment 2 of the present paper, they were presented between the modified positions, whereas in experiment 2 of Gobet and Simon (1994a), they were presented grouped as one-position trials. Our pooling seems however justified, as no significant difference was noted between the two samples for any of the variables examined in this section.

Methods

Subjects.

26 male subjects participated in this experiment. The ELO ratings ranged from 1680 to 2510, with a mean of 2078 and a standard deviation of 232.7. Subjects were divided 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 Methods.

The game positions used in this experiment come from the same sources as the ones described in the Section Materials of the second experiment of this paper. The random positions were created by re-assigning, using random numbers, pieces from game positions to new squares .

Results and discussion

Percentage of recall.

With game positions, Masters recalled 92%, Experts 57.1% and Class A players 32.2%. With random positions, the percentages of recall were, respectively, 19%, 13.8% and 12.4%. The analysis of variance yields the standard outcome of this type of experiment: a main effect of Skill [$F(2,23) = 44.41$, $p < .001$], a main effect of Type of position [$F(1,23) = 309.21$, $p < .001$] , and an interaction [$F(2,23) = 34.17$, $p < .001$].

Age.

Using Age as a predictor for the results of random and game position, we obtained the following regression lines: Random = 21.75 - .26 * Age ($r^2 = .127$, $p = .07$) and Game = 47.24 + .17 * Age ($r^2 = .003$, $p > .10$). Age then does not play any important role, and will be left out in the remaining analyses. That we did not find Charness' (1981) results (negative effect of age) may be explained by the small age range of our sample and by the fact that our oldest subject was only 49 years old.

Chunks.

For all chunks, the *largest chunk* per position offers clear-cut results. There is an effect of Skill [$F(2,23)=6.92$, $p<.005$], of Type of position [$F(1,23)= 62.57$, $p<.001$], and an interaction [$F(2,23)=9.42$, $p=.001$]. Masters produce larger chunks with game positions than Experts or Class A players (respective means: 16.8, 11.9 and 6.6 pieces). With random positions, the largest chunk does not differ significantly between groups (in average, 4.6 pieces). For the mean *number* of chunks per position, one finds no effect of Skill [$F(2,23)=2.76$, ns], a very significant effect of Type of position [$F(1,23)=68.77$, $p<.001$], and no interaction [$F(2,10)=1.65$, ns] (see Figure 7).

Insert Figure 7 about here

Errors.

As expected, there is an important difference in the number of error of *omission* between the normal and random positions [$F(1,23)=256.05$, $p<.001$]. Analysis of variance indicates also a main effect of Skill [$F(2,23)=20.42$, $p<.001$], as well as an interaction [$F(2,23)=34.87$, $p<.001$]. The number of errors of omission in random positons is high for all skill levels, and varies as a function of the skill level for the normal positions (see Table 8, upper panel).

Insert Table 8 about here

For the errors of *commission* (see Table 8, lower panel), one finds a main effect of Skill [$F(2,23)=3.71$, $p<.05$]. The main effects of Type of position [$F(1,23)=.017$] and interaction [$F(2,23)=2.83$] are not significant. As was the case for positions modified by mirror image, Experts commit the largest number of errors of commission.

Analysis of variance of the errors of commission shows a main effect of Skill for the following variables: Local Translation [$F(2,23)=3.62$, $p<.05$], Diagonal Translation [$F(2,23)=5.12$, $p<.05$], Non-explained Errors [$F(2,23)=3.48$, $p<.05$]; the main effect of Type of position is significant for Incorrect Color [$F(1,23)=6.94$, $p<.02$], Local Translation [$F(1,23)=11.34$, $p<.005$] and Vertical or Horizontal Translation [$F(1,23)=.001$]; interaction is found for Local Translation [$F(2,23)=3.69$, $p<.05$] and Vertical or Horizontal Translation [$F(2,23)=4.61$, $p<.05$]. In all these experimental conditions, Masters commit the fewest errors in every type of position. Finally, all skill levels being grouped, errors of commission occur more often in normal positions than in random positions, with the exception of Incorrect Color.

Inter-piece latency.

Analysis of variance of means of log-transformed times shows a very significant main effect of Type of position [$F(1,23)=25.75$, $p<.001$]; but the main effects of Skill [$F(2,23)=0.06$, ns] and interaction [$F(2,23)=0.05$, ns] are not significant. With game positions, the median latencies are, on average, 2.13 sec for Masters, 2.53 sec for Experts and 2.40 sec for Class A players. With random positions, the respective median latencies are 2.62, 3.04 and 2.96 sec. Players' latencies are about one-half second shorter in game positions than in random positions.

Discussion

In interpreting the results presented in this section, it is necessary to remember their *post-hoc* character. As a consequence, the order of presentation of normal and random positions has not been controlled, which suggests some caution in generalizing the findings.

Chunk analysis has shown that randomizing positions affects the number of chunks and the size of the largest chunk. As for errors, the two types of positions show different numbers of errors of omission, but not of commission. Within the errors of commission, subjects commit more (about twice as many) errors by close translation and by horizontal and vertical translation in game positions than in random positions. Finally, the results indicate that inter-piece latencies are longer with random than with game positions.

We have also shown that, contrary to the predictions of the Chase and Simon model (1973b), pieces are chunked in the recall of random positions (cf. discussion of size of largest chunk). It seems, then, plausible that some learning occurs, even during this short presentation time. An alternative explanation, proposed by Gobet and Simon (1994a), is that chessplayers use slots in STM to store descriptions of patterns on the board (e.g.,"| Three | White pawns | on a diagonal | starting from a1 |".) This question of learning in random positions will be taken up again in our discussion of the next experiment.

Recall of game and random positions as a function of presentation time

While random positions have been used mainly as a control of subjects' "general" memory capacity, some researchers have studied them for their own stake. We have mentioned in the introduction the use of semi-random positions by Holding and Reynolds (1982) to study problem solving in chess. In another study, Reynolds (1982) has shown that several degrees of randomness may be obtained by manipulating the quantity of control that the pieces have on the center.

Some attempts have been made to study the role of presentation time for game positions (Saariluoma, 1984; Lories, 1987). It has been shown, not surprisingly, that increase in presentation time facilitates recall. It is however difficult to estimate precise functional relations between presentation time and recall performance, because these studies have been limited to a few data points (at most four; cf. Saariluoma, 1984).

Data are even scarcer for random positions. Djakov and al. (1927) presented for 1 minute a "random" position to masters and to subjects in a non-chessplaying control group, and found that masters' recall was better than control group subjects'. Two difficulties hamper the interpretation of this study. First, the subjects of the control group did not play chess at all, which brings about the question of familiarity with the material. Second, the position was a chess *problem*. Chess problems constitute a branch of chess where the first goal is to construct esthetic positions and combinations. Although this species of chess is quite different from normal chess games, it produces positions that are far from random. Lories (1987) found an effect of skill with the one-minute presentation of the semi-random positions generated by Holding and Reynolds (1982). Finally, Saariluoma (1989) used a procedure similar to Chase and Ericsson's (1982) for the memory of digits, dictating positions at the pace of 2 or 4 seconds per piece. He found that strong players are better both in the recall of game and random positions.

We think that a systematic manipulation of the presentation time for game as well as random positions may shed useful light on the chunking processes in chess, and may offer important data for theorizing about skilled memory in chess. This next experiment aims at gathering and analysing such data.

Experiment 5

Methods

Subjects.

21 subjects participated in this experiment, with ratings ranging from 2595 to 1770[16]. One subject, rated at 2345, quit the experiment after about 15 minutes, complaining about concentration problems. The 20 remaining subjects were assigned to three skill levels: Masters (n=5, mean=2498, sd= 86.6), Experts (n=8, mean=2121, sd=100.8) and Class A players (n=7, mean=1879, sd= 69.8). The mean age was 32.9 years (sd = 11.6). The youngest subject was 20 years, the oldest 70 years.

[16]When possible, the International rating was used. For Swiss players without international rating, the Swiss rating was used. Finally, four American subjects were given a corrected rating: American rating - 50.

Subjects were recruited at the Fribourg (Switzerland) chess club, during the Biel Festival and in the CMU community. They were paid 10 SFr (20 SFr for players having a FIDE title) for their participation.

Materials.

19 positions were selected from Bronstein (1979), Euwe (1978), Lisitsin (1958). Moran (1989), Reshevsky (1976), Smyslov (1972) and Wilson (1976) with the same criteria as the positions used in experiment 1. 10 random positions were created as described in Experiment 1.

Design.

Subjects were first familiarized with the computer display and shown how to select and place pieces on the board. They then received 2 warm-up positions (1 game and 1 random position) presented for 5 sec each.

For each duration (1, 2, 3, 4, 5, 10, 20, 30, 60 sec), two game positions and one random position were presented, with the exception that Masters did not received the game positions with presentation times over 10 seconds, for they were expected to reach nearly perfect performance by that time. The presentation times were incremented from 1 second to 60 seconds for half of the subjects, and decreased from 60 seconds to 1 second for the others.

Results

Percentage of correct pieces.

1) Game positions[17]. The upper panel of Figure 8 shows the performance expressed as percentage of pieces replaced correctly. The Masters' superiority is obvious. In 1 sec, they are at about the same level of performance as Experts after 10 sec and perform only slightly worse than Class A players after 30 sec. Note also that, while Class A players and Experts improve their scores monotonically, Masters approach a ceiling rapidly, after about 2 sec. All the differences mentioned are highly significant statistically.

Insert Figure 8 about here

----

[17]One Class A player refused to recall positions (game or random) below 5 seconds. One Grandmaster refused to recall any random position. Their (partial) results are included in our analysis.

How do Masters achieve their superiority? Do they have only a perceptual advantage, already evidenced at short presentation times and produced by the availability of more and bigger chunks in STM, or are they also able to profit from the supplementary presentation time to encode information into LTM (a "learning" advantage)? To answer these questions, we have fitted some simple functions to the data. Analysis of the parameters of these functions can shed light on these questions. A power law (average r^2 for Experts and Class A players[18] = .67) and a logarithmic function (average r^2 = .65) fit the data reasonably well, better than a simple linear regression line (average r^2 = .58). However, the best fit was provided by a growth function,

$$P = 100 - Be^{-c(t-1)} \tag{1}$$

where P is the percentage of correct answers, (100-B) is the percentage memorized in 1 second, c a constant, and t the presentation time, in seconds. The average r^2 for experts and Class A players is .69.

This function supposes that the rate at which additional pieces are stored after one second is proportional to the number of pieces not already retained. Table 9 gives the parameters fitting the data best, for the three skill levels. One sees that both (100-B), the percentage learned in 1 sec, and the subsequent learning rate, c, increase with skill. The parameter c doubles from class A to Experts, and increases by a factor of 15 from class A to Masters. The parameter (100-B) increases by a factor of 1.3 from class A to Experts, and a factor of nearly 3 from class A to Masters.

Note however that the function does not account as well for the Masters' results as for the others', the reason being that the Masters have a relatively wide spread of scores at 10 sec, where they get an average of 92.4 %. As a matter of fact, two Masters performed at about 85%, while the three others were above 96%. Using only the data for the latter three players; we get a better fit (r^2 = .54), with B = 24.56 and c = 0.72 (the latter, more than 20 times the rate achieved by Class A players).

Insert Table 9 about here

<u>2) Random positions.</u> We have fitted the data obtained with the random positions with the same type of growth function (Figure 8, lower panel). The best fitting parameters are

[18]Because of the ceiling effect shown by the Masters, these functions do poorly with this skill group.

presented in Table 9, lower panel. Again, we find that skill levels differ both in the amount of information acquired quickly (100-B) and in the continuing rate of acquisition (c). The parameter c doubles from class A to Experts, and triples from class A to Masters. The parameter (100-B) increases by a factor of 1.5 from class A to Experts, and doubles from class A to Masters. Notice that the *relative* superiority after 1 second of Masters' over Experts and Class A players is almost the same for game positions and random positions; but the superiority of Masters in learning rate (percentage of remaining pieces learned) for longer intervals is much greater for game positions than for random positions.

These results indicate that strong players achieve higher percentages of recall both because (a) they perceive and encode a larger amount of information in the first 1 second of exposure and (b) they improve thereafter at a faster rate, because they recognize more and bigger chunks. Both the B and c parameters are higher for game than for random positions. For Experts and Class A players, c is about six times as large for game than for random positions, while (100-B) is a little more than 2 times as large. For Masters the ratio for c is nearly 25, and the ratio for (100-B) about 3.5, which confirms the view that chess Masters use their experience of the domain for recalling meaningful board positions.

Chunks.

We have analyzed some results from the distributions of chunks. As in the previous sections, chunks are defined as pieces whose successive corrected latency is less than 2 seconds. Pieces placed individually are not considered as chunks.

1) Game positions. Our first variable is the size of the largest chunk per position. Figure 9 shows that Masters reconstruct large chunks, even with presentation time of one second, and that the maximal size of chunks does not increase much with additional presentation time. Experts' largest chunks start with about 9-10 pieces, and increase logarithmically up to about 17 pieces at 60 seconds. A comparable increase may also been seen with class A players, who start, however, with smaller maximal chunks (about 4-5 pieces).

Insert Figure 9 about here

The second variable, the number of chunks, shows two different patterns, the first one displayed by Masters and Experts, the second one by Class A players. Masters' number of chunks increases from 2.9, at one second to 4.2 at four seconds, and then decreases to 3.3 at 10 seconds. Experts start with 2.2 chunks at a presentation time of 1 second, show a maximum with 4.5 chunks at 10 seconds, and then decrease to 2.9 at 60 seconds. Class A players increase the number of chunks from 2.2 at 1 sec to 5.4 at 60 seconds.

2) Random positions. For the *largest chunks*, we see (a) that all groups increase in a logarithmic fashion with additional time and (b) that the stronger players have larger maximal chunks than weaker players (see Figure 9). The standard deviation is high within each skill level (average sd = 2.00, 2.28 and 2.95 pieces for Masters, Experts and class A players, respectively). For all skill groups, the *number* of chunks increases as a function of presentation time. The number of chunks with a presentation time of 1 sec and 60 sec is 1.7 and 3.3, respectively, for Masters, 1.4 and 4.5 for Experts, and 1.2 and 4.3 for Class A players. The Skill levels do not differ significantly over the 9 presentation times [$F(2,12) = 0.28$, ns].

In summary, data on chunks in random positions show that stronger players place larger chunks and that the size of the largest chunk tends to increase with additional time. The number of chunks increases also with additional time, but there is no difference in number due to chess skill.

Errors.

1) Game positions.

The number of errors of omission correlates (negatively) with the presentation time ($r = -0.43$, $p<.001$). With 1 sec, Master miss 6.1 pieces by omission; with longer presentation times, they place the same number of pieces as in the stimulus position. For Experts, the number of errors of omission is 12.6 with 1 sec, then decreases exponentially to end up close to zero with 60 sec. Finally, Class A players commit 16.5 errors of omission with 1 sec and close to zero with 60 sec.

The three skill levels present different patterns for errors of commission. Masters make fewer than 3 errors from 1 sec to 4 sec, and then reduce this number to an average of 1.25 with 5 and 10 sec. For Experts the numbers of errors of commission decrease more or less exponentially from 5.5 with 1 sec to 0.5 with 60 sec. Finally, the number of error of commission is constant for Class A players from 1 second to 30 seconds (an average of 4.5). Even with 60 seconds, Class A players commit on average 2.8 commission errors.

2) Random positions.

For all skill levels, the number of errors of omission is high with short presentation times (21.4 pieces, on average, with 1 sec), and decreases exponentially with longer presentation time. Masters tend to produce fewer errors of omission than Experts, and Experts fewer than Class A players. These differences are larger with longer presentation times. For example, the numbers of errors of omission at 10 sec are 14.7, 15.1 and 16.5 for Master, Experts and Class A players, respectively, while the corresponding numbers at 60 sec are 4.5, 6.6 and 9.0. The negative correlation of errors of omission with percent correct is strong ($r = -0.74$, $p<.001$).

All skill levels make more errors of commission with longer presentation times. In general, Masters tend to make fewer errors of commission than Experts and Class A players, though the difference is not statistically significant. With 1 sec, the average number of errors of commission is 2 pieces; with 60 sec, it is 5.5 pieces.

Discussion

Chase and Simon (1973,a) found that, for game positions, stronger players place both larger chunks and more chunks (remaining within the 7±2 item size of the STM). Our data replicate this finding, with the qualification that the difference in number of chunks tends to disappear with long presentation times. With random positions, we found that strong players placed both larger chunks and more chunks.

Saariluoma (1989), using auditory presentation, the pieces being presented at a rate of 1 piece per 2 seconds, found that stronger players achieve a better recall of random positions. Lories (1987) found the same effect of skill with semi-random positions and visual presentation for 1 minute. Both results indicate that skilled players have better memory for random positions than weaker players when presentation time is sufficiently long. Our results confirm these findings. A surprising finding, however, is a difference in recall for random positions with rapid presentation.

We note first that this difference is small (about 20% between Masters and Class A players at a presentation time of five seconds) in comparison with the difference for game positions (60% percent in the data of this experiment). But still, there is a difference. To be sure that our data do not represent an anomaly, we have compiled in Table 10 the percentage of recall from various experiments in the literature that used recall of random positions as a control condition.

Insert Table 10 about here

We see that, with the exception of Chase and Simon's results (1973a), stronger players do remember somewhat more pieces in random positions. That these differences were in most cases not statistically significant may be explained both by the small size of the effect and the small number of subjects used in these experiments.

It is not surprising that the growth curve fits our data pretty well, for this type of curve has been shown to fit a wide variety of learning and memory data (Lewis, 1960). The difference of intercept between players at different Skill levels (the B parameter) is predicted by the Chase-Simon theory: strong players recognize larger chunks more rapidly, and is confirmed by our data on the size of the largest chunk. This difference means that strong players not only perceive more during the first seconds of exposition, but that they also either recognize more patterns or learn more patterns afterwards. The differential rate of change for random positions suggests that strong players are also able to *learn* new chunks more rapidly. One explanation of this state of affairs is that strong players possess a better representation of the chess board, accessing more rapidly the spatial description of a given square. This may give them an edge when chunking new bits of information, but does not explain how STM limits are overcome. In general, the between-skill difference in the rate of improvement with additional time contradicts Chase and Simon's theory, in particular with presentation times shorter than 10 sec, because it is not clear where the additional pieces or chunks are stored.

A possible explanation may be that, with practice in their domain, chess players develop a retrieval structure (cf. Chase and Ericsson, 1982), perhaps a schematic representation of the 64 squares of the board, that allows them to encode relatively rapidly the location of the pieces. When a game position is recognized (say, as a Tarrasch defense with White attack on the King's side), this default-value representation of the chess-board would be superseded by a much more precise representation, containing already instantiated information (for example the location of a few pieces). This accessibility to more powerful retrieval structures would explain why recall is easier for masters with game positions than with random positions (see Gobet and Simon, 1994,a, for a more detailed discussion of chess templates/schemas). We will say more about this theory in the next section.

General Discussion

In these experiments, we have examined three main phenomena: (a) differences between memory for normal game positions and for positions modified by reflection around an axis: horizontal, vertical, or both; (b) differences between memory for chess boards sampled from game positions and boards on which the same pieces are placed at random; (c) effect of presentation time on recall of game and random positions.

The experiments on boards modified by reflections around axes of symmetry were aimed at testing the claims of Holding (1985) that Simon and Gilmartin (1973) had overestimated the number of familiar chunks a player would have to hold in LTM to reconstruct a board. If a chunk were recognizable independently of the color of the pieces composing it and independently of its location on the board, then the same pattern, modified by change of color or location, would have to be represented only once in memory, and the total number of different patterns stored would be correspondingly reduced. The results of our experiments with modified boards call Holding's conclusions into question. Modifying the boards by reflection (hence altering the colors and positions of chunks) did decrease the number of pieces recalled, different degrees of modification producing different degrees of deficiency. The decrease in recall caused by reflections shows that the same chunks cannot be evoked to encode a group of pieces when the location of the group is altered. The effect was small however when only colors were swapped (reflection about the horizontal axis).

In the introduction of Experiment 1, we proposed a table illustrating the effect of various types of position distortions on the recall of chess positions. As our Experiments 1 and 2 allowed us to fill the missing cell, we can now present a complete table (see Table 11).

Insert Table 11 about here

Taken with Saariluoma's (1991) results, who used translation to modify his positions, these data lead us to conclude that the estimate of Simon and Gilmartin, that Grandmasters hold at least 50,000 familiar chunks in memory, is not excessive.

Our findings comparing recall for random versus normal positions replicate the findings of previous experiments. The substantial superiority in recall of high-rated over low-rated players that appears regularly when normal game positions are used as stimuli nearly disappears when random positions are used with a 5 second presentation time. The observed effects of presentation time on recall sharpen our understanding of the relation between game and random positions. First, Masters reach a very high level of recall after only one second when positions are taken from plausible games. Second, Masters keep a small but consistent superiority with random positions when the presentation time is short. This superiority increases with longer presentation times (up to one minute). Third, the same mathematical function accounts for the recall of random and game positions, apart from the ceiling effect for Masters recalling game positions. The parameters of the function indicate that strong players get better results in game as well as in random positions both because they perceive more at short presentation times and because they profit more from additional presentation time.

Finally, we found in Experiment 1 a low correlation among the typicality judgments given by five skilled players. This result, although obtained with a small sample, is interesting, as typicality has been often proposed as an important variable in chess knowledge (Goldin, 1978; Holding, 1985). If this result turns out to be robust, it would suggest that typicality varies from player to player, perhaps as a function of his or her style or of the openings he or she is playing.

Templates in Chess

The results obtained with the interfering task paradigm (Charness, 1974) and the recall of multiple boards paradigm (Cooke & al. 1993; Gobet & Simon, 1994a) have uncovered several weaknesses with the original formulation of the Chase and Simon (1973b) theory of chess expertise. We have proposed elsewhere (Gobet & Simon, 1994a) a theory that accounts for these results as well as for the phenomena explained by the earlier Chase and Simon theory. We believe that the principal findings in the experiments presented in this paper are compatible with the revised formulation of the theory. Before discussing how well the new theory squares with the data on mirror error reflection, recall of random positions and on the role of presentation time, we first summarize its main features.

Our modification of the Chase and Simon theory still takes chunking as the main mechanism by which chessplayers store information about the chess positions they have encountered in their games or have studied. In addition, it postulates that Experts and especially Masters use a mechanism that was earlier proposed to explain expert memory for lengthy number sequences (Chase & Ericsson, 1982). It hypothesizes that, in addition to modest-sized chunks of pieces, expert players also store in long-term memory *templates* of more or less "typical" positions corresponding mostly to different openings and variations, around the 15th to 20th move.

A Master could be expected to remember several thousand such templates, stored in a format that contains definite information about the location of, say, about 10 to 12 pieces. This fixed information in the template will be called its *core*. The templates also contain *slots*, serving as variables, where additional information can be inserted relatively quickly about any specific position belonging to the type represented by the template; say, information about three or four chunks of pieces. The slots could be of two kinds: slots containing default values that can be overridden by contrary information perceived on the current board, and simple variables whose values are undefined unless a chunk is recognized and inserted. Hence, what an expert can remember about a briefly-seen position is both the information contained in the fixed part of the template (information common to positions of this type), and also the information contained in a number of chunks that have replaced variable slots (default values or blanks) in the template. The template is evoked by recognizing the position as being of a certain familiar type.

The addition of templates to the simple chunking theory removes several criticisms of that theory: that it underplays the role of LTM in the recall task, and that the sizes of chunks were underestimated, perhaps because subjects had to hold all the chunk's pieces in their hands while placing them within the limit of the one-second interval between consecutive pieces (see Gobet and Simon, 1994b, for a detailed discussion on this question). The novelty of the template theory, in comparison with schema-theories proposed by Bartlett (1932) or Rumelhart & Norman (1981), is that the template acts as a *retrieval structure* (Chase & Ericsson, 1982). That is, variable values may be instantiated in a fraction of a second as has been shown to be the case with the retrieval structures employed by Chase and Ericsson and Staszewski to explain expert memory in recalling long digit strings. Learned retrieval structures with "slots" allow experts to encode information of specific kinds relating to their expertise rapidly into LTM.

Because templates are complex data structures, it takes a long time (of the order of several hours) to learn one. We therefore expect class A or weaker players to have few of them; Experts to have them only in some situations (perhaps for the positions occurring in the openings belonging to their repertoire), and professional players to have several thousand, even for types of positions they don't meet in their own tournament practice. When presented with a game position for a few seconds, a Master will first recognize a few chunks, which may evoke a template. During the next few seconds after retrieval of the template, default values may be corrected and then other slots instantiated. We will now apply this theory to the empirical findings we have presented in this paper.

Reflected positions

In the recall of positions modified by mirror image reflection, the template theory predicts, as Chase and Simon's does, that unmodified positions will be better recalled than reflected positions, the latter being likely to evoke fewer and smaller chunks and templates. Because the presence of free-variable slots may, however, allow Experts and Masters to overcome part of the deficiency, the template theory predicts that these modified positions will be relatively harder for weak players than for strong ones. In both experiments 1 and 2, this prediction is supported: the ratio of percentage correct for normal positions to the average percentage correct for the three other conditions is greater for weak players than for strong players (Experiment 1 : Masters, 1.18; Experts, 1.23; Class A 1.35; Experiment 2 : Masters, 1.04; Experts, 1.41; Class A, 1.30). The theory also predicts more errors of commission for modified positions. This was found only for Masters (in both experiments), perhaps because the other groups did not use templates often. Finally, it predicts that the largest chunks (corresponding to the template cores) will be bigger for unmodified positions than for modified positions. This prediction was supported in the first experiment, with the qualification that the differences between conditions were small, but not in the second.

Random positions

Our explanation for the differential recall of game and random positions is basically the same as the one proposed by Chase and Simon: the skillful players' superior performances depend on their recognizing familiar patterns of pieces, templates and chunks, in the game positions; the near-absence of these chunks from the boards with randomly placed pieces reduces this advantage. In particular, usually no template is accessible.

How explain, then, the Masters' superiority over weaker players for random positions, given enough time? It is unlikely that they find an appropriate template, as their recall is low even after 60 seconds (about 16 pieces). However, assuming 8 seconds to chunk two objects (Newell and Simon, 1972), about seven new chunks may be formed in 60 seconds, enough to explain Masters' performance with random positions. Yet, this does not explain the large chunks (4-5 pieces) recalled with presentation less than 10 seconds. Except for the hypothesis advanced in the discussion of Experiment 4, that players may use STM slots, not to encode pointers to chunks in LTM, but to encode descriptions of one single pattern (e.g. "Three | black | pawns | forming a triangle | with top on d4" would use 5 slots in STM), we have at present no explanation for this finding.

Presentation time

The duration of the presentation will determine: (a) accessibility of the template (core + uncorrected default values); this occurs in the same latency range as recognition processes (hundreds of msec); (b) ability to correct default values and instantiate other slots (a few seconds); (c) ability to elaborate the template (at least several minutes, possibly hours). On the basis of the assumed parameter values, we expect the following results for Masters, who are the most likely to possess many templates and chunks: the size of the largest chunks (corresponding to the templates) will increase during the first seconds and then will stay more or less constant. This prediction is supported by our data; the size of Masters' largest chunks increases slightly between presentation times of 1 second (13 pieces) and 4 seconds (16.5 pieces), and then seems to stay constant (16.1 pieces with 10 sec). Combining this with the nearly constant mean number of chunks, for Masters, for times less than 10 seconds (2.9 with 1 sec, 4.2 with 4 sec and 3.3 with 10 sec), we speculate that 1 sec does not allow additional chunks to be inserted in the template, but that a few additional seconds do.

Chunks and Templates: Compiled or Interpreted?

The first two experiments addressed the role of location encoding in chess knowledge. Our results, converging with Saariluoma's (1991), support the hypothesis that location is encoded. Why encode "white Knight on f6, black King on g8, black Queen on d7," when it would be sufficient to encode the relation as "Knight fork of King and Queen," and fill in the location for each occurrence? An answer is that it is more efficient to store the specific chunks, for chunks encoding location are recognized faster and easier than general-relation chunks, which require extra time for interpretation and instantiation. Chess masters surely possess some variable-free chunks (concepts like "fork" show that they do); but they also hold in memory many quite specific compiled chunks that allow a faster access to LTM information.

Chase and Simon proposed that, when a pattern is recognized it may suggest a move. Casual evidence shows that, with masters, some patterns elicit a precise move (for example, in several French defense positions often mishandled by Black, the move "white bishop takes black pawn h7 with check" is self-evident to Masters), while some patterns elicit variable-free generalized actions ("install a piece on a weak square"). That such a mechanism allows proposing reasonable moves was shown by Gobet and Jansen (1994), who describe a production system that triggers moves when recognizing patterns, using both compiled conditions and compiled actions.

We have proposed that, besides chunks encoding information about the specific location of pieces, chess players possess chunks with variables. Templates employ both kinds of chunks, compiled and interpreted. The core of the template is supposed to encode the precise location of pieces, while the slots may either encode the type of piece but not its location or encode a square but not the type of piece located on it. Templates offers then two features that are adaptatively important: they are rapidly accessible, and, if enough time is allowed, information may be encoded more precisely by correcting default-value slots or filling in slot having no default value.

The template theory predicts that, as positions are taken deeper in the game (instead after around 20 moves, as in most experiments reported in the literature), it will be harder for Masters to use a template and recall will decline. Masters no doubt do possess templates for late middle games and endgames, but as the game tree continues to branch, the ratio of templates to legal positions is much smaller in late than in early middle games. While it is known that players recall endgames more poorly than early middlegames (Chase & Simon, 1973a; Saariluoma, 1984), no empirical data have been reported about the recall of late middlegames (say after 40 moves).

The concept of typicality of positions, although common in the literature, has never been clearly defined. In the framework of the template theory, a position is seen as typical if it is sorted to a node that has evolved into a template. Note that the same position may be sorted differently on different occasions (it may be parsed into chunks differently, and the chunks sorted in different orders).

We may offer a few other speculations based on the template theory. The larger number of errors of commission made by Experts can result from their greater ability to recognize positions as belonging to particular types (templates), and their supplementing the pieces remembered accurately with typical (but, in this case, erroneous) chunks of pieces in positions of the recognized type. In comparison with Class A players, Experts have developed more templates; in comparison with Masters, however, their template types are too few to contain as much correct information about specific positions. Another speculation is that players at the beginning of their chess careers develop mainly conceptual chunks, and that perceptual chunks occur only at a high level of expertise. Thus, although class A players are able to play reasonable games and to remember positions well when given enough time to instantiate their conceptual chunks, they do not have enough time to do so with rapid presentation times, and then fail on the recall experiments below 30 seconds. Data on the development of chess expertise are needed to test this prediction.

In this paper, we have presented some results shedding light on the relation between skill in chess and the type of positions to be recalled: first, chess players' memory is disrupted by mirror image reflections of positions and second, Masters keep a superiority with random positions (with short as well as with long presentation times), but with a smaller absolute difference in percentage correct than for game positions with short presentation times. Our results also shed new light on chunking in chess: first, chessplayers do find some chunks in random positions and second, Masters' chunks are larger than was estimated by Chase and Simon (1973a). We have proposed that most of these results may be accounted for by the template theory we have developed in Gobet and Simon (1994a), which also explains how strong players are able to recall with considerable precision several boards presented briefly in succession. The most surprising finding is the as-yet unexplained fact that the size of the largest chunks is not affected by mirror image reflection. The results for random positions may be accounted for by the strategies subjects use and by the Masters' repertoire of large chunks. Finally, we have speculated on the role of compiled and non-compiled chunks in templates in particular and in chess memory in general.

Like the Chase and Simon theory, the template theory postulates that chess skill is built upon the presence of tens of thousands of chunks, stored in LTM and acquired over a long period of play and study. In addition, it proposes that some frequently-occurring chunks evolve into more complex structures, templates similar to Chase and Ericsson's retrieval structures.

Templates contain, in addition to information about the pattern of pieces, various other kinds of chess information, such as potential moves, plans, tactical motives, evaluation of the position, associations to similar positions, and so on. While Chase and Simon's chunks (still present in the template theory) are associated with local information, mainly about potential moves, templates allow players to evaluate whole positions, thus explaining expert players' abilities to judge and categorize positions rapidly, and to find potentially good moves and other relevant information. Our extension of the chunking theory is also compatible with Holding's (1985) SEEK theory, and it proposes a plausible way to implement the SEarch, Evaluation and Knowledge processes Holding postulated, still respecting the known speed and capacity limits of the human information processing system.

References

Bronstein, D. (1979). *Zürich International Chess Tournament 1953*. Translated from the second Russian edition by Jim Marfia. New York: Dover Publications.

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

Charness, N. (1981). Visual short-term memory and aging in chess players. *Journal of Gerontology, 1981b, 36*, 615-619.

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.

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.

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

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.

Djakow, I. N., Petrowski, N. W., & Rudik, P. A. (1927). *Psychologie des Schachspiels*. Berlin: de Gruyter.

Euwe, M. (1972). *Jugement et plan*. Paris: Payot.

Euwe, M. (1978). *The development of chess style*. New York: McKay.

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

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

Gobet F. & Jansen, P. (in press). Towards a chess program based on a model of human memory. In J. van den Herik (Ed.), *Advances in Computer Chess 7*.

Gobet, F. & Simon, H. (19894a). Chess memory: recall of several positions. Submitted for publication.

Gobet, F. & Simon, H. (1994b). Delimitation of chess chunks: a replication. Submitted for publication.

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

Goldin, S.E. (1978). Memory for the Ordinary: Typicality Effects in Chess Memory, *Journal of Experimental Psychology: Human Learning and Memory, 4*, 605-616.

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*. Paris: Payot.

Kolers, P. A. & Perkins, P. N. (1975). Spatial and ordinal components of form perception and literacy. *Cognitive Psychology, 7*, 228-267.

Krabbé, T. (1985). *Chess curiosities*. London: George Allen & Unwin.

Krogius, N. (1976). *Psychology in chess*. London: R.H.M Press.

Lewis, D. (1960). *Quantitative Methods in Psychology*. New York: McGraw-Hill.

Lisitsin, G. M. (1958). *Strategia i taktika shahmat.* Moskva: Fisikultura i sport.

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

Moran P. (1989). *A. Alekhine: agony of a chess genius*. Jefferson, N.C.: McFarland & Co.

Philidor, F. A. D. (1749). *L'analyse du Jeu des Echecs*. London.

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

Reshevsky, S. (1976). *The Art of Positional Play*. New York: McKay.

Reynolds, R. I. (1982). Search heuristics of chess players of different calibers. *American Journal of Psychology, 95*, 383-392.

Richman, H., Staszewski, J., & Simon, H.A. (1993). *Simulation of expert memory with EPAM*. Submitted for publication.

Saariluoma, P. (1984). *Coding problem spaces in chess: A psychological study*. Commentationes scientiarum socialium 23. Turku: Societas Scientiarum Fennica.

Saariluoma, P. (1989). Chess players' recall of auditorily presented chess positions. *European Journal of Cognitive Psychology, 1*, 309-320.

Saariluoma, P. (1991, July). *Location coding in Chess*. Paper presented at the International Conference on Memory, Lancaster (UK).

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

Smyslov, V. V. (1972). *My best games of chess*. New York: Dover.

Wilson, F. (1976). *Lesser-known chess masterpieces: 1906-1915*. New York: Dover.

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. Correspondence concerning this article should be addressed to Herbert A. Simon, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213.

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

Table 1.

Overall relations, location and recall performance as a function of the type of transformations imposed on positions.

Type of transformation from game position	Overall relations	Location	Recall performance
No transformation	same	same	Standard
?	same	different	?
Hybridization from 4 positions	different	same	Close to standard
Diagonal swapping	different	different	Close to random

Table 2.

Mean number of errors of omission (upper panel) and errors of commission (lower panel) as a function of category and type of modification (in parentheses, standard deviation).

Errors by omission

	NORMAL	HORIZONTAL	VERTICAL	CENTRAL
Masters	3.13 (4.40)	3.00 (2.96)	8.80 (4.85)	4.47 (5.69)
Experts	3.27 (2.75)	4.80 (6.12)	6.67 (5.60)	5.87 (5.05)
Class A	8.06 (3.44)	11.73 (5.71)	10.67 (5.59)	11.93 (4.61)

Errors by commission

	NORMAL	HORIZONTAL	VERTICAL	CENTRAL
Masters	4.74 (1.49)	4.83 (1.73)	5.13 (4.32)	5.56 (2.61)
Experts	8.39 (2.25)	7.55 (3.52)	7.92 (4.03)	8.88 (2.81)
Class A	5.81 (1.30)	4.42 (3.16)	5.70 (3.90)	5.33 (1.96)

Table 3.

Correlation matrix of typicality scores given by 5 judges to the positions used in experiment #1.

	1.	2.	3.	4.	5.
1. GM_1	1				
2. GM_2	.047	1			
3. M_1	.227	.28	1		
4. M_2	.501	.259	.575	1	
5. A	.213	-.101	.587	.509	1

Table 4.

Mean number of errors of omission (upper panel) and errors of commission (lower panel) as a function of category and type of modification (in parentheses, standard deviation).

Omission

	NORMAL	HORIZONTAL[1]	VERTICAL	CENTRAL
Masters	0.37 (0.53)	-0.12 (0.18)	1.87 (2.65)	1.87 (1.06)
Experts	5.55 (2.22)	7.20 (5.74)	8.65 (4.19)	9.85 (8.13)
Class A	14.71 (3.68)	15.79 (2.30)	17.71(0.91)	16.04 (4.34)

[1] A negative value indicates that subjects have placed in average *more* pieces that contained in the stimulus positions.

Commission

	NORMAL	HORIZONTAL	VERTICAL	CENTRAL
Masters	2.75 (0.00)	1.75 (0.71)	4.00 (2.47)	3.75 (1.06)
Experts	7.05 (2.72)	6.70 (4.93)	6.85 (3.47)	7.05 (4.50)
Class A	1.83 (1.02)	1.83 (1.09)	1.67 (0.74)	2.29 (2.28)

Table 5.

Placement of chunks and isolated pieces. Integers indicate the number of pieces by chunk (n=1 for isolated pieces).

	All pieces replaced	Correct Pieces only
M1		
Position # 2	5 -3-1-1	4-1
Position #11	7-1-2	6
Position #13	4-6	3-3
Position #22	2-2-1	2
Position #24	3-1-2	3
M3		
Position # 2	1	1
Position #11	1-1-1	1-1
Position #13	2-2	2-2
Position #22	2	2
Position #24	2	2
E6		
Position # 2	2-1-2-2-2-3-1-1-1	1-1-1
Position #11	1-2-1-1-2-1-3-1-2-1-1-2-1	2-1
Position #13	5-2-1-3-1-1-2-1-1-1-1-1-1-2	5-1
Position #22	4-1-2-2-1-1-1-1	-
Position #24	5-1-1-1-2	2-1
A2		
Position # 2	1-1	1
Position #11	4-1-1	4
Position #13	4-1-1	2-1
Position #22	2	-
Position #24	1-2-1-1-1	1-1-1

Table 6.

Detail of reconstruction for 4 subjects of the random positions depicted in Fig. 8 . Only correct pieces are mentioned. In parentheses, interpiece latency (in sixtieth of sec.).

	position #2	position #13
M1	Bc4 (107) Nd3 (55) Re7 (64) Qe3 (84) pc1 (256)	Qf7 (87) Rd6 (77) Bd5 (64) ke6 (148) pb1 (90) pb2 (71)
M3	Kh5 (147)	pa8 (103) pa1 (93) pb1 (201) pc1 (66)
E6	pc1 (61) ph4 (50) Pb7 (153)	pc1 (103) pb2 (62) pb1 (70) pa1 (68) Pa2 (71) rf8 (91)
A2	kf7 (202)	pb1 (281) pc1 (102) qb8 (825)

Table 7.

Errors of commission in normal and random positions (mean for one position).

	normal positions	random positions
Incorrect Type	1.60	0.83
Incorrect Color	0.03	0.63
Close Translation	9.23	5.03
Diagonal Translation	1.97	2.05
Vertical or Horizontal Translation	7.87	3.91
Symmetrical Exchange	0.13	0.22
Unexplained errors	0.37	0.57

Table 8.

Mean number of errors of omission (upper panel) and of commission (lower panel) as a function of category and type of position.

Omission

	NORMAL	RANDOM
Masters	0.39	18.27
Experts	5.51	17.59
Class A	14.40	18.28

Commission

	NORMAL	RANDOM
Masters	1.61	1.98
Experts	5.22	3.96
Class A	2.55	3.62

Table 9.
Recall percentage as a function of time presentation.
Parameter estimation of the function $P = 100 - Be^{-c(t-1)}$.

Game positions

	Parameter	Estimate	ASE[1]	95% Confidence Interval Lower	Upper
Class A	B	75.176	3.227	68.713	81.639
	c	0.033	0.005	0.023	0.042
Experts	B	66.414	3.154	60.124	72.704
	c	0.074	0.011	0.051	0.097
Masters	*B*	*29.240*	*4.851*	*19.303*	*39.178*
	c	*0.435*	*0.169*	*0.089*	*0.781*

Random positions

	Parameter	Estimate	ASE[1]	95% Confidence Interval Lower	Upper
Class A	B	90.617	1.566	87.481	93.753
	c	0.006	0.001	0.004	0.008
Experts	B	85.338	2.157	81.035	89.641
	c	0.012	0.002	0.008	0.015
Masters	B	80.539	2.706	75.040	86.038
	c	0.018	0.003	0.012	0.024

[1]Asymptotic Standard Error

Goodness of fit

	r^2 using all data points		r^2 using group means	
	Games	Random	Games	Random
Class A	.66	.51	.98	.95
Experts	.72	.50	.99	.87
Masters	.35	.69	.71	.93

Table 10.

Number of pieces correctly replaced for random positions as a function of skill, in various studies. Except Frey and Adesman (1976), who used a presentation time of 8 sec., and Gold and Opwis (1992), who used a presentation time of 10 sec., all study use a presentation time of 5 sec.

Source	N of subjects	Rating (in Elo points)[a]			
		<1600	1600-2000	2000-2400	>2400
[c] This paper, experiment 1	12		3.0	3.8	3.3
[c] This paper, experiment 2	13		3.2	3.0	4.0
[b] Frey & Adesman (1976)	13	2.0	2.5		
[c] Saariluoma (1984), experiment 3	4	2.3	4.0		
[c] Saariluoma (1984), experiment 4	4	2.5		5.0	7.3
[c] Saariluoma (1990), experiment 1	12	4.8		10.0	
[cd] Saariluoma (1990), experiment 2	9	2.4	3.4	4.8	
[b] Chase & Simon (1973a)	3	3.5	1.5		3.0
[cd] Gold & Opwis (1992) 10 sec	40	3.0	4.3		
[c] Gobet & Simon (1994a), exp. 2	13		3.0	4.0	5.2

[a] The group mean rating is used for classification

[b] USCF rating is used

[c] International rating, or equivalent, is used

[d] The difference between skill levels is significant at the .05 level.

Table 11.

Overall relations, location and recall performance as a function of the type of transformations imposed on positions.

Type of transformation from game position	Overall relations	Location	Recall performance
No transformation	same	same	Standard
Mirror image modifications	same	different	Slighly impaired
Hybridization from 4 positions	different	same	Close to standard
Diagonal swapping	different	different	Close to random

Figure captions

Figure 1. Example of Saariluoma's (1991) position modification by swapping two quadrants: (a) before the swapping; (b) after the swapping.

Figure 2. Example of the kinds of positions used in experiment 1a. The same position is presented (a) under its normal appearance; (b) after reflection about the horizontal axis; (c) after reflection about the vertical axis and (d) after reflection about the central axes.

Figure 3. Mean percentage of pieces correct as a function of chess skill and type of position. Mean percentage with random positions is shown for comparison sake.

Figure 4. Frequency histogram of pieces placed independently and of chunks. Upper panel: for all pieces, correct and incorrect; lower panel : only correct pieces.

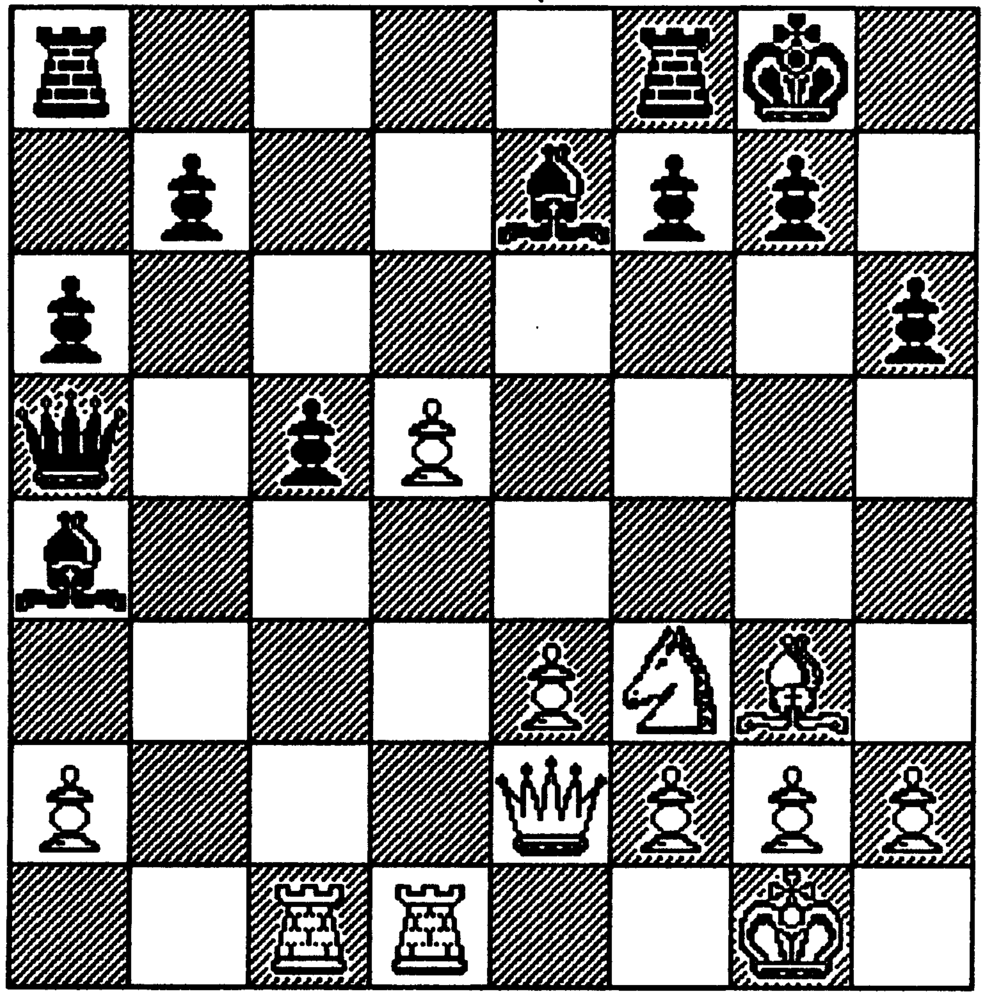
Figure 5. Mean percentage of correct pieces as a function of chess skill and type of position. Random positions are shown for comparison.

Figure 6. Two random positions used in experiment 3.

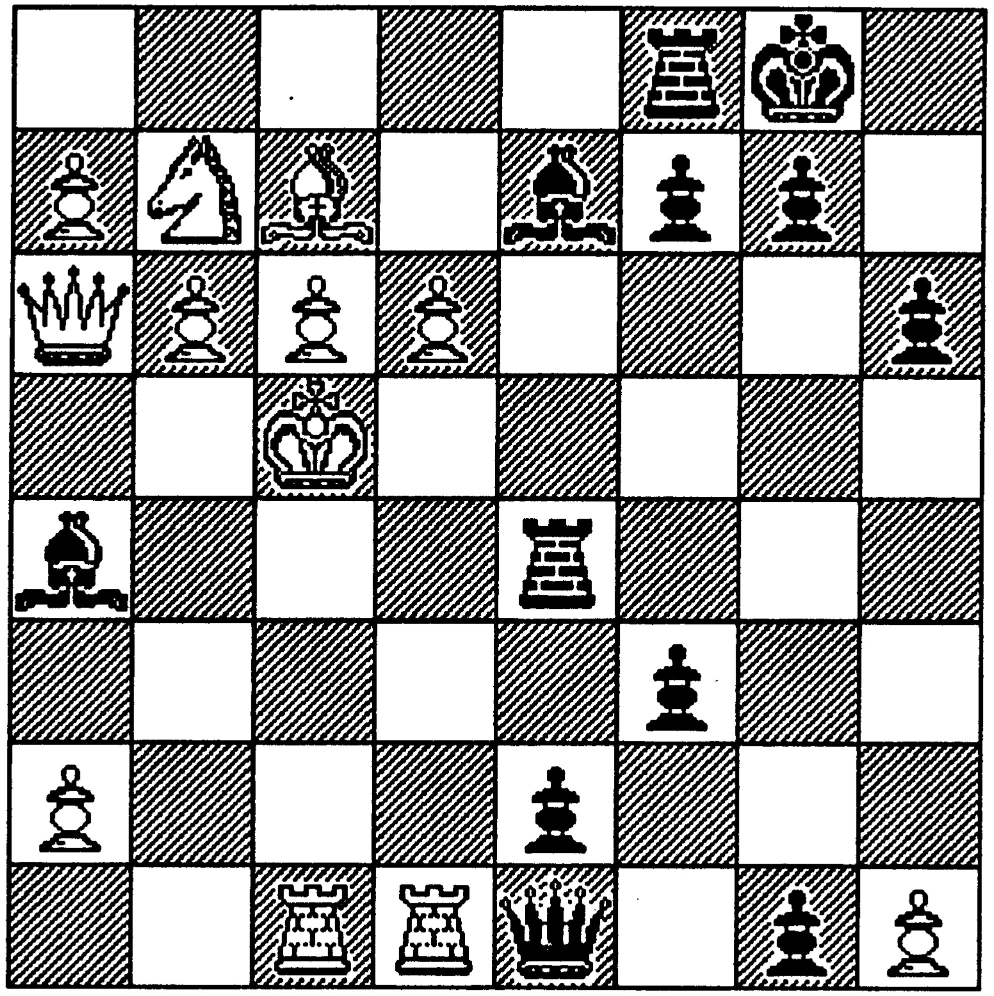
Figure 7. Mean number of chunks per position as a function of chess skill and type of position.

Figure 8. Percentage of correct pieces as function of presentation time and chess skill for game positions (upper panel) and random positions (lower panel). The best fitting exponential growth function is also shown for each skill level.

Figure 9. Largest chunk (in pieces) per position as function of presentation time and chess skill for game positions (upper panel) and random positions (lower panel).

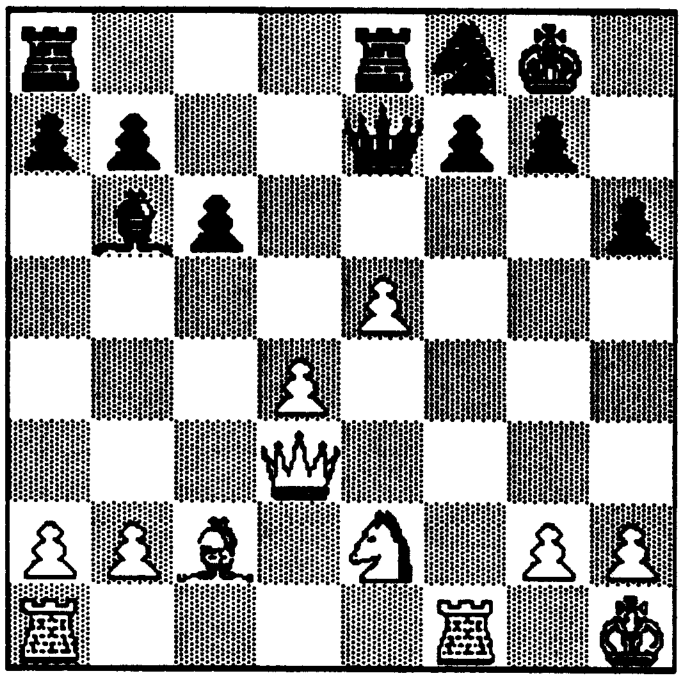


Original position.

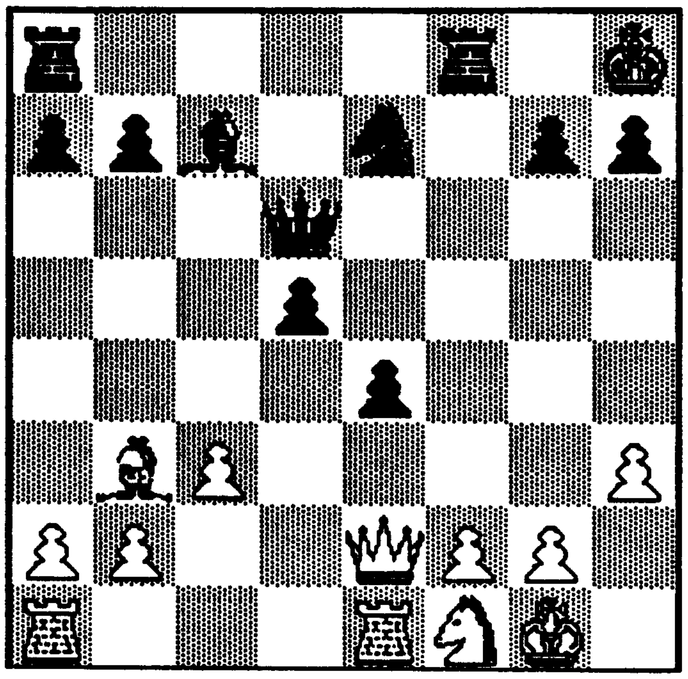


Position obtained after swapping the upper left quadrant with the lower right quadrant.

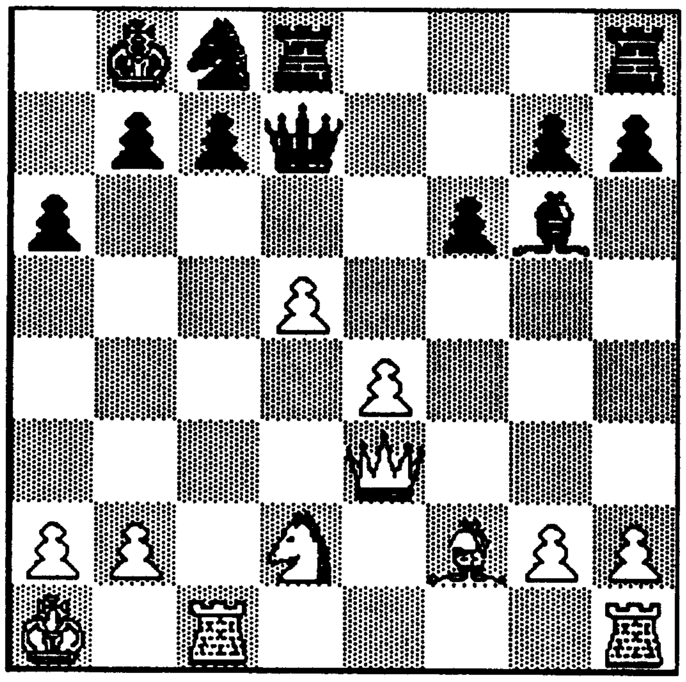
Figure 1



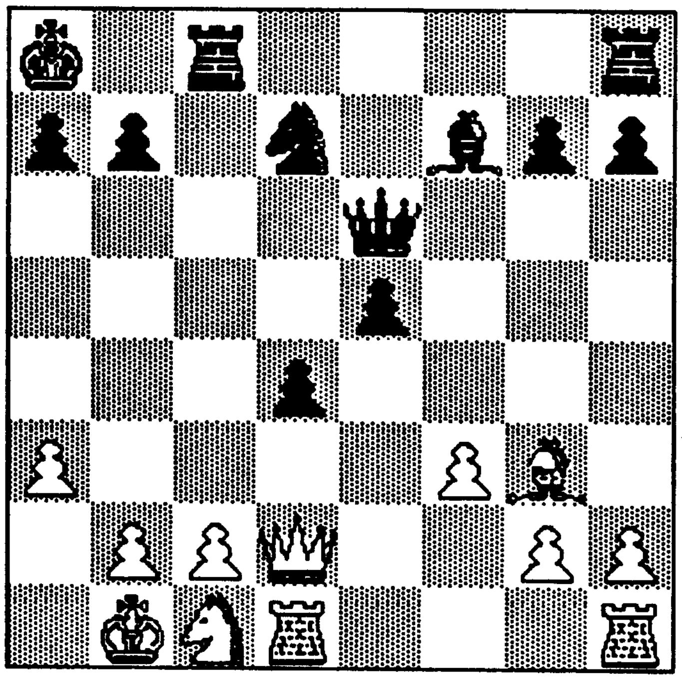
a) original position (normal)



b) after horizontal axis reflection



c) after vertical axis reflection



d) after central axes reflection

Figure 2

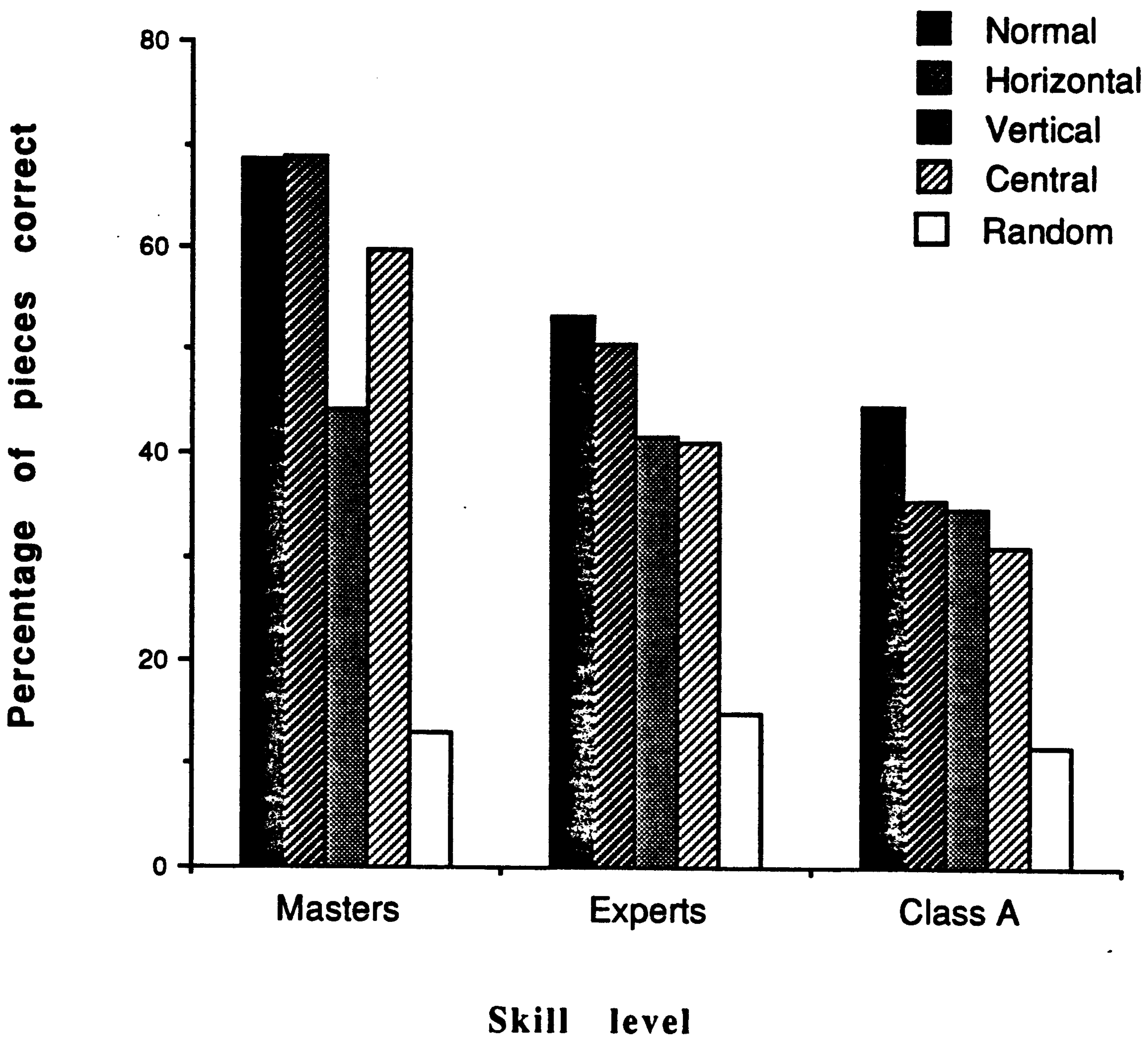
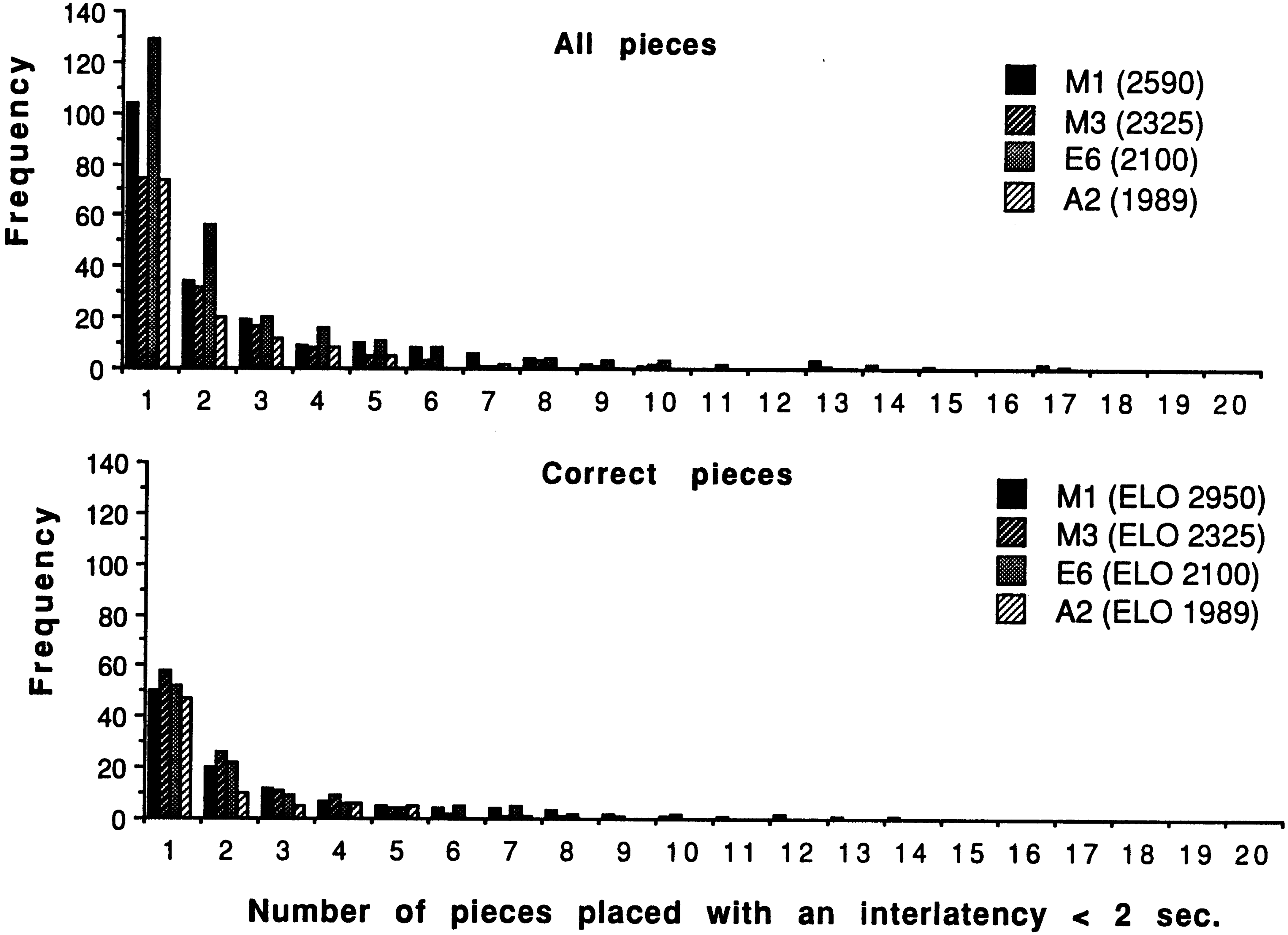
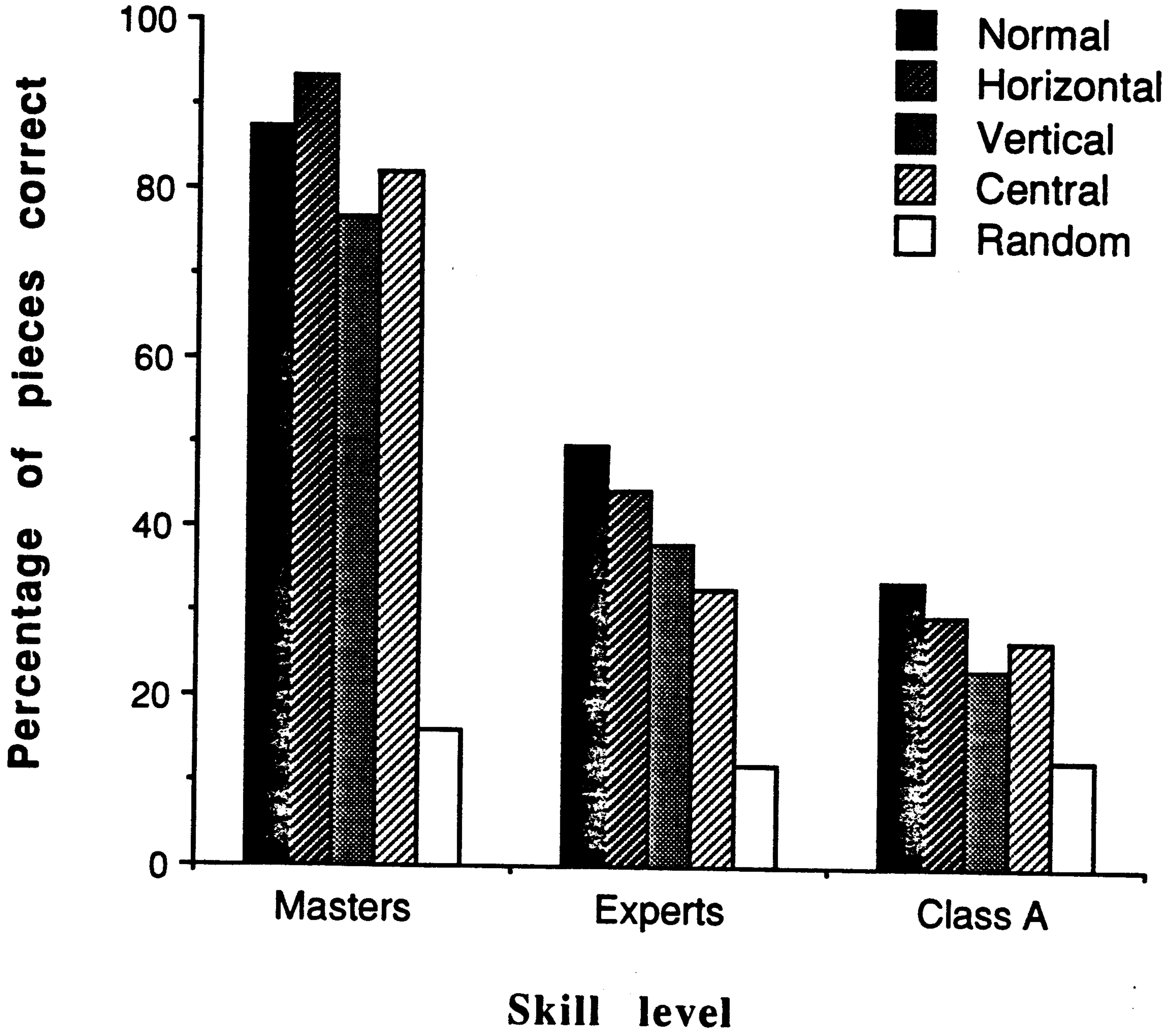

Normal
Horizontal
Vertical
Central
Random
80
60
40
20
0
Percentage of pieces correct
Masters
Experts
Class A
Skill level

Figure 3



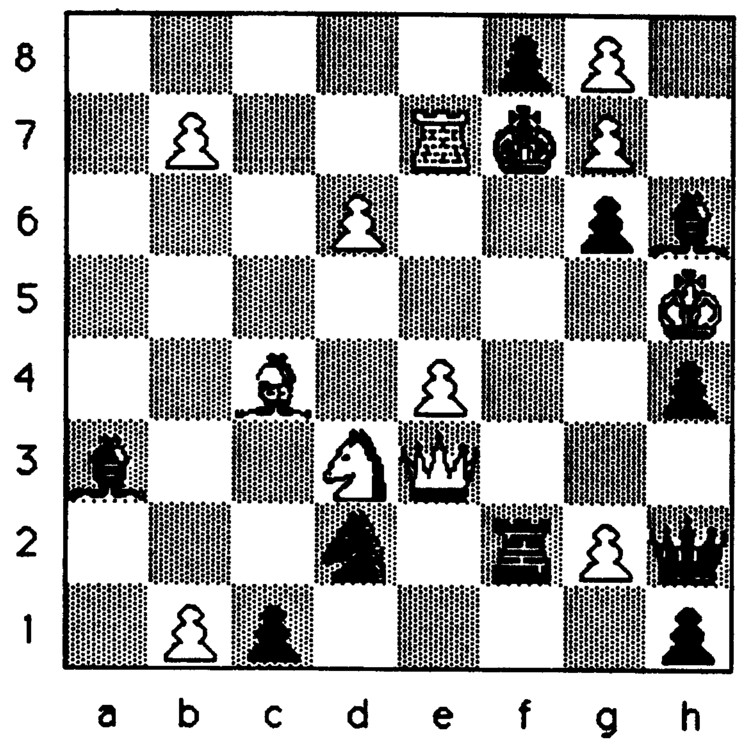
All pieces
Frequency
140
120
100
80
60
40
20
0
M1 (2590)
M3 (2325)
E6 (2100)
A2 (1989)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Correct pieces
Frequency
140
120
100
80
60
40
20
0
M1 (ELO 2950)
M3 (ELO 2325)
E6 (ELO 2100)
A2 (ELO 1989)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Number of pieces placed with an interlatency < 2 sec.

Figure 4

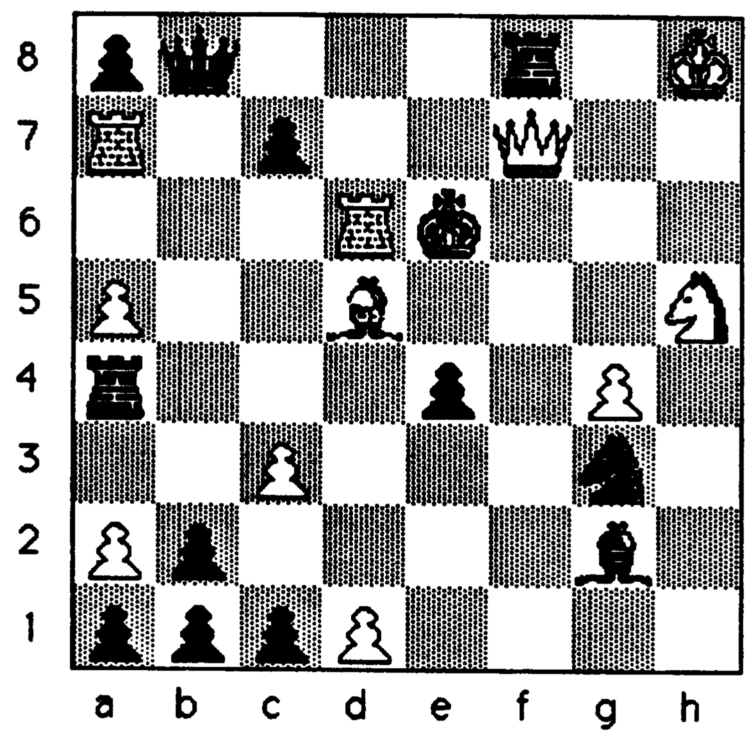


100
80
60
40
20
0
Percentage of pieces correct
Normal
Horizontal
Vertical
Central
Random
Masters
Experts
Class A
Skill level

Figure 5



Position #2



Position #13

Figure 6

Figure 7

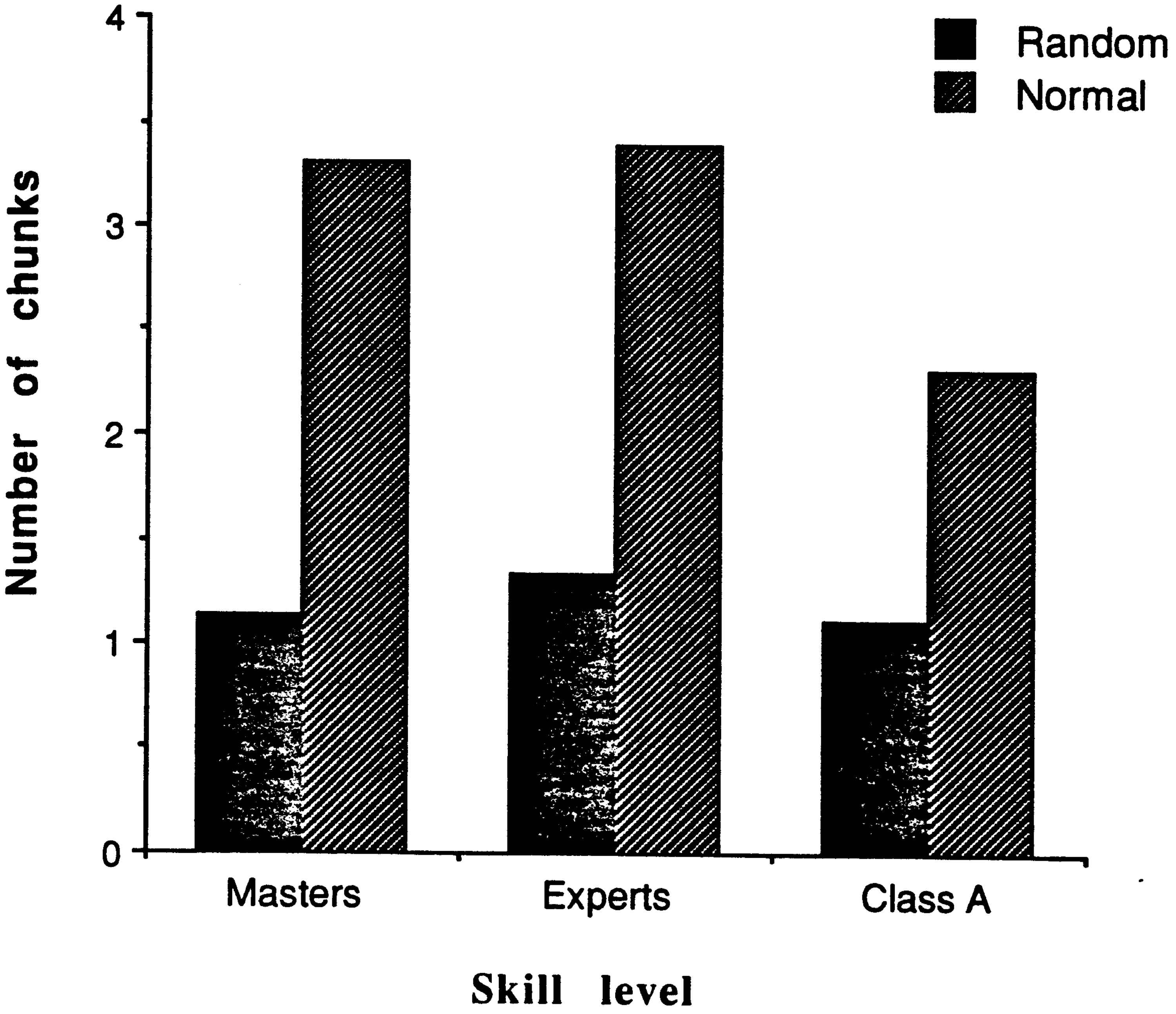
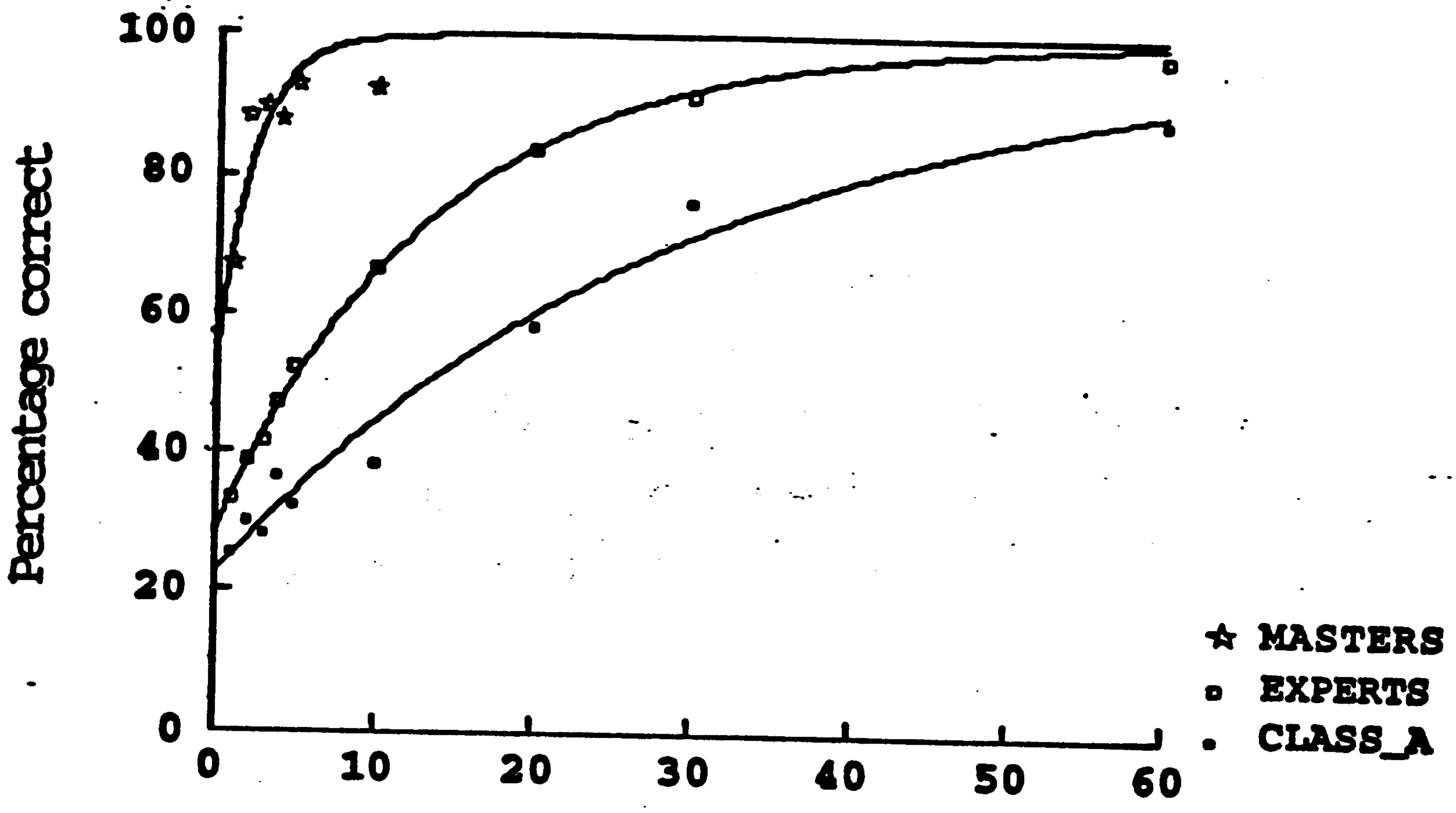


Figure 7

Game positions



Random positions

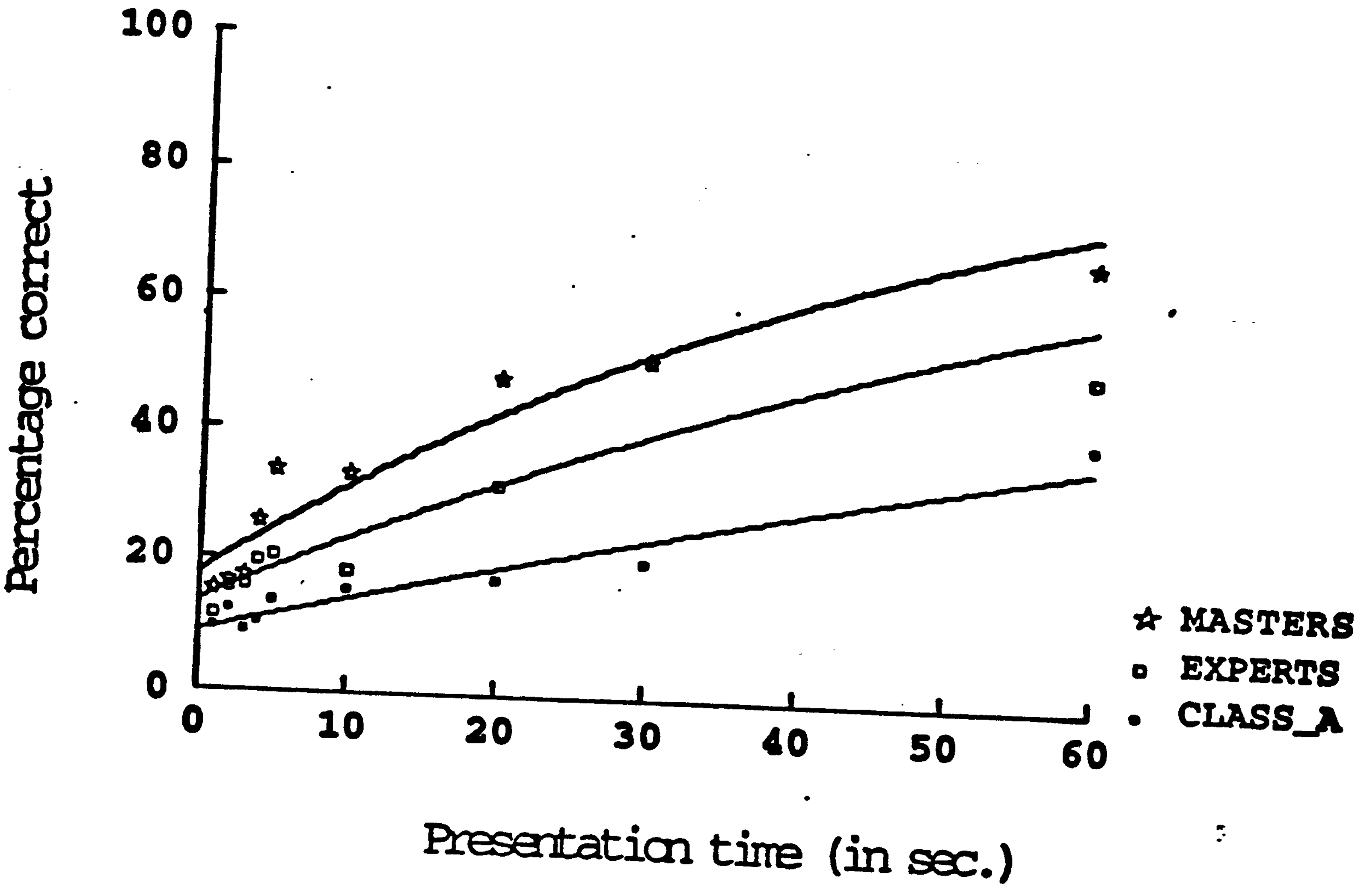
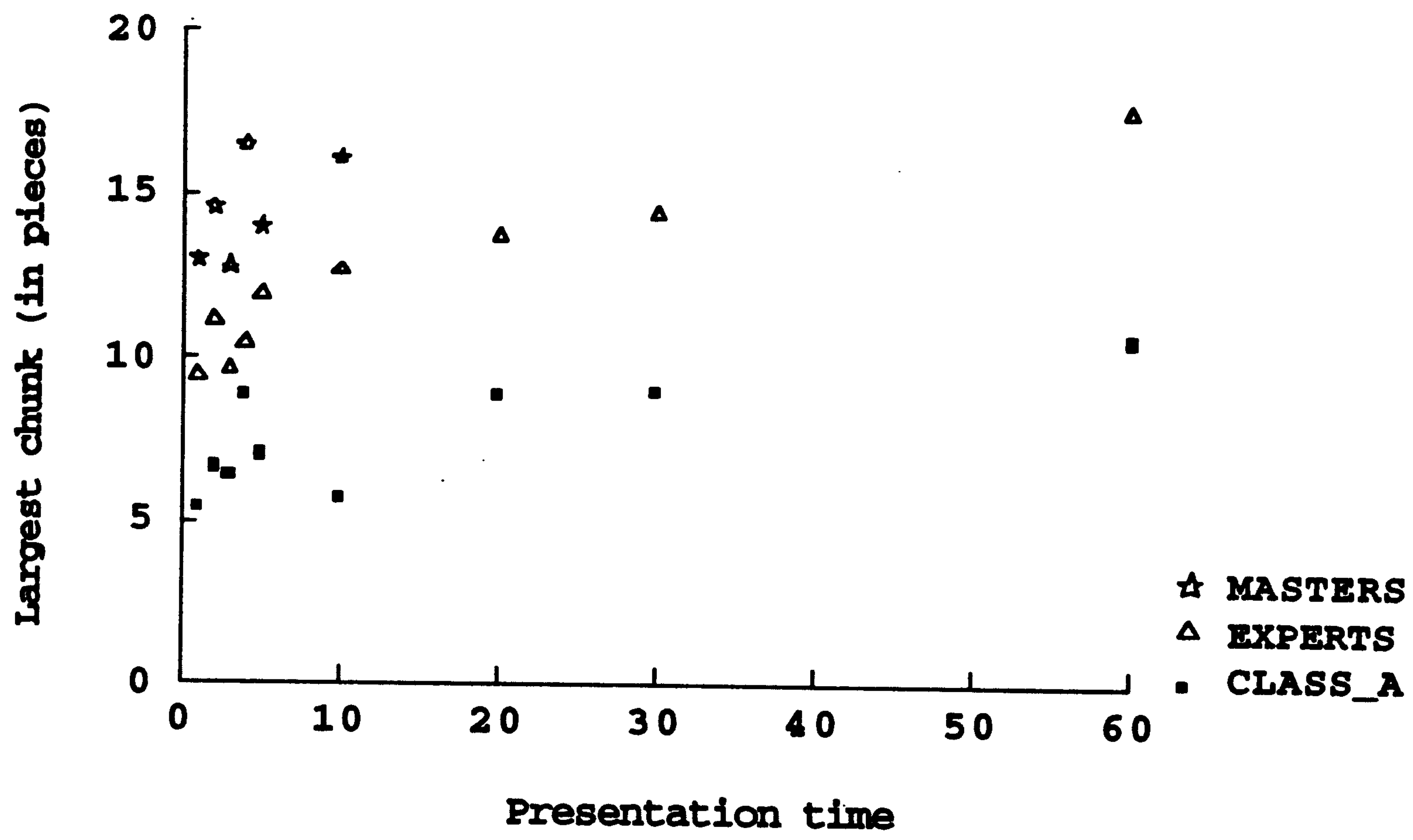


Fig. 8

Game positions



Random positions

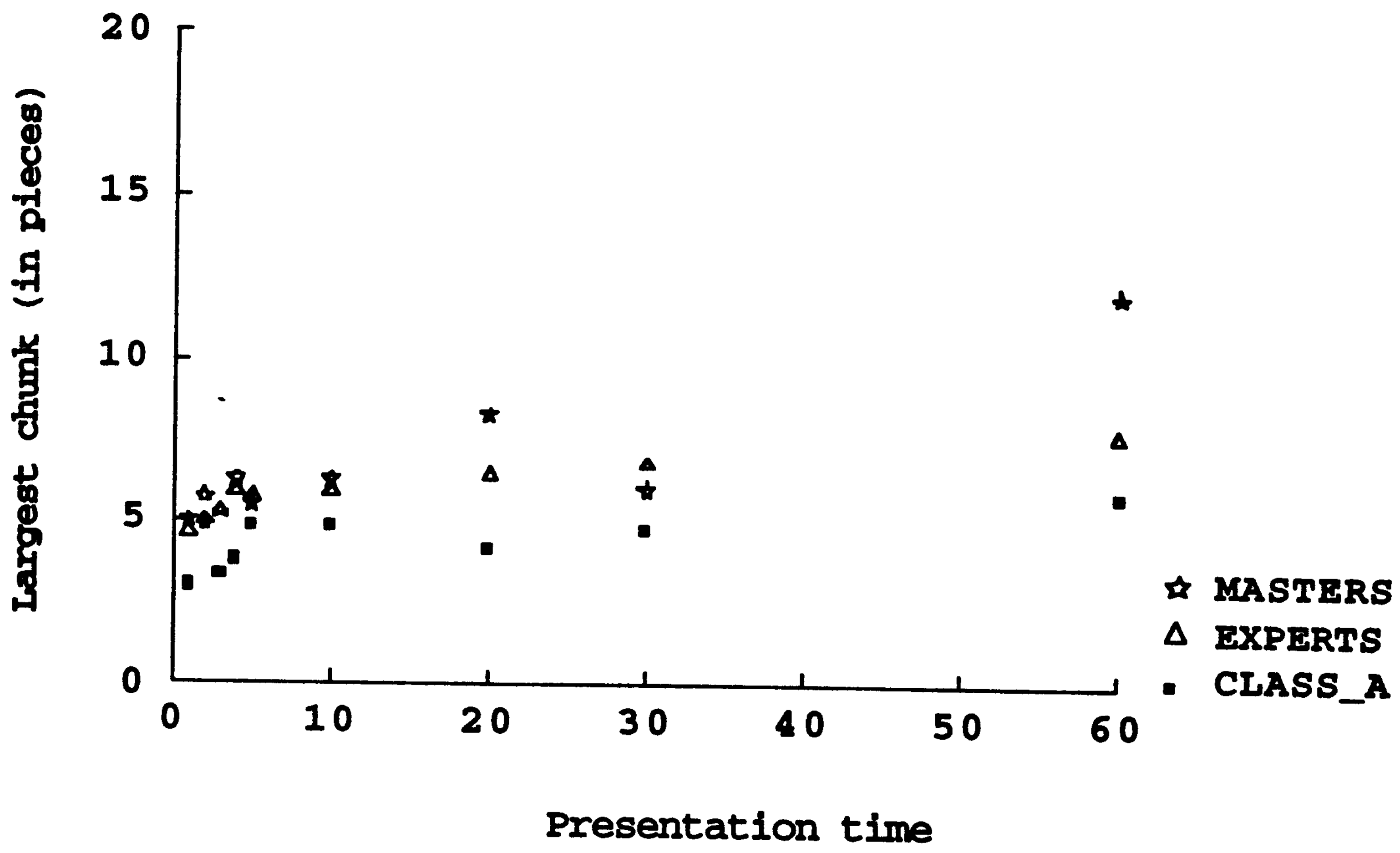


Figure 9