Recall of Random and Distorted Chess Positions: Implications for the Theory of Expertise

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

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

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

Running head: Recall of Distorted and Random Chess Positions

Abstract

This paper explores two important questions that have been raised about chunking with experts: (a) the number of chunks held in Long-Term Memory; and (b) the chunking mechanisms used in random positions. In two experiments, we investigate whether it is possible to reduce substantially the usual estimate that chess Masters store some 50,000 chunks by assuming that the same chunks can encode patterns under various mirror image reflections. As we find that recall is impaired when positions are modified by such mirror image reflection, it is dubious that several patterns can be represented by the same chunk. Additional analyses investigate the recall of random positions and the structural properties of chunks recalled from such positions. The role of interpretation and compilation of chunks is discussed, and it is shown that a good account of the results of the two experiments is given by the template theory proposed by Gobet and Simon (1994a).

Recall of Random and Distorted Chess Positions: Implications for the Theory of Expertise

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 (see Ericsson & Smith, 1991, for reviews).

Simon and Gilmartin (1973) and Chase and Simon (1973b) proposed, as an order-of-magnitude estimate, the often-cited figure of 50,000 chunks -- familiar patterns of pieces -- in the memories of chess Masters and Grandmasters, roughly comparable to natural language vocabularies of college-educated people. This number has been recently challenged by Holding (1985, p. 109; 1992), who 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[1] 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 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

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

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, 1994, 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). To take a less extreme example, the Knight in the pattern <white Knight c3 and Pawns c4 and d4> controls eight squares, but only four when the pattern is shifted two squares to the left. Needless to say that two such patterns have totally different roles in the semantics of chess.

At a more general level, and going beyond chess, 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 that evoke from memory information about their significance. 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, in press).

The sensitivity of perception to transformations of stimuli has long been a topic of research in psychology. For example, Ss experience considerable difficulty in reading upside-down printed text, 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.

Saariluoma (1984, 1994) addressed this question by manipulating the location of chunks. In one experiment, he constructed positions by first dividing the original position 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 Ss ranking from Class C to Expert level. Results of the recall task show that Ss 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 Ss 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 (1994) 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 real 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.[2] In a recall task, Saariluoma found that Ss recall these positions about as well as game positions, from which he concludes that encoding maintains location information. These results show moreover that Ss 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 by Saariluoma 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 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 positions, but the locations of chunks have changed. 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. Under the hypothesis that chunks encode relations of proximity, defense and attack between pieces, but not their location, 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

[2]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.

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).

The correctness of this hypothesis of invariance is not obvious, as players may feel at ease in certain positions but not in the corresponding positions with Black and White reversed, 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. In particular, given the fact that White has the advantage of tempo and so dictates the game, one should expect, on average, that White builds up attacking positions while Black has to content himself with choosing defensive set-ups.[3] 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 around horizontal, vertical and central axes of symmetry.

Four points should be mentioned. First we used a transformation by *reflection*, and not by *translation* as in Saariluoma's swapping experiment. Second, our transformations do not break up any relation between the pieces in the position. In consequence, if a location-free chunk is present in the non-modified version of the position, it is present in the three other permutations. Third, although our transformations keep the relations between pieces intact, they may change the up-down and/or left-right orientation of these relations. Regrettably, there is no transformation that manipulates location while keeping at the same time the overall chess relations intact and their orientation unchanged. Fourth, 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).

[3]This tendency is illustrated by the name traditionally given to openings. Variations arising from a white node are termed "Attack" or "Opening", while variations arising from a black node are dubbed "Defense".

Because Holding does not relate his remarks on chunks to a detailed theoretical model replacing MAPP, it is difficult to draw predictions from his position. In this paper, we will pit an extreme version of Holding's assertion -- that chunks encode only information on relations, and not on locations -- against an extreme version of Chase and Simon (1973b): chunks always encode information on location. As will be argued in the conclusion, it is possible that both types of encoding occur to some extent simultaneously. We now test the respective predictions, first with computer simulations (Experiment 1), and then with human Ss (Experiment 2).

Experiment 1

In order to gain a better understanding of the role of mirror image reflections in chess, we have conducted a few computer simulations. Because we are interested here in the statistical distribution of patterns in LTM, and not in the details of STM, attention mechanisms and retrieval structures, we have used a simplified version of CHREST (Gobet, 1993a,b), a model of chess players' memory and perception from the EPAM family (Feigenbaum & Simon, 1984; Simon & Gilmartin, 1973).

Methods

Material

A database of several thousand positions, taken from recently played grandmaster games, was used for the learning phase. Fifty new positions, each appearing in the four different permutations, were used during the testing phase. The positions of the first permutation were kept unchanged; those of the second permutation were modified by taking the mirror image with respect to the horizontal axis of the board; the positions of the third permutation were modified with reflection about the vertical axis. The positions of the fourth permutation 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

Procedure

The simplified version of CHREST builds up a discrimination net containing chess chunks by scanning the database positions. During the learning phase, the model randomly fixated twenty squares in each position, and sorted the pieces within a range of two squares from the fixated square through the discrimination net. Information is encoded with indication of the location on the board. For example, an instance of short-castle position, a common pattern in chess, is encoded as <Pf2, Pg2, Ph2, Kg1, Nf3>, with P standing for Pawn, K for King and N for Knight. During the recall task, the patterns on the board are sorted through the net, possibly giving access to nodes already stored in LTM and encoding similar information.

For the simulation of the recall task, the program was tested after learning each 10,000 nodes (more often in the early stages of learning). Learning was halted during the tests. The discrimination nets were extended up to 70,000 nodes. For each position, as during learning, the model randomly fixated twenty squares on the board, and sorted the pieces within a range of two squares from the fixated square through the discrimination net. The percentage of pieces correct is the number of pieces belonging to the stimulus position replaced in the correct location (erroneous placements were not penalized).

Results

Our main interest is in the relative performance on the different types of positions. As can been seen in Figure 3, with a net of 70,000 the normal positions are slightly better recalled than the horizontally mirrored ("Horizontal") positions (respective means: 65.4% vs. 63.2%). The difference is reliable [$F(1,13) = 19.80$, $p<.005$]. When pooled, Normal and Horizontal positions are better recalled than Vertical and Central positions pooled [$F(1,13) = 363.92$, $p< 10^{-9}$]. The recalls of Vertical and Central positions, respectively 53.3% and 52.5%, on average, do not differ reliably [$F(1,13) = 4.10$, ns]. The figure also depicts, using the variable delta, the difference in recall between the Normal and Horizontal conditions, combined, as compared with the Vertical and Central conditions, combined. This difference, averaged over all memory nets, is 11.4%. Delta increases as a function of the number of nodes in the early stages of learning, until the fourth net (number of chunks = 2500), but then remains stable. In general, the percentage of recall is a monotonically increasing function of the number of nodes. The function, Percentage = a + b * log[Number_Nodes] accounts in all four conditions for more than 98% of the variance. Finally, Figure 3 shows that the recall of random positions improves slightly with the number of in nodes, up to 23.4%.

Insert Figure 3 about here

Discussion

In these simulations, mirror image reflection, in particular around vertical axis, makes the recall of chess positions harder for our chunking model. In increasing the number of chunks in its net, the model learns some patterns that can be applied in any permutation, thus allowing a general improvement. The model also learns very specific patterns that are unlikely to be recognized when the positions is modified around the vertical axis, in particular with castled positions. Hence the increasing superiority of normal and horizontal positions over vertical and central positions.

With respect to experiment 2, simulation data suggest main effects of Skill and of Type of position. They also predict a weak interaction, if sufficiently weak players (number of postulated chunks less than 2500) are included in the experiment. In contrast, Holding's assumption, in its extreme version, would predict no difference in the recall of the various conditions. Our alternative hypothesis, based on analysis of the chess environment and the computer simulations, 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 axis (vertical symmetry) and 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.[4] Finally, the simulations predict that recall of positions modified by central symmetry (reflection about both axes) should not differ from recall of positions modified by vertical symmetry.

In summary, after modification of the position, 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. We now test whether chess players behave as predicted by the computer model.

---

[4]In most chess games, both players' Kings castle on the King's side. From a database of 10,500 recently played games, we have computed White's and Black's King locations after 20 moves. Ten percent of the White Kings were located on the Queen's side (ranks "a", "b", "c") , 8% in the center (ranks "d", "e") and 82% on the King's side (ranks "f", "g", "h"). The respective percentages for Black Kings are 6%, 9% and 85%. Thus, for most positions, vertical and central modifications will send the Kings to the Queen's side, a location they occupy in only about 8% of games.

Experiment 2

This experiment was run in two different locations, with a slightly different material (see below). As ANOVA detects no interaction of location (taken as a between-subject variable) with the variables discussed below, we have pooled the data.

Methods

Subjects

One female and 24 male chess players volunteered for this experiment. Their ratings ranged from 1680 to 2540 ELO.[5] Ss were classified in three groups: Masters (n = 5, mean ELO = 2401, sd = 109), Experts (n = 11, mean ELO = 2127, sd = 73) and Class A Players (n = 9, mean ELO = 1878, sd = 97). Their ages varied from 17 to 45, with mean = 28 and standard deviation = 9. In the component experiment 1, 12 Ss were recruited in New York's Manhattan Chess Club, and were paid $10 for their participation ($ 20.- for the players having a FIDE title). In the component experiment 2, 13 Ss were recruited from the Fribourg (Switzerland) Chess Club and from players participating in the Nova Park Zürich tournament, and were paid as the New York players. The New York Ss also participated in experiment 2 of Gobet (1993a), on the recall of multiple boards. The Swiss Ss also participated in the copy task experiment reported in Gobet and Simon (1994b).

Control Task

In order to check against the possibility that the strong players had superior memory capacities, we constructed random positions by assigning the pieces from a normal game (mean number of pieces=25) position to squares on the chessboard according to random numbers provided by a computer. Ss of component experiment 1 received five random positions, inserted randomly among the experimental positions. Ss of component experiment 2 received three random positions, presented at the beginning of the experiment.

---

[5]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. USCF ELO have been corrected in accordance.

Material

Experiment component 1. 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 (normal, horizontal, vertical and central groups), according to the 4 permutations described in Experiment 1. The groups were comparable as to numbers of pieces and position typicality (as judged by the first author, whose rating is about 2400 ELO). Positions were presented in random order. The set of positions and their order was the same for all Ss.[6] Positions were presented on the screen of a Macintosh SE/30, and Ss had to reconstruct them using the mouse (For a description of the experimental software, see appendix in Gobet & Simon, 1994b).

Experiment component 2. Sixteen positions were selected from Lisitsin (1958), Wilson (1976), Reshevsky (1976), Euwe (1978), Moran (1989) and Smyslov (1972), with the same criteria as were used in component 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. Each subject thus received the positions in random order and with random assignment to type of modification.

Procedure and design

Ss received instruction on the goal of the experiment, and could familiarize themselves with the functioning of the program and (if necessary) were instructed on how to

---

[6]The order was C R N H V N H N V C R H R V C H N C V H V R N R C, where N tands for Normal, H for Horizontal, V for Vertical C for Central, and R for Random.

use the mouse to reconstruct the positions.[7] Ss of the component experiment 1 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. Ss of the component experiment 2 received, in order, the copy task (described in Gobet & Simon, 1994b), the control task (recall of random positions) and the mirror image reflection recall task.

Each position appeared for 5 seconds; the screen was then black during 2 seconds (5 seconds for Ss of component experiment 2) 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.

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 replaced correctly, the mean number and mean largest size of chunks, and the number and type of errors. We first report on the mirror image manipulation results, and then on the random positions.

Results

Mirror-Image Modifications

No significant correlation was found between the dependent variables and age or time to perform the task. Hence we will not include these variables in the following analyses.

Percentage of pieces correct. No subject recognized the types of modification to which the positions had been subjected. Figure 4 shows the results for the experimental positions. (Random positions are also shown, for comparison). Analysis of variance indicates a main effect of Skill [$F(2,22)=24.52$, $p<.001$], of Type of modification [$F(3,66)=20.85$, $p<.001$], and an interaction [$F(6,66)= 2.18$, $p < .05$]. The interaction is due to the relatively high recall of horizontal positions by Masters and of central positions by Masters and class A players. Contrast analysis shows that positions modified around the

---

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

vertical axis differ reliably from positions not modified around this axis [$F(1,22)=96.79$, $p < .001$]. For normal and horizontal modifications together, the mean percentages of pieces correct are 77.3%, 49.7% and 34.5%, respectively, for Masters, Experts and Class A players. For vertical and central modifications together, the respective means are 62.9%, 38.6% and 27.5%, respectively. Note that delta (the difference of vertical&central positions from normal&horizontal positions) increases with skill, as predicted by the computer simulation. The means are 14.4%, 11.1% for 7.0% for Masters, Experts and Class A players, respectively. This interaction is statistically significant [$F(2,22) = 3.48$, $p < .05$].

Insert Figure 4 about here

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. Throughout this discussion, we define a chunk as a sequence of at least two pieces whose mean inter-piece (adjusted) latency is less than or equal to 2 sec. 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. Using the same computer apparatus and correcting latencies in the same way for mouse time, we have replicated elsewhere (Gobet & Simon, 1994b) the main results of Chase and Simon's (1973a) copy and recall tasks, including the distributions of within- and between-chunk inter-piece latencies and the pattern of correlation between latencies and probabilities of chess relations. In the following analyses, chunks are defined as including correct as well as incorrect pieces.

For the size of the largest chunk, there is no significant effect of type of position [$F(3,66)=1.50$, ns], although the pattern of means is in the right direction: the largest chunks are bigger in the normal and horizontal conditions (means=7.7, 7.5, respectively) than in the vertical and central conditions (means=7.0 and 7.2, respectively). Contrast analysis shows that positions modified around the vertical axis tend to differ from positions not modified around this axis [$F(1,22)= 3.63$, $p=.07$]. There is a statistically significant difference between skill levels [$F(2,22)=4.70$, $p < .05$]. The average of the largest chunk per position is 10.1 for Masters, 7.1 for Experts and 6.1 for Class A players. No interaction is found [$F(6,66)=0.67$, ns].

An ANOVA, performed on the number of chunks per position, finds no main effect of the Type of modification [$F(3,66)= 0.77$, ns], although the pattern of means is in the predicted direction. For all skill levels together, the mean number of chunks per position is 3.6, 3.4, 3.3 and 3.3 for the normal positions, horizontal, vertical and central conditions respectively. There is a main effect of Skill [$F(2,9)=9.03$, $p= .001$]. The mean number of chunks per position is, pooling the 4 conditions, 3.8 for Masters, 4.1 for Experts and 2.3 for Class A players. No interaction is found [$F(6,66)= 0.06$, ns].

Error analysis. We have divided errors into 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 are the pieces placed wrongly by the subject.

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 vertical symmetry positions).

Insert Table 2 about here

With errors of omission, ANOVA indicates a main effect of Skill [$F(2,22)=13.40$, $p<.001$] and a main effect of Type of modification [$F(3,66)=8.54$, $p<.001$]. No interaction is present [$F(6,66)=0.95$, ns]. Note the inverted-U shaped variation of errors of commission with skill: Experts commit more errors of commission than Masters and Class A, who do not differ substantially. The difference is significant [$F(2,22)=7.65$, $p<.005$]. Although the patterns of means show that Masters make more errors of commission with positions modified by a reflection around the vertical axis, no main effect of Type of modification nor interaction is found [$F(3,66)=1.47$, ns and [$F(6,66)=0.22$, ns]. It is therefore reasonable to conclude that mirror image reflections affect mainly the number of omissions, and not the number of errors of commission.

Game vs. Random positions

Although the random positions in Experiment 2 were used primarily as a control task, it is instructive to examine the behavior of our Ss 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 chess-semantic structure may offer useful information on the way skilled players cope with material that does not allow them to apply their chess knowledge. As suggested by the simulations of Experiment 1, 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.

Percentage of pieces correct. The control task (random boards) shows the classical recall superiority for game positions vs. random positions [$F(1,22)=291.51$, $p<.001$] and the classical interaction [$F(2,22)=16.50$, $p<.001$]. There is also a significant skill difference [$F(2,22)=14.38$, $p<.001$]. For random positions, the mean percentage of correct pieces among the three skill levels are: Master=14.3, Expert=13.6 and Class A=12.5; $F(2,22)=0.18$, ns. The mean percentage across the three skill levels is 13.3, that is, about 3.3 pieces.

This result agrees with those found in the literature. 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. (See Gobet & Simon, 1995) for a discussion of this skill difference in the recall of random positions.)

Chunks. When no attention is paid to the correctness of pieces building up a chunk, the means of the largest chunks are bigger for game positions than for random positions (means with game positions: 11.2, 7.4 and 6.2 pieces for Master, Experts and class A players, respectively; means with random positions: 4.1, 4.3 and 4.1 pieces; $F(1,22)=105.1$, $p<.001$). There is also an interaction between skill and type of position [$F(2,22)=11.57$, $p<.001$]: 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 [$F(2,22)=2.47$, ns].

For the number of chunks, one finds a main effect of Type of position [$F(1,22)= 98.44$, $p < .001$], a main effect of Skill [$F(2,22)=4.36$, $p<.05$] and an interaction [$F(2,22)= 4.35$, $p<.05$]. Respective mean number of chunk per positions are, for Masters, Experts and Class A players, 4.0, 4.3, 2.5 for game positions and 1.2, 1.7, 1.8 for random positions. Fewer chunks are elicited in random than in game positions, and Experts propose more chunks than the players of either higher or lower skill in game positions ($p<.05$), but not in random positions.

Errors. As expected, there is an important difference in the number of errors of omission between the normal and random positions [$F(1,22)=278.88$, $p<.001$]. Analysis of variance indicates also a main effect of Skill [$F(2,22)=7.82$, $p<.005$], as well as an interaction [$F(2,22)=21.89$, $p<.001$]. The number of errors of omission in random positions is high for all skill levels (respectively 19.0, 16.1, 17.9 for Masters, Experts and Class A players), and varies as a function of the skill level for the normal positions (respectively 2.03, 4.3, 12.5).

We have seen that the type of mirror image reflection had no effect on the errors of commission in game positions. What about the random positions, which can be considered as extreme modifications of normal positions? ANOVA detects a main effect of Skill [$F(2,22)=6.85$, $p=.005$]; the main effect of Type of position [$F(1,22)=3.25$, $p < .10$] and interaction [$F(2,22)= 3.38$, $p<.10$] are only marginally significant. The respective means for Masters, Experts and Class A players are 3.9, 7.8 and 3.1 for game positions, and 2.4, 5.5 and 3.9 for random positions). As was the case for positions modified by mirror image, Experts commit the largest number of errors of commission.

Discussion

In this experiment, we have found that, for all skill levels, our Ss 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. The average difference in recall performance between Normal and Horizontal positions, combined, and Vertical and Central positions, combined, is 10.3%. This is in close agreement with the difference found in the computer simulations of Experiment 1 (on average, 11.4 %). We also found that stronger players have better recall than weak players in all four conditions. Chunk size analysis gave a (non-significant) indication that the number of chunks is reduced and that the largest chunks contain more pieces in the unmodified and horizontally modified conditions than in the others. Finally, the number of omission errors is sensitive to the experimental manipulation, whereas number of errors of commission is not.

These results, and their close correspondence with the simulations, in which location was specified for all patterns that were stored, suggest that chess knowledge is generally encoded in such a way as to retain information about the precise location of the pieces. Conceptual knowledge just 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.

It follows that the deterioration of the Ss' performances with mirror image reflections of the positions, in close quantitative agreement with the deterioration observed in the simulations, taken together with Saariluoma's (1994) results, throws doubt on Holding's (1985, p. 109) hypothesis that chunks are recognized independently of the colors of the pieces or their locations on the board. Transformation of the positions affected mainly the number of errors of omission. It appears that chess information, or at least much of it, encodes both the color and the precise location of the pieces. Because Holding's calculation of the number of chunks required in LTM to attain nearly perfect recall is based on the now undermined assumption that chunks are not location-specific, we must reject his conclusion that the number of chunks in an expert's memory is much smaller than the 50,000 estimated by Chase and Simon.

As for skill differences, we found that stronger players commit fewer errors of omission, and that results show an inverted U-curve for the errors of commission, experts committing most such errors. A similar inverted U-curve was found for the number of chunks per position. For all skill levels, the number of chunks is well within the postulated number of visual chunks, four (Zhang & Simon, 1985). We also found that, in general, Masters replace large chunks, even exceeding 10 pieces.

In the second set of results, related to random positions, we have seen that randomizing positions affects the number of errors of omission, but not the number of errors of commission. There are important differences in the size of the largest chunk recalled between the recall of random and game positions, respectively. In addition, a striking feature of the recall of random positions is the presence of numerous chunks, occasionally as large as 6 or 7 pieces, for stimuli supposed to be devoid of any semantic organization. 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 Ss prefer to construct a small number of new chunks rather than to memorize a list of pieces without any relation among them. Alternatively, chessplayers use slots in STM to store descriptions of patterns on the board (e.g., "| Three | White pawns | on a diagonal | starting from a1 |".) It is unclear whether this strategy is the most efficacious (scores are lower than what Chase and Simon's [1973b] model would predict: seven pieces, that is 28% correct), but the approach is probably close enough to the normal activity of chess players.

General Discussion

In these experiments, we have examined two 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.

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 do not support Holding's claims. 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). These effects are of about the same magnitude in both experiments and correspond closely with the effects produced by the computer simulations, using location-specific chunks, presented in ~~the Introduction~~ Experiment 1.

In the introduction, we presented a table illustrating the effect of various types of position distortions on the recall of chess positions, with one missing cell. Experiment 2 allows us to fill the missing cells: mirror image reflection, which keeps the overall relations between pieces, but not their location, produces a slight impairment in the recall performance. Taken with Saariluoma's (1994) 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.

In general, the results presented in this paper are compatible with the template theory proposed by Gobet and Simon (1994a), an amended and extended version of the earlier Chase and Simon model. This theory still takes chunking as the main mechanism by which chessplayers store information in long-term memory about the chess positions they have encountered in their games or have studied. In addition, it postulates that some chunks evolved into more complex data structures, called templates, which describe more or less typical positions and allow rapid storage of task-related information far beyond short-term memory capacity. Templates contain fixed information (their <u>core</u>) about a dozen of pieces, and <u>slots</u>, 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. Template operate with a mechanisms similar to those postulated for retrieval structures (Chase & Ericsson, 1982; Richman & al., in press).

Templates are evoked for use in recalling a particular position by recognizing the position as being of a certain familiar type. 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. 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 occurring often in their games, and professional players to have several thousand, even for types of positions they seldom meet in their own tournament practice.

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. This was found to be the case in Experiment 2. The template theory also predicts that the largest chunks (corresponding to the template cores) will be bigger for unmodified positions than for modified positions. This prediction was only weakly supported. The 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 templates and chunks from the boards with randomly placed pieces reduces this advantage. In particular, usually no template is accessible. With respect to skill differences, the template theory offers predictions similar to Chase and Simon's model. It predicts that percentage of correct pieces and size of the largest chunk are positively correlated with strength. It also predicts, because larger chunks are expected to be found, that the number of omissions should be less for strong players. All these predictions are verified. As for the errors of commission, the template theory proposes that, as the subjects in our experiment were requested not to guess systematically the location of pieces, such errors are caused by discrepancies between the image (the internal representation) of the board and the board itself. It predicts that players of high skill commit few such errors (they can use the template slots to encode the type and location of pieces either absent from or wrongly encoded in other chunks) and that weak players also commit few such errors (they recognize few chunks). At intermediate skill levels, some of the chunks recognized may encode incorrectly the location of a few pieces, leading to errors of commission that would not occur if fewer templates were recognized. This could account for the fact, found in our data, that Experts made more errors of commission than both the more-highly skilled Masters and the less-skilled Class A players.

This paper's results, consistent with Saariluoma's (1994), 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 chunks, which require extra time for interpretation and instantiation. Chess masters surely possess some generalized chunks (concepts like "fork" show that they do); but the experimental evidence strongly indicates that 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. Patterns may elicit generalized actions ("install a piece on a weak square") or precise moves. 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 -- i.e., evoked by recognizing the weakness created by Black. 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.

In this paper, we have presented some findings that shed light on the relation between skill in chess and the type of positions to be recalled: first, chess players' memory is diminished by mirror image reflections of positions. Second, Masters' chunks are larger than was estimated by Chase and Simon (1973a). Third, chessplayers do find some chunks in random positions. Most of these results can 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 results for random positions may be accounted for by the strategies Ss use and by the Masters' repertories of unusual as well as common chunks. Finally, we have speculated on the role of fixed and variable chunks in templates in particular and in chess memory in general.

References

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16). New York: Academic Press.

Chase, W. G., & Simon, H. A. (1973a). Perception in Chess. *Cognitive Psychology, 4*, 55-81.

Chase, W. G., & Simon, H. A. (1973b). The mind's eye in chess. In W. G. Chase (Ed.) *Visual information processing*. New York: Academic Press.

Ericsson, K. A., & Smith, J. (1991). *Studies of Expertise : Prospects and Limits*. Cambridge: Cambridge University Press.

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

Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8*, 305-336.

Gobet F. & Jansen, P. (1994). Towards a chess program based on a model of human memory. In H. J. van den Herik, I. S.Herschberg, & J.E. Uiterwijk (Eds.), *Advances in Computer Chess*. Maastricht: University of Limburg Press.

Gobet, F (1993a). *Les mémoires d'un joueur d'échecs*. [Chess players' memories]. Fribourg (Switzerland): Editions Universitaires.

Gobet, F. (1993b). A computer model of chess memory. *Proceedings of 15th Annual Meeting of the Cognitive Science Society*, (pp. 463-468).

Gobet, F. & Simon, H. (1994a). *Templates in Chess Memory: A Mechanism for Recalling Several Boards*. Complex Information Paper #513, Carnegie Mellon University, Pittsburgh, PA 15213.

Gobet, F. & Simon, H. (1994b). *Expert Chess Memory: Revisiting the Chunking Hypothesis*. Complex Information Paper #515, Carnegie Mellon University, Pittsburgh, PA 15213.

Gobet, F. & Simon, H. (1995). Role of presentation time in Recall of Game and Random Chess Positions. Submitted for publication.

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.

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

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

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

Lisitsin, G. M. (1958). *Strategia i taktika shahmat*. [Strategy and Tactics in Chess]. Moskva: Fisikultura i sport.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

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

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

Richman, H., Staszewski, J., & Simon, H.A. (1995) (in press). Simulation of expert memory with EPAM IV. *Psychological Review*.

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

Saariluoma, P. (1994). Location coding in Chess. *The quarterly journal of experimental psychology, 47A*, 607-630.

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.

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

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 this research as well as to Pertti Saariluoma and two anonymous reviewers for helpful comments on drafts of this paper. We are grateful to Murray Campbell and Peter Jansen for making available to us the database of chess positions used in experiment 1.

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	2.0 (3.5)	1.7 (2.7)	6.0 (5.3)	3.0 (4.5)
Experts	4.3 (2.7)	5.9 (5.8)	7.6 (4.8)	7.7 (6.3)
Class A	12.5 (4.7)	14.4 (3.9)	15.4 (4.5)	14.7 (4.6)

Errors by commission

	Normal	Horizontal	Vertical	Central
Masters	3.9 (1.5)	3.6 (2.1)	4.7 (3.4)	4.8 (2.2)
Experts	7.8 (2.4)	7.2 (4.0)	7.4 (3.6)	8.0 (3.6)
Class A	3.2 (2.2)	2.7 (2.2)	3.0 (2.8)	3.3 (2.6)

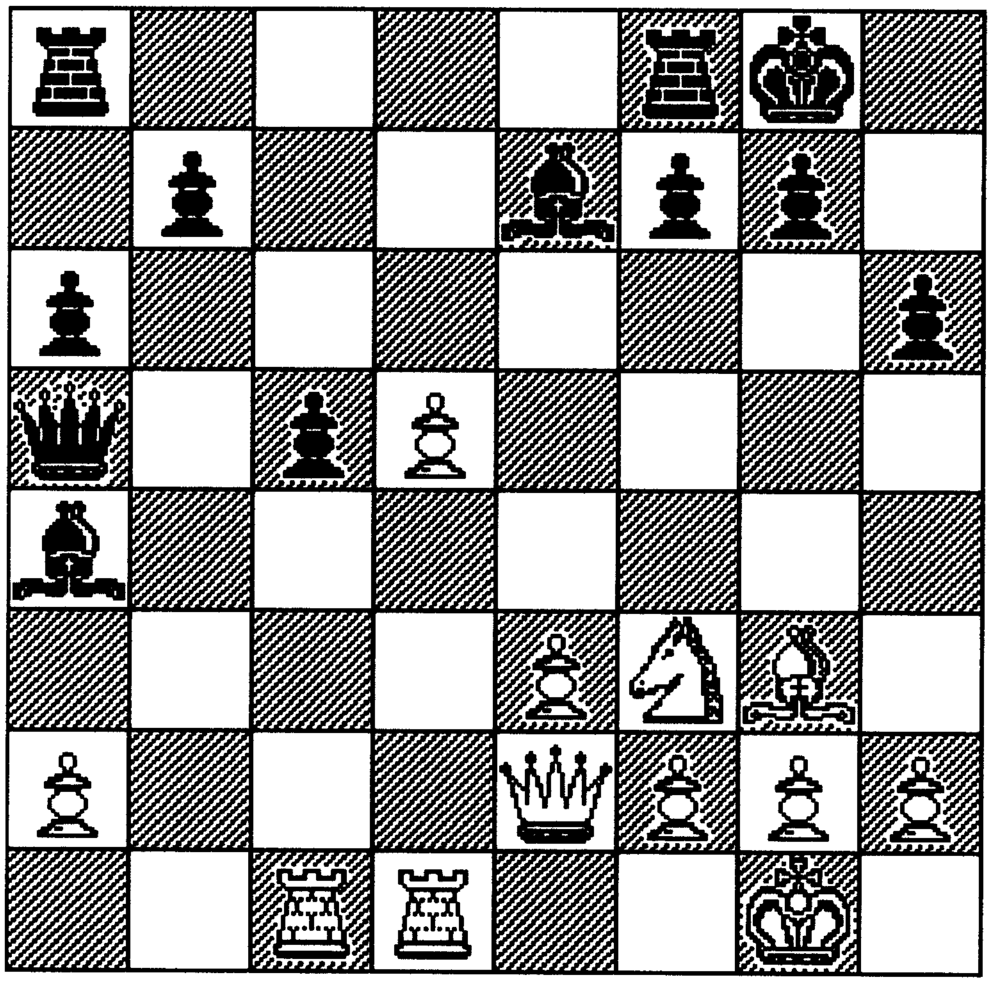
Figure captions

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

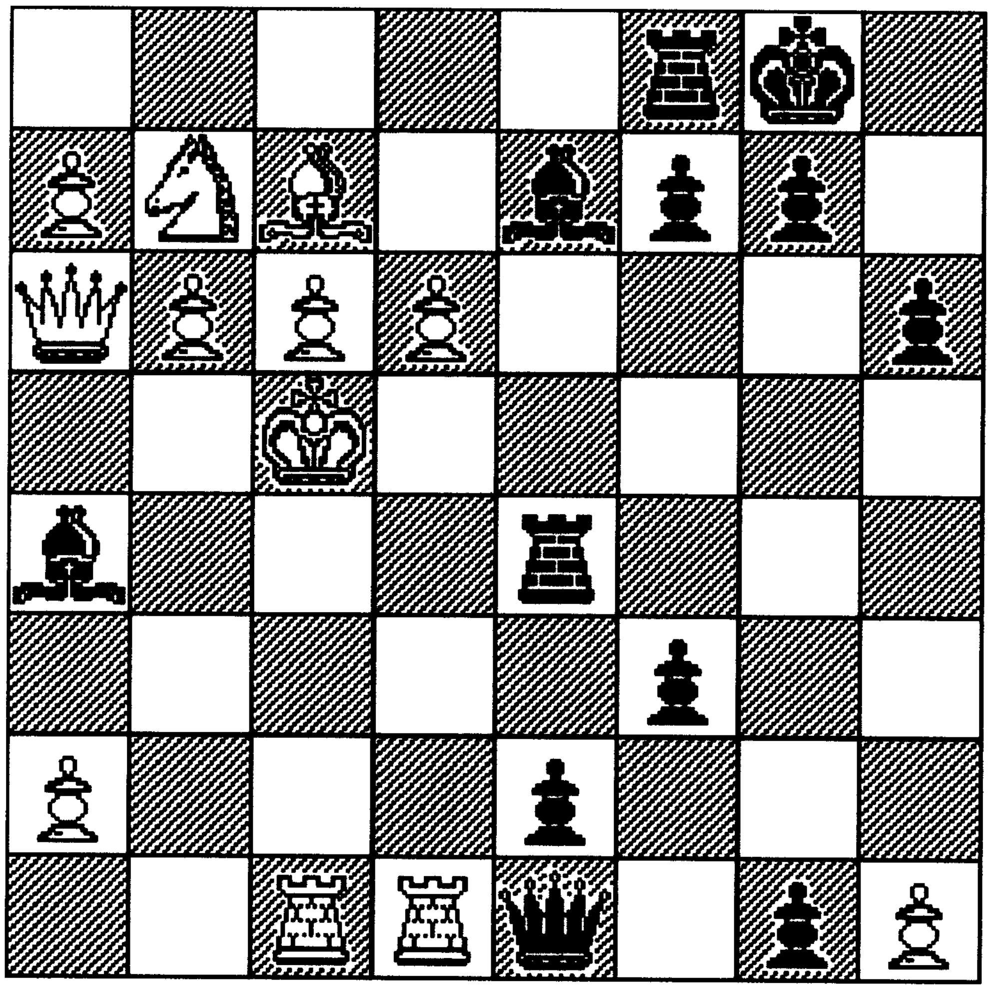
Figure 2. Example of the kinds of positions used in experiment 1. The same position is presented (a) under its normal appearance; (b) after reflection about the horizontal axis; (c) after reflection about the vertical axis and (d) after reflection about the central axes.

Figure 3. Computer simulations showing the recall percentage of game, horizontal, vertical and central position as a function of the number of nodes in the discrimination net.

Figure 4. 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.

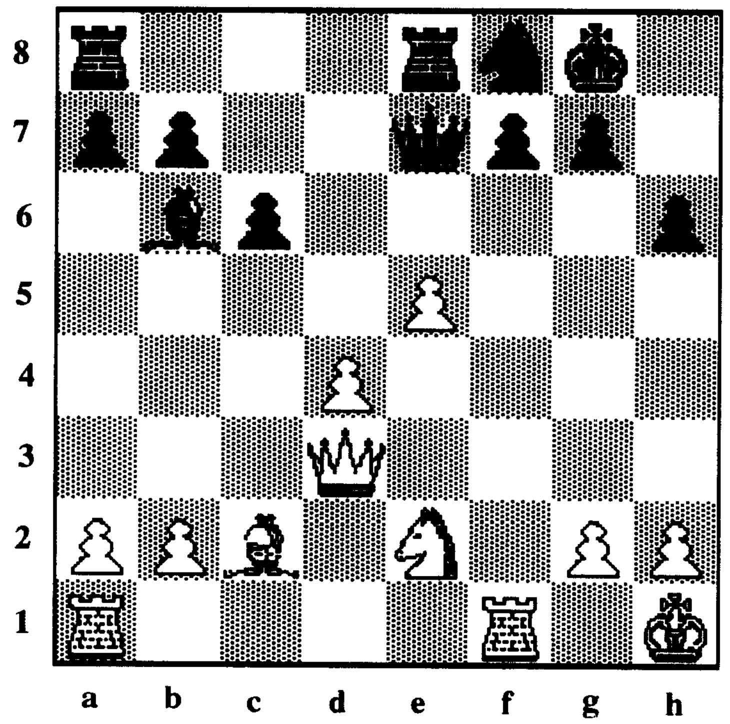


Original position.

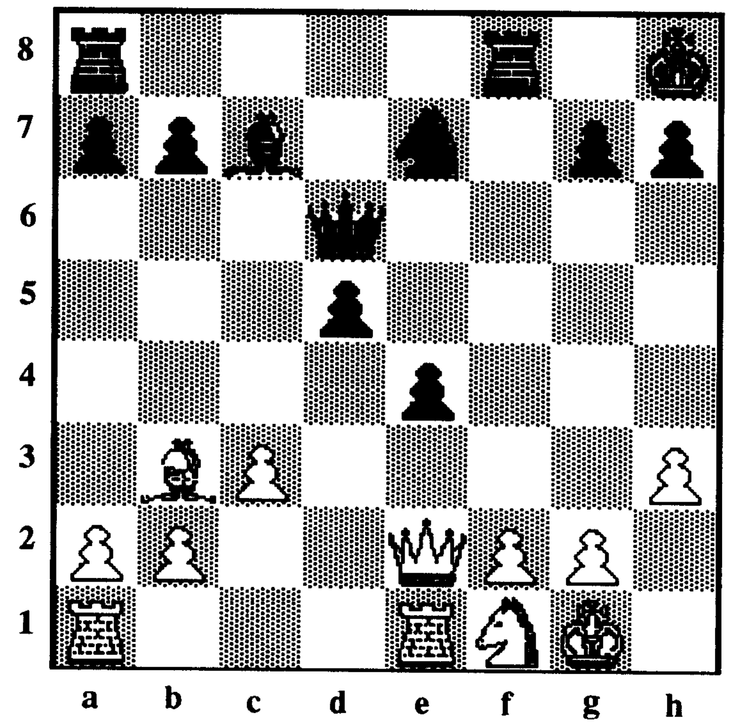


Position obtained after swapping the upper left quadrant with the lower right quadrant.

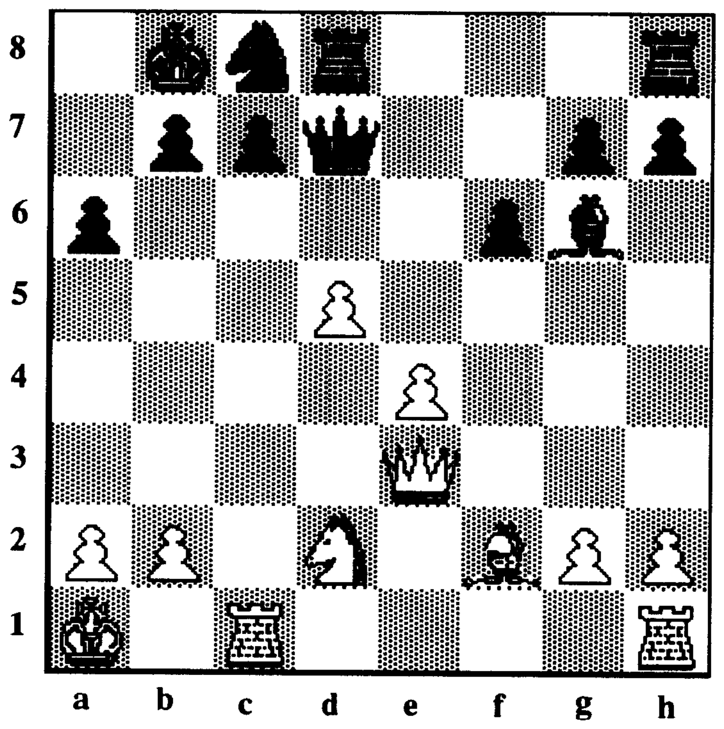
Figure 1



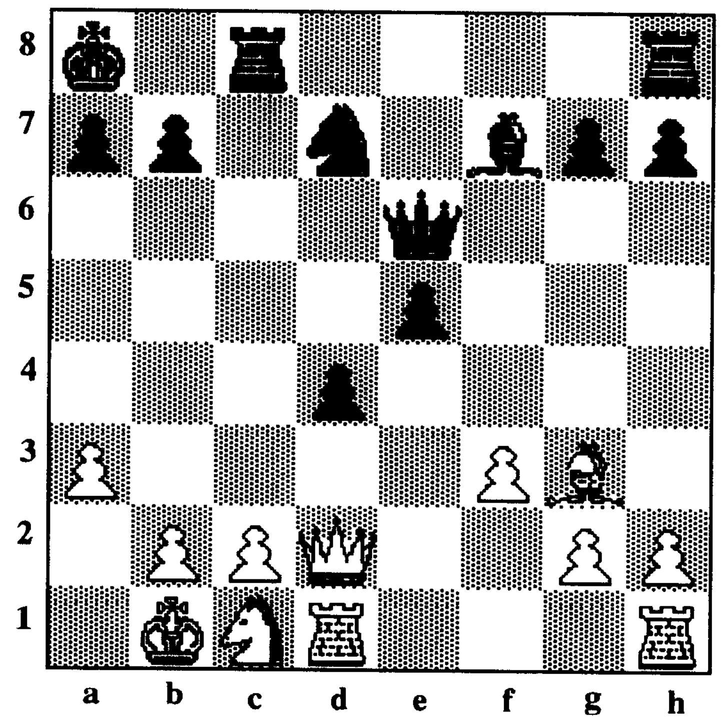
a) original position (normal)



b) after horizontal axis reflection

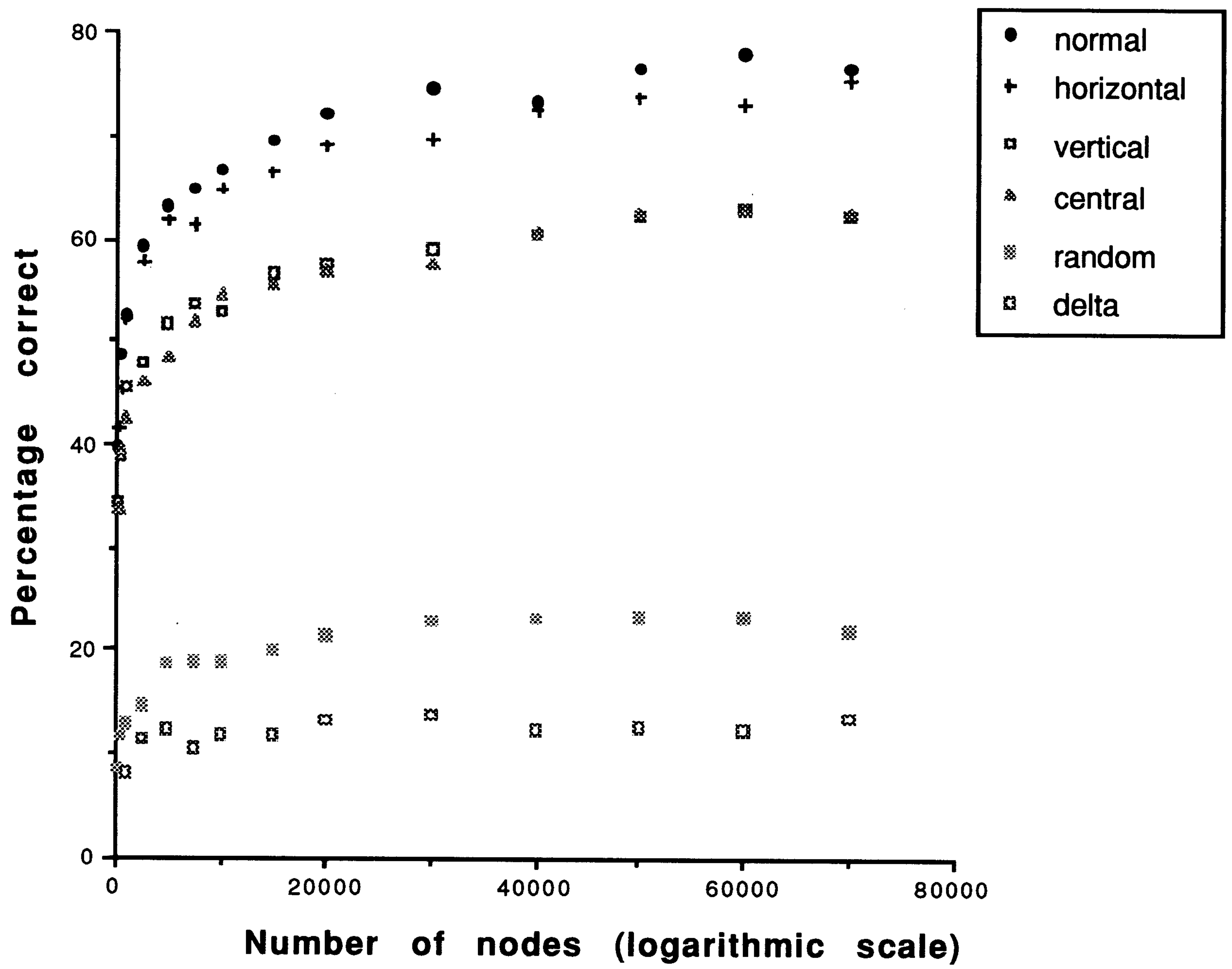


c) after vertical axis reflection



d) after central axes reflection

Figure 2



normal
horizontal
vertical
central
random
delta
Percentage correct
80
60
40
20
0
0
20000
40000
60000
80000
Number of nodes (logarithmic scale)

Figure 3

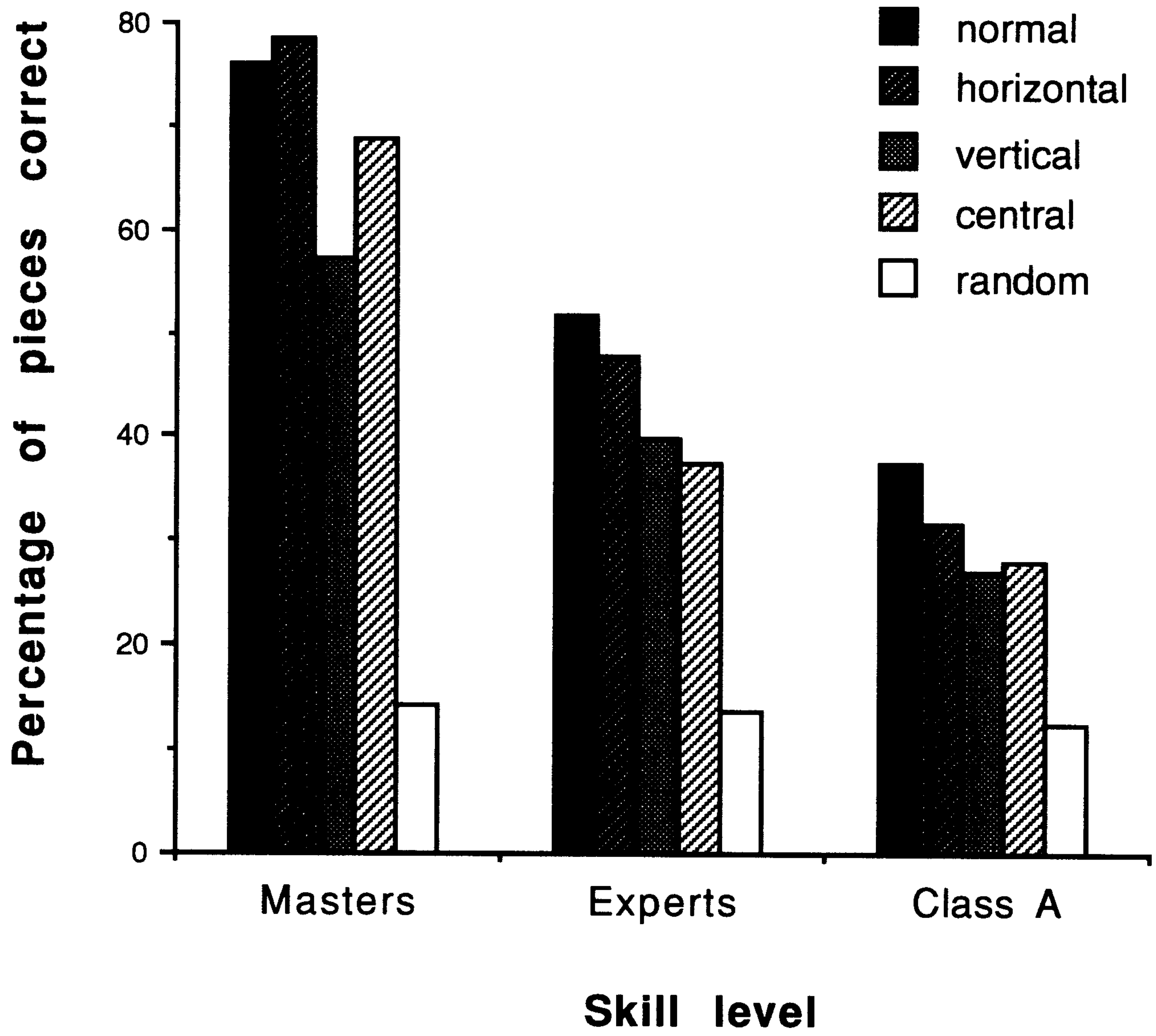

80
60
40
20
0
Percentage of pieces correct
normal
horizontal
vertical
central
random
Masters
Experts
Class A
Skill level

Figure 4