

Gobet, F. & Simon, H. A. (1996). Recall of random and distorted positions: Implications for the theory of expertise. *Memory & Cognition*, 24, 493-503.

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 the question, important to the theory of expert performance, of the nature and number of chunks that chess experts hold in memory. It examines how memory contents determine players' abilities to reconstruct (a) positions from games, (b) positions distorted in various ways and (c) and random positions. Comparison of a computer simulation with a human experiment supports the usual estimate that chess Masters store some 50,000 chunks in memory. The observed impairment of recall when positions are modified by mirror image reflection, implies that each chunk represents a specific pattern of pieces in a specific location. A good account of the results of the experiments is given by the template theory proposed by Gobet and Simon (in press) as an extension of Chase and Simon's (1973a) initial chunking proposal, and in agreement with other recent proposals for modification of the chunking theory (Richman, Staszewski & Simon, 1995) as applied to various recall tasks.

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 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 (common words, for example, are familiar items). The evidence is also strong that experts in a given domain store large numbers of chunks of information that can be accessed quickly, when relevant, by recognition of cues in the task situation. Memory is organized as an indexed data base where recognition makes available stored information of meanings and implications relevant to the task at hand. Many studies of expertise, a domain 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 a review).

Simon and Gilmarin (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, a magnitude roughly comparable to that of natural language vocabularies of college-educated people. This number has been challenged by Holding (1985, p. 109; 1992), who has suggested that the number could be reduced by half by assuming that the same chunk represents constellations of either White or Black pieces¹ and further reduced by assuming that constellations shifted from one part of the board to another are encoded by the same chunk.

As we interpret Holding's view, chunks could be seen as schemas encoding abstract information like: "Bishop attacking opponent's Knight from direction \underline{x} , which is protected by a Pawn from direction \underline{y} ," where the exact location on the board is not encoded. The alternative to his hypothesis is that chunks do encode precise piece locations, and therefore that different chunks would be activated upon recognition of a White pattern

and the identical (except for color) Black pattern, 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, 1994, for a similar view). In order to replace a chunk correctly on the board, information must be available, in one form or another, about the exact location of the chunk.

Quite apart from the task of reconstructing positions, 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).² 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, 1995).

The sensitivity of perception to transformations of stimuli (an aspect of the phenomenon of transfer) has long been a topic of research in psychology. M. Wertheimer (1982) reports children's difficulties in transferring the demonstration of the area of a parallelogram when the figure used during the demonstration is flipped and rotated by 45°. In addition, subjects 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 (Kolars & Perkins, 1975). After a substantial number of hours of practice, however, subjects' speed increases to approximately the level for normal text. 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 locations 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). (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.³ 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 those for 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 noticed because they do not fit the color distribution normally found in chess

positions). In a second set of experiments, Saariluoma (1994) removed 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.⁴ In a recall task, Saariluoma found that subjects recall these positions about as well as game positions. From this experiment he concludes that encoding maintains location information (the chunks within the quadrants appear in the same locations as they would in game positions). 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, which we will later refer to as a template) is not available.

Insert Table 1 about here

Table 1 summarizes the results obtained in experiments on 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 the issue of specificity directly: in this case, the chess relations (mainly attack, defense and proximity) are the same between 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 places on the chess board. Under the hypothesis that

chunks encode relations of proximity, defense and attack between pieces but not their specific location on the chess board, 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 long-term memory (LTM). The same chunk could then also encode constellations like [King on b1 + Pawns on a2-b2-c2] and [King on b8 + Pawns on a7-b7-c7].

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, or with the location of the chunks shifted (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 initiative of the first move, one should expect, on average, that White builds up attacking positions while Black has to choose defensive set-ups, so that different chunks will occur for White and Black pieces, respectively.⁵ We will shed some light on the question by using normal game positions and game positions that have been modified by taking mirror images around horizontal or vertical axes of symmetry, or around center of symmetry.

Four points about our transformations should be mentioned. First we use a transformation by reflection, and not by translation as in Saariluoma's swapping experiment. Second, our transformations do not break up any relations 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 also 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, no transformation manipulates location while keeping both 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).

Because Holding (1985, 1992) does not relate his remarks on chunks to a detailed theoretical model replacing Chase and Simon's model, it is difficult to draw predictions from his views. 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 subjects (Experiment 2).

Experiment 1 (Simulation)

In order to gain a better understanding of the role of mirror image reflections in chess, we have conducted some computer simulations of the reconstruction process, using a simplified version of CHREST (Gobet, 1993a,b), a model of chess players' memory and perception from the EPAM family (Feigenbaum & Simon, 1984; Simon & Gilmarin, 1973).

Methods

Material

A database of several thousand positions from recent Grandmaster games was used as a source of chunks for the learning phase. Fifty new positions, each appearing in the four different permutations, were used for the recall task. In condition 1 of the tests, the position was unchanged (Normal position); in condition 2, it was modified by taking the mirror image with respect to the horizontal axis of the board (Horizontal position); in condition 3, it was modified by reflection about the vertical axis (Vertical position). In condition 4, it was subjected to both modifications simultaneously, that is, reflected through the center of symmetry of the board (Central position). Figure 2 illustrates these four conditions for a particular position.

Insert Figure 2 about here

Procedure

The simplified version of CHREST builds up a discrimination net containing chess chunks from 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, enlarging the net as new patterns were found. Patterns were encoded with indication of their locations on the board. For example, an instance of a short-castled position, a common pattern, was 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 noticed on a board were 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 each 10,000 nodes had been added by learning (more often in the early stages of learning). Learning was halted during the tests. The discrimination nets were progressively extended up to 70,000 nodes. For each position, as during learning, the model randomly fixated twenty squares (twenty fixations take human subjects about five seconds; see De Groot & Gobet, in press) on the board, and sorted the pieces within a range of two squares from the fixated square through the discrimination net. Once the twenty fixations finished, the program compared the contents of the chunks recognized (the internal representation of the chunks) with the stimulus position. The percentage of pieces correct for a trial was the number of pieces belonging to the stimulus position also found, in the correct location, in at least one chunk (erroneous placements were not penalized).

Results

Our main interest is in the relative performance on the different types of positions. As can be seen in Figure 3, the normal positions are slightly better recalled than the horizontally

mirrored ("Horizontal") positions (respective means, averaged over the 14 nets: 65.4% vs. 63.2%). The difference is reliable [$F(1,13) = 19.80$, $MS_e = 3.45$, $p < .005$]. When pooled, normal and horizontal positions are better recalled than vertical and central positions pooled [$F(1,13) = 363.92$, $MS_e = 19.93$, $p < 10^{-9}$]. The recalls of vertical and central positions, respectively 53.3% and 52.5%, on average, do not differ reliably [$F(1,13) = 4.06$, $MS_e = 1.95$, 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 increases monotonically with the number of nodes. The function, $\text{Percentage} = a + b * \log[\text{Number_Nodes}]$ 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, especially around the vertical axis, makes the recall of chess positions harder for the model. In increasing the number of chunks in its net, the model learns some patterns that can appear 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.

The simulation data predict that the identical experiment with human subjects will show 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.⁶ 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, 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.

Experiment 2 (Human subjects)

This experiment was run in two different sites, with slightly different material (see below). As ANOVA detects no interaction of site (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.⁷ Subjects were classified in three groups: Masters (n = 5, mean ELO = 2395, sd = 108), Experts (n = 11, mean ELO = 2146, sd = 69) and Class A Players (n = 9, mean ELO = 1890, sd = 92). Their ages varied from 17 to 45, with mean = 28 and standard deviation = 9. One contingent of players, 12 subjects, were recruited in New York's Manhattan Chess Club, and were paid \$10 for their participation (\$20 for the players having a FIDE title). A second contingent, 13 subjects, 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 subjects also participated in experiment 2 of Gobet (1993a), on the recall of multiple boards. The Swiss subjects also participated in the copy task experiment reported in Gobet and Simon (1994).

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 position (mean number of pieces=25) to squares on the chessboard according to random numbers provided by a computer. Subjects in the first contingent received five random positions, inserted randomly among the experimental positions. Subjects in the second contingent received three random positions, presented at the beginning of the experiment.

Material

First contingent. Twenty positions were randomly selected from various chess books, using the following criteria: (a) the position was reached after about 20 moves; (b) White was to move; (c) the position was "quiet" (i. e. is not in the middle of a sequence of exchanges); (d) the game was played by (Grand)masters, but was obscure. The mean number of pieces was 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 subjects.⁸ Positions were presented on the screen of a Macintosh SE/30, and subjects had to reconstruct them using the mouse. Subjects placed a piece by first selecting it in a rectangular box located on the right of the board and displaying the 6 different kinds of White and Black pieces, and then by clicking it on the appropriate square. This process had to be repeated for each new placement of piece. (For a more detailed description of the experimental software, see appendix in Gobet & Simon, 1994).

Second contingent. Sixteen positions were selected with the same criteria as were used with the first contingent. 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

Subjects 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 use the mouse to reconstruct the positions.⁹ Subjects of the first contingent 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. Subjects of the second contingent received, in order, the copy task (described in Gobet & Simon, 1994), 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 subjects of contingent 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 omit these variables in the following analyses.

Percentage of pieces correct. Post-experimental questioning does not indicate that any 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 [$F(2,22)=24.52$, $MS_e = 401.57$, $p<.001$], of Type of modification [$F(3,66)=20.85$, $MS_e = 44.95$, $p<.001$], and an interaction [$F(6,66)= 2.41$, $MS_e = 44.95$, $p < .05$]. 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 vertical axis differ reliably from positions not modified around this axis [$F(1,22)=96.79$, $MS_e = 108.56$, $p < .001$]. 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. The interaction Skill x Type of position is statistically significant [$F(2,22) = 3.48$, $MS_e = 108.56$, $p < .05$]. This is illustrated in the Figure 4 by the fact that delta (the difference of vertical¢ral positions from normal&horizontal positions) increases with skill, as predicted by the computer simulation. Finally, normal positions do not differ reliably from horizontal positions, nor vertical positions from central positions.

Insert Figure 4 about here

Chunk analysis. As the chunking hypothesis plays an important role in memory models, we analyze in some detail the potential effects of our modifications on the number and size of chunks. Our hypothesis is that the modifications decrease the likelihood of evoking chunks in LTM, affecting 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, 1994) 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 per position, there is no significant effect of type of position [$F(3,66)=1.56$, $MS_e = 2.46$, ns], although (insignificantly) 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 [$F(1,22)= 3.63$, $MS_e = 10.48$, $p=.07$]. There is a statistically significant difference between skill levels [$F(2,22)=4.70$, $MS_e = 22.13$, $p < .05$]. 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 [$F(6,66)=0.67$, $MS_e = 2.46$, ns].

An ANOVA, performed on the number of chunks per position, finds no main effect of the Type of modification [$F(3,66)= 0.77$, $MS_e = 0.65$, 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 [$F(2,22) = 9.03$, $MS_e = 3.99$, $p=.001$]. 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 [$F(6,66)= 0.06$, $MS_e = 0.65$, 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 [$F(2,22)=13.40$, $MS_e = 72.53$, $p<.001$] and a main effect of Type of modification [$F(3,66)=8.54$, $MS_e = 5.50$, $p<.001$]. No interaction is present [$F(6,66)=0.95$, $MS_e = 5.50$, 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 [$F(2,22)=7.65$, $MS_e = 28.69$, $p<.005$]. 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 [$F(3,66)=1.47$, $MS_e = 2.15$, ns] and [$F(6,66)=0.22$, $MS_e = 2.15$, 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 briefly the behavior of our subjects with this material, because the literature does not offer very much information on this topic.

Percentage of pieces correct. Results show the classical recall superiority for game positions vs. random positions [$F(1,22)=291.51$, $MS_e = 66.37$, $p<.001$] and the classical interaction Skill x Type of position [$F(2,22)=16.50$, $MS_e = 66.37$, $p<.001$]. Stronger players tend to recall random positions better, though the effect is not significant [$F(2,22)=0.18$, $MS_e = 32.71$, ns]. Almost all published results show the same pattern: the best players recall slightly more pieces than weaker players (See Gobet & Simon, 1995).

Chunks. The means of the largest chunks are clearly bigger for game positions than for random positions (means for Master, Experts and class A players, respectively: with game positions: 11.2, 7.4 and 6.2 pieces; with random positions: 4.1, 4.3 and 4.1 pieces), and skill differences are found only with game positions. Respective mean number of

chunk per position 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, 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). The corresponding means in errors of commission for Masters, Experts and Class A players are 2.4, 5.5 and 3.9.

Discussion

In this experiment, for all skill levels, subjects have somewhat more difficulty in recalling positions modified by vertical or central reflection than positions modified by horizontal reflection or unmodified positions. None of the modifications decreases the recall percentage nearly 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 correspond closely with those obtained in the simulations, in which location was specified for all patterns that were stored. This suggests strongly 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, an empirically supported explanation of the processes involved.

The deterioration of the subjects' performances with mirror image reflections of the positions, in close quantitative agreement with the deterioration observed in the simulations, and taken together with Saariluoma's (1994) results, run counter to Holding's (1985, 1992) 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 its pattern-recognizing component, 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 refuted 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 per reconstruction is well within the postulated number of visual chunks, four (Zhang & Simon, 1985). We also found that, in general, Masters replace large chunks, sometimes 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.

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, by chance, be those that occur in normal games, and their recall may therefore be explained by an access to

LTM. However, inspection shows that such an explanation does not hold for all chunks, and we must conclude that chessplayers may use special strategies to recall pieces on a board that is almost bare of familiar patterns. For example, they may use multiple slots in short-term memory (STM) to store descriptions of patterns on the board (e.g., "| Three | White Pawns | on a diagonal | starting from a1 |"), which would account for the fact that, on average, only one or two chunks, as defined by the 2-second boundary, are stored per position. In any event, total recall of random positions is no more than if subjects stored information about one distinct piece in each STM slot.

General Discussion

In these experiments, we have examined differences between memory for normal game positions and for positions modified by reflection around an axis (horizontal, vertical, or both) as well as 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 whether 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 human experiment and computer simulation with modified boards do not indicate the presence of location-free or color-independent chunks. 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 general, the experiment with chessplayers and that with computer simulations, using only location-specific chunks, yielded effects of about the same magnitude.

The fact that recognition for recall depended on these location-specific chunks does not imply that all information retrieved using such cues is equally specific. A number of different cues may signal the appropriateness, in a chess game or other task, of the same general response: an attack on the King's position, say, or the advance of pawns. To explore these possibilities one would have to present the subjects with different tasks than simple recall or position reconstruction tasks.

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 cell: mirror image reflection, which retains the overall relations between pieces but not their locations, produces a small 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.

Gobet and Simon (in press), building on the research on expertise of Chase, Ericsson and Staszewski (e.g., Chase & Ericsson, 1982), have modified the earlier Chase and Simon model by including templates among the chunks that chessplayers store in long-term memory. Templates are simply chunks having slots in which some additional information can be stored rapidly. Templates, typically describing positions that arise out of common chess openings and therefore familiar to Masters and (to a lesser extent) Experts, contain fixed information (their core: about a dozen chess pieces), and slots,

serving as variables. Additional information can be inserted relatively quickly in slots about a specific position belonging to the type represented by the template (say, information about three or four chunks of pieces). Templates employ the same mechanisms as are postulated for retrieval structures, structures for which there is substantial experimental evidence in other memory tasks (Chase & Ericsson, 1982; Richman et al., 1995).

Templates are evoked when a position is recognized as being of a certain familiar type (the Panov-Botvinnik Attack, say, in the Caro-Kann Defense). When shown a game position for a few seconds, a Master will first recognize a few chunks, which will usually evoke a template. After retrieval of the template, default values may be rapidly corrected and then other slots instantiated. Because templates are complex data structures, it takes a long time (perhaps of the order of 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 Masters to have several thousand, even for types of positions they seldom meet in their own tournament practice.¹⁰

In the recall of positions modified by mirror image reflection, the template theory predicts, as Chase and Simon's theory does, that unmodified positions will be better recalled than reflected positions, the latter being likely to evoke fewer and smaller chunks (and templates), and consequently to cause more errors by omission. 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 in the game positions; the near-absence of these patterns from the boards with randomly placed pieces reduces this advantage.

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 templates recognized may encode incorrectly the location of a few pieces, but, because they do not possess a sufficient number of slots to correct them, this may lead 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. A probable reason 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 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, 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 subjects use and by the Masters' repertoires 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. These findings about expert memory in chess are consonant with other recent models of expert memory in a variety of tasks.

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.
- De Groot, A. D. & Gobet, F. (in press). *Perception and memory in chess: Heuristics of the professional eye*. Assen: Van Gorcum.
- Ericsson, K. A., & Smith, J. (Eds.). (1991). *Studies of expertise : Prospects and limits*. Cambridge: Cambridge University Press.
- Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science*, 8, 305-336.
- Glickman, M. E. (1994). *Report of the USCF ratings committee*. US Chess Federation.
- 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. & 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. & Simon, H. (in press). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*.

Gobet, F. & Simon, H. (1994). *Expert chess memory: Revisiting the chunking hypothesis*.
CIP #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.

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.

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.

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

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

Wertheimer, M. (1982). *Productive thinking* (enlarged edition). Chicago: The University of Chicago Press.

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

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.

Footnotes

¹Holding 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).

²Following standard chess practice, squares are designated by a letter for the column and a number for the row, from the viewpoint of the player of the White pieces. The columns (files) are lettered from left to right, the rows (ranks) numbered from front to back. Thus c1 is the third column, first row -- the square where White's Queen's Bishop stands at the beginning of the game.

³**Chess players are classified into skill levels according to the Elo rating, a system internationally used. Grandmasters are usually rated above 2500 Elo, International Masters above 2400, Masters between 2200 and 2400, Experts between 2000 and 2200, Class A players between 1800 and 2000, Class B players between 1600 and 1800, and so on.**

⁴Note 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.

⁵This 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".

⁶In 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.

⁷USCF ratings are in general higher than international ELO ratings. We have used the table proposed by Glickman (1994) to convert the USCF ratings into international ratings.

⁸The 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 stands for Normal, H for Horizontal, V for Vertical C for Central, and R for Random.

⁹One Expert, who had difficulties in manipulating the mouse, used algebraic notation to dictate the positions to the experimenter, who handled the mouse.

¹⁰Handbooks on the openings, which Masters study assiduously, contain thousands of lines of play that can be stored as templates.

Table 1

Overall Relations, Location and Recall Performance as a Function of the Type of Transformation 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 and Errors of Commission as a Function of Skill Level and Type of Modification

Skill level	Type of position			
	Normal	Horizontal	Vertical	Central
Errors by omission				
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				
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)

Note. Standard deviations are given in parentheses.

Figure captions

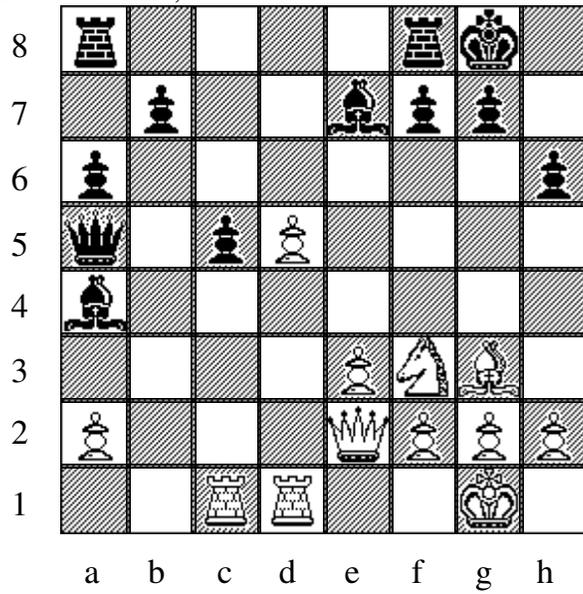
Figure 1. Example of Saariluoma's (1991) position modification by swapping two quadrants: (a) before the swapping; (b) after the swapping.

Figure 2. Example of the kinds of positions used in experiment 1 and 2. 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, central and random positions as a function of the number of nodes in the discrimination net. Delta indicates the difference of vertical¢ral positions from normal&horizontal positions.

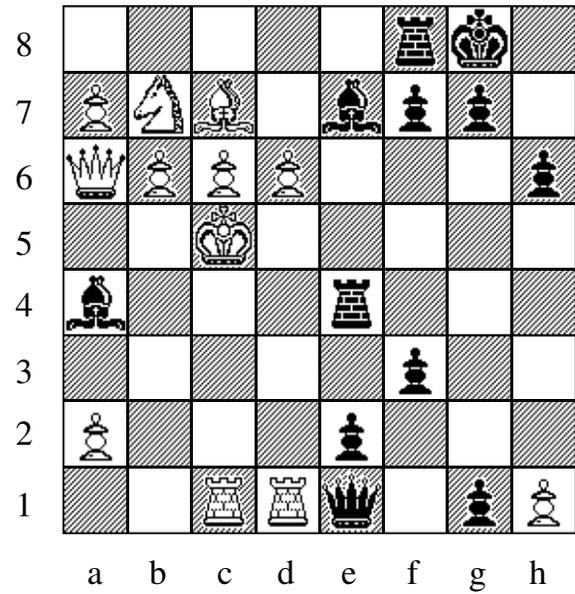
Figure 4. Mean percentage of pieces correct as a function of chess skill and type of position. Mean percentage with random positions is shown for comparison sake. Delta indicates the difference of vertical¢ral positions from normal&horizontal positions.

November 24, 2007

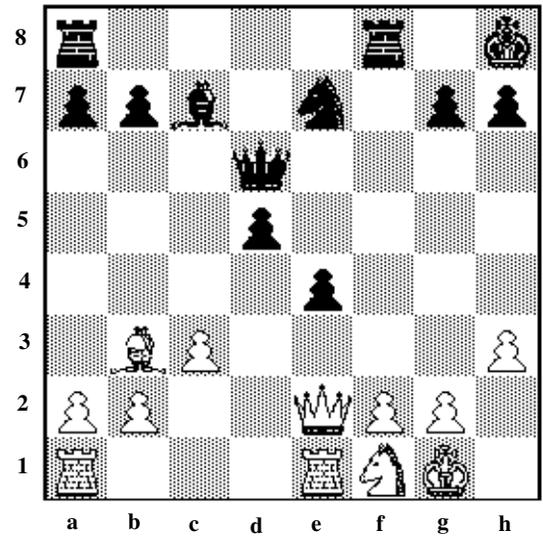
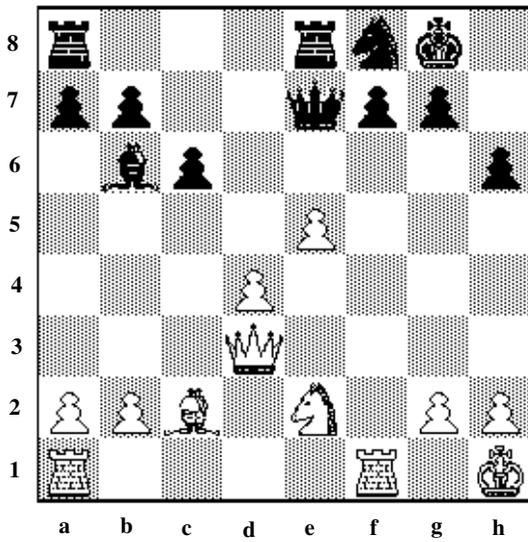


a) Original position.

33

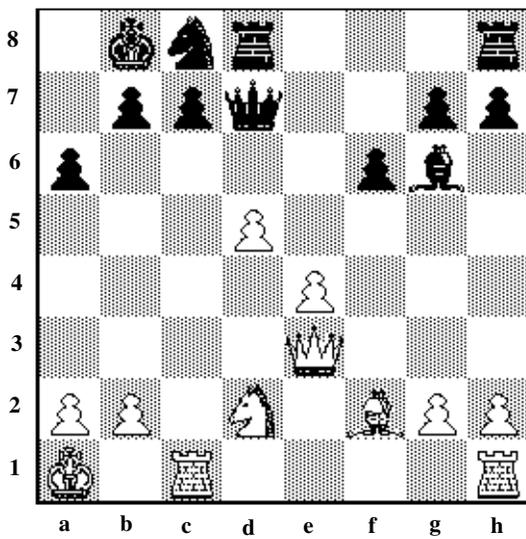


b) Position obtained after swapping the upper left quadrant with the lower right quadrant.

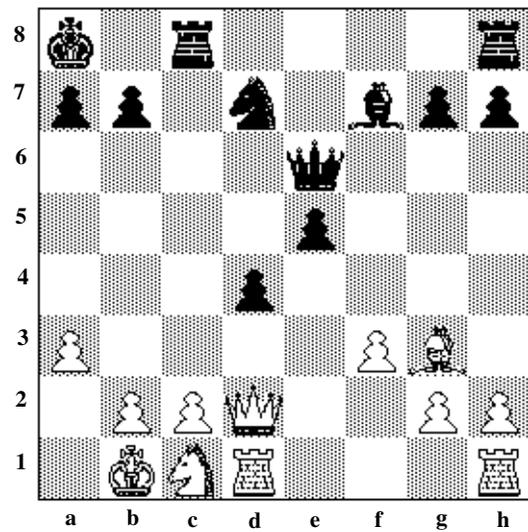


a) original position
(normal)

b) after horizontal axis reflection



c) after vertical axis reflection



d) after central axes reflection

