Five Seconds or Sixty?

Presentation Time in Expert Memory

[in press, *Cognitive Science*]

Fernand Gobet

ESRC Centre for Research in Development

Instruction and Training

Department of Psychology

University of Nottingham

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: Presentation time in expert memory

Abstract

The template theory presented in Gobet and Simon (1996a, 1998) is based on the EPAM theory (Feigenbaum & Simon, 1984; Richman et al., 1995), including the numerical parameters that have been estimated in tests of the latter; and it therefore offers precise predictions for the timing of cognitive processes during the presentation and recall of chess positions. This paper describes the behavior of CHREST, a computer implementation of the template theory, in a task when the presentation time is systematically varied from one second to sixty seconds, on the recall of both game and random positions,

and compares the model to human data. As predicted by the model, strong players are better than weak players with both types of positions. Their superiority with random positions is especially clear with long presentation times, but is also present after brief presentation times, although smaller in absolute value. CHREST accounts for the data, both qualitatively and quantitatively. Strong players' superiority with random positions is explained by the large number of chunks they hold in LTM. Strong players' high recall percentage with short presentation times is explained by the presence of templates, a special class of chunks. The model is compared to other theories of chess skill, which either cannot account for the superiority of Masters with random positions (models based on high-level descriptions and on levels of processing) or predict too strong a performance of Masters with random positions (long-term working memory).

Five Seconds or Sixty?

Presentation Time in Expert Memory

The study of expert behavior offers a unique window into human cognition. By looking at extreme performance, cognitive scientists determine what parameters of the human information-processing system can change with extensive practice, what parameters are relatively stable, and what strategies can be used to overcome the limits imposed by these stable parameters. As experts must abide by the known limits of cognition, the information gained from them can be generalized to non-experts, who, except perhaps for individual variations in the stable parameters, could become experts given enough practice and study in a domain.

Ever since the seminal research of De Groot (1946/1978), expertise in chess has provided cognitive science with a wealth of empirical data and theoretical developments. Several factors explain the important role played by chess research, including: complexity of the task, which requires nearly a decade of training to reach professional level; flexibility of the chess environment, allowing many experimental manipulations; presence of a quantitative scale to rate players (the Elo rating; Elo, 1978); and cross-fertilization with research in artificial intelligence. As several of the phenomena first encountered in chess research, such as selective search, importance of pattern recognition, progressive deepening and experts' remarkable memory (Chase & Simon, 1973a, 1973b; De Groot, 1946/1978) have been found to generalize to other domains of expertise, the chess domain offers strong external validity (Charness, 1992; Gobet, 1993a). Chess research, which employs players who have spent thousand of hours practicing their art, also offers strong ecological validity. Even Neisser (1998, p. 174), who has repeatedly emphasized the importance of studying ecologically valid tasks, notes that "I have always regarded the study of chess and similar skills as part of the naturalistic study of cognition (Neisser, 1976, p. 7)."

In this paper, we provide new empirical evidence about the rates at which highly expert and less expert chess players acquire information about chess positions that they observe, and how their learning rates depend on a position's resemblance to those encountered in actual games between skilled players. We show how these empirical data can be closely approximated by a computer model, CHREST, which is a variant of the EPAM model of perception and memory with memory contents relevant for chess. We believe that the phenomena we report, and the mechanisms we propose to account for them are, for the reasons mentioned above, distinctly relevant to understanding the role of expertise, not only in chess, but in other domains of human expert professional activity as well.

The Chunking Theory

Experts in various domains vastly outperform novices in the recall of meaningful material coming from their domain of expertise. To account for this result, Chase and Simon (1973a, 1973b) proposed that experts acquire a vast database of chunks, containing, as a first estimate, 50,000 chunks. When presented with material from their domain of expertise, experts recognize chunks and place a pointer to them in short-term memory (STM). These chunks, each of which contains several elements that novices see as units, allow experts to recall information well beyond what non-experts can recall. Aspects of the

chunking theory, which derives from the EPAM theory of memory and perception (Feigenbaum & Simon, 1962; 1984), were implemented in a computer program by Simon and Gilmartin (1973).

Intensive research in skilled memory has shown that parts of the original chunking model were wrong. For example, in contrast to the usual assumptions about STM, chess masters are relatively insensitive to interference tasks (Charness, 1976; Frey & Adesman, 1976) and can recall several boards that have been presented successively (Cooke, Atlas, Lane & Berger, 1993; Gobet & Simon, 1996a). In addition, Chase and Ericsson (1982) and Staszewski (1990) have shown that highly trained subjects can memorize up to 100 digits dictated at a brisk rate (1 second per digit). Because an explanation of this performance based on chunking requires learning far too many chunks, Chase, Ericsson and Staszewski proposed that these subjects have developed structures ("retrieval structures") that allow them to encode information rapidly into long-term memory (LTM). Such structures have been used at least since Roman antiquity, when orators used such mnemonics as the method of loci to memorize their speeches (Yates, 1966), and there is direct evidence, both from historical materials and from Chase, Ericsson and Staszewski's experiments that retrieval structures are an essential aid to expert memory performance.

Richman, Staszewski and Simon (1995) have proposed a detailed model of the way DD, one of the subjects studied by Chase, Ericsson and Staszewski, developed his digit-span skill. The model adds to the EPAM theory, which accounts for a wide range of data in memory and perception (Feigenbaum & Simon, 1984; Richman & Simon, 1989), the idea of a retrieval structure. The augmented EPAM model was able to simulate in detail most aspects of DD's learning and performance.

Alternatives to the Chunking Theory

Several explanations have been proposed to account for the "anomalies" in chess recall—the lack of interference by interposed tasks and the simultaneous recall of several boards. Charness (1976) has proposed that Masters use more sophisticated codes to store positions; Lane and Robertson (1979) argue that Masters process positions more deeply, in the sense of Craik and Lockhart (1972); Cooke et al. (1993) propose that Masters encode positions using high-level descriptions, such as the type of opening the position is likely to stem from.

These explanations are of dubious explanatory or predictive value for two reasons. First, they are stated in very vague terms, and it is unclear how they relate to a general model of (chess) expertise. No detailed mechanisms are specified, nor time parameters for the processes, nor precise capacity parameters for the memories. Their ambiguity makes it easy to apply them, without awareness, in an inconsistent fashion to different situations. Hence, it is often quite unclear what behavior they predict in specific experimental paradigms.

Second, Gobet and Simon's review of the literature (1996c) shows that, contrary to a widely held belief, strong players outperform weaker players to some extent in the recall of random positions presented for a short time (10 seconds and less). It is unclear how the three explanations mentioned above could explain this superiority without postulating any *ad hoc* addendum. By contrast, Chase and Simon's (1973b) modified theory, as simulated by Gobet and Simon (1996b), offers a ready explanation: the larger databases of chunks that strong players possess allow them to recognize more often the familiar chess patterns (typically of two or three pieces) that sometimes occur by chance in random chess positions.

Random positions are important because they offer a stronger test for theories of expert memory than game positions. It is relatively easy to propose explanations for experts' superiority with meaningful material, because this relates to what Newell (1990) calls the knowledge level. It is harder to come up with explanations about the weaker effects found with random positions, which tap both the knowledge level and the symbolic level, and their realization in an architecture. In particular, making predictions about the small effects found with random positions requires a theory able to make quantitative predictions.

As noted above, research on expert memory in recent years has revealed another mechanism besides

simple chunking that is available to experts. Experts can, by practice and training, acquire retrieval structures in LTM that contain not only the usual kinds of information found in chunks, but also contain, as components of their chunks, "slots" or variables with which new information can be associated at a rate of about one item per quarter second. Because of the speed with which slots can be filled, these retrieval structures, acquired through extensive experience in the task domain, provide what is essentially additional STM capacity for tasks in the domain of expertise. Two theories using such a mechanism have been proposed to account for the anomalies in chess recall.

Ericsson and Kintsch (1995), building on Chase and Ericsson's (1982) theory of skilled memory, have proposed a particular version of the notion of retrieval structures as applied to chess. They postulate specifically that experienced chess players acquire in LTM a hierarchical retrieval structure that is mapped onto the 64 squares of a chess board. Retrieval cues can be stored in the squares of the retrieval structure, or at higher levels of the hierarchical structure, allowing a rapid encoding into LTM. As argued in Gobet and Simon (1996a), however, such a model predicts a much too high recall of random positions (as individual pieces could rapidly be associated with individual squares). In addition, the theory, not being incorporated in an operable simulation model, leaves unspecified many details of the mechanisms used to store and retrieve information and the times necessary to carry out their processes (Gobet, 1998a), and therefore cannot make clear quantitative predictions or sometimes even qualitative ones.

The other approach involving retrieval structures, the template theory (Gobet & Simon, 1996a), uses the same framework as the EPAM theory (Richman et al., 1995), with which it is compatible. In its application to chess, it is implemented in the CHREST program, providing a means for testing it unequivocally and unambiguously against experimental data. As in the earlier chunking theory of Chase and Simon (1973b), it assumes that chess experts develop a large EPAM-like net of chunks during their practice and study of the game. In addition, it assumes that some chunks, which recur often during learning, develop into more complex schematic retrieval structures (templates) having slots for variables that allow a rapid encoding of chunks or pieces.

Templates, which generally represent familiar opening positions after 10 or 15 moves, can store more pieces than the chunks hypothesized in Chase and Simon's model, which seldom exceeded about four or five pieces. When a game position is recognized (say, as a King's Indian defense), the corresponding stored representation of the chess board provides specific information about the location of a number of pieces (perhaps a dozen) together with slots which may possess default values ("usual" placements in that opening) that may be quickly revised (see Figure 1 for an example of the kind of templates created by CHREST). This access to more powerful retrieval structures explains the large initial chunks that are, in fact, recalled by Masters (and, to a lesser extent, Experts) in game positions (Gobet & Simon, 1998). For templates, cued by salient characteristics of the position, would be likely to be recognized early and enlarged quickly by filling slots and altering incorrect default values. (The model will be presented in more detail later, in the introduction to the computer simulations.)

Figure 1. A template acquired by CHREST (with a net of 300,000 chunks). The core contains constant information and the slots contain variable information. Instantiated values are in angle brackets. The diagram on the left shows a stimulus position, and the diagram on the right shows what information could be provided by the template above (the exact instantiation of the slots depends on the order of eye movement and of the time available to store values). The marked squares indicate that slots are used to store information.

Core:

Pc4 Pe4 Pf3 Pg2 Ph2 Be3 Nc3

pc6 pd6 pf7 pg6 nc5 bg7

Slots:

Square-Slots   Piece-Slots

c5 : -   white rook : <e1>

d4 : -   white bishop : <f1>

g1 : <white king>   white knight : -

c3 : -   black bishop : -
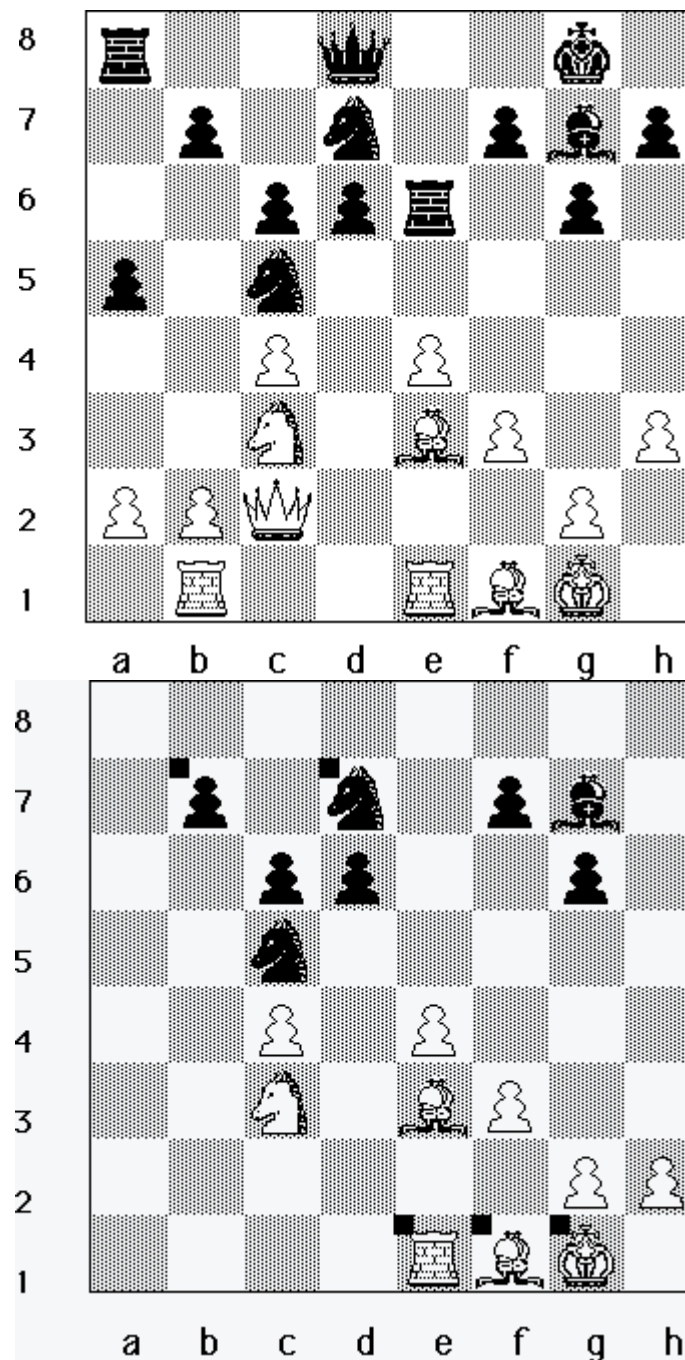
e5 : -   black knight : <d7>

e3 : -   black pawn : <b7>

f7 : -

The model accounts for the lack of interference from intervening tasks by assuming that subjects either keep a pointer to one template (or chunk) in STM during the completion of the interfering task or, given sufficient time (about 8 s), learn an additional retrieval path to this template in the EPAM net. It thereby accounts for the recall of multiple boards (and subjects' difficulties with more than four or five such boards) by the fact that new cues (new branches in the discrimination net) can be learned only slowly, and that pointers to templates have to be kept in STM, thereby interfering with the processing of new positions. Finally, the model does not predict the same over-performance in random positions as Ericsson and Kintsch's, because patterns of pieces complex enough to match templates are almost never present in random positions (see Gobet & Simon, 1996b, c).

A note about terminology may be appropriate here. We use "schema," consistently with its use in cognitive psychology since Bartlett (1932), as a general term for any LTM cognitive structure. Structures whose names can be stored in STM as units are called "chunks." Schemas include "retrieval structures," which are learned with the specific intention of providing sets of "slots" (variable places) in which specific chunks of information can be stored rapidly. "Templates" are schemas implicitly learned in the process of acquiring substantive knowledge, which also contain slots that can be used for rapid augmentation.

The use of a computational implementation offers three advantages with respect to the concept of schema and its specializations: (a) the concepts have precise definitions; (b) mechanisms are proposed to explain how schemata are created; and (c) mechanisms and time constraints are proposed to explain how schemata are used in memory tasks.

Role of Presentation Time in Chess Memory

Most research on chess memory has focused on the recall of positions presented for a short interval of time and relatively little is known about the effect of varying presentation time. With game positions Charness (1981), Saariluoma (1984) and Lories (1987) show that an increase in presentation time facilitates recall, as was found with recall of other materials (Kintsch, 1970). However, it has been difficult to fit mathematical functions to the relation between presentation time and recall percentage because these studies provided only a few data points (three in Charness, 1981; at most four in Saariluoma, 1984; two in Lories, 1987). With random positions, Saariluoma (1989), using auditory presentation at a rate of 2 seconds per piece (a total of about 50 seconds per position), found that stronger players achieve better recall than weaker players. Lories (1987) found the same effect of skill with semi-random positions and visual presentation for 1 minute.

The EPAM model, which is at the basis of the template theory, has been used extensively in simulations of data on verbal learning (Feigenbaum, 1963; Feigenbaum & Simon, 1962). It is important to see whether EPAM's assumptions about learning— especially, that 8 seconds are needed to learn a new chunk—also lead to correct predictions about chess experts' memory. In addition, data on short-term learning in chess offer fine-grained observations that may tease apart the component mechanisms of the process. In particular, we are interested in the detailed predictions of the template theory. Roughly, this theory predicts that recognition of chunks in game positions should insure strong players' superiority even with short presentation times (say, one second), and that strong players should reach their maximal recall rapidly, owing to the possibility of filling slots in applicable templates. In addition, it predicts that, in random positions, strong players should recall more than weaker players, because they are more likely to recognize (small) chunks in such positions. This superiority should increase with presentation time, because longer times allow masters to combine chunks of pieces recursively, while weaker players have only the opportunity to chunk individual pieces. Finally, we are interested in the sizes of chunks produced by the model, in game as in random positions, as compared with human players.

In the remainder of this paper, we compare the human data on the recall of game and random positions presented for 1 to 60 seconds with the predictions of the template theory. We first present the results of human subjects, ranging from Class A players to Masters. Then, after describing CHREST, the implementation of the template theory, in some detail, we report how the simulations were run and the

results obtained. To anticipate the conclusion of the paper, the human data match the predictions of the CHREST model very well.

Human Experiment

Methods

Subjects

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

Subjects were recruited at the Fribourg (Switzerland) chess club, during the Biel Festival and in the Carnegie Mellon University community. The subjects were paid the equivalent of $10 ($20 for players having an official title) for their participation.

Materials

Nineteen game positions were selected from various chess books with 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 was 25. Ten random positions were created by assigning the pieces from a game position to random squares on the chessboard.

Positions were presented on the screen of a Macintosh SE/30, and subjects had to reconstruct them using the mouse. (For a detailed description of the experimental software, see Gobet & Simon, 1998). The screen was black during 2 seconds preceding display of the blank chessboard on which the subject was to reconstruct the position. No indication was given of whether White or Black was playing the next move, and no feedback was given on the correctness of placements.

Design and Procedure

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

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

Results

Percentage of Correct Pieces

1) <u>Game positions</u>. The upper panel of Figure 2 shows the performance expressed as percentage of pieces replaced correctly. The Masters' superiority is obvious. In 1 s, they are at about the same level of performance as Experts are after 10 s, and they perform only slightly worse than Class A players after 30 s. Note also that, while Class A players and Experts continue to improve their scores, Masters approach a ceiling rapidly, after about 3 s. The three skill levels differ statistically significantly at presentation times of 10 s or less: $\underline{F}(2,16) = 27.63$, $\underline{p} < .001$; so do Experts and Class A players with presentation times above

10 s: $\underline{F}(1,13) = 10.92$, $\underline{p} < .01$.

<u>Figure 2.</u> Percentage of correct pieces as a function of presentation time and chess skill for game positions (upper panel) and random positions (lower panel). The best fitting exponential growth function is also shown for each skill level.
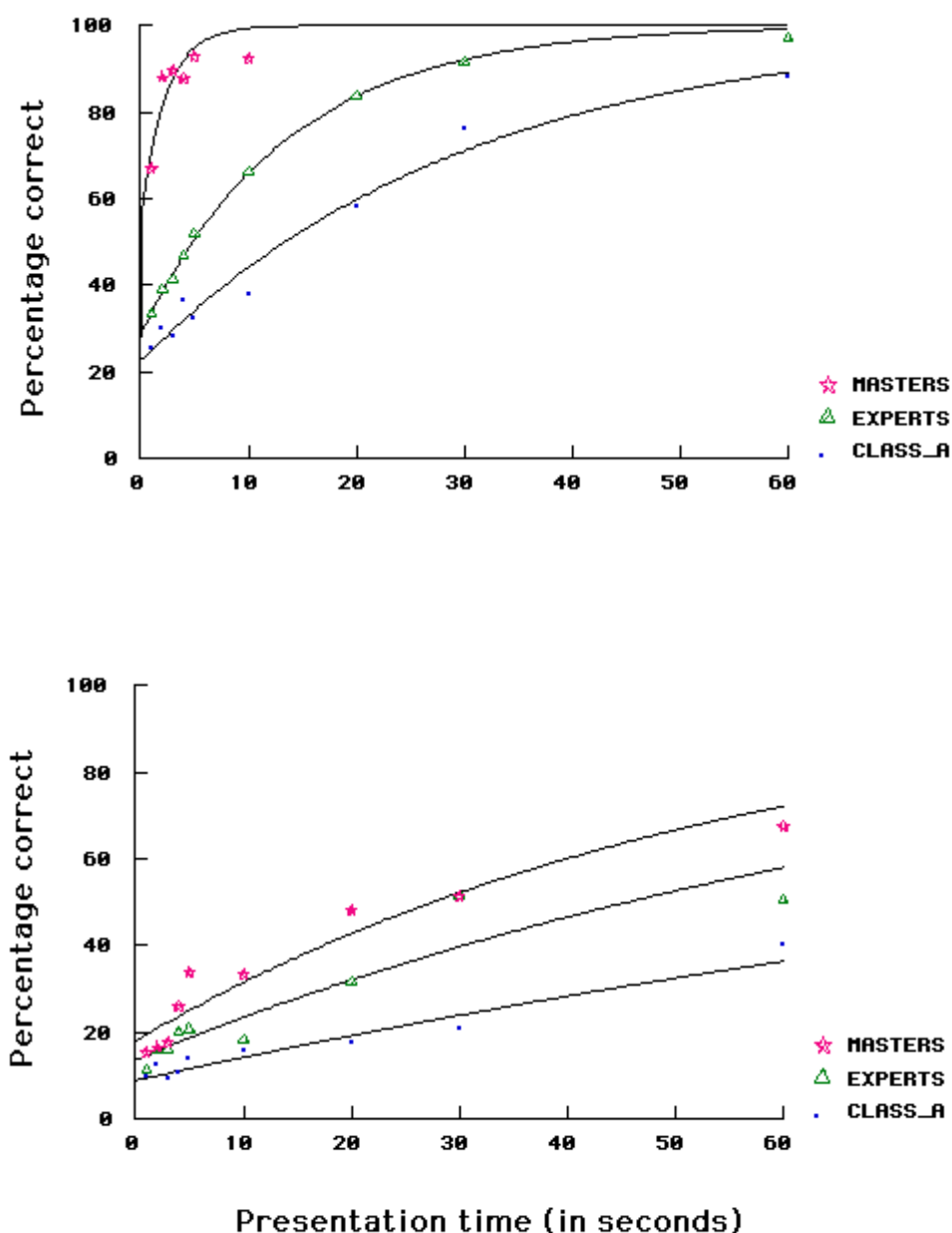
How do Masters achieve their superiority? Do they have only a perceptual advantage, already evidenced at short presentation times and produced by the availability of more and bigger chunks in LTM for recognition, or are they also able to profit from the supplementary presentation time to learn new chunks and to encode new information into the variable slots in LTM (a learning advantage)? We have fitted some simple functions to the data whose parameters can shed light on these questions. A power law (average r2 for Experts and Class A players = .67) and a logarithmic function (average r2 = .65) fit the data reasonably well, better than a simple linear regression line (average r2 = .58). However, the best fit was provided by the logistic growth function,

$$P = 100 - Be_{-c(t-1)} \quad (1)$$

where P is the percentage of correct answers, (100-B) is the percentage memorized in 1 second, c a

constant, and t the presentation time, in seconds. The average r^2 for Experts and Class A players is .69. The r^2 for Masters is .35.

With this function, the rate at which additional pieces are stored after one second is proportional to the number of pieces not already stored. Table 1 gives the parameters fitting the data best, for the three skill levels, and Table 2 indicates the goodness of fit. One sees that both (100-B), the percentage learned in 1 s, and the subsequent learning rate, c, increase with skill. The parameter c increases by a factor of 2 from class A to Experts, and by a factor of 15 from class A to Masters. The parameter (100-B) increases by a factor of 1.3 from class A to Experts, and a factor of nearly 3 from class A to Masters.

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

Table 1.

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

		Game positions			
		95% Confidence Interval			
	Parameter	Estimate	ASE1	Lower	Upper
Class A	B	75.176	3.227	68.713	81.639
	c	0.033	0.005	0.023	0.042
	100 – B	24.824			
Experts	B	66.414	3.154	60.124	72.704
	c	0.074	0.011	0.051	0.097
	100 – B	33.586			
Masters	*B*	*29.240*	*4.851*	*19.303*	*39.178*
	c	*0.435*	*0.169*	*0.089*	*0.781*
	100 – B	70.760			

		Random positions			
		95% Confidence Interval			
	Parameter	Estimate	ASE1	Lower	Upper
Class A	B	90.617	1.566	87.481	93.753
	c	0.006	0.001	0.004	0.008
	100 – B	9.383			
Experts	B	85.338	2.157	81.035	89.641
	c	0.012	0.002	0.008	0.015
	100 – B	14.662			
Masters	B	80.539	2.706	75.040	86.038
	c	0.018	0.003	0.012	0.024

	100 – B	19.461			

1Asymptotic Standard Error

Table 2.

Recall percentage as a function of time presentation (human data). Goodness of fit obtained with the function P = 100 - Be-c(t-1) and the parameters estimated in Table 2.

<div align="center">

Goodness of fit

r2 using all data points r2 using group means

</div>

	Games	Random	Games	Random
Class A	.66	.51	.98	.95
Experts	.72	.50	.99	.87
Masters	.35	.69	.71	.93

2) <u>Random positions.</u> The three skill levels differ significantly for short presentation times of random positions (ten seconds or less): $\underline{F}(2,15) = 7.74$, $\underline{p} < .005$, as well as for long presentation times (more than ten seconds): $\underline{F}(2,16) = 14.36$, $\underline{p}<.001$. We have fitted the data from the random positions with the same logistic growth function (Figure 2, lower panel). The parameters for best fit are shown in Table 1, lower panel, and the goodness of fit in Table 2. Again, skill levels differ both in the amount of information acquired quickly (100-B) and in the continuing rate of acquisition (c). The parameter c doubles from class A to Experts, and triples from class A to Masters. The parameter (100-B) increases by a factor of 1.5 from class A to Experts, and doubles from class A to Masters. Notice that the <u>relative</u> superiority after 1 second of Masters over Experts and especially Class A players is slightly greater for game positions than for random positions; and the superiority of Masters in learning rate (percentage of remaining pieces learned per second) for longer intervals is much larger for game positions than for random positions.

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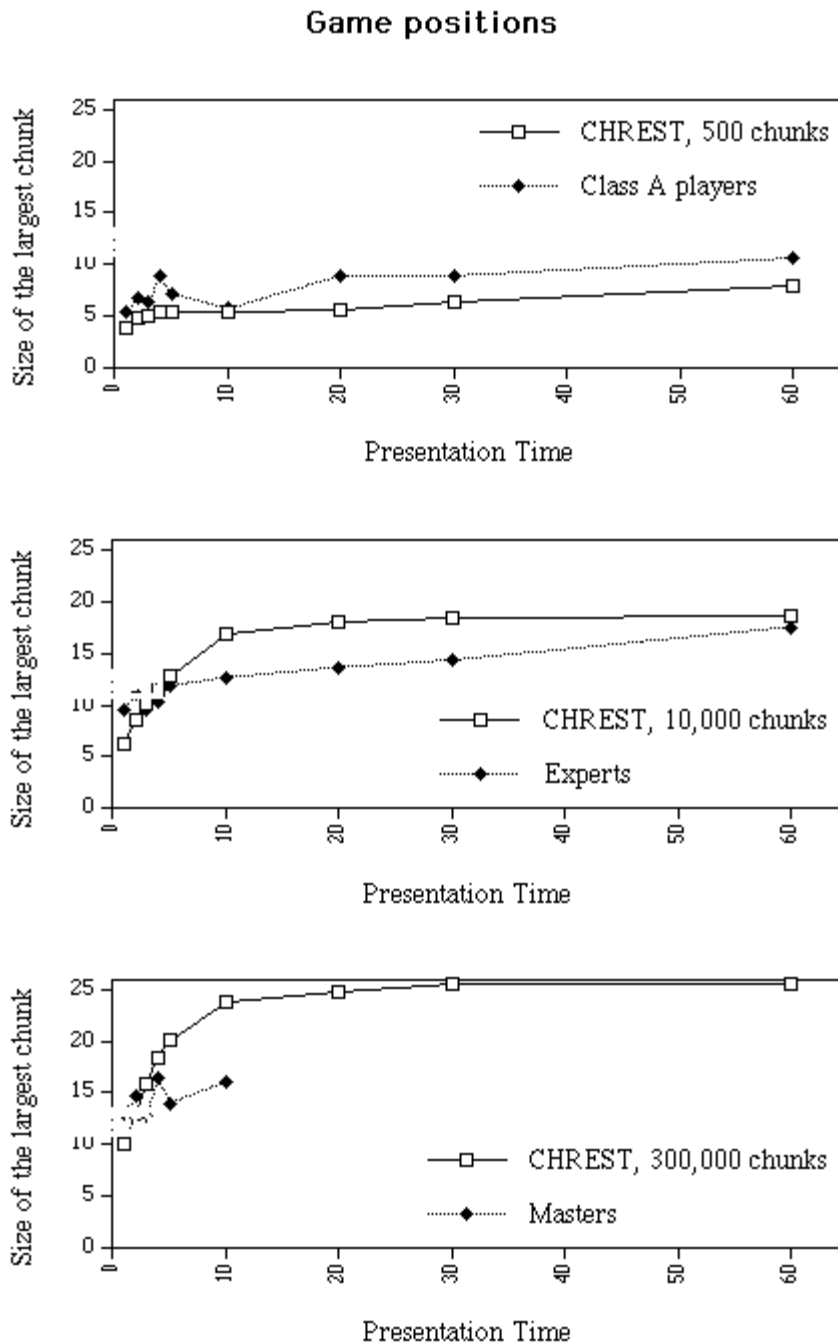
Chunks

We have also analyzed the distributions of chunk sizes. As in Chase and Simon (1973a), chunks are defined as sequences of pieces having corrected latencies of less than 2 seconds between successive pieces (see Gobet & Simon, 1998, for an in-depth discussion of the operational definition of chunk and the method for correcting latencies). Pieces placed individually are not classified as chunks, as they often are the product of guessing. Chase and Simon (1973a), as well as Gobet and Simon (1998), have shown that the 2-second boundary is strongly supported by converging evidence both from the distribution of latency times in a copy task and from the contrast of the many within-chunk chess relations with the relatively few between-chunk chess relations. We use two variables describing chunks, which are both theoretically important: (a) size of the largest chunk, which has direct implication for evaluating the template theory, and (b) number of chunks, which allows us to estimate the size of visual STM.

1) <u>Game positions.</u> Our first variable is the size of the <u>largest chunk</u> per position. Figure 3a shows that Masters reconstruct large chunks, even after a one-second view of the board, and that the size of the largest chunk does not increase much with additional presentation time (the model's data, plotted for

comparison, will be explained below). Experts' largest chunks start with about 9-10 pieces, and increase almost linearly up to about 17 pieces at 60 seconds. A comparable increase may also be seen with class A players, who start, however, with smaller maximal chunks (about 4-5 pieces). Statistical tests indicate that the three groups differ with each other for the presentation times of 10 seconds or less: $\underline{F}(2,16) = 6.30$, $\underline{p} = .01$. Experts still differ from Class A players at longer presentation times: $\underline{F}(1,13) = 6.11$, $\underline{p} < .05$.

<u>Figure 3.</u> Largest chunk (in pieces) per position as a function of presentation time and chess skill (or number of chunks in the net) for game positions and random positions, for human data and computer simulations.
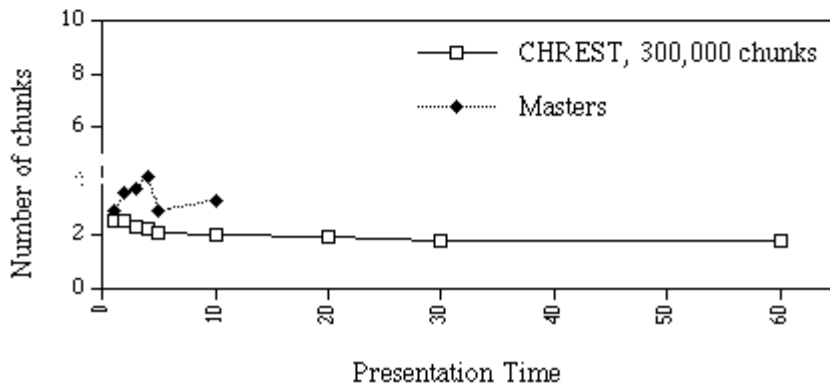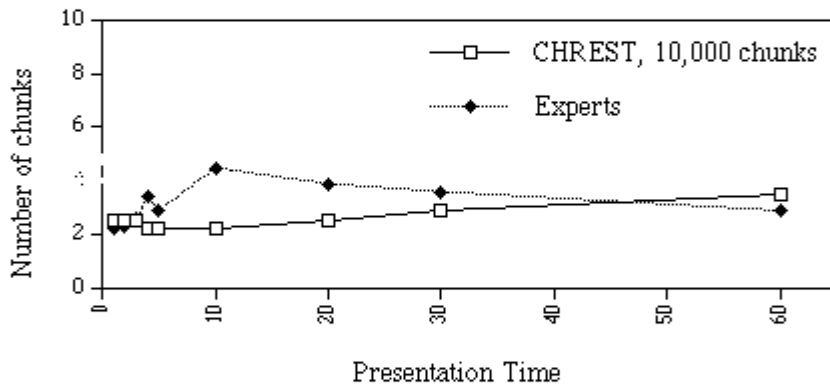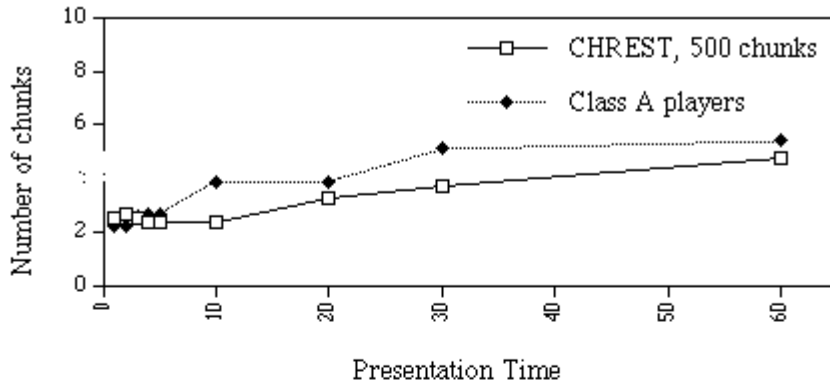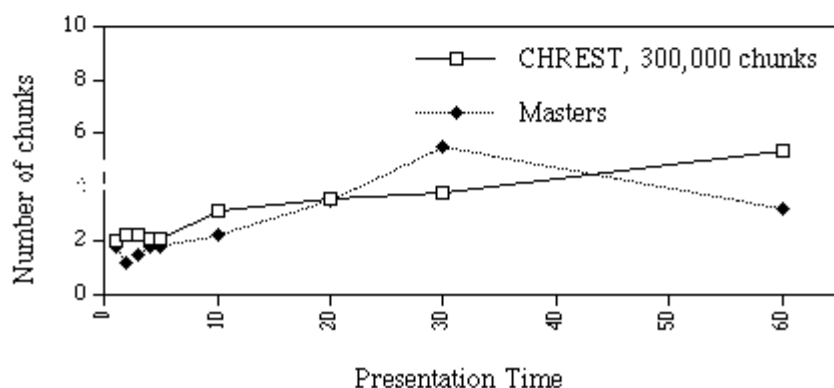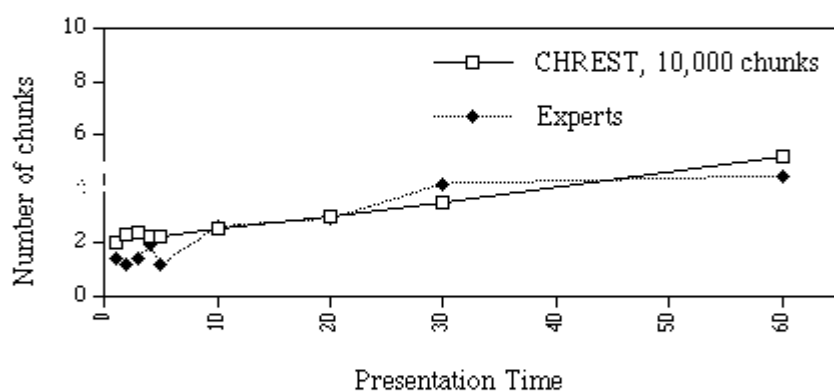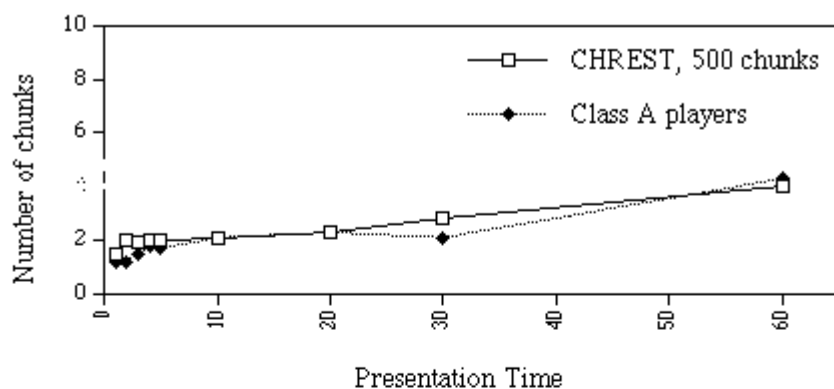
Random positions
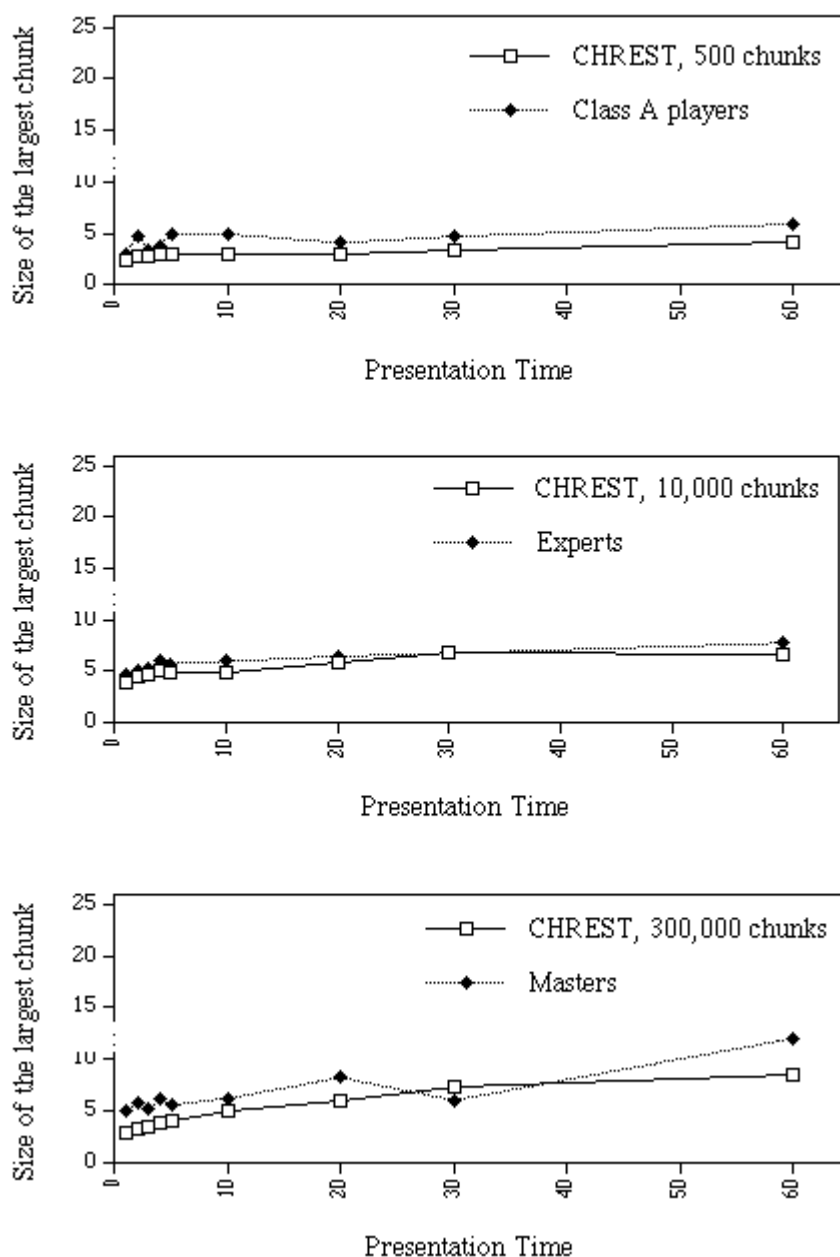
The second variable, the <u>number</u> of chunks, shows two different patterns, the first displayed by Masters and Experts, the second, by Class A players (see Figure 4a). Masters' numbers of chunks increase from 2.9, at one second to 4.2 at four seconds, and then decrease to 3.3 at 10 seconds. Experts start with 2.2 chunks at a presentation time of 1 second, show a maximum with 4.5 chunks at 10 seconds, and then decrease to 2.9 at 60 seconds. Class A players increase the number of chunks from 2.2 at 1 s to 5.4 at 60 seconds. Numbers of chunks for the three skill levels do not differ reliably at presentation times of 10 seconds or less [$F(2,16) = 0.40$, ns.]. By contrast, Experts differ from Class A players at presentation times longer than 10 seconds [$F(1,13) = 8.34$, $p < .05$].

<u>Figure 4.</u> Number of chunks per position as a function of presentation time and chess skill (or number of chunks in the net) for game positions and random positions, for human data and computer simulations.

Game positions

Random positions

2) <u>Random positions.</u> For the <u>largest chunks</u>, we see (a) that <u>maximum size</u> increases with additional time nearly linearly for all groups and (b) that the stronger players have larger maximal chunks than weaker players (see Figure 3b). The difference between the three skill levels is not statistically significant with presentation times equal to or below 10 s [$\underline{F}(2,15) = 2.05$, ns], and is marginally significant with presentation times longer than 10 s [$\underline{F}(2,16) = 3.41$, $\underline{p} = .058$]. The standard deviation is high within each skill level (average sd = 2.0, 2.3 and 2.9 pieces for Masters, Experts and class A players, respectively). For all skill groups, the <u>number</u> of chunks increases as a function of presentation time. The number of chunks with a presentation time of 1 s and 60 s is 1.7 and 3.3, respectively, for Masters, 1.4 and 4.5 for Experts, and 1.2 and 4.3 for Class A players (see Figure 4b). The skill levels do not differ significantly over the 9 presentation times [$\underline{F}(2,15) = 0.32$, ns].

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

Errors

We are interested in two types of errors, errors of omission and errors of commission. The number of errors of omission is 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 incorrectly by the subject.

1) <u>Game positions.</u> The number of <u>errors of omission</u> correlates negatively with the presentation time ($r = -0.43$, $p < .001$). At 1 s, Masters miss 6.1 pieces by omission; with longer presentation times, they place as many pieces as are in the stimulus position. For Experts, the number of errors of omission is 12.6 at 1 s, then decreases almost linearly to come close to zero at 60 s. Finally, Class A players commit 16.5 errors of omission at 1 s and close to zero at 60 s. The three skill levels differ reliably at presentation times less than or equal to 10 s [$F(2,16) = 7.62$, $p = .005$]. The difference between Experts and Class A players reaches significance with presentation times longer than 10 s [$F(1,13) = 4.49$, $p = .054$].

The three skill levels present different patterns for <u>errors of commission</u>. Masters make fewer than 3 errors from 1 s to 4 s, and then reduce this number to an average of 1.25 at 5 and 10 s. For Experts the numbers of errors of commission decrease more or less linearly from 5.5 at 1 s to 0.5 at 60 s. Finally, the number of errors of commission is constant for Class A players from 1 s to 30 s (an average of 4.5). Even with 60 s, Class A players make on average 2.8 commission errors. The skill differences are not statistically significant at presentation times less than or equal to 10 s [$F(2,16) = 1.98$, ns]; Experts differ from Class A players at presentation times longer than 10 s [$F(1,13) = 8.35$, $p < .05$]. Notice that, expressed as percentages of pieces placed, the errors of commission decrease rapidly as the player's strength increases.

2) <u>Random positions.</u> For all skill levels, in random positions the number of <u>errors of omission</u> is high with short presentation times (21.4 pieces, on average, with 1 s), and decreases logarithmically with longer presentation time. Masters tend to produce fewer errors of omission than Experts, and Experts fewer than Class A players. These differences are larger with longer presentation times. For example, the numbers of errors of omission at 10 s are 14.7, 15.1 and 16.5 for Masters, Experts and Class A players, respectively, while the corresponding numbers at 60 s are 4.5, 6.6 and 9.0. The three groups do not differ at presentation times equal to or less than 10 s [$F(2,15) = 1.12$, ns], they differ with presentation times longer than 10 s [$F(2,16) = 4.17$, $p < .05$]. Finally, the negative correlation of errors of omission with percent correct is strong ($r = -0.74$, $p < .001$).

All skill levels make more <u>errors of commission</u> with longer presentation times. With 1 s, the average number of errors of commission is 2 pieces; with 60 s, it is 5.5 pieces. In general, Masters tend to make fewer errors of commission than Experts and Class A players, though the difference is not statistically significant [$F(2,16 = 1.14$, ns]. As a percentage of pieces placed on the board, there is again a strong negative correlation between errors of commission and the players' strength.

<u>Discussion</u>

Saariluoma's (1989) and Lories's (1987) results indicate that skilled players have better memory for random positions than weaker players when presentation time is sufficiently long. Our results confirm these findings and show also that, surprisingly, there is a difference in recall for random positions with rapid presentation. This difference is small (about 20% between Masters and Class A players at a presentation time of five seconds) in comparison with the difference for game positions (60% percent in the data of this experiment). But still, there is a difference.

The logistic growth curve, which has been shown to fit a wide variety of learning and memory data (Lewis, 1960), fits our data quite well. The difference of intercept between players at different skill levels (the 100 - B parameter) is predicted by the Chase-Simon theory: strong players recognize larger chunks more rapidly, and is confirmed by our data on the size of the largest chunk. Chase and Simon (1973a) found that, for game positions, stronger players place both larger chunks and more chunks, sometimes as many as 7 or more chunks. Our data replicate this finding, with the qualifications that the difference in

number of chunks tends to disappear with longer presentation times; and that the chunk size of our subjects was larger than theirs but the number of chunks correspondingly smaller.

Evidence available today places the STM <u>visual</u> memory capacity in the range of 3-4 chunks (Zhang & Simon, 1985), and we know now (Gobet & Simon, 1996b, 1998) that the apparent number of chunks in experiments using physical chess boards instead of computer displays was inflated and their average size reduced by physical limits (subjects' inability to grasp more than about a half dozen pieces in the hand at once). In replacing large chunks, subjects had to pause within the chunk for more than two seconds to pick up additional pieces, and single chunks were thus counted as two or more. Chase and Simon's players sometimes placed as many as 7 or more chunks, while in our current experiment we have seen that the usual limit was 3 or 4, consistent with other evidence on limits of STM visual memory. With random positions, we found that strong players placed both larger chunks and more chunks than did weaker players.

The difference in the c parameter in the game positions means that strong players not only perceive (and retain) more information during the first seconds of exposure, but that they also recognize and/or learn more or larger patterns afterwards. As the total numbers of chunks held and recalled, at all skill levels and for all presentation times, appear to be generally within the short-term memory limit of 3-4, it does not appear that many new chunks are acquired in LTM during the experiment. (New chunks would increase the apparent STM capacity limit, for they could be held in LTM and indexed either by names held in auditory STM or by retrieval cues in the environment.)

The rapid increase in recall during the first seconds (perhaps 2 or 3 seconds for Masters and 5 or 10 for Experts) most likely reflects the time required to notice successive patterns on the board. As we know that a single act of recognition of a complex pattern requires 0.5 s or more (Woodworth & Schlossberg, 1961, p. 32 to 35), it is likely that subjects, after the initial recognition of a pattern, spend some time scanning the board searching for additional familiar patterns that are present, the less practiced subjects taking longer to discern patterns than the more expert. The initial differences among the various skill levels, reflected in the values of (100 - B), would then largely reflect the sizes of their initial chunks, while the differences in c would largely reflect the rates at which new chunks were noticed, or additional pieces were added to chunks.

Beyond the first 5 or 10 s, there is enough time available to fixate some new chunks or to augment chunks that are already familiar. Estimates of fixation times in verbal learning experiments suggest that it may take about 8 s to add an element to an existing chunk or to form a new chunk (Simon, 1976). However, to account for the differences in the magnitude of c at different skill levels, we have to assume that experts and Masters do not usually add single pieces to chunks, but combine smaller chunks into larger. This would account for the gradual decrease in the number of chunks recalled by Masters after 4 s and by Experts after 10 s, while the average chunk sizes continued to increase.

The template mechanism also explains why the advantage of Masters over weaker players is much more pronounced in game positions, which typically evoke large templates, than in random positions, which do not correspond to any familiar templates and contain only small chunks at best (see Gobet and Simon, 1996a, for a more detailed discussion of chess templates). Notice, in Table 1, that the values of 100-B and c (and especially the latter) for Masters in random positions are substantially lower than the values for class A players in game positions.

We now turn to computer modeling to test whether the regularities formulated above can be captured by a coherent, process theory.

The Computer Simulations

In the simulations, we used CHREST, a computer program originally based on PERCEIVER (Simon & Barenfeld, 1969) and on MAPP (Simon & Gilmartin, 1973), which was itself based on EPAM (Feigenbaum & Simon, 1962, 1984), and later expanded to include the idea of templates. Earlier accounts

of CHREST, given in De Groot and Gobet (1996) and Gobet (1993a, 1993b), assumed a retrieval structure somewhat similar to the one proposed by Ericsson and Kintsch (1995), with the qualification that only chunks fixated upon in the fovea could be encoded rapidly in the slots provided by the retrieval structure. This generic retrieval structure has now been replaced by multiple templates (see Gobet & Simon, 1996a, for a discussion of this change). Gobet, Richman, and Simon (in preparation) present a detailed description of the present version of CHREST and report simulations of eye movements, of the interference task, the recall of multiple boards, the recall of positions modified by mirror-image and by translation, and other phenomena.

Although employing the basic discrimination net and learning processes of EPAM, CHREST has a somewhat more veridical process to simulate eye movements on the chess board; somewhat simpler tests in the net, specialized for sorting patterns in terms of pieces and their locations; and uses schemas in the form of templates for chess patterns. It is an EPAM slightly specialized to the task of recall in chess, the changes representing domain-specific memory structures and processes and an elaboration of the perceptual system.

Overview of the Model

The model consists of four components: (a) an LTM, (b) an STM, with a capacity of three visual chunks (preliminary simulations, described in Gobet, 1998b, showed that this value derived from Zhang & Simon, 1985, and Richman et al., 1995, gave a fit at least as good as larger values), (c) a discrimination net, which is an index to LTM, and (d) a "mind's eye", which stores visuo-spatial information for a short time. STM is a queue, with the exception of the largest chunk met at any point in time (henceforth named the hypothesis), which is kept in STM until a larger chunk is met. As noted above, templates, which include slots where variable information can be stored, are a special case of chunks.

The main processes are (a) eye movement mechanisms ; (b) mechanisms managing information encoding and storage in STM; (c) mechanisms allowing information to be learned in LTM; and (d) mechanisms updating information in the mind's eye.

During the presentation of a position to be recalled, the program fixates squares following the saccades specified by the eye movement module. Each fixation delineates a visual field (all squares within +/- two squares from the square fixated), and the pieces belonging to this visual field are sorted through the discrimination net. If a chunk (a pattern already familiar to the discrimination net) is found, a pointer to it is placed in STM, or, when possible, the chunk is used to fill one slot of a template. Given enough time, the program learns in three different ways. First, it can either add a new branch to access a node by a novel path or create a new node (and a branch leading to it) in the discrimination net; this discrimination process takes 8 s. Second, it can add information to an already existing chunk; this familiarization process takes 2 s. Third, it can fill in a template slot ; this process takes 250 msec.

Detailed Description of CHREST

Eye fixations

The program first attempts to use information provided by the hypothesis (the largest chunk met so far) to fixate a new square, by following a branch below the chunk and fixating the square associated with this branch. If this is not possible (for example, because there is no branch below the hypothesis or because this mechanism was already used for the previous fixation), the program uses several alternative mechanisms: fixation of perceptually salient pieces, fixation of a square following a relation of attack or defense, fixation of a region of the board that has not been visited yet, or, finally, a random fixation in the peripheral visual field. A full description of these mechanisms is given in De Groot and Gobet (1996).

Learning Chunks and Creating Templates.

The attention of the program is directed by eye movements. For each new fixation, the model sorts the pieces found in the visual field through the discrimination net. The visual field is defined as the squares located at most two squares away from the fixation point (see De Groot & Gobet, 1996, for empirical data justifying this parameter). Learning of chunks is essentially the same as in the EPAM model, with the qualification that only one type of test is carried out in CHREST: "What is the next item in the visual space?", while EPAM allows for testing various features of objects.

Taking pieces and squares as the basic perceptual features, rather than more primitive components of these units, allows the system time and space to grow large nets. Presumably human players also carry out other tests besides tests of the location of chunks of pieces, such as tests dealing with threats, plans, and other concepts, but these are not performed by CHREST. A chunk is encoded as a list of the pieces on their squares (POS), sorted in an arbitrary order (see De Groot & Gobet, 1996) ; for example: (Kg2, Re1, Pf2, Pg3, Ph4). A chunk can consist of a single POS (for example, (Kg2)).

The net is grown by two learning mechanisms, familiarization and discrimination. When a new object is presented, it is sorted through the discrimination net. When a node is reached, the object is compared with the image of the node, which is the internal representation of the object. If the image underrepresents the object, new features are added to the image (familiarization). If the information in the image and the object differ on some feature or some sub-element, a new node is created (discrimination).

In addition to these two learning mechanisms, CHREST also creates "similarity links" between nodes and templates. When learning, the model compares each chunk coming into STM with the largest chunk already residing there. If the two chunks are sufficiently similar, so that more than k POS's are present in both (in the simulation, k was fixed at 1), a similarity link is created between the chunks. During the recognition phase, the model can move via a similarity link from the node reached by sorting to the similar chunk (see below). This mechanism facilitates the recognition of templates.

Templates are chunks that possess at least one slot where variable information can be stored. In the domain of chess, template slots are created when the number of nodes below a given node that share identical information (either a square, a type of piece, or a chunk) is greater than a parameter, arbitrarily set to 3 in the simulations. Another constraint in the creation of a template is that a chunk should contain at least 5 elements. Slots can encode information only referring to their type: that is, either to kind of piece, location, or chunks having a common sub-chunk.

Access of Chunks during Recognition

The object (group of POS's) is sorted through the discrimination net. When a node is reached, the program checks whether the node has similarity links. If it does, the program chooses the link leading to the node having the most "information value," where information value is defined as:

number of POS's in chunk image + number of slots (if the chunk is a template). The chunk accessed is then placed in STM or, if templates have already been recognized, in a slot. It takes 50 ms to encode a chunk into STM and 250 ms to encode a chunk into a template slot.

On-line Learning in LTM

During the simulation of the presentation of a position, learning can occur in three ways: (a) by adding a new branch to access an existing chunk (which means that episodic cues allow this node to be accessed in the future, even if it is not in STM), (b) by combining two chunks (which is adding a chunk as a component of another chunk), and (c) by familiarization. The first two learning mechanisms each require 8 s (see Simon, 1976, for a discussion of this parameter), and the third 2 s. With long presentation times, it is assumed that players use the strategy of attending to a region of the board from which more information can be obtained, and then applying one of the learning mechanisms.

Simulations: Learning Phase

During the learning phase, the program scans a database of several thousands of chess positions taken from masters' games. It fixates squares with simulated eye movements, and learns chunks using the discrimination and familiarization processes. Templates and similarity links are also created at this time.

Three nets were created, estimated to correspond roughly to the recall percentages of Class A players, experts, and masters in experiment 1 with the five-second presentation of a game and a random position (two degrees of freedom were therefore lost here for each net). These nets had respectively 500 nodes, 10,000 nodes, and 300,000 nodes, and will be referred to as 1/2k, 10k, and 300k nets. Slightly better matches of human to simulation data could likely be obtained by fitting data from nets of different sizes directly to the human data, but of course with a loss of degrees of freedom.

Simulations: Performance Phase

For each presentation time, the program was tested with 50 game positions, randomly selected from a database of positions not used during the learning phase, and 50 random positions, created by the same procedure as in experiment 1.

During the reconstruction of a position, the program first uses the information stored in STM, and then information stored in LTM to place pieces sequentially. It is assumed that POS's stored in a template are replaced on the board as a single chunk. A piece already replaced from a previous chunk is ignored. Possible conflicts (e.g., a square containing several pieces) are resolved sequentially, based on the frequency with which each placement is proposed. The program may thereby change its mind about the location of a given piece or about the contents of a square, as human players do. Finally, the program uses some weak heuristics to limit placements, such as the fact that only one white king can be replaced in a given position.

As with human data, we will discuss the effect of presentation time on the percentage of pieces recalled correctly, the sizes and number of chunks, and the numbers and type of errors.

Percentage of Pieces Correctly Replaced

Figure 5 shows the results of the simulations for game positions (upper panel) and random positions (lower panel). Table 3 gives the B and c parameters of the logistic function in the six conditions, as estimated for the computer simulations. The logistic functions extract most of the information about percentage correct, showing an excellent goodness of fit (all r_2 are larger than .95). Comparison with Table 1 again shows high agreement of simulations with subjects: both for game and random positions, the (100-B) and c parameters increase with skill, in good quantitative agreement with the human data. An exception is the group of human masters in the game positions, where the program underestimates performance somewhat, and where the simulations yield smaller estimates of B and c than do the human data.

Figure 5. Percentage of correct pieces as a function of presentation time and number of chunks in the net for game positions (upper panel) and random positions (lower panel). The best fitting exponential growth function is also shown for each number of chunks.
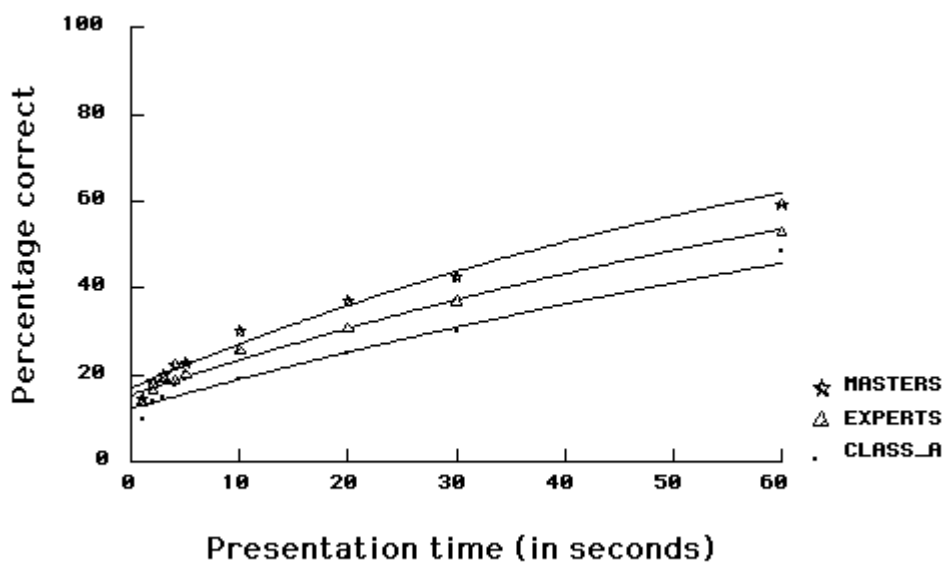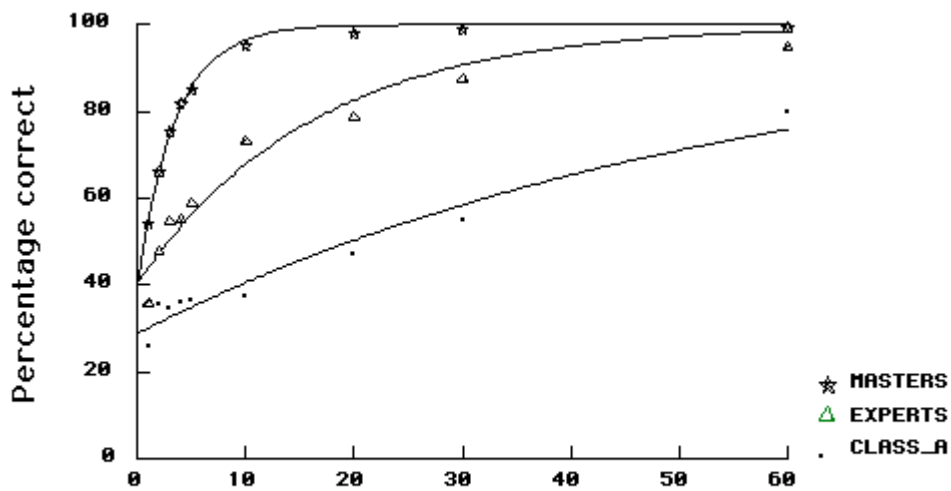
Table 3.

Recall percentage as a function of time presentation (simulations). Parameter estimation of the function P = 100 - Be-c(t-1).

Game positions					
95% Confidence Interval					
# chunks	Parameter	Estimate	ASE1	Lower	Upper
500	B	69.821	1.734	65.720	73.921
	c	0.018	0.002	-0.022	-0.013
	100 – B	30.179			
10k	B	56.050	2.968	49.033	63.067
	c	0.062	0.012	-0.089	-0.034
	100 – B	43.950			
300k	B	45.123	1.003	42.752	47.494

		Estimate	ASE1		
	c	0.287	0.014	-0.321	-0.253
	100 – B	54.877			
		Random positions			
		95% Confidence Interval			
# chunks	Parameter	Estimate	ASE1	Lower	Upper
500	B	87.057	0.980	84.739	89.375
	c	0.008	0.001	-0.010	-0.007
	100 – B	12.943			
10k	B	83.760	0.663	82.193	85.327
	c	0.010	0.000	-0.011	-0.009
	100 – B	16.240			
300k	B	81.887	0.961	79.614	84.159
	c	0.013	0.001	-0.014	-0.011
	100 – B	18.113			

[1]Asymptotic Standard Error

Chunks

We will not report the r_2 between simulations and data for chunks and errors, because in these cases, where some variables do not change much from one presentation time to another (e.g., number of chunks, or errors of commission), r_2 is almost bound to be small even if the fit of the linear parameters is good. Instead, we will compare the intercept and slopes of the variables with those estimated from human data. In most cases, these variables change linearly, or nearly so, with presentation time.

The program captures the two main characteristics of the number of chunks: the number is less than 3 with short presentation times, and increases with additional time (see Table 4 and Figure 4). However the slight "bulge" (increase followed by decrease) in number of chunks for human subjects in the Master-game and Expert-game conditions is not captured by the simulations.

As the Table 4 and Figure 3 show, the sizes of the largest chunks and the increases in size with presentation time observed with human subjects are also observed in the simulations, and the predicted increases are in the right range. (The largest chunk of the Masters studied in Gobet & Simon, 1998, was 17 pieces in the game condition, which matches the model's prediction better than the Masters described here).

Table 4.

Statistics on size of the biggest chunk and number of chunks (mean, intercept, slope, and amount of variance accounted for by linear regression) for game and random positions and for human data and simulations.

Human   Simulated

			Mean	intercept	slope	r2	Mean	intercept	slope	r2
		A	7.6	6.5	0.07	.64	5.5	4.7	0.05	.87
	Game	E	12.3	10.4	0.13	.90	13.5	10.8	0.18	.56
		M*	14.5	13.3	0.30	.38	19.6	16.5	0.21	.50
SIZE										

		A	4.4	3.9		0.03	.70	3.0	2.6		0.02	.86
	Random	E	5.9	5.2		0.04	.85	5.2	4.6		0.04	.73
		M	6.6	5.1		0.10	.82	4.9	3.5		0.10	.91
		A	3.4	2.6		0.06	.82	3.0	2.3		0.04	.94
	Game	E	3.1	3.0		0.01	.04	2.6	2.3		0.02	.81
		M*	3.4	3.5		-0.01	.00	2.1	2.3		-0.01	.60
NUMBER												
		A	2.0	1.3		0.05	.90	2.3	1.7		0.04	.96
	Random	E	2.4	1.5		0.06	.84	2.8	2.0		0.05	.99
		M	2.5	1.8		0.05	.42	2.9	2.1		0.06	.95

Note: A = Class A players (or, for the simulations, net with 500 nodes); E = Experts (or net with 10,000 nodes); M = Masters (or net with 300,000 nodes).

* The master game condition has only 6 data points

Errors

The program captures the numbers of <u>omissions</u> well (see Table 5), except perhaps in the Master-Game condition where, because of the generally stronger performance of the human Masters than of the simulation in replacing pieces, the simulation slightly overestimates omissions until about 10 s have elapsed. As might be expected from the differences in correct replacements, the number of errors of omission (both for human players and simulations) declines strongly from Class A players to Masters in game positions, but declines more moderately in random positions.

Table 5.

Statistics on errors of omission and commission (mean, intercept, slope, and amount of variance accounted for by linear regression) for game and random positions and for human data and simulations.

Human Simulated

			Mean	intercept	slope	r2	Mean	intercept	slope	r2
		A	9.5	13.5	-0.27	.82	12.5	16.0	-0.23	.98
	Game	E	5.8	8.7	-0.19	.61	6.2	9.5	-0.22	.74
		M*	0.6	1.5	-0.06	.26	1.3	3.2	-0.13	.44
OMISSION										
		A	16.9	19.7	-0.19	.70	18.2	20.1	-0.12	.91
	Random	E	14.6	17.8	-0.21	.76	14.3	17.3	-0.20	.98
		M	13.6	17.5	-0.26	.78	14.2	18.5	-0.27	.98
		A	4.5	4.8	-0.02	.43	1.7	1.3	0.03	.82
	Game	E	4.2	5.5	-0.09	.87	2.8	2.9	-0.01	.28

		M*	2.2	2.8		-0.15	.37	2.8	3.0	-0.01	.59
COMMISSION											
		A	3.5	2.4		0.07	.76	1.6	1.9	-0.02	.35
	Random	E	3.7	2.8		0.06	.60	4.6	4.0	0.04	.78
		M	2.1	1.4		0.04	.50	3.2	1.8	0.10	.99

Note: A = Class A players (or, for the simulations, net with 500 nodes); E = Experts (or net with 10,000 nodes); M = Masters (or net with 300,000 nodes).

*The master game condition has only 6 data points

The program does not predict as well the numbers of errors of commission, which are highly variable among the human subjects (Table 5, lower half). It underestimates errors of commission by the Class A players, most severely in the random condition, but generally overestimates such errors by the Experts and Masters. The main reason for the program committing errors of commission is that the image of a chunk may contain information about the locations of more pieces than just those used to recognize the pattern. If one or more of the pieces not tested is shown in the wrong location, it will be replaced there. We would expect the simulation to make more such errors in the Master and Expert conditions than in the Class A conditions, as it does.

What the simulation does not include is motivation to find locations on the board for the pieces that remain unplaced, even when the player has not stored reliable information to guide the placement. As Class A players have many more "uncertain" or "unknown" pieces than Experts, and Experts than Masters, such motivation would cause the number of the human players' errors of commission to vary in the same fashion, as in fact they do. From the trend in errors of commission with presentation time (not shown here), we find that Masters and Experts tend to correct their errors over time in game conditions, but the number of errors of commission grows in random conditions. Among the Class A players, there is no strong trend in numbers of errors of commission with game positions as presentation time increases.

We believe that, to capture these phenomena, the simulations would require an additional mechanism to balance the risks of errors of commission and omission when a choice is made under high uncertainty. From the data it appears that the balance of choices, on a percentage basis, is not strongly influenced by level of chess skill. The design of such a mechanism might be approached from the viewpoint of signal detection theory. There is a large body of psychological evidence, deriving from the original proposal of Tanner and Swets (1954), showing that substantial strategy differences among subjects can be expected in the trade-off between errors of omission and commission.

Role of Templates

Templates play a key role in explaining the high recall of human Masters with short presentation times, and also the large chunk size shown by Masters and Experts with short presentation times (with long presentation times, chunk size may also be increased by learning by familiarization). For example, the net with 300,000 nodes yields the following percentages correct when access to template slots is turned off (based on 50 game positions): 58.4, 64.9, 64.1, 64.3, 63.5, 64.5, 80.7, 88.9, 99.1, for 1, 2, 3, 4, 5, 10, 20, 30, and 60 s, respectively. The plateau from 2 s to 10 s probably reflects the STM limit of chunks, with perhaps recognition of a single large template (without filled slots), then having to await noticing one or more other fairly large chunks. The level after 10 s, 64.5% of pieces replaced, is about one-quarter less than the value of 84%, in Figure 5, where the simulation had access to the template slots, and compares even less favorably with the performance of the human Masters, which, at 10 s, ranged from about 85% to above 96%.

Discussion of the Simulations

Overall CHREST captures the human data well, taking into account that the model predicts behavior for three specific levels of memory size and particular sets of templates stored at each of these levels. There is no reason to suppose that all the subjects in each group had exactly the same numbers of templates, much less identical template repertories. Moreover, small variations in focus of attention during eye movements could produce individual differences in recall.

The main problem is with the simulation of the Masters' performance, where the program performs slightly less well than humans. Several kinds of changes would improve the program's fit to the data. One possible change is to increase the number of nodes in the net. The program underestimates human performance with the 5 s presentation (by 8%), which may simply signal bad calibration, for we have no independent strong evidence of the actual sizes of the discrimination nets of human Masters: they could be larger than the estimate of 300K nodes that we used. Even a modestly increased number of templates, acquired through study of opening variations, could make an important difference.

Alternatively, we could assume that nodes that have been held long enough in STM are sufficiently activated to be reached subsequently by search in LTM, a mechanism that has been explored in EPAM's simulation of DD's memory performance, and gains some credibility from the successful experience with it in that context. At the moment, we prefer to postpone further speculation until we can devise additional independent measures for estimating the key parameters.

The second most serious problem, the differences between human and simulated errors of commission, is harder to assess, due to the high variability of human data on these errors. The variability can very well derive from differences in strategy—in the relative weights that subjects attach to making errors of commission and omission respectively.

Discussion and Summary

In our experiment with chess players at three levels of skill, where the presentation time was systematically varied from one second to 60 seconds, we found a skill difference in recall performance both for game and random positions. The absolute differences in percentage correct between skill levels are smaller for random than for game positions, but still statistically significant. Although the difference for random positions was larger with long presentation times, it was also present with short presentation times.

We have been able to explain these and the other experimental findings in terms of the mechanisms postulated by the CHREST model, using discrimination nets of 1/2K, 10K, and 300K chunks for the three skill classes, Class A player, Expert and Master, respectively. The simulation accounts for the effects of the skill previously acquired and incorporated in LTM by the discrimination and familiarization processes that lie at the heart of EPAM and CHREST. It also accounts for additional learning during the presentation of the positions; for the duration of the presentation will determine: (a) access to the templates (cores + uncorrected default values), which occurs in the same latency range as the process of recognition (hundreds of msec); (b) ability to correct default values and instantiate other slots (requiring a fraction of a second each); (c) ability to create a new node by discrimination (requiring about 8 s) and to familiarize the image of a node (requiring about 2 s); (d) and ability to elaborate the template itself in LTM (requiring at least several minutes, possibly much more) — again by the processes of discrimination and familiarization.

We showed that logistic growth with duration of presentation fits the recall performance of all three classes of players (and of the simulations), both for game and random positions, with different parameters for each class of player. Strong players differ from weaker players both in the percentage of correct pieces they recall in one second and the rate at which asymptote is approached over longer time intervals.

The simulation model employs a special kind of domain-specialized chunk, the template, which is a LTM retrieval structure with slots (variables) that can be filled rapidly in each application (thus resembling STM elements). Chunks, including templates, are accessed in LTM through a discrimination net. Thus,

the templates are simply large chunks describing patterns that are met frequently in Masters' practice, especially common opening variations, which evolve into these more complex structures. The template theory embedded in CHREST is a modification and important extension of the chunking theory (Chase and Simon, 1973b).

The theory gives templates an important role in Masters' recall of chess positions taken from games. On the other hand, the superiority of strong players in recall of random positions is very largely explained by their larger repertoire of small chunks, some of which will match the adventitious patterns that appear on the random boards.

An alternative proposal (Ericsson & Kintsch, 1995) to account for the unusual memory capabilities of Masters employs a retrieval structure that is quite different from the templates of CHREST. We have shown that this alternative structure makes several predictions of performance that are contradicted by the empirical data. None of the other mechanisms (especially notions of "depth of processing") that have been proposed to account for these phenomena make predictions that are quantitatively, or even qualitatively, compatible with the chess data or with the performance, in other experiments, of memory experts who consciously build retrieval schemes in LTM.

The template theory gives a generally good account of the effect of presentation time upon the recall of game and random chess positions, explaining, in particular, the small advantage of experts even in recall of random boards. This, together with CHREST's ability to predict correctly a wide range of other quantitative and qualitative phenomena of chess perception and memory, shows it to be an accurate process theory for explaining these phenomena. Moreover, its generality is seen through its close kinship with the EPAM model, which has been tested successfully in a wide array of other task domains, using the same values of basic parameters that have been used with CHREST.

Theories, however, are more than just devices for prediction and explanation. Often, their most valuable contribution is to throw new light on unexplained conundrums and paradoxes in the observed phenomena. In the present case, we see this especially in the puzzles posed by the Chase-Simon experiments (the non-constancy in the number of chunks, the apparent violation of the three-chunk limit of visual STM capacity) and by Charness's discovery of experts' remarkable abilities to retain information about briefly viewed positions after performing intervening tasks. These puzzles motivated Chase, Ericsson and Staszewski's experiments in mnemonics, and the idea of retrieval structures; and retrieval structures suggested, in turn, the "slotted" templates and the CHREST model that account for these anomalies.

Of course the process is not at an end. In complex task environments like chess, settling one set of issues almost always discloses new questions and renews the process of modifying and extending theory to answer them. This is the kind of cumulative process for growing science that Allen Newell (1973) called for in his famous "Twenty Questions" paper: a process that leads us gradually toward a broad and deep theory of the information processing mechanisms that underlie perception, discrimination, and learning and that enable expert performance. There will undoubtedly be many more experiments, and many modifications and enhancements of theory along that path.

Finally, we should like to remark again that the characterization of chess memory represented by CHREST is not special to expertise in chess, but, as the more general EPAM architecture demonstrates, is generalizable to other domains of expert performance, where templates and/or retrieval structures appropriate to a domain can account for the abilities of experts to acquire information with great rapidity about the special kinds of situations they encounter in their professional activity.

References

Bartlett, F. C. (1932). *Remembering*. Cambridge: Cambridge University Press.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 641-653.

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

Charness, N. (1992). The impact of chess research on cognitive science. *Psychological Research, 54,* 4-9.

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

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

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

Cooke, N. J., Atlas, R. S., Lane, D. M., & Berger, R. C. (1993). Role of high-level knowledge in memory for chess positions. *American Journal of Psychology, 106*, 321-351.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-681.

De Groot, A. D. (1978). *Thought and choice in chess*. The Hague: Mouton Publishers. First Dutch edition in 1946.

De Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess. Heuristics of the professional eye*. Assen, NL: Van Gorcum.

Elo, A. (1978). *The rating of chess players, past and present*. New York: Arco.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245.

Feigenbaum, E. A. (1963). The simulation of verbal learning behavior. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill.

Feigenbaum, E. A., & Simon, H. A. (1962). A theory of the serial position effect. *British Journal of Psychology, 53*, 307-320.

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

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

Glickman, M. E. (1994). *Report of the USCF ratings committee.* US Chess Federation.

Gobet, F. (1993a). *Les mémoires d'un joueur d'échecs [The memories of a chess player]*. Fribourg: 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. (1998a). Expert memory: A comparison of four theories. *Cognition, 66,* 115-152.

Gobet, F. (1998b). Memory for the meaningless: How chunks help. *Proceedings ot the 20th Annual Meeting of the Cognitive Science Society.*

Gobet, F., Richman, H., & Simon, H. A. (in preparation). Chess players' memory and perception: A

unified process model.

Gobet, F., & Simon, H. A. (1996a). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31,* 1-40.

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

Gobet, F., & Simon, H. A. (1996c). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review, 3,* 159-163.

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

Kintsch, W. (1970). *Learning, memory, and conceptual processes.* New York: John Wiley.

Lane, D. M., & Robertson, L. (1979). The generality of the levels of processing hypothesis: An application to memory for chess positions. *Memory and Cognition, 7*, 253-256.

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

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

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.

Neisser, U. (1976). *Cognition and Reality. Principles and implications of cognitive psychology*. San Francisco: Freeman & Company.

Neisser, U. (1998). Stories, selves, and schematas: A review of ecological findings. In M. A. Conway, S. E. Gathercole & C. Cornoldi, *Theories of memory: Volume II*. Hove, UK: Psychology Press.

Newell, A. (1973). You can't play 20 questions with nature and win. In W. G. Chase (Ed.), *Visual information processing.* New York, NY: Academic Press.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Richman, H. B., & Simon, H. A. (1989). Context effects in letter perception: Comparison of two theories. *Psychological Review, 3*, 417-432.

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

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

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

Staszewski, J. (1990). Exceptional memory: The influence of practice and knowledge on the development of elaborative encoding strategies. In F. E. Weinert & W. Schneider (Eds.), *Interactions among aptitudes, strategies, and knowledge in cognitive performance*, (pp. 252-285). New York: Springer.

Tanner, W. P. and Swets, J. A. (1954). A decision making theory of visual detection. *Pychological Review, 61*, 401-409.

Woodworth, R. S., and Schlossberg, H. (1961). *Experimental psychology* (revised edition). New York: Holt, Rinehart and Winston.

Yates, F. A. (1966). *The art of memory*. 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