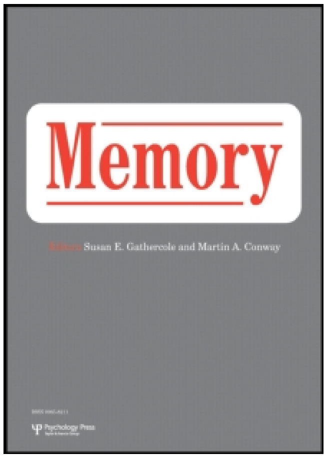


Memory
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683358

Chunks in expert memory: Evidence for the magical number four … or is it two?

Fernand Gobet [a]; Gary Clarkson [a]
[a] University of Nottingham, UK.

Online Publication Date: 01 November 2004
To cite this Article: Gobet, Fernand and Clarkson, Gary (2004) 'Chunks in expert memory: Evidence for the magical number four … or is it two?', Memory, 12:6, 732 - 747
To link to this article: DOI: 10.1080/09658210344000530
URL: http://dx.doi.org/10.1080/09658210344000530

MEMORY, 2004, *12* (6), 732–747

Chunks in expert memory: Evidence for the magical number four . . . or is it two?

Fernand Gobet and Gary Clarkson
University of Nottingham, UK

This study aims to test the divergent predictions of the chunking theory (Chase & Simon, 1973) and template theory (Gobet & Simon, 1996a, 2000) with respect to the number of chunks held in visual short-term memory and the size of chunks used by experts. We presented game and random chessboards in both a copy and a recall task. In a within-subject design, the stimuli were displayed using two presentation media: (a) physical board and pieces, as in Chase and Simon's (1973) study; and (b) a computer display, as in Gobet and Simon's (1998) study. Results show that, in most cases, no more than three chunks were replaced in the recall task, as predicted by template theory. In addition, with game positions in the computer condition, chess Masters replaced very large chunks (up to 15 pieces), again in line with template theory. Overall, the results suggest that the original chunking theory overestimated short-term memory capacity and underestimated the size of chunks used, in particular with Masters. They also suggest that Cowan's (2001) proposal that STM holds four chunks may be an overestimate.

When De Groot (1946/1965) investigated chess players' mental processes in a problem-solving task, he found no large skill differences in the depth of their search, the number of moves considered, or the search heuristics employed. But when he examined memory for briefly presented positions taken from Master games, he found that Masters demonstrated a vast superiority over weaker players. De Groot concluded that the key to expertise is not in any superior general processing abilities, but in domain-specific knowledge. Further research has confirmed that experts are highly selective in their search behaviour and that they can handle a much greater volume of domain-specific information than novices; this ability is observed in the presence of a normal cognitive capacity and in a number of domains, including games, mnemonics, music, sciences, and sports (Ericsson, Chase, & Faloon, 1980; Ericsson & Lehman, 1996; Gobet, 1998; Saariluoma, 1995; Thompson, Cowan, Frieman, Mahadevan & Vogl, 1991; Vicente & Wang, 1998; Wilding & Valentine, 1997).

For a long time, the main explanation for this skill effect has been Chase and Simon's (1973) chunking theory, which centres around the concept of a *chunk*—long-term memory (LTM) information that has been grouped in a meaningful way and that is remembered as a single perceptual unit. According to this theory, experts have acquired a large number of such chunks, which reflect the statistical structure of their environment (Simon & Gilmartin, 1973). These chunks can be used to encode information rapidly and act as the condition parts of productions, explaining phenomena such as the almost instantaneous identification of a good move in a chess position. Recently, however, there has been intense theoretical debate about how best to explain experts' performance, with four contend-

Correspondence should be addressed to Fernand Gobet, Department of Human Sciences, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK. Email: fernand.gobet@brunel.ac.uk

The authors thank Brian Hayward for his help in locating participants for the experiment, as well as Guillermo Campitelli, Peter Cheng, Peter Lane, Iain Oliver, David Peebles, Judith Reitman-Olson, Gareth Williams, and two anonymous referees for useful comments.

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210344000530

ing theories involved. While Chase and Simon's chunking theory is still considered by some as one of the best contenders (e.g., Gobet, 1998), others have criticised its account of the empirical data (e.g., Charness, 1976; Ericsson & Kintsch, 1995; Frey & Adesman, 1976; Holding, 1985; Saariluoma, 1995; Vicente & Wang, 1998). To address such criticisms, Gobet and Simon (1996a) have expanded the chunking theory by adding mechanisms to automatically acquire high-level schemas, known as templates. A different account is put forward by Ericsson and Kintsch (1995), who, in their theoretical framework of long-term working memory (LTWM), propose that experts acquire encoding and retrieval mechanisms to adapt to the demands that their environment makes upon working memory. Finally, Vicente and Wang (1998), in their constraint attunement theory, propose that experts become attuned to goal-relevant constraints in the material of their domain of expertise and that these constraints are critical in recall experiments; they also propose that it is necessary to analyse these goal-relevant constraints within the structure of the environment before proposing process theories of experts' behaviour. The various positions are discussed at length in Ericsson, Patel, and Kintsch (2000), Simon and Gobet (2000), and Vicente (2000).

In a different line of research, psychologists have used the concept of a chunk to estimate the capacity of short-term memory (STM). In a highly influential paper, Miller (1956) proposed that people can remember about seven chunks; later estimates suggested a more limited capacity, such as three chunks (Broadbent, 1975) and four chunks (Coltheart, 1972; Cowan, 2001). As discussed at great length by Cowan (2001), the concept of an STM capacity limit has been controversial, and some authors have proposed disposing with this idea altogether (e.g., Ericsson & Kirk, 2001; Meyer & Kieras, 1997). However, it is also a powerful hypothesis, unifying data across a variety of domains; indeed, Cowan (2001) was able to amass a remarkable amount of evidence pointing to an STM capacity of around four chunks.

The concept of a chunk has thus proved important for understanding both expert behaviour and STM capacity. The goal of this paper is to bring together these two research traditions, and to evaluate the evidence, both old and new, that supports limits of visual STM capacity in experts and novices. As will be expanded upon below, both chunking and template theories make clear predictions about STM capacity and chunk size; an added advantage is that these theories have been implemented as computer programs, offering a clear-cut definition of the concept of a "chunk", which has not often been the case in the literature (Cowan, 2001; Lane, Gobet, & Cheng, 2001). The other two main theories of expert memory are less explicit about chunk size and number.[1] First, Ericsson and Kintsch's (1995) LTWM theory, although it postulates the presence of patterns and schemas, which presumably include chunks as defined by Chase and Simon, does not give enough detail about these concepts to make quantitative predictions. In general, however, LTWM assumptions seem at variance with a fixed STM capacity (Ericsson & Kintsch, 1995; Ericsson & Kirk, 2001), and evidence of such a capacity limit, in particular with experts who are supposed to flexibly store information in LTWM, should count as negative evidence for this theory. Second, the concept of a chunk is beyond the scope of Vicente and Wang's (1998) theory, which, being a product theory, does not include assumptions about internal mechanisms or structures.

While we will oppose the different predictions of chunking and template theories, our main interest is in refining our knowledge of the exact nature of chunks, and in particular their quantitative attributes. An important task of any mature science, but all too often neglected in psychology (Grant, 1962; Meehl, 1967; Simon, 1974), is to carry out experiments narrowing down the quantitative estimates of the parameters of its theories. In this article, we will attempt to refine the theoretical value of the number of chunks in visual STM, and also to gather further information about the size of chunks in chess Masters. Given that only two theories of expertise (chunking theory and template theory) make quantitative predictions about these parameters, this paper will concentrate on them. Finally, the paper will also test Cowan's claim of an STM capacity of four items.

We first review in some detail Chase and Simon's (1973) main experiments and describe the chunking theory, which was developed to account for their results. After briefly reviewing some of

[1] As far as we know, other theories of expertise, such as Holding's (1985) or Saariluoma's (1995), do not make specific predictions about these two variables either.

the empirical support for the concept of a chunk, we consider some data that do not fit the predictions of the chunking theory. We then discuss the template theory (Gobet & Simon, 1996a, 2000), which was developed to remedy these weaknesses while keeping the strengths of the original chunking theory. This brings us to a discussion of the size and number of chunks in chess memory experiments, and prepares the path for the experimental part of the paper, where the diverging predictions of the chunking and template theory will be tested.

CHASE AND SIMON'S (1973) EXPERIMENTS AND THEORY

Building on De Groot's (1965) work, Chase and Simon (1973) studied the perceptual and memory structures employed by chess players of varying strength, and used two experimental paradigms. The *recall task* used the same method as De Groot (1965). Participants were allowed to inspect a position for 5 seconds before it was removed from view and subsequently attempted to reconstruct as much of the position as they could recall. In the *copy task*, participants reconstructed a stimulus board position onto an empty board, while the stimulus board remained in view. The stimulus and the reconstruction boards could not be fixated simultaneously, so glances between the boards could be used to detect the chunks (collections of pieces) held in memory. Based on the similarity of the latency distributions in the recall and copy tasks, and the assumption that glances could be used to define chunks in the copy task, Chase and Simon hypothesised that pieces placed with less than 2 seconds' interval belong to the same chunk; conversely, pieces placed with an interval of more than 2 seconds belong to different chunks.

In the recall task, stronger players demonstrated greater recall, confirming De Groot's (1965) finding. Analysing the latencies and the number of semantic relations (colour, defence, attack, proximity, and kind) between successive piece placements in the recall and copy tasks, Chase and Simon drew some inferences about the memory structures used to mediate experts' superior performance. They found that, on average, the number of chess relations between successive pieces belonging to a chunk is much greater than the number of relations between successive pieces not belonging to a chunk.

Referring to Miller (1956), Chase and Simon (1973) proposed that experts are limited by the same STM limits as non-experts (about seven items), but that they use chunks to encode the positions into a small number of units. The crucial feature of chunks is that they are a single storage unit, retrievable from LTM in one act of recognition. Chunks are acquired over years of practice and study within a domain, at a relatively slow rate (about 8 s to create a new chunk, and 2 s to add information to an extant chunk). Once learnt, chunks allow chess experts to rapidly recognise known (parts of) positions, and to access information about potential moves. Chunks are indexed by a discrimination network, where critical features of perceptual stimuli are tested. Such an organisation allows perceptual stimuli to be rapidly categorised, thus enabling experts to extract the salient elements of a position quickly. It also allows stronger players to perceive the positions as collections of familiar configurations of pieces rather than a collection of individual pieces, as novices do. Hence, experts can memorise an entire position in spite of the limits of STM.

EVIDENCE SUPPORTING THE CONCEPT OF A CHUNK

Chase and Simon's (1973) experimental technique is not the only paradigm providing evidence supporting the existence of chunks in chess (pointers to the extensive empirical support for chunking in other domains are given in Gobet et al., 2001). Several techniques, reviewed in Gobet and Simon (1998), have brought converging evidence for the psychological reality of chunks, as defined either by latency in placement or by number of relations between pieces. These techniques include sorting tasks (Gruber & Ziegler, 1990), guessing tasks (De Groot & Gobet, 1996; Gruber, 1991), recall tasks (Frey & Adesman, 1976), and hierarchical cluster analysis of piece placements (Gold & Opwis, 1992). Several of these studies also support De Groot's (1965) suggestion of the involvement of higher-level knowledge in chess; skilled players explicitly refer more to high-level, abstract knowledge and less to the types of chunks proposed by Chase and Simon (1973).

Of particular interest for the present study is the use of the *partitioning technique*, first used by Reitman (1976) with Go positions, in which participants are required to separate positions into clusters by circling groups of pieces that they feel belong together. In a study of knowledge structures and age variations in chess, Chi (1978) presented participants with a recall task similar to Chase and Simon's (1973), but also incorporated a partitioning task in the experiment. Comparing the latencies found in the recall task with the clusters identified in the partitioning task, Chi found that the average amount of time participants took to place pieces that cross cluster boundaries was longer (about 3 s) than that for pieces belonging to the same cluster (about 1.5 s). Chi also observed that some clusters overlapped. Freyhoff, Gruber, and Ziegler (1992) conducted a study in which participants divided chess positions into (non-intersecting) groupings that made sense to them; subsequently they were asked to combine groups into larger ones and also divide them into smaller ones. This procedure allowed positions to be represented as a hierarchy of clusters. Freyhoff et al. (1992) found that Masters gave larger groupings of pieces at all three levels of partitioning and that the average number of pieces in the clusters increased as the position became more typical.

TEMPLATE THEORY

In spite of the empirical support for the hypothesis of chunking, which we have just reviewed, a number of empirical findings have challenged Chase and Simon's (1973) chunking model. The most damaging evidence is perhaps the small effect of interfering stimuli between presentation and recall of chess positions (Charness, 1976; Cooke, Atlas, Lane, & Berger, 1993; Frey & Adesman, 1976; Gobet & Simon, 1996a). The lack of interference effects, which suggests rapid storage in LTM, is problematic because the chunking theory proposes that, in the recall task, learning is relatively slow and information is only encoded into STM. Another important finding undermining the theory is that a variety of empirical techniques, as we have just seen, suggest that Masters perceive chess positions at a higher level than proposed by Chase and Simon; in short, the data suggest that Masters' chunks should be larger than Chase and Simon predicted.

The *template theory* (Gobet & Simon, 1996a) was proposed as a refinement of the chunking theory. It retains the idea that chunks, which are recursively made of (sub)chunks, are indexed by a hierarchical discrimination network, but suggests that frequently encountered chunks develop into higher-level structures (templates) with slots allowing rapid LTM encoding. (Note that this rapid LTM encoding happens only for filling in slots; otherwise, learning takes the same time as proposed by the chunking theory). Slots are created when there is variable information for parts of positions belonging to the same class; this information may include chunks.[2] Thus, the idea of a hierarchical organisation (e.g., Cooke et al., 1993; De Groot, 1965; Gobet, 1993, Saariluoma, 1995) is captured both by the basic structure of the discrimination network and the possibility of encoding chunks into templates. Templates also hold pointers to potentially good moves and other templates. Finally, based on work by Zhang and Simon (1985), visual STM is limited to three items.

Aspects of the template theory have been implemented in a computer program, CHREST (Chunk Hierarchy and REtrieval STructures), which accounts for a number of data, such as the overlap between chunks, the pattern of eye movements during the first seconds of the presentation of a position, and the role of presentation time on recall performance (De Groot & Gobet, 1996; Gobet & Simon, 2000). In the context of this article, it is important to note that CHREST, as did MAPP, a partial computer implementation of the chunking theory (Simon & Gilmartin, 1973), closely simulates the semantic relations shared by two pieces belonging to the same chunk (Gobet, 2001).[3] As shown by the simulations in Gobet and Simon (2000), a key prediction of CHREST is that the size of the largest chunk in the recall of a position should be substantially bigger than proposed by Chase and Simon (1973); this is due both to the recursive fashion with which chunks are acquired and to the fact that templates contain slots that can be filled rapidly.

[2] The information in the core and in the slots of a template forms a whole; even though some slots may contain chunks, just as chunks may contain (sub)chunks, their placement is considered as belonging to that of the template itself. Thus, in simulations with CHREST, the placement of a template is counted as the placement of a single (large) chunk.

[3] The relations observed in chunks differ in interesting and systematic ways from the *a priori* relations observed in game positions—the structure of the environment (see Chase & Simon, 1973, Gobet & Simon, 1998, and Gobet, 2001, for a detailed discussion). These differences, and the fact that experts use larger chunks than novices, strongly suggest that chunks are memory structures, and not only a measure of the elements of chess.

SIZE AND NUMBER OF CHUNKS

A number of reasons limit the generalisability of Chase and Simon's (1973) study. First, Chase and Simon only used three players in their experiment and their Master was out of practice and in his forties, two factors that may have affected the results. Second, as mentioned above, subsequent experimental data have suggested that stronger players have a higher-level perceptual and memory organisation than Chase and Simon proposed. Third, the number and size of chunks may have been an artefact of the limited capacity of the hand for holding chess pieces.

To address these questions, Gobet and Simon (1998) replicated Chase and Simon's (1973) study using a computer display instead of physical chessboards, and with a large sample including Masters, Experts, and Class A players. They found an important difference in comparison to Chase and Simon's (1973) results: Their Masters replaced larger and fewer chunks than in the original study, which matches the predictions of the template theory. This led Gobet and Simon to suggest that Chase and Simon's study underestimated Masters' chunk sizes and overestimated their number, and that the size of the hand was a limiting factor of chunk sizes in the earlier study, making it hard to pick up more than 4–5 pieces.[4] However, Gobet and Simon's (1998) results do not irrefutably establish a larger chunk size, as their subjects were given only the computer task. To fully disentangle the issue, it is necessary to report data from the same set of subjects using both Chase and Simon's (1973) and Gobet and Simon's (1998) procedures. Given the importance that the concept of a chunk has played in research into cognition in general and into expertise in particular, it is crucial to establish the nature of the differences found in the two studies.

OVERVIEW OF THE EXPERIMENT

As previously described, Chase and Simon's (1973) and Gobet and Simon's (1998) studies employed two different media for the presentation of the same task: physical board and pieces, and computer presentation, respectively. The current study combines these two media within the same group of participants in an attempt to establish which of the conflicting results are due to the different methods used, and which are due to random variations in sampling. A partitioning task, similar to that used by Reitman (1976) and Chi (1978), is also incorporated into the design, with the aim of producing converging evidence for chunks as defined by the latencies between placements. Using the three different experimental techniques conjointly will remove any between-subjects variance that may have affected results of recall percentage, chunk size, and number of chunks held in STM.

While the chunking and template theories obviously share a number of characteristics, as the latter derives from the former, they also differ on several counts, in particular with respect to STM capacity and chunk size. The chunking theory predicts a visual STM capacity of around seven chunks while, according to the template theory, this number should be three (close to the four proposed by Cowan, 2001). With both presentation media, chunking theory predicts that Masters' chunks should not exceed four or five pieces. The template theory makes different predictions as a function of the presentation medium: the chunk sizes obtained by the physical-board method should be smaller than those obtained using the computer-presentation method; in the first case, they should be limited by hand capacity, and include at most four or five pieces. In the second case, they should be relatively large, and sometimes include more than fifteen pieces (Gobet & Simon, 1996a, 2000). Finally, based on previous experiments (Chi, 1978; Gruber & Ziegler, 1992), the partitioning task is expected to show similar clusters of pieces to the computer and physical-board methods.

METHOD

The copy and recall tasks were given to all participants using both the original medium (physical chess pieces and board) and the computer-display medium. The order in which the copy and recall tasks were presented was counterbalanced for skill level, medium of presentation, and position type (random or game positions). All participants completed the partitioning task as the final component of the experiment.

[4] The objection that the computer display artefactually leads to an overestimation of chunk size is addressed in detail, and refuted, in Gobet and Simon (1998).

Participants

Two females and ten males were recruited either from the University of Nottingham or from chess clubs in the local area. Participants were grouped in three skill levels based on BCF (British Chess Federation) ratings:[5] Masters (n = 4; mean BCF = 202), Class B players (n = 4; mean BCF = 143) and novices (n = 4; all could play chess but had no BCF rating). The players were paid for participation based on their skill level (£30 for Masters, £10 for Class B players, and £6 for the novices). The mean age was 22.5 years (SD = 5.8), ranging from 15 years up to 35 years.

Apparatus and materials

Each participant was randomly allocated 28 positions, which comprised 20 game positions (5 per experimental condition) and 8 random positions (2 per experimental condition).[6] The game positions were randomly selected from a database of Master-level games (5000 positions), after Black's twentieth move. Similarly, the random positions were taken from a database of 1000 such positions, which were created by shuffling the piece location of game positions. In both the game and random stimuli, the average number of pieces per position within a task was 25 ± 1, and the range was from 22 to 28 pieces. Each participant had a different random order and random assignment of positions to conditions. To remove the possible confound of practice effects, we used a set of eight different practice positions, constant across participants, one for each of the conditions.

Physical-board display. All experiments were run using two standard competition chessboards (40.5 × 40.5 cm) and two full sets of standard pieces. A wooden sliding partition was used to allow control over the time during which participants could view the stimulus board position. A standard video camera was used to film the participants. Videocassettes were analysed frame-by-frame using an editing suite accurate to one frame (4 ms). The second author coded the data by hand from the videotapes to a computer file including: the time of piece placements, the piece and its position in each placement, the pieces that were removed, and switches in glances between the two boards in the copy task. Once the data had been coded, the resulting files were checked for typos and inconsistencies by an *ad hoc* program. After their correction, the video data were all checked once again.

Computer display. The positions were presented on the screen on an Apple Macintosh II. The chessboard was 9 × 9 cm and the pieces were of standard shape. During presentation of the positions the background to the board remained black. The reconstruction board was presented in the left-hand side of the display. To the right of the board the pieces to be used in reconstructing the position were presented in a rectangular box containing the six types of chess pieces, both white and black. A white box containing the text "OK" was displayed in the top left corner of the screen, used by the participants to progress to the next stimulus presentation when the current reconstruction was completed to their satisfaction or ability. A piece was placed by positioning the cursor over the desired kind in the box on the right-hand side and clicking the mouse button. After a piece had been selected in this manner, the participants selected a square to place it in and again clicked the mouse button when the cursor was appropriately positioned. Placing each successive piece required participants to select another from the rectangular box of pieces. During the copy task, two numbers were displayed in white buttons just above the board. These numbers were used to switch between the position on the stimulus board and the board for reconstructing the position. The log-files stored the time between piece placements, the pieces that were selected, the positions where pieces were placed, pieces that were removed, and switches between the two different views of the boards in the copy task. (See Gobet & Simon, 1998, for more detail about the software used.)

Partitioning task. All 28 experimental positions were transferred to paper with the same appearance as in the computer task. These positions were used for the partitioning task in the same order as during the experimental tasks. The

[5] The BCF (British Chess Federation) rating is an interval scale ranking competitive chess players, similar to the Elo rating, a more widely used rating system. BCF ratings may be converted into Elo points with the following formula: Elo = (8 * BCF) + 600. See Holding (1985), for more details on chess rating systems.

[6] Previous studies (e.g., Gobet & Simon, 2000) have shown that chess players are not very keen to recall random positions; hence, to keep participants' motivation high, we used fewer random positions than game positions.

partitions drawn by the participants were coded and stored in a computer file.

Design

A mixed factorial design was used. The between-group independent variable was the skill level of the participants, with three levels (Masters, Class B players, and novices). The within-group independent variables were the task type (recall or copy), the medium of the task (physical board or computer display), and the type of position (game or random). Not counting the partitioning task, there were therefore eight experimental conditions for each skill level: task (2) × medium (2) × position (2).

The dependent variables of the copy task were the mean maximum chunk size and the mean number of chunks. The dependent variables of the recall task were the percentages of pieces correctly recalled, the mean maximum chunk size, and the mean number of chunks. The dependent variables of the partitioning task were the size and number of partitions.

Procedure

Standard instructions were presented prior to each of the experimental tasks. The participants performed a practice trial for each of the eight experimental conditions.

Physical-board presentation

The two boards were placed side by side and separated by a sliding screen that, when closed, obscured the board to the participants' left (*stimulus board*) whilst leaving the right-hand board (*reconstruction board*) in view. The participants were seated in front of the right-hand board and had to reconstruct the board to their left. After every trial, the pieces were removed and placed on the right side of the reconstruction board in a predefined organisation constant throughout the study. A fixed video camera mounted on a tripod was set up to film the reconstruction board and the participants' head to allow later analysis of the behaviour during each reconstruction.

Copy task. A position was set up on the stimulus board when the dividing screen was obscuring it from the participants' view. When the screen was removed, the participants had to copy the position onto the reconstruction board. The stimulus position was kept in view throughout each trial so that participants could switch their glance between it and the reconstruction board. The two boards were arranged so that only one could be seen at any particular time and switches in the direction of glance could show which board was currently being fixated. Participants were encouraged to perform the task as quickly as possible.

Recall task. The recall experiment was the same as the copy experiment except that the stimulus board was only in view for 5 s, during which time no pieces were allowed to be placed. Participants could start reconstructing the stimulus position from memory as soon as the screen was closed again; there was no time limit.

Computer presentation

The participants were seated in front of the computer and allowed to familiarise themselves with the software by selecting, placing, removing, and overwriting pieces; all of these actions were performed using the mouse. After each trial the participants could decide when to proceed to the next trial by clicking the "OK" button in the top left of the display.

Copy task. Two buttons labelled "1" and "2" displayed at the top of the screen were used to select which board (stimulus or reconstruction) was viewed, as only one of the boards was presented on screen at a time. The participants could switch views between the stimulus board and the reconstruction board as often as they wished.

Recall task. The recall experiment was as the copy experiment but the stimulus board was only presented for 5 s, during which time no pieces could be placed. Participants could start reconstructing the stimulus position from memory as soon as the reconstruction board was displayed on the screen.

Partitioning task

The partitioning task was always carried out as the final part of the experiment. Participants were presented with the 28 positions they had seen during the other tasks and were instructed to group the pieces into groups that made sense to them, using rings that were drawn on to the board positions. Caution was taken to ensure that no suggestions were made about how the pieces should be grouped together, i.e., whether groupings could overlap or be nested, or the size that the groupings could be (see the Appendix for the exact wording of the instructions).

RESULTS AND DISCUSSION

In order to check that the physical-board and computer displays offer converging results on as many variables as possible, and to verify their concordance with previous studies, we first analyse the distribution of latencies in the copy task, followed by the percentage correct in the recall task. We then present the analysis of the chunk data, where the key predictions of this paper are addressed, and finish with a discussion of the partitioning task.

Distribution of latencies between piece placements in the copy task

Participants behaved differently in the computer and the physical-board conditions, which affected chunk size and number. In the physical-board condition participants studied the position for a short moment before placing a few pieces and repeated this until the entire position was copied, whereas in the computer condition participants studied the stimulus board for a considerably greater time before commencing reconstruction, using fewer but larger chunks (see Figures 1, 2, and 3).

Latencies in the copy task are important in validating the 2-second cut-off that will be used to define a chunk throughout this study. The copy task data were coded for the analysis of two ways of placing pieces: *within-glance placements* (WGP), in which successive pieces are placed without reference to the stimulus board; and *between-glance placements* (BGP), in which successive placements are interrupted by a glance at the stimulus position. In this analysis, as in Gobet and Simon (1998), all of the participants' results were pooled into four groups defined by the method of placement (WGP/BGP) and the task display (computer/physical board). The reader is referred to Chase and Simon (1973) and Gobet and Simon (1998) for more detail about this methodology.

Physical-board presentation

As in Chase and Simon (1973), BGP and WGP latencies show very different distributions; WGP latencies have a mean of 0.64 s and a median of 0.44 s, and are highly skewed to the right, whereas BGP latencies have a mean of 2.43 s, and a median of 2.20 s, and show little skewness. The range of WGP latencies (3.80 s; from 0.0 s to 3.80 s)[7] is smaller and at lower values than the BGP latencies (range = 14.72 s, from 0.28 s to 15.00 s). Of the WGP latencies, 95% are less than 2 s compared with 42.6% of the BGP latencies and 98.3% are less than 2.5 s compared with 63.7% of the BGP latencies.

Overall, the results compare well with those of Chase and Simon, with the difference that our participants were faster. The median of the WGP latencies was 1.00 s for Chase and Simon, and 0.44 s in our data; the mean of the BGP latencies was 3.00 s and 2.43 s, respectively. Given that perceptual and motor abilities are known to decline with age (Birren & Schaie, 1996) these differences are likely due to the fact that our sample was much younger than Chase and Simon's.

Computer presentation

As in Gobet and Simon (1998), the latencies were corrected by subtracting from the time between two placements the time needed to move the cursor to the destination square once a piece had been selected. The participants' WGP latencies had a mean of 1.20 s and a median of 1.03 s, compared to a mean of 10.06 s and a median of 8.20 s for the BGP latencies. The computer presentation showed a greater range than the physical-board condition of 8.37 s (0.35 s–8.72 s) and 42.91 s (2.22 s–45.13 s) for WGP and BGP latencies respectively. Of the WGP latencies, 93.7% are below 2 s and 96.6% below 2.5 s. Again BGP latencies show a different distribution, with 0% of the latencies being less than 2.5 s.

Overall, the results show the same patterns as those of Gobet and Simon (1998), and the absolute values are reasonably close. For example, the median of the WGP was 1.37 s in Gobet and Simon, and 1.03 s in the present study; the median of the BGP latencies was 7.30 s and 8.20 s, respectively.

Summary

The qualitative differences between the distributions are the same as with the previous studies; on average WGP latencies are shorter than BGP latencies, with a greater proportion of them below 2 s. The differences in the distributions between

[7] A latency of zero seconds indicates that two pieces were placed simultaneously.

the computer and the physical-board display are the same as between Chase and Simon (1973) and Gobet and Simon (1998). Mean latencies are longer in the computer condition; they were more than twice as large for WGP latencies and more than four times as large for BGP latencies.

A greater proportion of BGP latencies are below 2 s in the physical-board condition than in the computer-display condition; this is likely to be attributable to the strategy employed by the participants. The physical-board presentation allows participants to carry out tasks in parallel; glances can be made when pieces are being transported. When participants move a piece to a position they can check that the intended placement position is correct and in doing so make the subsequent placement seem to be between glances. This behaviour is not possible in the computer presentation. The high percentage of within-glance placements and low percentage of between-glance placements below 2 s add support to the validity of using this boundary for defining chunks.

Recall task: Percentage correct

As there is no main effect of display type ($F < 1$), the results with the two presentation media were pooled. The percentages of pieces correctly recalled in the game positions are 70.0%, 45.0%, and 17.0% for Masters, Class B players, and novices, respectively. The corresponding percentages for random positions are 22.3%, 16.9%, and 14.4%. The ANOVA reports main effects of position type, $F(1, 9) = 64.72$, $MS_e = 0.87$, $p < .001$, and skill level, $F(2, 9) = 325.40$, $MS_e = 0.33$, $p < .001$, and a significant interaction, $F(2, 9) = 12.14$, $MS_e = 0.17$, $p < .01$.

Participants recall more pieces correctly in the game positions than in the random positions and, in the former case, recall percentage increases reliably with skill level; post-hoc analysis shows that each skill level was statistically different from the others, with a linear relationship between skill and recall. The pattern of means suggests a skill effect with random positions, although the differences are not statistically significant. The results then replicate the robust effect of skill on recall of game positions, as well as confirming Gobet and Simon's (1996b) finding that Masters do typically show a slight (but often statistically non-significant) advantage for random positions.

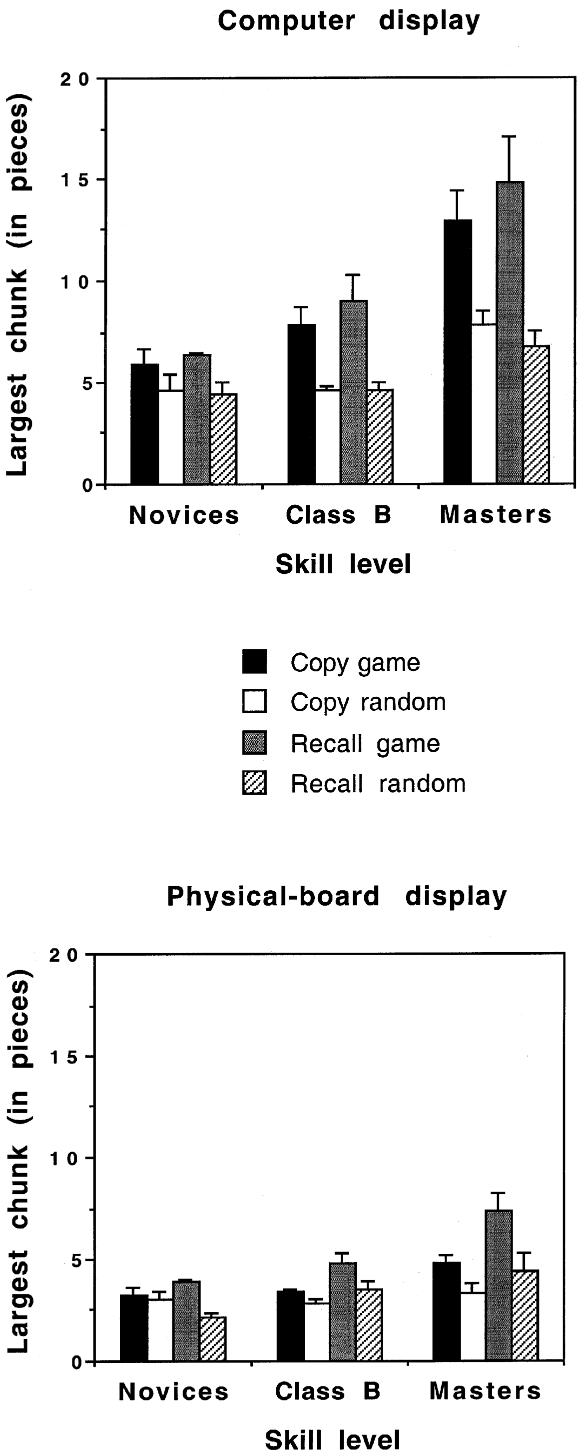
Analysis of chunks

An analysis of the number and the size of chunks is critical in addressing the predictions of the chunking and template theories. In particular, the chunking theory predicts that chunk size should not exceed four or five pieces, and that STM capacity should be around seven chunks. By contrast, the template theory predicts the presence of much larger chunks and a STM capacity of three. As in previous studies, chunks are defined as groups of pieces placed with an interpiece latency of less than 2 s.

Chunk size. Chase and Simon (1973) found a rather small difference in chunk size between skill levels in the recall task, the median largest chunk of their Master being only five pieces. As mentioned in the introduction, the template theory predicts that chunks should be larger than predicted by the original chunking theory (Gobet & Simon, 1996a). This is what Gobet and Simon (1998) found, their Masters obtaining chunks as large as 16.8 pieces in the recall task. The interest in this section is whether the discrepancy in chunk size between the two studies can be explained by the different presentation media.

As the skewness of the data makes the arithmetic mean unsuitable, we focus on the maximum chunk size, which has the added advantage of directly addressing theoretical predictions. (For completeness, Figure 2 shows the median chunk sizes.) The data were calculated by taking the median size of the largest chunk from each experimental presentation to each participant; the median was then averaged within each skill level. Chunks are defined using the 2-second boundary between piece placements as the limit for pieces within a chunk. This includes correctly and incorrectly placed pieces—both are psychologically the same, because both may arise from chunks in memory (Gobet & Simon, 1998), although some placements may be due to guessing, in particular with novices. Figure 1 shows the means obtained for the average maximum chunk size in the various conditions, for the computer and physical-board presentations, respectively.

The ANOVA shows a main effect of skill, $F(2, 9) = 31.42$, $MS_e = 112.59$, $p < .001$, with Masters having larger chunks than both Class B players and novices, a main effect of position type, $F(1, 9) = 74.38$, $MS_e = 175.0$, $p < .001$, with larger chunks for the game condition, and a main effect of display type, $F(1, 9) = 76.16$, $MS_e = 308.17$, $p <$



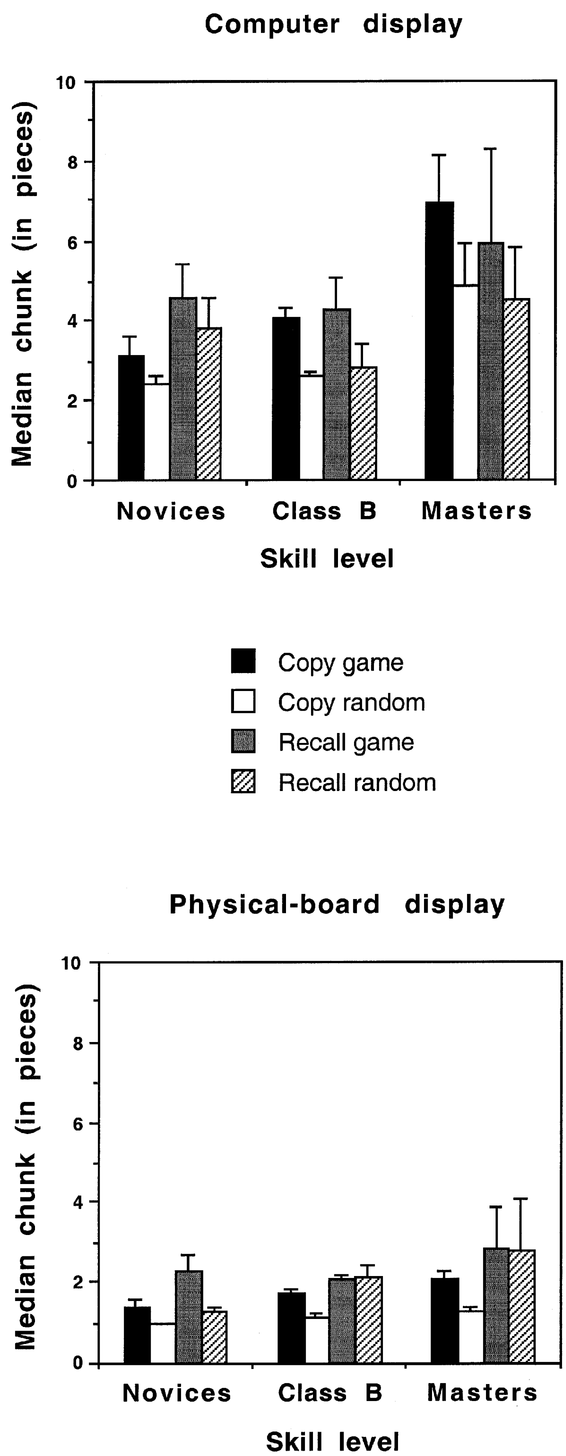


Figure 1 (top left and bottom left). Mean of median largest chunk as a function of skill level, task, and type of position. Upper panel: computer display. Lower panel: physical-board display. Errors bars indicate standard errors of the mean.

Figure 2 (top right and bottom right). Mean of median chunk size as a function of skill level, task, and type of position. Upper panel: computer display. Lower panel: physical-board display. Errors bars indicate standard errors of the mean.

.001, with larger chunks in the computer condition. Significant interactions are shown between display and skill, $F(2, 9) = 6.37$, $MS_e = 25.76$, $p < .05$, with Masters showing a greater increase over the physical board in the computer display than both Class B players and novices; position and skill, $F(2, 9) = 8.38$, $MS_e = 39.41$, $p < .01$, Masters showing a larger difference in chunk size between game and random positions than Class B players or novices; and display and position, $F(1, 9) = 47.87$, $MS_e = 41.34$, $p < .001$, with a greater increase in chunk size for game positions over random positions with computer display.

In game positions, the Masters' median maximum chunk size is the largest with the computer display—14.8 pieces during recall and 12.9 pieces when copying. Masters' maximum chunks are smaller with physical-board displays: 7.5 and 3.8 pieces for recall and copy respectively. Class B players and novices both showed larger chunk sizes in the computer condition: 9.0 pieces and 6.3 pieces respectively in the recall task and 7.8 pieces and 5.9 pieces respectively in the copy task. In random positions, with the computer display, Masters show a slightly greater chunk size (7.9 pieces for copy and 6.7 pieces for recall) than the other skill levels (4.6 pieces for copy and 4.5 pieces for recall); whereas chunk size in random positions remains constant across skill levels in the real board display (3.1 pieces for copy and 3.3 pieces for recall).[8]

The similarity of the relative differences between display types within the current study (computer presentation showing larger chunks than physical-board presentation) and between the experiments of Chase and Simon (1973) and Gobet and Simon (1998), provides further evidence that the early chunking theory underestimated the size of chunks used. However, a possible objection to our analysis of the copy task is that the BGP latencies include viewing time, which could be used to learn larger chunks than those already stored in LTM, for example by concatenating two chunks, or by filling in the slots of templates. (Indeed, this is exactly what is predicted by the template theory; see the computer simulations of the role of the presentation time in a recall task; Gobet & Simon, 2000). This objection seems particularly appropriate in the case of the computer display, which includes BGP latencies as long as 45 s.

We addressed this question by looking at how chunk size varies as a function of the viewing time. Two predictions are made by theories based on chunks: first, there should be some large chunks with short viewing times, and the proportion of these chunks should increase with skill level; second, with long viewing times there should be a correlation between size of chunks and study time, as this may reflect learning processes; by contrast, no such correlation should be present with shorter times, as chunks are assumed to be already stored in LTM. We focus on the data from copying game positions in the computer condition; we used 3 s as threshold (there were too few observations below 2 s) and a size of at least four pieces as definition of a ''large'' chunk. For the first prediction we found that, with viewing times less than 3 s, 67%, 41%, and 12% of the chunks, for Masters, Class B players, and novices respectively, contain at least four pieces. With a Class B player, one chunk was as large as 11 pieces with a viewing time of 1.83 s. With respect to the second prediction, we found that all correlations between study time and chunk size were non-significant with viewing times less than 3 s (the correlations were actually negative), but that they were all significantly positive with viewing times more than 3 s: .71, .52, and .63, for Masters, Class B players, and novices respectively (all $p < .01$). These results support the psychological reality of chunks, but also suggest that the copy task may overestimate chunk size, as participants can take advantage of the time available to acquire larger chunks; this is less likely in the recall task, where the viewing time is limited to 5 s.

Number of chunks. A drawback of using the 2-second boundary for defining chunks is that, in recall tasks, pieces placed individually are often incorrect and seem to be the product of guesswork (Chase & Simon, 1973; Gobet & Simon, 1998; Gobet & Jackson, 2002). As an example from our experiment, consider the recall of random positions displayed on the computer screen. On aggregate, players placed 194 pieces; out of these, 21 pieces were placed individually (i.e., less than one piece per position, on average). Six of these 21

[8] At first blush, it may be surprising to find chunks as big as 6.7 pieces in the recall of random positions. However, CHREST does predict relatively large chunks (up to five pieces with Masters) in this condition, because the 5-second presentation allows chunks to be augmented by the familiarisation mechanism, which takes 2 s, and because the chunks reached in the discrimination net may contain additional but incorrect information, leading to errors of commission, as observed in the human data (Gobet & Simon, 2000).

pieces were placed correctly (28.6%), and 15 incorrectly (71.4%). As counting these pieces would obviously affect the estimation of STM span, which is the purpose of this section, we used the further requirement that chunks contain at least two pieces.

The results (see Figure 3) show a main effect of position type, $F(1,9) = 6.80$, $MS_e = 3.78$, $p < .05$, and of task type, $F(1,9) = 217.9$, $MS_e = 397.5$, $p < .001$, as well as interactions of display type and skill, $F(2,9) = 5.75$, $MS_e = 12.73$, $p < .05$, position and display type, $F(1,9) = 92.94$, $MS_e = 26.30$, $p < .05$, position and task, $F(1,9) = 55.78$, $MS_e = 18.07$, $p < .001$. Interactions are also seen between display type, position type, and skill level, $F(2,9) = 8.35$, $MS_e = 2.36$, $p < .01$, and position, task and skill, $F(2,9) = 8.73$, $MS_e = 2.83$, $p < .001$.

During the copy task, the two presentation methods produce a different pattern of results, which as was mentioned earlier, appears to be due to differences in strategies. Even so, in both presentation media, participants produce fewer chunks in the recall task than in the copy task, because the copy task does not require them to store multiple chunks in STM at the same time. Since one of the hypotheses tested in this paper centres on STM capacity, we will focus on the *recall* task. With the computer display, the number of chunks recalled is always fewer than three, thus supporting the prediction of the template theory. In spite of the small numbers involved, the number of chunks is significantly larger with Masters and Class B players than with novices in the recall of game positions. All groups place between 1.5 and 1.9 chunks when recalling random positions.[9] With the physical boards, random positions, the number recalled is between 1.3 and 2.1 chunks; however, in the recall of game positions, it reaches up to 4.8 chunks with experts, closer to the 7 chunks proposed by Chase and Simon. This is also what was predicted from the template theory with the additional hypothesis that some large chunks get broken down due to the limited size of the hand grasp. (This hypothesis is further supported by the fact that the percentage correct is the same with both presentation media.)

In the present analysis the number of chunks used during the computer presentation is very

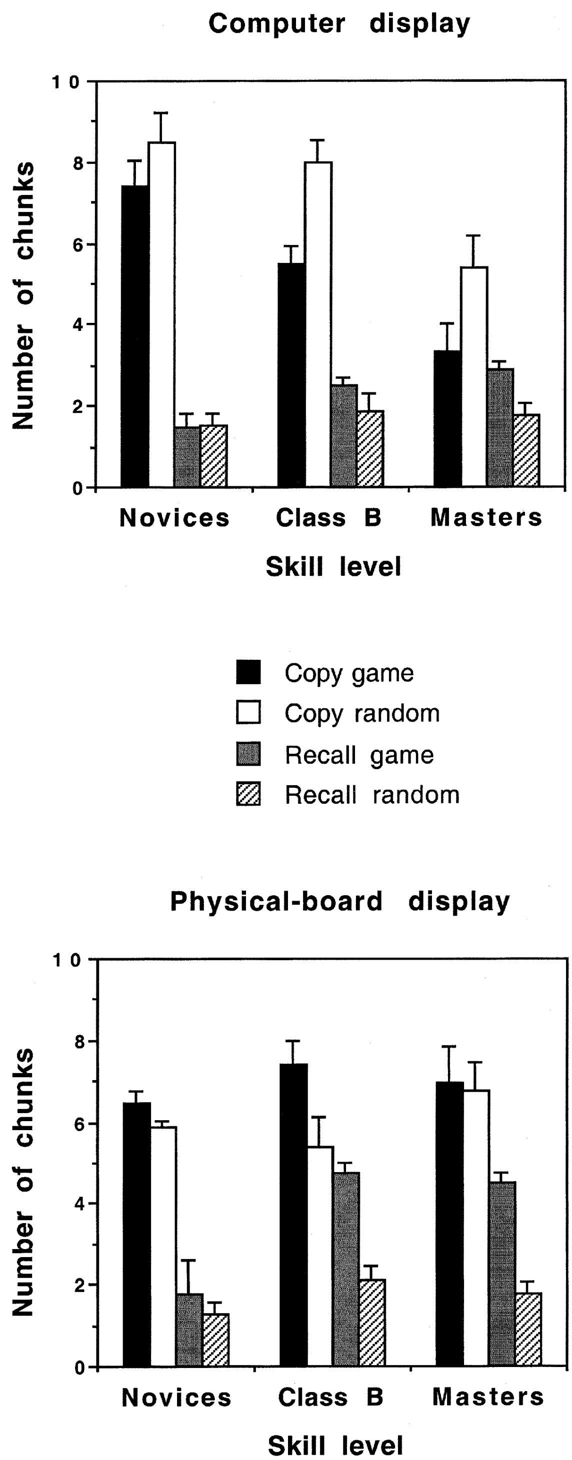


Figure 3. Number of chunks as a function of skill level, task, and type of position. Upper panel: computer display. Lower panel: physical-board display. Errors bars indicate standard errors of the mean.

[9] In the recall tasks, Masters replaced many more pieces (correctly or incorrectly) in the game positions (on average, 87%) than in the random positions (on average, 35%). The difference was smaller for the Class B players (62% vs 31%) and the novices (31% vs 21%).

close to Gobet and Simon's (1998) data; Masters use fewer chunks in the copy task than the other skill levels and recall of random positions shows no differences. The number of chunks in the recall of game positions differs, in that there is a small increase in the number of chunks recalled with skill level, although the differences are non-significant. The physical-board presentation shows a pattern of results similar to Chase and Simon's (1973), Masters and Class B players recalling more chunks than novices with game positions.

Summary of chunk analysis. Both the computer presentation and the physical-board presentation demonstrate that it is the size of chunks, and not the number of chunks that are stored in STM, that mediates skilled players' advantage in the recall task. Overall, the data support the predictions of the template theory rather than those of the chunking theory. With the computer presentation, STM capacity, as estimated by the number of chunks recalled, was below three chunks with all skill levels, and the largest chunks reached 15 pieces with the Masters. With the physical-board presentation of game positions, the number of chunks increased and, accordingly, the size of the largest chunk decreased, as predicted by Gobet and Simon's (1998) hypothesis that hand capacity would bound the number of pieces. Supporting the supposition that this increase of chunk number is due to the breaking down of large chunks, the number of chunks stayed below three with random positions. By contrast, in both conditions, the chunking theory predicted a STM capacity of around seven chunks, and chunks of no more than about five pieces. Finally, the results suggest that Cowan's (2001) estimate of a STM capacity of four chunks may be too high: four out of the six recall conditions yield an estimate around two chunks; the two exceptions (4.8 chunks with Class B players and 4.5 chunks with Masters with game positions in the physical-board condition) can readily be explained by the above hypothesis that these numbers have been inflated by the breakdown of chunks, itself occasioned by the limited hand capacity.

Partitioning-task data

Analysis of the partitioning data posed a number of problems, in particular due to inconsistencies in the types of response between participants. Despite the use of standardised instructions and a high degree of caution to prevent demand effects, participants grouped pieces in the partitioning task in widely varying ways, both within and between participants and skill levels. They rarely put every piece from a position into groupings, often only grouping a small number of pieces, even in the game positions. The following example provides a good indicator of the types of groupings that were observed. One novice grouped most positions in a few small groupings, but on two positions grouped the entire board into two clusters (white and black pieces). There is a very low chance that this participant recognised the position and saw the board as two meaningful collections of pieces beyond the shared colour of the pieces within the groupings. As a consequence of this variability, the results show few of the expected effects.

Two 3 × 2 (Skill level × Position type) mixed ANOVAs were conducted with the following dependent variables: number of clusters and maximum cluster size. A main effect of position type is shown on the maximum cluster size, $F(1, 9) = 24.94$, $MS_e = 16.43$, $p < .001$. Game positions relative to random positions show larger average maximum cluster sizes (5.1 vs 3.4 pieces). These results are consistent with what was found about the chunks in the copy and recall tasks. Contrary to what we expected, however, there was no main effect of skill level or interactions of skill level with position type.

Comparisons were also made between clusters and chunks. The chunks that participants identified in the copy task and the clusters from the partitioning task were matched for each of the positions. The matching process was carried out in both directions (from chunks to clusters and from clusters to chunks) to account for the possibility that chunks and clusters may be subsets of one another. A match is defined if the intersection of an identified chunk with a cluster from the partitioning task is equal to the chunk size or one piece less; the reverse process is used to match clusters to chunks.

The average size of chunks is much greater than that of clusters. For example, for the game position, the sizes are 9.8 and 3.9 pieces respectively. Accordingly, a high percentage of clusters are matched to chunks (mean = 79%) relative to the percentage of chunks matched to clusters (mean = 35%). A similar result is found with the random positions, where a larger proportion of clusters are matched to chunks (mean = 87%) than chunks matched to clusters (mean = 26%). Random positions, as might be expected, were often not grouped at all, suggesting that the participants saw no meaningful relationships between the pieces.

Earlier chess research using partitioning (Chi, 1978; Freyhoff et al., 1992) has found that participants group pieces in clusters similar to the chunks defined by a 2-second latency. The results were inconclusive in the present study. As mentioned, a possible explanation for this outcome is that participants varied more in their grouping methods than in the previous studies, in spite of our instructions, which were carefully controlled to remove any suggestions of how the pieces should be grouped. Presumably different participants understood ''meaningful collections of pieces'' in very different ways. The previous studies do not give the details of the instructions used, so our conclusions must remain very tentative. Another plausible reason for the lack of fit to chunks is based on observations of the participants during the experiment. In order not to confound the recall and copy tasks, the partitioning task was always carried out at the close of the experiment, and many of the participants had begun to lose interest in the experiment by this stage. Also, the number of positions to be partitioned was quite sizeable (28 positions). These two factors combined may have led participants to perform the task hastily and thus to group few pieces, as observed.

GENERAL DISCUSSION

While the chunking hypothesis (Chase & Simon, 1973) has dominated the expertise literature for almost two decades, several alternative theories have recently been proposed to account for top-level performance in domains such as chess. These theories include LTWM theory (Ericsson & Kintsch, 1995), constraint attunement theory (Vicente & Wang, 1998), and template theory (Gobet & Simon, 1996a). In this study, we wanted to put together research into STM capacity (e.g., Cowan, 2001; Miller, 1956) and research into expert behaviour. However, among theories of expertise, only the chunking and the template theory make clear-cut predictions about chunk size and number during a recall task. As a consequence, the paper has concentrated on the divergent predictions of these theories.

In their extension of Chase and Simon's (1973) copy and recall experiments, Gobet and Simon (1998) used a computer display instead of a physical-board display, with the aim of removing the problems related to the limit of hand grasp and to the parallelism of actions. The two methods led to several differences in the estimated chunk sizes and numbers used by Masters; to some extent, these differences paralleled the differential predictions of both theories: an STM capacity of seven chunks and relatively small chunks (up to five pieces) for the chunking theory, and an STM capacity of three chunks and large chunks (up to fifteen pieces) for the template theory. In this paper, the two presentation methods were used with the same participants in order to directly test these predictions.

With respect to recall accuracy, it was found that the display types produced identical levels of performance, suggesting that the relatively poor performance of Chase and Simon's (1973) Master was not due to the presentation method itself. The bulk of this paper centred on the way pieces were chunked by the participants. Chase and Simon's 2-second boundary used to (approximately) define chunks was supported by the analysis of the latencies between piece placements in the copy task. With both types of display, but more so with the computer-presentation method, most placements within one glance at the stimulus board, considered as delimiting a chunk, were less than 2 seconds, whereas few between-glance placements, considered as delimiting separate chunks, were less than 2 seconds.

As predicted by the template theory (supplemented by the hypothesis that limits in hand capacity will break chunks down), but not by the chunking theory, Masters used much larger chunks in the computer condition than in the physical-board condition. As a consequence, Masters also showed greater relative chunk sizes in the game positions over random positions in the computer task. According to the template theory, larger chunks are created by the combination of smaller chunks. However, the method used in this study was not sensitive enough to precisely pinpoint such subcomponents. One possibility to address this question in further research is to observe how novices acquire chunks with practice, thus offering direct evidence that subcomponents are combined in higher structures (e.g., Gobet & Jackson, 2002).

In general, the number of chunks used in the recall task was closer to the three predicted by the template theory than the seven predicted by the chunking theory. The exceptions were with the physical-board presentation of game positions, where Masters and Class B players used 4.5 and 4.8 chunks, respectively—significantly more than the novices. We have argued that these exceptions flow directly from the hypothesis that chunks are split due to handgrasp limits. These results are supported by an experiment of Gobet and Jackson (2002), who, over 15 sessions, trained two novices to memorise chess positions presented on a computer, and found that the number of chunks in a recall task was consistently equal to or less than three. Thus, while the chunking theory was correct that larger chunks, not an increased STM capacity, mediate Masters' superior recall performance, it underestimated the size of chunks and overestimated STM capacity. The results also suggest that confounding factors, such as age and lack of practice, rather than the method used, may have affected Chase and Simon's Master performance. Overall, these results about the size and number of chunks provide strong support for the template theory. The small number of STM chunks is not predicted by LTWM theory (Ericsson & Kintsch, 1995), according to which rapid and flexible LTM storage does not put any constraint in the way the board is reconstructed, and by the constraint attunement theory (Vicente & Wang, 1998), which is silent about internal structures and mechanisms.

The clusters identified in the partitioning task provided a relatively poor match to the chunks obtained in the copy task, in disagreement with previous literature (Chi, 1978; Gruber & Ziegler, 1990). Analysis demonstrated that the clusters showed qualitative differences with the chunks obtained by computer and physical-board presentation; in general, chunks were larger than clusters. Matching of the two groupings for each position showed a much lower proportion of chunks matched to clusters than clusters matched to chunks, which is a direct result of the larger size of chunks. However, the partitioning data failed to provide converging evidence for the definition of chunks; a number of reasons for this outcome were discussed. A task for future research will be to investigate the partitioning task more closely with particular attention to the amount of guidance given about the type of possible groupings.

Beyond chess and expertise, the results are of importance for the estimation of STM capacity. Taken at face value, the numbers of chunks found in most recall conditions, which were below three and even close to two, suggests that both Cowan's (2001) ''number four'' and template theory's ''number three'' overestimate STM capacity, perhaps because the numbers found in various experiments may have been inflated for methodological reasons; a possibility is that some chunks get split into several smaller chunks, as was presumably the case in the physical-board condition of our experiment. By contrast, it could also be argued that some large chunks may reflect smaller chunks grouped during output (e.g., Lane et al., 2001). The template theory, which has time parameters attached to all the cognitive operations it postulates, including the creation and the enrichment of chunks, can beneficially be used to make predictions about the number of chunks in STM and about when and how these chunks may be grouped or split during output, and thus to inform empirical research into STM capacity.

Manuscript received 4 July 2002
Manuscript accepted 16 July 2003
PrEview proof published online 26 November 2003

REFERENCES

Birren, J. E., & Schaie, K. W. (Eds.). (1996). *Handbook of the psychology of aging* (4th ed.). San Diego, CA: Academic Press.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3–18) Chichester, UK: Wiley.

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

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

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

Coltheart, M. (1972). Visual information-processing. In P. C. Dodwell (Ed.), *New horizons in psychology, II*. Harmondsworth, UK: Penguin Books.

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.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114.

de Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton. [First edition in Dutch, 1946.]

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

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182.

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

Ericsson, K. A., & Kirk, P. (2001). The search for fixed generalizable limits of ''pure STM'' capacity: Problems with theoretical proposals based on independent chunks. *Behavioral and Brain Sciences*, *24*, 120–121.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273–305.

Ericsson, K. A., Patel, V., & Kintsch, W. (2000). How experts' adaptations to representative task demands account for the expertise effect in memory recall: Comment on Vicente and Wang (1998). *Psychological Review*, *107*, 578–592.

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

Freyhoff, H., Gruber, H., & Ziegler, A. (1992). Expertise and hierarchical knowledge representation in chess. *Psychological Research*, *54*, 32–37.

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

Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition*, *66*, 115–152.

Gobet, F. (2001). Is experts' knowledge modular? *Proceedings of the 23rd meeting of the Cognitive Science Society* (pp. 336–431). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Gobet, F., & Jackson, S. (2002). In search of templates. *Cognitive Systems Research*, *3*, 35–44.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I. et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*, 236–243.

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 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.

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*, *24*, 651–682.

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

Grant, D. A. (1962). Testing the null hypothesis and the strategy and tactic of investigating theoretical models. *Psychological Review*, *69*, 54–61.

Gruber, H. (1991). *Qualitative Aspekte von Expertise im Schach*. Aachen: Feenschach.

Gruber, H., & Ziegler, A. (1990). Expertisegrad und Wissensbasis. Eine Untersuchung bei Schachspielern. *Psychologische Beitrage*, *32*, 163–185.

Holding, D. H. (1985). *The psychology of chess skill*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Lane, P. C. R., Gobet, F., & Cheng, P. C. H. (2001). What forms the chunks in a subject's performance? Lessons from the CHREST computational model of learning. *Behavioral and Brain Sciences*, *24*, 128–129.

Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*, 103–115.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, *104*, 3–65.

Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–93.

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

Saariluoma, P. (1995). *Chess players' thinking: A cognitive psychological approach*. London: Routledge.

Simon, H. A. (1974). How big is a chunk? *Science*, *183*, 482–488.

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

Simon, H. A., & Gobet, F. (2000). Expertise effects in memory recall: A reply to Vicente and Wang. *Psychological Review*, *107*, 593–600.

Thompson, C. P., Cowan, T., Frieman, J., Mahadevan, R. S., & Vogl, R. J. (1991). Rajan – a study of a memorist. *Journal of Memory and Language*, *30*, 702–724.

Vicente, K. J. (2000). Revisiting the constraint attunement hypothesis: Reply to Ericsson, Patel, and Kintsch (2000) and Simon and Gobet (2000), *Psychological Review*, *107*, 601–608.

Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review*, *105*, 33–57.

Wilding, J., & Valentine, E. (1997). *Superior memory*. Hove, UK: Psychology Press.

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

APPENDIX

Instructions for the Partitioning Task

In the final stage of the experiment you are required to group together chess pieces from all of the board-positions presented throughout the experiment. The chess positions should be divided into groups that indicate a meaningful unit of pieces. Groupings can be made in any way you see fit (as many or as few as you like, on each board) and should represent collections of pieces that you feel to share some relation with one another. A paper representation of all of the board positions will be presented and you should indicate related pieces by encompassing them with a pencil ring. Any further questions about the task will be answered during the experiment.