

mals. Chimpanzees are good candidates for investigating the origins of our cognitive evolution, being the closest relatives of humans among living creatures.

Recently, we found that a numerically trained chimpanzee had a memory for numbers in several aspects similar to humans (Kawai & Matsuzawa 2000b). The chimpanzee called Ai, has more than 20 years of experimental experience. Prior to the memory test, Ai learned to count dots on computer monitor or real objects and to select the corresponding Arabic numerals on a touch-sensitive monitor (Matsuzawa 1985). Ai also learned to order the numbers from zero to nine in sequence, regardless of the inter-integer distance (Biro & Matsuzawa 1999; Tomonaga & Matsuzawa 2000). Utilizing her numerical skills, we set up a memory task. A set of numbers (e.g., 1, 3, 4, 6, and 9) was spatially distributed on a screen. Ai was required to touch the numbers in an ascending order. Immediately after the selection of the lowest number (i.e., 1), all the remaining numbers were masked by a white square. Hence Ai had to memorize the numbers (now masked) accurately to select the correct sequence. She reached more than 95% correct with four numbers and 65% with five, significantly above chance in each case (17 and 4%, respectively). This indicates that she could memorize the correct sequence of any five numbers (Kawai & Matsuzawa 2000b).

The most interesting result concerned Ai's response time. The longest response times were obtained for the first number of the sequence. Response times were shortest for the other numbers, and did not differ from one number to another. Thus, her mean reaction time of first response to a set size of five was 721 msec, and then 446, 426, 466, and 41, respectively, for the remaining four, (now masked) numbers. This pattern of responses is similar to that of humans. For example, mean reaction times of five adult humans in the test were 1,430, 524, 490, 741, 672 msec, for each response, suggesting that both Ai and humans memorize the numbers and their locations before the first response (Biro & Matsuzawa 1999; Kawai, in press).

One may argue that both Ai and humans might use a rehearsal strategy during the longest reaction time preceding the first choice. The accuracy of humans decreased when the numbers were masked 750 msec after the initiation of the trial. However, Ai's reaction times were almost half of those of humans, and they remained approximately the same for the masked and unmasked trials. Although rehearsal constitutes a major cause of compound STM estimate, rehearsal was impossible because Ai's fast reaction times seem incompatible with rehearsal. Thus, Ai's performance in memorizing five items may reflect a "pure capacity limit."

There is a possibility that other mnemonic strategies were involved in the task. For instance, one might suspect that Ai used the configuration of the numbers as possible spatial cues for responding. The procedure ruled out this possibility however, because locations were randomized across trials and all trials were unique in each session, thereby demonstrating that long-term memory did not contribute to the performance. In addition, a detailed analysis of error trials confirmed that neither Ai nor the humans used spatial cues. The majority of errors (84.1% for Ai, 84.5% for humans) consisted in skipping one number only (e.g., selecting 1-3-6 instead of 1-3-4-6-9). The remaining errors were also independent of spatial factors. Most (87.5% for Ai, and 82.0% for humans) consisted of selecting the highest number in the sequence (i.e., 1-9 or 1-3-9 instead of 1-3-4-6-9), regardless of the spatial arrangement of the numbers on the screen. These trials were regarded as showing a recency effect because the highest numbers were the last to be processed in the pre-planned sequence. Even more interesting, the frequency of these last numbers was proportional to the size of the greatest number: more errors of this type were made when the sequence contained a 9 as the last number than when it contained an 8, and that tendency remained for the lowest numbers. All these results suggest that, like humans, Ai built up a linear representation of numbers, from 0 to 9, and referred to it in performing the task. Similarity in the error patterns for the two species moreover suggests that their

memory systems may share homologous mechanisms (Kawai & Matsuzawa 2000a; in press).

Because masking occurred at the first touch, one might argue that the memory span of the chimpanzee is four instead of five. As demonstrated above, both Ai and the humans had already planned their response at the onset of each trial: there is thus the possibility that the first number was included in the memorized sequence. Even more, according to our recent test for six numbers Ai's performance was about 30%, significantly higher than the chance level (0.8%). We do not deny the possibility that the "pure capacity limit" of the chimpanzee might be less than four. Further experimental studies will be required to determine this. Nevertheless, because comparable data were obtained in Ai and the humans, our study strongly suggests that if there are any "pure capacity limit" differences between the two species, they should be quantitative rather than qualitative. The essential point is that humans share quite a similar memory process with chimpanzees.

### What forms the chunks in a subject's performance? Lessons from the CHREST computational model of learning

Peter C.R. Lane, Fernand Gobet, and Peter C-H. Cheng

ESRC Centre for Research in Development, Instruction and Training,  
School of Psychology, University of Nottingham, Nottingham NG7 2RD,  
United Kingdom. {pcl; frg; pcc}@psychology.nottingham.ac.uk  
[www.psychology.nottingham.ac.uk/staff/{Peter.Lane; Fernand.Gobet; Peter.Cheng}](http://www.psychology.nottingham.ac.uk/staff/{Peter.Lane; Fernand.Gobet; Peter.Cheng})

**Abstract:** Computational models of learning provide an alternative technique for identifying the number and type of chunks used by a subject in a specific task. Results from applying CHREST to chess expertise support the theoretical framework of Cowan and a limit in visual short-term memory capacity of 3-4 looms. An application to learning from diagrams illustrates different identifiable forms of chunk.

Cowan's theoretical framework (sect. 2) assumes that the "focus of attention is capacity-limited," and that "deliberately recalled [information] is restricted to this limit in the focus of attention." This framework is compatible with the EPAM/CHREST family of computational models, and this commentary highlights the role that a model of learning can play in clarifying the nature of chunks. CHREST (Chunk Hierarchy and REtrieval STRucture) is a computational model of expert memory in chess players (Gobet 1998; Gobet & Simon, in press), and is based on the earlier EPAM model (Feigenbaum & Simon 1984) of perceptual memory. CHREST possesses an input device (simulated eye), a visual short-term memory (STM) for storing intermediate results (equivalent to the focus of attention), and a long-term memory (LTM) based around a discrimination network for retrieving chunks of information. Each chunk is learnt from information in the visual field, using the STM to compose information across one or more eye fixations.

The classic recall task (Chase & Simon 1973; Cowan, sect. 3.4.1; De Groot 1946; 1978) has been used to show that subjects recall information in chunks. The task requires the model/subject to observe a display for a set time period, and then reconstruct the stimulus from memory; in simulations, the chunks within the model's STM are used as the reconstructed response. In a study of chess expertise, Gobet (1998) showed how the accuracy of the reconstructed position depends on the number and size of chunks which the model identifies; the size of chunk depends on the level of expertise, but the number can be systematically varied, and a value of 3 or 4 was found to best match the performance of different levels of player, providing further empirical support for the findings of Cowan. Also significant is that the better performance of experts is explained by their use of larger chunks (typically, master chess players recall chunks of twice the size of average club

players), and the number and content of these chunks may be extracted from the model (see also Gobet & Simon 1998; in press)

Chase and Simon (1973) did, however, find that expert chess players appeared to recall more chunks than novices. As discussed in Gobet and Simon (1998), these findings do not contradict the existence of a fixed capacity limit, because additional factors affect the subject's performance; in this case, the number of pieces which the player can pick up. So, are the chunks observed in the subject's performance due to previously learnt information or to other factors relating to the task or cognitive performance? This question may be answered through a simulation of the learning process. The role of learnt knowledge in producing chunks in performance is currently being explored in a problem-solving version of CHREST (Lane et al. 2000a) which learns a diagrammatic representation for solving electric circuit problems. In Lane et al. (2000b) different computational models were analysed based on their respective representational, learning, and retrieval strategies for handling high-level information. From these two studies, it is clear that chunks observed in the model's performance may arise from a number of causes. Three of the more apparent are as follows:

(1) A chunk may be observed in the output because of an explicit representation in the system's LTM, which is the underlying representation used in the EPAM/CHREST family of computational models. For example, Richman (1996) describe a chunk as "any unit of information that has been familiarised and has become meaningful."

(2) A chunk may be observed in the output because the input has matched a stored chunk based on some similarity-based criterion; this is familiar from neural network approaches.

(3) A single chunk may be observed although it is based on a functional composition/decomposition of the stimulus and its sub-components. For example, subjects may retrieve and store multiple chunks within their STM, but the performance based on these multiple chunks may then give the appearance of a single chunk.

The presence of three distinct processes yielding chunk-like behaviour in such models clarifies how the observational characteristics of chunks inter-relate with learnt knowledge, and hence clarifies the connection between observed and learnt chunks. This connection assists in developing a deeper understanding of the capacity limit, especially in areas where the subject is continuously learning new chunks for composite objects. Most importantly, only by modelling the entire learning history of each subject can we really attempt to probe the content and format of chunks manipulated in STM, and thereby estimate STM capacity.

## The focus of attention across space and across time

Brian McElree<sup>a</sup> and Barbara Anne Doshier<sup>b</sup>

<sup>a</sup>Department of Psychology, New York University, New York, NY 10003;

<sup>b</sup>Cognitive Science, University of California, Irvine, CA 92717;

bdm@psych.nyu.edu bdoshier@uci.edu

www.psych.nyu.edu/dept/faculty/mcelree/research.html

www.aris.ss.uci.edu/cogsci/personnel/doshier/doshier.html

**Abstract:** Measures of retrieval speed for recently presented events show a sharp dichotomy between representations in focal attention and representations that are recently processed but no longer attended. When information is presented over time, retrieval measures show that focal attention and rapid privileged access is limited to the most recently processed unit or chunk, not the last 3–5 chunks that Cowan estimates from various recall procedures.

Cowan presents a diverse array of evidence to support the claim that the focal attention has a capacity of 3–4 chunks. Much of this evidence comes from studies examining processing limits in multi-element displays in which all elements are simultaneously displayed. These studies may provide good evidence for the claim

that there is a 3-to 4-item limit on the simultaneous coding and re-production of elements, at least in some domains. Cowan believes the same limit holds for sequentially displayed elements, namely, elements distributed over time rather than space. However, Cowan's estimates are based largely on indirect measurements. Crucially, measures of retrieval speed from studies using sequential presentation provide direct evidence for a distinct representational state associated with the focus of attention that is limited to the most recently processed unit. These measures indicate, contra Cowan, that only one chunk is maintained across a dynamically changing environment. One possibility is that the capacity of focal attention differs for simultaneously available elements arrayed in space, and for representations encountered over time. If Cowan's analysis is correct, perhaps we can attend to more than one simultaneously presented element; however, we do not appear to be able to process more than one temporally extended event.

**Cowan's evidence.** Cowan forwards, as an estimate of the capacity of focal attention, findings that the number of recalled items often converges on 3–4. However, recall performance is determined by a confluence of factors other than the capacity of focal attention. Undoubtedly, these estimates partly reflect the recall of representations outside focal attention, analogous to the way that serial position functions were classically argued to reflect output from both long-term and short-term. Further, recall is limited by forgetting that occurs over the learning phase and during the recall process (e.g., Doshier & May 1998). The number of items recalled, even when the preconditions enumerated by Cowan (sect. 1.2) are met, provides at best an indirect estimate of the capacity of focal attention, and is equated with focal attention primarily by assumption.

**Retrieval speed.** The claim that focal attention is distinct from more passive memory representations implies that information in focal attention is accessed more immediately than information in a passive state. Measures for access speed can provide direct evidence for distinct representational states if access speed can be measured for memories with different strengths (or probability of access). Unfortunately, RT does not provide pure estimates of retrieval speed because it is affected by memory strength (e.g., Doshier 1984; McElree & Doshier 1989; Wickelgren et al. 1980). However, retrieval speed can be directly measured with the response-signal speed-accuracy tradeoff (SAT) procedure. In this procedure, subjects are cued to respond at some time after the onset of a test probe. With a suitable range of cue times, the full time course of retrieval is evaluated, providing measures of when information first becomes available, the rate at which information accrues over retrieval time, and the asymptotic level of observed performance. The asymptote reflects the probability of retrieval, and provides an estimate of memory strength. When accuracy departs from chance, the rate at which it grows to asymptote jointly measure retrieval speed. More accessible information should be associated with an earlier intercept or faster rate, irrespective of differences in asymptotic accuracy.

Wickelgren et al. (1980) used a probe recognition task to examine SAT time-course profiles for accessing representations in a list of 16 sequentially-presented items. Asymptotic accuracy decreased monotonically with the decreasing recency of the tested item, indicating that memory strength systematically declines as time or activity is interpolated between study and test. Crucially, however, retrieval speed was constant across all serial positions save the last, most recently studied position. Retrieval speed was 50% faster when no items intervened between study and test. The most recently studied item received privileged access. This finding has been replicated with different procedures and materials, including a Sternberg task (McElree & Doshier 1989), a forced-choice recognition task (McElree & Doshier 1993), a paired-associate recognition task (Doshier 1981), and even when the task required judging whether a test item rhymed or was synonymous with a studied item (McElree 1996). Related effects are found in judgments of recency (McElree & Doshier 1993) and the *n*-back task (McElree, in press).