

Chunking mechanisms in human learning

Fernand Gobet, Peter C.R. Lane, Steve Croker, Peter C-H. Cheng, Gary Jones, Iain Oliver and Julian M. Pine

Pioneering work in the 1940s and 1950s suggested that the concept of 'chunking' might be important in many processes of perception, learning and cognition in humans and animals. We summarize here the major sources of evidence for chunking mechanisms, and consider how such mechanisms have been implemented in computational models of the learning process. We distinguish two forms of chunking: the first deliberate, under strategic control, and goal-oriented; the second automatic, continuous, and linked to perceptual processes. Recent work with discrimination-network computational models of long- and short-term memory (EPAM/CHREST) has produced a diverse range of applications of perceptual chunking. We focus on recent successes in verbal learning, expert memory, language acquisition and learning multiple representations, to illustrate the implementation and use of chunking mechanisms within contemporary models of human learning.

Recent work in perception, learning and cognition has uncovered substantial evidence for a unifying information-processing mechanism, known as 'chunking'. Such a mechanism was initially proposed by De Groot^{1,2}, based on studies of problem solving, and by Miller³, based on studies of perception and memory. Miller's key contribution was to propose an information measure for cognitive systems based upon the concept of a 'chunk', where each chunk collects a number of pieces of information from the environment into a single unit. The use of chunks explains how greater knowledge can lead to an increased ability to extract information from the environment, in spite of constant cognitive limitations. Since Miller's work, researchers in cognitive science have established chunking as one of the key mechanisms of human cognition, and have shown how chunks link the external environment and internal cognitive processes. This article summarizes the major sources of evidence for chunking within humans, and illustrates how chunking has been incorporated into computational models of human learning.

Defining and observing chunks

The literature on chunking encompasses many different areas of research, and the concept of a chunk has consequently diversified in its meaning. The literature itself can be divided into two broad areas, based on how and when chunking is assumed

to occur: the first assumes a deliberate, conscious control of the chunking process (*goal-oriented chunking*), and the second a more automatic and continuous process of chunking during perception (*perceptual chunking*). In spite of the surface variety among descriptions, a common definition of a chunk is possible: a chunk is a collection of elements having strong associations with one another, but weak associations with elements within other chunks^{4–6}. In this article, we address the questions of how chunks might be represented within a cognitive system, and what effect they will have on that system's behaviour.

'Some of the clearest evidence for perceptual chunking is found in how primitive stimuli are grouped into larger conceptual groups, such as letters into words'

Methodologically, a number of approaches have been taken to identify the presence of chunks, and to pinpoint the underlying mechanisms by which chunks are created, stored, retrieved and used. In this brief overview, we focus on a specific family of computational models (EPAM/CHREST) that have been closely associated with perceptual chunking. The importance of these models for investigating chunking mechanisms in human learning is threefold. First, the strong similarities between models lead to a consistent theoretical framework being applied in multiple domains. This parsimony strengthens the claims that chunking underlies many aspects of human learning. Second, detailed computational models provide the only realistic forum for identifying and assessing the major factors in learning from large, noisy and changeable sources of information. Third, the use of computational models enables chunks to be predicted, instead of simply used to explain behaviour *post hoc*.

Some of the clearest evidence for perceptual chunking is found in how primitive stimuli are grouped into larger conceptual groups, such as the manner by which letters are grouped into words, sentences or even paragraphs⁶. This grouping leads to memory and behavioural effects, in which the latencies for the output of items comprising a chunk are shorter than when those items comprise a number of smaller chunks^{4,7–12}. We illustrate this type of chunking effect in the chess domain (Box 1), as well as in a more complex problem-solving domain, involving multiple representations in physics.

A further line of evidence considers the detailed construction and retrieval of perceptual chunks within memory. The manner in which chunks are constructed affects the types of generalizations made, and so predicts typical errors or successes. We

Fernand Gobet*

Peter C.R. Lane

Steve Croker

Peter C-H. Cheng

Gary Jones

Iain Oliver

Julian M. Pine

ESRC Centre for Research in Development, Instruction and Training, School of Psychology, University Park, Nottingham, UK NG7 2RD.
e-mail: frg@psychology.nottingham.ac.uk

Box 1. Expert memory for chess positions

Much of what is known about expertise goes back to De Groot^a and Chase and Simon^b. One of De Groot's enduring contributions was to demonstrate the existence of clear differences between levels of player in a memory task, involving the brief presentation of a position taken from a tournament game. Typically, players at and above master level recall the entire position almost perfectly, but weaker players perform poorly (see Fig. I). However, Chase and Simon found no difference in recall of *random* positions between their three

subjects: a master, a class A player, and a novice. This uniformly poor recall of random positions, taken together with the superior performance of masters and grandmasters on game positions, presented such a vivid illustration of the principle that knowledge is the key to expertise that it has become a classic finding, widely cited in textbooks of cognitive psychology and in papers on expertise.

However, an earlier version of CHREST, a re-implementation and extension of MAPP^c, made contrary predictions about the recall of random positions. In the chess simulations, CHREST is trained from a database of master games, identifying patterns of pieces in these positions. As expected, the model's ability to remember game positions improved as the number and average size of its chunks increased. However, the model also showed a small, but robust increase in recall with random positions. The skill differences in recall were the result of an easily explained mechanism: simply by chance, a larger discrimination network is likely to include patterns found in random positions. A systematic review of experiments that asked chess players to recall random positions^d yields 12 studies in which masters demonstrated some superiority, and only one, Chase and Simon's study^b, where the master actually did worse than novices. Although the skill differences were not significant in most studies because of lack of statistical power, the effect becomes clear when the various studies are pooled (see Fig. Ib). The fact that perceptual chunking provides masters with an advantage even in random positions offers strong support for chunk-based theories, and is hard to explain for theories of expertise based upon high-level knowledge or schemata^e.

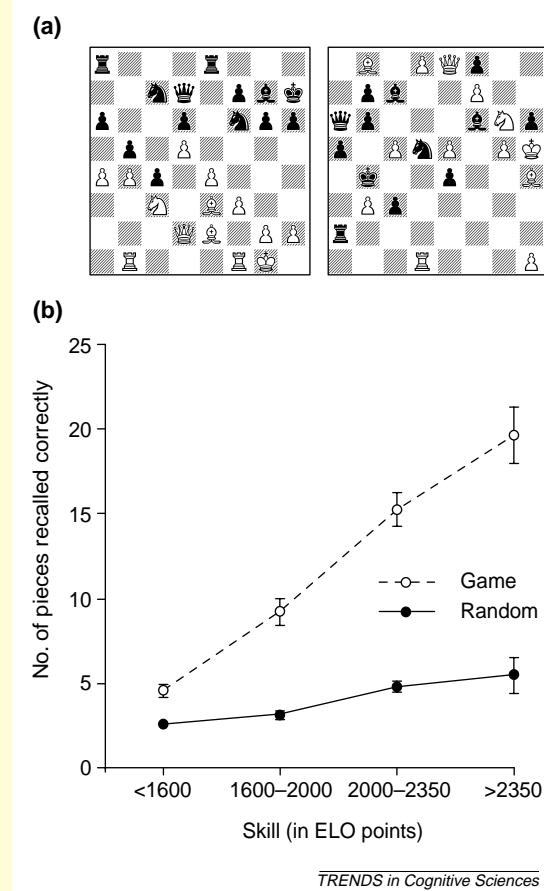


Fig. I. (a) Types of positions typically used in chess memory research. A game position taken from a masters' game (left), and a random position obtained by shuffling the piece locations of a game position (right). (b) Mean number (averaged over 13 studies) of pieces placed correctly as a function of position type (game or random) and skill level. Positions had 25 pieces on average, and the presentation time was ≤ 10 s. Error bars indicate standard errors of the means. (Adapted from Ref. d.)

describe some research in language learning where the mechanisms of chunk combination and generalization are used to predict typical errors in the linguistic development of children. Box 2 similarly illustrates how specific mechanisms of chunk retrieval and learning capture details of the learning curve within a verbal learning experiment.

References

- a De Groot, A.D. (1946) *Het Denken van den Schaker*, Noord Hollandsche
- b Simon, H.A. and Chase, W.G. (1973) Skill in chess. *Am. Sci.* 61, 393–403
- c Simon, H.A. and Gilmartin, K.J. (1973) A simulation of memory for chess positions. *Cognit. Psychol.* 5, 29–46
- d Gobet, F. and Simon, H.A. (1996) Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bull. Rev.* 3, 159–163
- e Gobet, F. (1998) Expert memory: a comparison of four theories. *Cognition* 66, 115–152

EPAM

Shortly after Miller's 1956 paper, Feigenbaum and Simon began to develop a pure and direct implementation of chunking mechanisms, known as EPAM (Elementary Perceiver and Memorizer)^{13–15}. Learning is simulated by the growth of a discrimination network, where internal nodes test

Box 2. Verbal learning

Ebbinghaus, in his 1885 monograph on learning and memory, initiated the area of experimental psychology known as verbal learning^a. Although predominant in the 1940s–1960s, this area of research is still highly productive, involving a number of key questions based around a simple experimental format. For example, in the paired-associate task, subjects are presented with a list of stimulus–response pairs, such as: <DAG-BIF> <DOZ-TEK> etc. Then, subjects are prompted with the stimulus only, and asked to recall the appropriate response (e.g. <DAG->); the target list is presented as many times as required until some level of success is attained. A number of important questions in learning can be investigated using this paradigm, including the serial position effect, the constant fixation time effect, the Von Restorff effect, oscillation and retroactive inhibition, the effects of similarity, familiarity and meaningfulness, and one-trial learning^{b-d}.

EPAM was, and arguably still is, the best explanation of these phenomena using a single theory^d (an illustration of how EPAM explains oscillation during learning is given in Fig. 1). EPAM mechanisms have also been used to explain other phenomena in verbal learning, such as: the tip-of-the-tongue

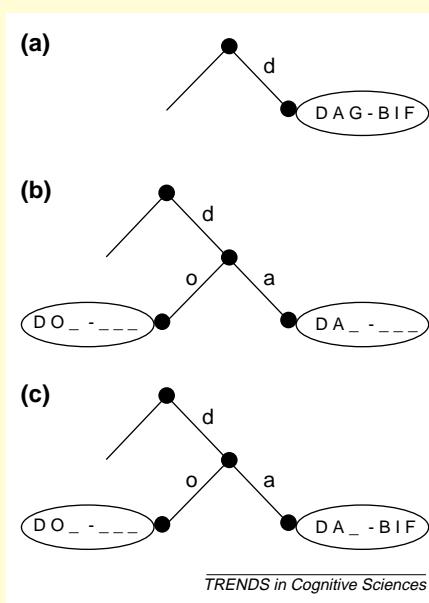


Fig. 1. Verbal learning. This figure illustrates how EPAM can forget an already learnt response; such 'forgotten' responses can be relearned on re-presentation of the target stimulus–response pair, leading to the phenomenon of 'oscillation'. (a) After several presentations of a given stimulus–response pair, such as <DAG-BIF>, the network is as shown. (b) Next, the system encounters <DOZ-TEK>. Discrimination leads the network to be modified as shown. Note that, in discriminating, only those elements known to be features are included in the node image, and so the target response is lost, hence the network has forgotten something it knew before. (c) On seeing <DAG-BIF> again, the network can relearn the correct response.

phenomenon^e, and the learning of spelling by children^f. Also of historic interest is that EPAM illustrates how the information-processing 'revolution' of the 1950s, far from rejecting phenomena identified by (neo-)behaviourism, actually provided mechanisms that explained them. Neo-behaviourists considered that complex verbal skills could be seen as chains or hierarchies of elementary stimulus–response associations, a view that is not too far from chunking – with the qualification, though, that information-processing models postulate internal mechanisms and structures.

References

- a Ebbinghaus, H. (1964) *Memory: A Contribution to Experimental Psychology*, Dover
- b Feigenbaum, E.A. and Simon, H.A. (1962) A theory of the serial position effect. *Br. J. Psychol.* 53, 307–320
- c Simon, H.A. and Feigenbaum, E.A. (1964) An information processing theory of some effects of similarity, familiarity, and meaningfulness in verbal learning. *J. Verb. Learn. Verb. Behav.* 3, 385–396
- d Feigenbaum, E.A. and Simon, H.A. (1984) EPAM-like models of recognition and learning. *Cognit. Sci.* 8, 305–336
- e Brown, R. and McNeill, D. (1966) The 'tip of the tongue' phenomenon. *J. Verb. Learn. Verb. Behav.* 5, 325–337
- f Simon, D.P. and Simon, H.A. (1973) Alternative uses of phonemic information in spelling. *Rev. Educ. Res.* 43, 115–137

for the presence of perceptual features, and leaf nodes store 'images', the internal representation of external objects. The learning mechanisms, detailed below, support the addition of information to leaf nodes and the addition of new tests. Although seemingly a simple model, EPAM has replicated a large number of empirical results in verbal learning and expertise research.

Components

The original version of EPAM consisted of a finite short-term memory (STM), a discrimination network, and attention mechanisms, although the core mechanisms are most apparent in the later version, EPAM-III¹⁵ (see Fig. 1). The discrimination network consists of a hierarchical sequence of tests, with each test located at an internal node of the network (see Fig. 2a). The leaf nodes in the network contain an internal representation of the external object (an 'image'), and might index further procedural or semantic information in long-term memory (LTM). EPAM assumes (but does not implement) mechanisms for extracting features from its input stimuli.

Information flows through EPAM as follows. First, a stimulus is perceived and converted into a set of features. Second, these features are sorted by the tests of the discrimination network, to retrieve a pointer to a node within LTM; this pointer is then stored within STM. Third, depending on an internal comparison process, learning may or may not occur within the network. Finally, an action might be taken by the system, or else the next stimulus is retrieved from the environment. Note that some mechanisms, such as eye movement and learning, can operate in parallel.

Mechanisms of learning

The extent of the system's knowledge about a given stimulus is indicated by the leaf node reached after sorting that stimulus through the discrimination network. Learning occurs by comparing the information held in the leaf node with that in the stimulus. If there is a perfect match, no learning occurs. If the image is a subset of the stimulus, additional features are added to the image (Fig. 2b). Finally, if there is a mismatch, the network is augmented: the leaf node becomes an internal node,

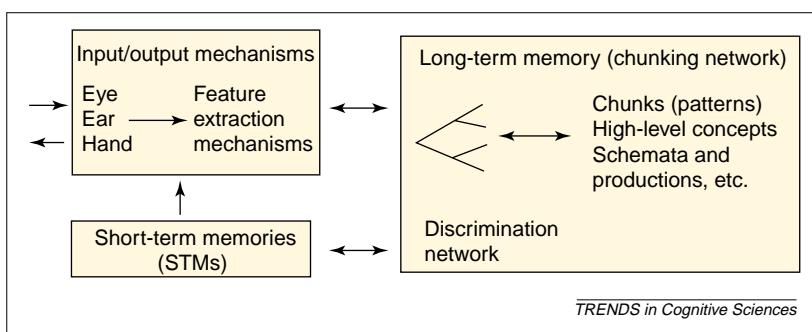


Fig. 1. Overview of the EPAM/CHREST architecture (see text for details).

with a test for the mismatch, and the stimulus and old image are used as the basis of new leaf nodes (Fig. 2c).

Time parameters

One of the strongest features of EPAM is that it includes approximate but absolute parameters that allow quantitative predictions to be made about the rate of learning and information retrieval. Some of these parameters characterize capacity limits¹⁶ (e.g. between 3 and 7 items in STM), but most quantify the time required for key cognitive processes: traversing a node during sorting takes about 10 ms, creating a new node takes about 10 s, and adding information to an extant node takes about 2 s. These parameters have been derived from both empirical data and architectural considerations^{15–17}.

Domains of application

EPAM has applications to various phenomena in verbal learning (Box 2). EPAM has also been used to

explain the role of context in letter perception¹⁸, the role of strategies in concept formation¹⁹, and the acquisition of chess expertise^{20,21}. In particular, the EPAM mechanisms for the acquisition of chess expertise led to the development of the chunking theory itself^{4,22}, which has had a substantial impact on research into expertise in general²³. Some recent work on chess expertise is illustrated in Box 1.

The main limitations of early versions of EPAM include the slow storage of knowledge into LTM and the lack of specific mechanisms for creating semantic knowledge. The former assumption seems valid enough for domains where individuals have a low level of expertise, such as in verbal learning. However, research in expert behaviour has shown that experts can rapidly store material from their domain of expertise^{24,25}. The next section describes CHREST, which is one of two recent extensions to EPAM designed to remedy these shortcomings (similar changes have been included in the other extension, EPAM-IV; Ref. 26).

CHREST

CHREST (Chunk Hierarchy and REtrieval STructures)^{17,27–30} features a number of additions to EPAM's basic learning mechanisms, providing a greater degree of self-organization and adaptation to complex data. This section summarizes some of the new mechanisms within CHREST, before describing some applications.

Components

The general organization of CHREST is similar to that of earlier versions of EPAM (see Fig. 1). In addition, all the major mechanisms of EPAM have been retained within CHREST: information in LTM is indexed through a discrimination network; LTM learning occurs through the processes of discrimination and familiarization; information must be stored within STM before it can be compared or used; and all time parameters for the learning and retrieval mechanisms are retained. One small difference is that every node within CHREST's discrimination network can contain an image (compare Figs 2a and 3a). The major changes are in the form of additional mechanisms for creating lateral links between nodes, and for elaborating information within chunks to form more complex schemata; these changes improve the richness of semantic memory without affecting important properties of the previous simulations.

Lateral links. A lateral link is a semantic association between two nodes within the discrimination network³¹; some examples are illustrated in Fig. 3. Learning a lateral link can only occur when the system's STM contains a pointer to the relevant nodes within the discrimination network; this constraint ensures that links between nodes are based only on a spatial or temporal contiguity, thus

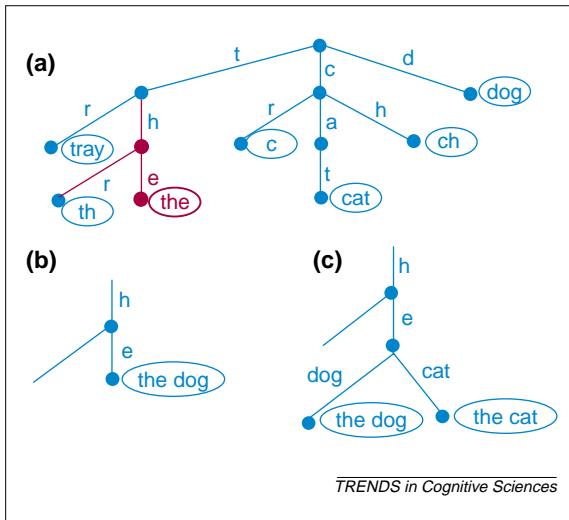


Fig. 2. EPAM-mechanisms. (a) An example discrimination network. The dark circles represent the nodes within the network; solid lines represent the test links between pairs of nodes, with the test letters shown. Ellipses contain images for the leaf nodes. Note that images may contain more or less information than is contained in the path of tests from the root node. (b) Presenting 'the dog' leads to 'familiarization'; that is, information is added to the current node. Note that the entire word 'dog' can be added to the image, as it appears elsewhere in the network. (Only the part of the network shown in red is altered.) (c) Subsequently presenting 'the cat' leads to discrimination; that is, extra links and nodes are added to the network. Note that the link can use whole words as tests if they appear elsewhere in the network.

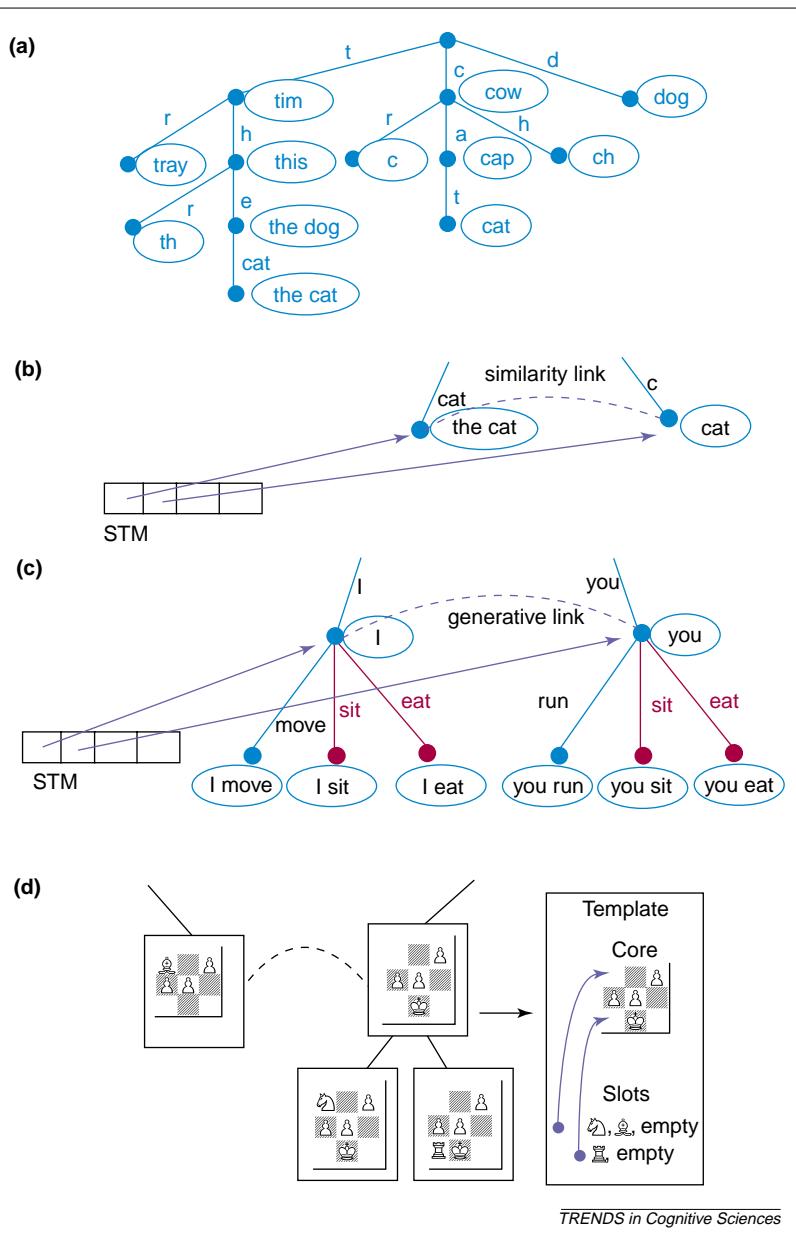


Fig. 3. (a) An example CHREST discrimination network. This differs from the EPAM network in that every node has an image; thus, during discrimination, only one new node need be added, as highlighted for 'the cat'. (b) Forming a 'similarity' link. Such links can be formed when pointers to two nodes of the discrimination network are within STM, and the images at these nodes overlap (e.g. in the case of 'the cat' and 'cat' in the above network). 'Equivalence' links and 'production rules' are formed in an identical fashion, with two nodes referred to in STM satisfying a condition of spatial or temporal contiguity, instead of a simple overlapping of the node images. (c) Forming a 'generative' link. Generative links associate two nodes that have similar descendant test links; for example, the nodes for 'I' and 'you' here agree in the highlighted verbs. The network might follow a generative link to create novel utterances, not found in its training data. (d) Forming a template. Templates represent a consolidation of information from a number of separate nodes. They are formed when a node in STM satisfies certain conditions relating to the degree of overlap and variety of information in its child nodes and any nodes associated with it through similarity links. When these conditions are met, the node is converted into a template, with a 'core' based on the image of the original node, and 'slots' based on the variations in the images contained in the child and similar nodes. In the template illustrated, the slots contain pieces that could occupy particular squares.

preserving an essential property of perceptual chunks. The nodes within STM are then compared using a similarity function, which compares the images or descendant tests at the nodes. Broadly speaking, lateral links can be divided into two

categories: one category is where the two nodes match the similarity function, and the second category is where they do not. Applications of CHREST have used several kinds of lateral link, including: *similarity links*, to connect nodes whose images are similar (see Fig. 3b); *generative links*, to connect nodes whose descendant tests are similar (Fig. 3c); *production links*, to form condition–action rules; and *equivalence links*, to link multiple representations of the same object.

Schemata. CHREST also proposes mechanisms for the creation of 'retrieval structures', which are stable structures of retrieval cues; various forms of specific and generic retrieval structures are described in the literature^{28,32–37}. The specific retrieval structures used in CHREST are known as 'templates' and are created automatically during pattern recognition^{27,34}. Templates are created when an internal node meets specific criteria relating to its connectivity with other nodes within semantic memory; Fig. 3d illustrates a simple example. The template provides a form of slotted schema, with both fixed and variable information, and can be referenced within STM as a single chunk. Templates allow specific details of a stimulus to be plugged into appropriate slots (an operation requiring only around 250 ms), facilitating rapid recall. It is the presence of templates that can account for much of experts' superior memory skills^{28,34}.

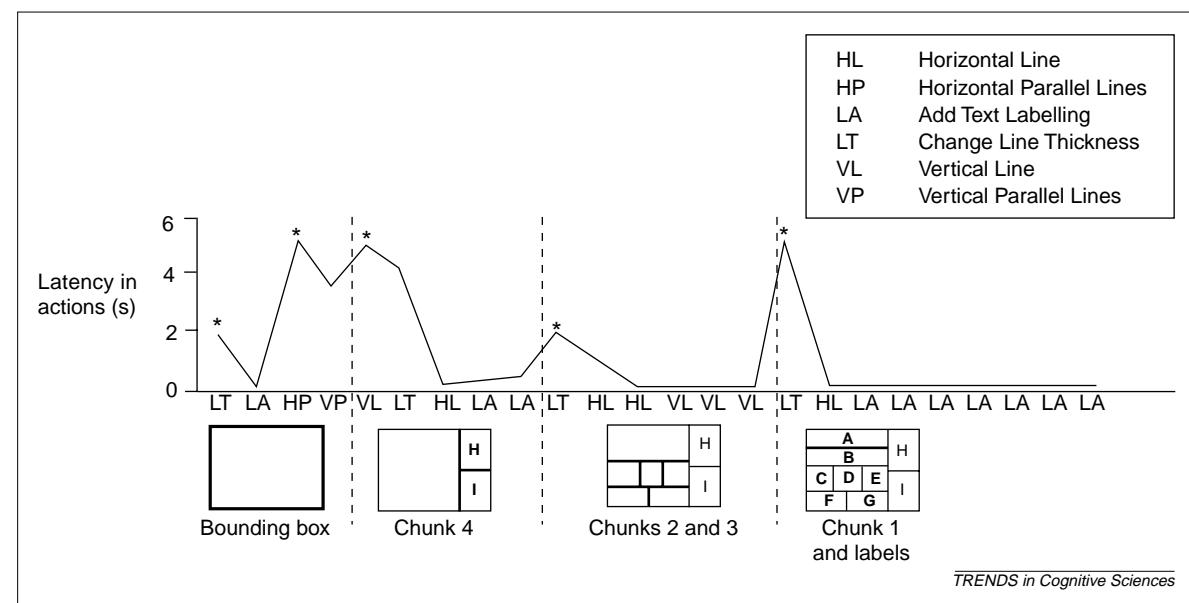
Applications

CHREST has been used in several domains to explore the role and impact of perceptual chunking. In most of these applications, the model learns from naturalistic input, and its performance is directly compared with that of human subjects. We focus here on expert behaviour in chess, the acquisition of multiple representations in physics, and the acquisition of syntactic categories in language learning. Other applications of CHREST include the Piagetian balance beam task³⁸, the acquisition of vocabulary³⁹, and expert knowledge of computer programs (Gobet, F. and Oliver, I., unpublished data).

Expert memory for chess positions

CHREST's first application was to chess expertise, continuing earlier work by Barenfeld, Gilman and Simon^{20,21}. CHREST learns a discrimination network by scanning its simulated eye across positions taken from a large database of masters' games. Pieces within the visual field are passed to the discrimination network for learning, and information across successive fixations is combined in STM. Several sets of empirical data have been accounted for by CHREST, including the recall of random positions (see Box 1), the role of presentation time, the effect of various board modifications (e.g. by mirror image), and eye movements during the first seconds in which a position is presented^{17,28–30,40,41}.

Fig. 4. The performance of one subject when solving an electric circuit problem using a diagrammatic representation. The progressive solution is illustrated below the graph, which plots the latency between each drawing action. The features added at each stage are indicated by bold lines in the schematic solution. The peaks in the graph (highlighted by *) divide the solution into stages; these stages correspond to the predicted chunks derived from the model.



TRENDS in Cognitive Sciences

Three key assumptions account for most of these results: slow encoding time, unless templates are accessed; limited STM capacity; and learning mechanisms that pick up patterns that often recur in game positions. Various proposals for using chunking mechanisms in actual chess playing have also been made^{22,42–44}.

Learning multiple representations in physics
In a similar way to its application in chess, it is possible to apply the chunking theory to more general areas of education, as has been proposed by Glaser and others^{45,46}. Our own research group has tackled the learning of electric circuits, a topic often taught in introductory physics courses. In particular, we were interested in how students learn to combine and use multiple representations of electric circuits^{47–49}, producing one form of representation based on the other. As illustrated in Fig. 4, the production of the diagrammatic solution can be subdivided, based on latencies in the drawing actions, into a number of stages. The stages are consistent across different subjects and correspond with the model's predictions based on the identification of familiar chunks in the given problem. These results show that chunking effects arise during complex problem solving, and are analogous to those found in the pure recall of chunks in chess and alphabet recitation¹⁰. Such results have immediate potential for application in the development of educational curricula, where the model can be used as a subject '*in silico*' for investigating the effectiveness of different presentational content and ordering.

Acquisition of syntactic categories

We have developed a variant of CHREST, called MOSAIC, that simulates the early acquisition of syntactic categories by children (aged 2–3 yrs)^{50,51}.

MOSAIC learns in a similar manner to CHREST, taking input from maternal utterances recorded in a play situation⁵². Where MOSAIC differs from the other applications of CHREST is in its use of an auditory input, and its ability to construct generative links, which facilitate the generation of novel utterances (Fig. 3c). Words associated by generative links form groups, which approximate more formal syntactic categories.

'In a similar way to its application in chess, it is possible to apply the chunking theory to more general areas of education'

MOSAIC can be used to predict the kinds of errors made in combining chunks acquired during early language experience; the pattern of errors within MOSAIC conforms to that found in young children, demonstrating that the chunking mechanisms within MOSAIC (and therefore within CHREST) conform to similar learning mechanisms in humans. Other work has confirmed that these ideas also apply to vocabulary acquisition³⁹.

Other computational approaches to chunking

Various computational approaches adopt the idea of chunking, either as a central or incidental feature. Chunking arises naturally within symbolic models of cognition, where elements of information are combined into single units. In this article, we have emphasized the use of a discrimination network to index long-term memory using the EPAM/CHREST family of models, although other such families exist^{53,54}. An alternative approach relies on a production-rule representation for long-term memory. The two major examples of this are Soar⁵⁵

Questions for future research

- What is the neurophysiological plausibility of computational models based on chunking mechanisms?
- How can chunking mechanisms best be linked directly to sensors and effectors? And what impact will this have on the primitive elements on which chunks are based?
- How can perceptual and goal-oriented chunking mechanisms best be integrated?
- Can a theory based purely on chunking mechanisms lay any claims towards being a universal theory of cognition?
- Is language acquisition just like the acquisition of any other type of knowledge (as implied by the common mechanisms across domains of the CHREST/MOSAIC models)?
- Can the modelling work be extended to include chunking mechanisms in non-human primates and other animals?

and ACT-R⁵⁶, which differ in the amount of information contained within their rules, and in what information may be incorporated into learning new rules.

Within Soar, chunks are the production/action rules themselves, which include explicit goals in their descriptions. Learning new chunks takes advantage of the dependency of subgoals observed during problem solving. Within ACT-R, chunks include all types of declarative knowledge, which can take various forms. Note that the EPAM/CHREST class of model most clearly exemplifies the notion of perceptual chunking, whereas this second class incorporates more features relevant to goal-oriented chunking. In recent years though, both the ACT-R and Soar communities have added perceptual and motor extensions to their architectures, providing a basis for including perceptual chunking within their models. An unrealized goal for EPAM/CHREST-type models has been to include goal-oriented chunking.

Acknowledgements

The authors would like to thank Daniel Freudenthal, the anonymous referees, and the editor of Trends in Cognitive Sciences for their helpful comments on an earlier version of this article. This research was funded by the UK Economic and Social Research Council and the Leverhulme Trust.

References

- 1 De Groot, A.D. (1946) *Het Denken van den Schaker*, Noord Hollandsche
- 2 De Groot, A.D. (1978) *Thought and Choice in Chess*, Mouton Publishers
- 3 Miller, G.A. (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97
- 4 Chase, W.G. and Simon, H.A. (1973) Perception in chess. *Cognit. Psychol.* 4, 55–81
- 5 Cowan, N. (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24
- 6 Simon, H.A. (1974) How big is a chunk? *Science* 183, 482–488
- 7 Broadbent, D.E. (1975) The magic number seven after fifteen years. In *Studies in Long-term Memory* (Kennedy, A. and Wilkes, A., eds), pp. 3–18, John Wiley & Sons
- 8 Chase, W.G. and Ericsson, K.A. (1981) Skilled memory. In *Cognitive Skills and their Acquisition* (Anderson, J.R., ed.), pp. 141–189, Erlbaum
- 9 Gobet, F. and Simon, H.A. (1998) Expert chess memory: revisiting the chunking hypothesis. *Memory* 6, 225–255
- 10 Klahr, D. et al. (1983) Structure and process in alphabetic retrieval. *J. Exp. Psychol. Learn. Mem. Cognit.* 9, 462–477
- 11 McLean, R.S. and Gregg, L.W. (1967) Effects of induced chunking on temporal aspects of serial recitation. *J. Exp. Psychol.* 74, 455–459
- 12 Reitman, J.S. and Rueter, H.H. (1980) Organization revealed by recall orders and confirmed by pauses. *Cognit. Psychol.* 12, 554–581
- 13 Feigenbaum, E.A. and Simon, H.A. (1962) A theory of the serial position effect. *Br. J. Psychol.* 53, 307–320
- 14 Feigenbaum, E.A. (1963) The simulation of verbal learning behavior. In *Computers and Thought* (Feigenbaum, E.A. and Feldman, J., eds), pp. 297–309, McGraw-Hill
- 15 Feigenbaum, E.A. and Simon, H.A. (1984) EPAM-like models of recognition and learning. *Cognit. Sci.* 8, 305–336
- 16 Simon, H.A. (1969) *The Sciences of the Artificial*, MIT Press
- 17 De Groot, A.D. and Gobet, F. (1996) *Perception and Memory in Chess: Heuristics of the Professional Eye*, Van Gorcum
- 18 Richman, H.B. and Simon, H.A. (1989) Context effects in letter perception: comparison of two theories. *Psychol. Rev.* 3, 417–432
- 19 Gobet, F. et al. (1997) Goals, representations, and strategies in a concept attainment task: the EPAM Model. *Psychol. Learn. Motiv.* 37, 265–290
- 20 Simon, H.A. and Barenfeld, M. (1969) Information processing analysis of perceptual processes in problem solving. *Psychol. Rev.* 7, 473–483
- 21 Simon, H.A. and Gilmarin, K. (1973) A simulation of memory for chess positions. *Cognit. Psychol.* 5, 29–36
- 22 Simon, H.A. and Chase, W.G. (1973) Skill in chess. *Am. Sci.* 61, 393–403
- 23 Charness, N. (1992) The impact of chess research on cognitive science. *Psychol. Res.* 54, 4–9
- 24 Charness, N. (1976) Memory for chess positions: resistance to interference. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 641–653

Non-symbolic computational models represent a further alternative. The essential idea behind these models, including many connectionist networks, is that information should be distributed across multiple units, and that no unit should be dedicated to a specific function. This, almost by definition, rules out chunking mechanisms, although simplified forms are possible. For example, Elman has argued for the value of a staged learning cycle in work on language learning with simple recurrent networks (SRNs)^{57,58}. The idea is that a complex recursive grammar is too hard for the SRN to learn all at once, and so SRNs are first trained to identify the local dependencies within the grammar, and later to extend these to longer dependencies. In this case, the local dependencies can be interpreted as chunks. Further non-symbolic algorithms that could similarly be interpreted as learning chunks include Kohonen networks⁵⁹ and the Adaptive Resonance Theory⁶⁰.

Conclusion

This article has described and summarized chunking mechanisms in human learning, focusing on the EPAM/CHREST family of computational models and their applications. We have identified two broad classes of chunking: goal-oriented and perceptual chunking. From the diversity of available empirical evidence, the general notion of chunking appears to be a robust and important one in contemporary cognitive science. The lesson to be taken away from the EPAM/CHREST examples we have described is that perceptual chunking is a valuable element of distributional accounts in many areas, such as verbal learning, expertise, problem solving and language acquisition. Future work is likely to see better integration between the goal-oriented and perceptual forms of chunking, leading to more comprehensive applications and cognitive theories.

- 25 Chase, W.G. and Ericsson, K.A. (1982) Skill and working memory. *Psychol. Learn. Motiv.* 16, 1–58
- 26 Richman, H.B. *et al.* (1995) Simulation of expert memory with EPAM IV. *Psychol. Rev.* 102, 305–330
- 27 Gobet, F. and Simon, H.A. (2000) Five seconds or sixty? Presentation time in expert memory. *Cognit. Sci.* 24, 651–682
- 28 Gobet, F. (1998) Expert memory: a comparison of four theories. *Cognition* 66, 115–152
- 29 Gobet, F. and Simon, H.A. (1996) Recall of random and distorted positions: implications for the theory of expertise. *Mem. Cognit.* 24, 493–503
- 30 Gobet, F. (1993) A computer model of chess memory. In *Proc. Fifteenth Annu. Cognit. Sci. Soc.*, pp. 463–468, Erlbaum
- 31 Gobet, F. (1996) Discrimination nets, production systems and semantic networks: elements of a unified framework. In *Proc. Second Int. Conf. Learn. Sci.*, pp. 398–403, Northwestern University
- 32 Ericsson, K.A. and Kintsch, W. (1995) Long-term working memory. *Psychol. Rev.* 102, 211–245
- 33 Ericsson, K.A. and Kintsch, W. (2000) Shortcomings of generic retrieval structures with slots of the type that Gobet (1993) proposed and modelled. *Br. J. Psychol.* 91, 571–590
- 34 Gobet, F. and Simon, H.A. (1996) Templates in chess memory: a mechanism for recalling several boards. *Cognit. Psychol.* 31, 1–40
- 35 Gobet, F. (2000) Some shortcomings of long-term working memory. *Br. J. Psychol.* 91, 551–570
- 36 Lane, P.C.R. *et al.* (2000) Learning-based constraints on schemata. In *Proc. Twenty Second Annu. Cognit. Sci. Soc.*, pp. 776–781, Erlbaum
- 37 Saariluoma, P. and Laine, T. Novice construction of chess memory. *Scand. J. Psychol.* (in press)
- 38 Gobet, F. (1999) Simulations of stagewise development with a symbolic architecture. In *Dynamics, Synergetics and Autonomous Agents* (Dauwalder, J.P. and Tschacher, W., eds), pp. 143–158, World Scientific
- 39 Jones, G. *et al.* (2000) Learning novel sound patterns. In *Proc. Third Int. Conf. Cognit. Modelling*, pp. 169–176, Universal Press
- 40 Gobet, F. (1993) *Les mémoires d'un joueur d'échecs [Chess players' memories]*, Editions Universitaires
- 41 Gobet, F. and Simon, H.A. (1996) Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bull. Rev.* 3, 159–163
- 42 Chase, W.G. and Simon, H.A. (1973) The mind's eye in chess. In *Visual Information Processing* (Chase, W.G., ed.) pp. 215–281, Academic Press
- 43 Gobet, F. and Jansen, P. (1994) Towards a chess program based on a model of human memory. In *Advances in Computer Chess* (Vol. 7) (van den Herik, H.J. *et al.*, eds), pp. 35–60, University of Limburg Press
- 44 Gobet, F. (1997) A pattern-recognition theory of search in expert problem solving. *Think. Reason.* 3, 291–313
- 45 Baxter, G.P. and Glaser, R. (1998) Investigating the cognitive complexity of science assessments. *Educ. Measure. Issues Practices* 17, 37–45
- 46 Gobet, F. and Wood, D.J. (1999) Expertise models of learning and computer-based tutoring. *Comput. Educ.* 33, 189–207
- 47 Cheng, P.C-H. (1998) A framework for scientific reasoning with law encoding diagrams: analysing protocols to assess its utility. In *Proc. Twentieth Annu. Cognit. Sci. Soc.*, pp. 232–235, Erlbaum
- 48 Lane, P.C.R. *et al.* (2000) CHREST: investigating how humans learn to solve problems using diagrams. *Artif. Intell. Simul. Behav. Quart.* 103, 24–30
- 49 Lane, P.C.R. *et al.* (1999) Learning perceptual schemas to avoid the utility problem. In *Proc. Nineteenth SGES Int. Conf. Knowledge Based Syst. Appl. Artif. Intell.*, pp. 72–82, Springer-Verlag
- 50 Croker, S. *et al.* (2000) Modelling the optional infinitive phenomena: A computational account of tense optionality in children's speech. In *Proc. Third Int. Conf. Cognit. Modelling*, pp. 78–85, Universal Press
- 51 Jones, G. *et al.* (2000) A process model of children's early verb use. In *Proc. Twenty Second Annu. Cognit. Sci. Soc.*, pp. 723–728, Erlbaum
- 52 Theakston, A.L. *et al.* (1999) The role of performance limitations in the acquisition of 'mixed' verb-argument structure at stage 1. In *New Directions in Language Development and Disorders* (Perkins, M. and Howard, S., eds), pp. 119–128, Plenum Press
- 53 Gennari, J.H. *et al.* (1989) Models of incremental concept formation. *Artif. Intell.* 40, 11–61
- 54 Ling, C.H. and Marinov, M. (1994) A symbolic model of the nonconscious acquisition of information. *Cognit. Sci.* 18, 595–621
- 55 Newell, A. (1990) *Unified Theories of Cognition*, Harvard University Press
- 56 Anderson, J.R. (1983) *The Architecture of Cognition*, Harvard University Press
- 57 Elman, J.L. *et al.* (1996) *Rethinking Innateness. A Connectionist Perspective on Development*, MIT Press
- 58 Elman, J.L. (1993) Learning and development in neural networks: the importance of starting small. *Cognition* 48, 71–99
- 59 Kohonen, T. (1995) *Self-organising Maps*, Springer-Verlag
- 60 Carpenter, G.A. *et al.* (1991) ARTMAP: supervised real-time learning and classification of nonstationary data by a self-organising neural network. *Neural Netw.* 4, 565–588

TICS online – making the most of your personal subscription

- High quality printouts (from PDF files)
- Links to other articles, other journals and cited software and databases

All you have to do is:

- Obtain your subscription key from the address label of your print subscription
- Then go to <http://www.trends.com>
- Click on the '**Claim online access**' button at the bottom of the page
- You should see a BioMedNet login screen. If you see this, please enter your BioMedNet username and password*. If you are not already a member please click on the '**Join Now**' button and register. You will then be asked to enter your subscription key.
- Once confirmed you can view the full-text of *TICS*

*If you get an error message please contact Customer Services (info@current-trends.com) stating your subscription key and BioMedNet username and password. Please note that you do not need to re-enter your subscription key every time as BioMedNet 'remembers' your subscription. Institutional online access is available at a premium. If your institute is interested in subscribing to print and online please ask them to contact ct.subs@qss-uk.com