APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **19**: 183–204 (2005)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1110

Chunking Models of Expertise: Implications for Education

FERNAND GOBET*

Brunel University, UK

SUMMARY

Chunking models offer a parsimonious explanation of how people acquire knowledge and have been validated in domains such as expert behaviour and the acquisition of language. In this paper, we review two computational theories based on chunking mechanisms (the chunking theory and the template theory) and show what insight they offer for instruction and training. The suggested implications include the importance of perception in learning, the cost of acquiring knowledge, the significance of segmenting and ordering instruction material, the role of the variability of the instructional material in acquiring schemata, and the importance of taking individual differences into account. Copyright © 2005 John Wiley & Sons, Ltd.

While our understanding of learning and instruction has improved substantially during the last century, there are still a number of key questions that denote islands of ignorance. Among the most pressing of these questions, one may mention: What are the mechanisms of learning? What is the exact role of feedback? To what extent does transfer exist, and how can we facilitate its presence? Is the order in which a curriculum is taught important? How can instruction produce learning with understanding, as opposed to only skill accretion? How can we harness the new resources offered by advances in technology to improve the education we provide to our children?

The first thesis of this article is that the principles and mechanisms derived from research into expertise can inform research in education. By this, we do not mean to imply that training techniques used by top experts should be imported without change into the classroom, but rather that an understanding of how experts reach high levels of performance could suggest general learning mechanisms, and thus help design better teaching methods, both at high and low levels of expertise.

The second thesis concerns the medium that best embodies theories of learning, in particular when they address the acquisition of complex skills. Becoming an expert or assimilating the contents of a curriculum at school (say, in mathematics) engages a number of cognitive processes that will interact with each other, with the teaching environment (teachers and other pupils) and the subject matter to be learned. Given the complexity and the dynamic character of these processes and interactions, their detailed and veridical

*Correspondence to: Dr F. Gobet, School of Social Sciences & Law, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK. E-mail: fernand.gobet@brunel.ac.uk

description requires the use of theories implemented in computer programs. Alternative media are likely to be unsatisfactory as they lose a significant amount of information: verbal theories will not do, as they cannot capture the constraints provided by the learner and its environment, and mathematical theories will require too many simplifying assumptions.

The third thesis argues that theories based on chunking mechanisms, and in particular the chunking and the template theories, not only provide a powerful explanation of expert behaviour, but also give critical guidance about how learning occurs in the classroom. While the educational principles that can be derived from these theories sometimes clash with the principles advocated by other theories in education, they also lead to testable predictions.

This article first highlights some of the most striking phenomena in research into expertise. Then, it outlines two closely linked theories of expertise (the chunking theory and the template theory), and shows how they can account for many of these phenomena. Based on an analysis of the mechanisms underpinning these two theories, we then infer several principles that are of direct relevance to education. Finally, we identify some limits in our approach and suggest directions for further research.

KEY PHENOMENA OF EXPERTISE

Some of the most striking phenomena characterizing expert behaviour were first discovered by De Groot (1978) and Chase and Simon (1973) in their work on chess, and later replicated by researchers in other domains. There are at least two reasons why chess was chosen. Its competitive nature makes it possible to precisely differentiate between skill levels, which is not the case in many domains of expertise studied in psychology, such as medicine and physics. Chess also offers a nice balance between simplicity of the rules and complexity of the space of possible games; this enables both the design of elegant experiments and a fruitful cross-fertilization with mathematics and artificial intelligence.

A first obvious result is that experts, by surpassing normal cognitive and physical limits, can obtain performances that seem extraordinary to laypeople; among other things, experts develop mechanisms that overcome novices' limited learning rates and memory capacities. Experts are often able to recognize the key features of a problem rapidly, using perceptual cues. This perceptual expertise enables them to be highly selective in their search and to solve routine problems without exploring many alternatives. Indeed, in some domains such as fire-fighting or nursing, experts explore only one option (Klein, 1998). This 'professional eye' seems to be made possible by the extensive knowledge that experts have of their domain. The presence and role of such knowledge has been experimentally shown by experts' ability to memorize briefly presented material much better than novices. A further empirical generalization is that it takes a long time of intense dedication to become an expert—at least ten years in many domains. Finally, transfer seems to be minimal from one domain of expertise to another.

An extensive discussion of these results is provided by Gobet, de Voogt, and Retschitzki (2004) for chess and other board games. A broader coverage of expertise is offered by several edited collections (Ericsson, 1996; Ericsson & Smith, 1991; Hoffman, 1992; Zsambok & Klein, 1997), as well as by the articles of this special issue.

THE CHUNKING THEORY

Chase and Simon (1973) proposed their chunking theory in order to explain the phenomena we have just highlighted. Their starting point was a computer program called EPAM (Elementary Perceiver and Memorizer; Feigenbaum & Simon, 1962), which postulates that learning occurs through the incremental growth of a discrimination network giving access to long-term memory (LTM). With this network, external stimuli are sorted to the appropriate chunk[1] through a sequence of *perceptual tests*. Each learning process takes a definite amount of time (e.g. 8 s to create a new chunk). The chunking theory was also influenced by PERCEIVER, a program Simon and Barenfeld (1969) had developed to explain aspects of attention and perception in chess.

To account for the differences between novices and experts in memory tasks, Chase and Simon added assumptions about the capacity of short-term memory (STM), which they proposed was limited to seven items. They hypothesized that chess players perceive a chess position as chunks, which can be defined as groups of pieces forming perceptual and semantic units. Pointers to these chunks are placed in STM, and the recall of the position consists in unpacking the information contained in these chunks. Given that chess masters have acquired more and larger chunks than weaker players, they can memorize briefly presented chess positions better. These ideas were implemented in a computer program called MAPP (Memory-aided Pattern Perceiver; Simon & Gilmartin, 1973), which led to the estimate that it takes about 50,000 LTM chunks to reach chess masters' performance in this recall task.

To account for masters' ability to find good moves without carrying out extensive search, Chase and Simon postulated that chunks are linked to suggestions for plans, moves, and other types of information. Perceptual chunks thus act as conditions to actions, and the chunking theory realizes a simple production system (Newell & Simon, 1972). Chase and Simon also proposed mechanisms explaining how look-ahead search can be carried out in the mind's eye; in particular, they emphasized that recognition mechanisms occur both using information in the external board and information in the imagined positions. Thus, contrary to a common misconception (e.g. Holding, 1985), the chunking theory's recognition mechanisms explain both masters' ability to find moves 'intuitively', that is, almost instantly, in routine positions, and their ability to carry out selective search when necessary.

Chase and Simon's theory was able to explain a number of phenomena in chess expertise using simple mechanisms, and these mechanisms also apply, at least as a first approximation, to other domains of expertise as well. These include expertise in arts, sports, science, and the professions (Richman, Gobet, Staszewski, & Simon, 1996). However, in spite of this popularity, researchers have also identified a number of weaknesses with this theory. These weaknesses relate primarily to its emphasis on STM, its assumption that LTM encoding times are slow even with experts (Charness, 1976; Chase & Ericsson, 1982), its relatively weak emphasis on high-level, schematic knowledge (Freyhoff, Gruber, & Ziegler, 1992; Holding, 1985), and the fact that the computer simulations did not replicate masters' performance. In recent years, the theory has been revised to remedy these shortcomings.

[1]In EPAM, chunks are nodes in the discrimination network.

REVISIONS OF THE ORIGINAL CHUNKING THEORY

The data posing problems to the chunking theory mainly come from memory research. They may be classified under two umbrellas: domains where the explicit goal is to improve one's memory, that is, mnemonics; and domains where memory improvement comes as a side effect of the intrinsic goals of the domain (e.g. winning in games and sports). The two revisions of the chunking theory addressed each of these domains in turn.

EPAM-IV (Richman, Staszewski, & Simon, 1995) shows how the basic chunking mechanisms, supplemented with the idea of a retrieval structure, can account in great detail for the behaviour of a mnemonist in the digit-span task, including the overall shape of performance improvement over 3 years and the way digits are temporally grouped during recall. Retrieval structures refer to structures that enable domain-specific material to be rapidly indexed and encoded in LTM (Chase & Ericsson, 1982). Like earlier versions of EPAM, EPAM-IV is comprised of three main components: an STM (which now consists of auditory and visual subcomponents), an LTM (which now includes a semantic and a procedural component), and a discrimination network, which acts as an index to LTM (see Figure 1). In this model, retrieval structures are schemata in semantic LTM and are considered as deliberately acquired. Richman et al. fully specify all the cognitive mechanisms used by their model—as required to run computer simulations. Each process is associated with a time cost, which makes it possible to derive quantitative predictions from the model about the time course of behaviour. Finally, Richman et al. clearly indicate the aspects of their theory that are considered as fixed, and those aspects that can be modified either by learning or strategies.

The template theory (Gobet & Simon, 1996c, 1998) aims to explain expert behaviour in domains where memory improvement is not the explicit goal. It is implemented as a computer program, called CHREST (Chunk Hierarchy and REtrieval STructures; De Groot & Gobet, 1996; Gobet & Simon, 2000).[2] CHREST, which incorporates mechanisms for perception, learning and memory management, is largely based on EPAM, although it incorporates some important extensions and improvements. Consistent with the earlier

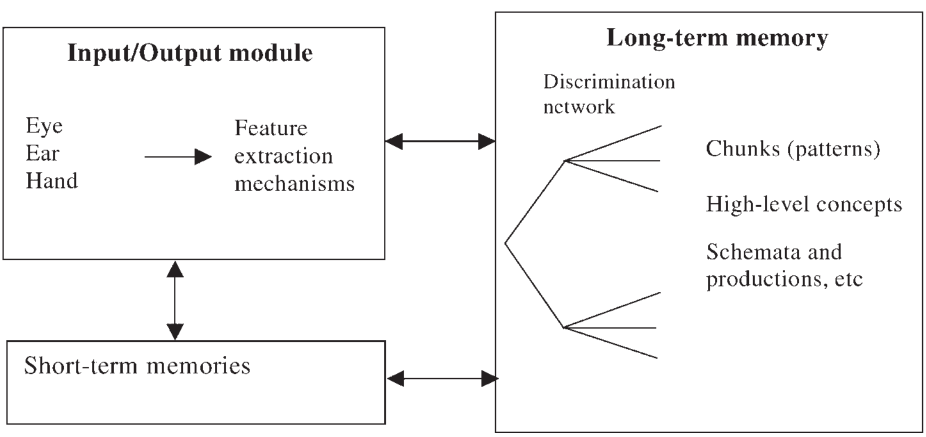

Figure 1. The overall architecture of EPAM and CHREST. The discrimination network can be seen as an index to long-term memory

[2]A CHREST tutorial is now regularly offered in major cognitive-science conferences (Gobet & Lane, 2004).

models, the emphasis is on the limits of cognition, both with respect to capacity (e.g. visual STM can hold only four chunks) and learning rate parameters, which are essentially those used by EPAM).

Like EPAM, CHREST simulates the acquisition of knowledge as the growth of a discrimination network, which develops both as a function of the current state of the system and the input from the environment (we thus have a self-organizing system). However, learning is more sophisticated in CHREST than in EPAM, and the network contains action chunks in addition to perceptual chunks. While the mechanisms leading to the creation of new nodes and the addition of information to existing nodes were already present in the earlier model, two new mechanisms have been added to CHREST (for details, see Gobet, 1996; Gobet et al., 2001). The first mechanism adds a new link between existent nodes. This *lateral* link may either indicate that the two nodes share some similarity, or that the two nodes form a production, with one node acting as the condition of the other (see Figure 2). The second mechanism, which is perhaps the main novelty in CHREST, enables schemata (called *templates* in this context) to be acquired incrementally and automatically. Templates are made of two components: (a) a *core*, which is similar to a chunk in that it encodes fixed information; and (b) *slots*, which can encode variable information and are based upon the information of the nodes linked to a given node. It is assumed that information can be encoded rapidly in the slots (250 ms), which is faster than adding information to a node (2 s) or creating a new node (8 s). The importance of templates is that they show how higher-level structures can be built from chunks and provide mechanisms for the rapid LTM encoding shown by experts. Thus, the concept of templates helps address two of the main criticisms levelled against the chunking theory.

A further improvement is that eye movements are simulated in a more plausible fashion than in the earlier PERCEIVER and MAPP models (see De Groot & Gobet, 1996, for a discussion). In particular, that which is perceived determines what is being learnt, and that which is being learnt determines what will be perceived in the future. Hence, there is a close coupling between perception and memory, which has been extensively documented in several fields of psychology, including the psychology of expert behaviour. Eye movements are also important for problem solving, as shown by Lane, Cheng, and Gobet's (2000) work on multiple diagrammatic representations (see Figure 3).

Finally, chunks are learned incrementally and on-line, which was not the case with MAPP, where chunks were designed by the programmer. This enables CHREST to pick up the statistical distribution of the environment naturally, a feature that turns out to be critical for simulating expert behaviour and the acquisition of language. Thus, by providing a more complete implementation than MAPP, CHREST can explore in much more detail how the constraints of the environment affect the development of expertise.

DOMAINS OF APPLICATION

Together, EPAM-IV and CHREST have addressed most of the weaknesses identified with the chunking theory. The simulations carried out with EPAM-IV on the digit-span task show how a model based on chunking can deliberately create and use retrieval structures that enable a rapid encoding of information. They account for a variety of results, both at an aggregate level (e.g. the overall learning curve over 3 years of practice), and at a more detailed level (e.g. the time needed to mentally scan the retrieval structure from one digit group to an adjacent group).

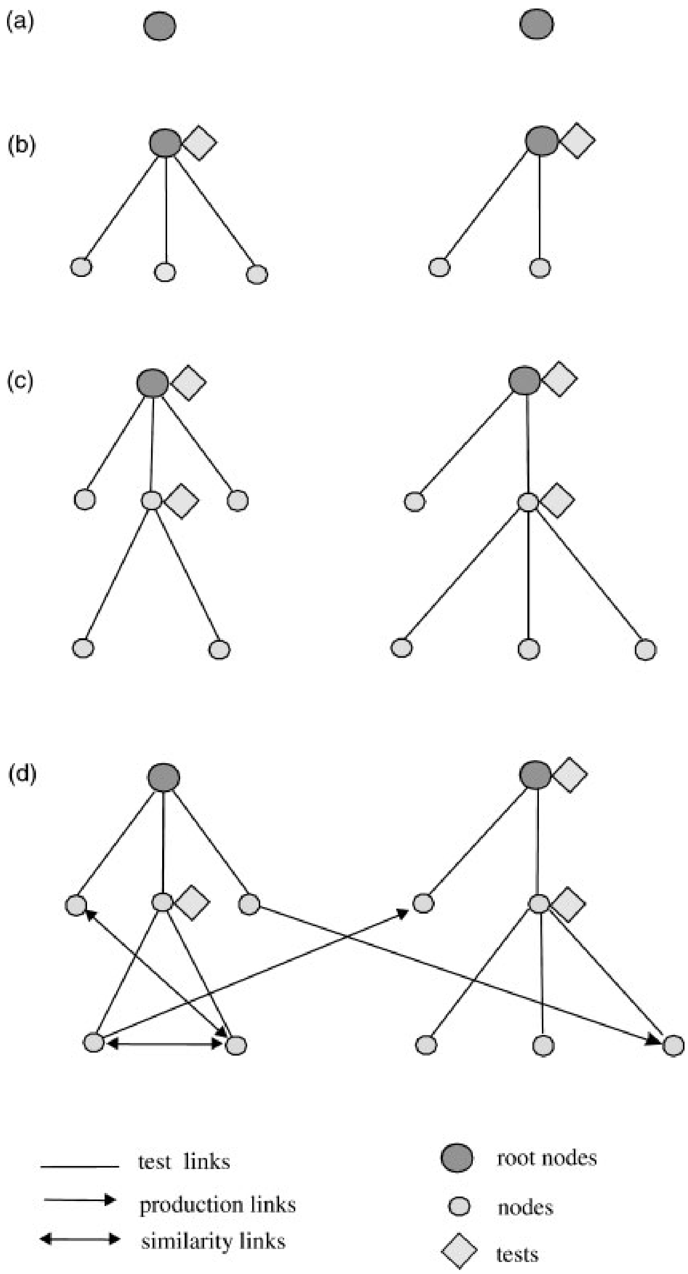

Figure 2. Diagrammatic representation illustrating learning and the types of links used with CHREST. (a) Two discrimination nets (one for perceptual information, on the left, and the other for actions, on the right) start both with a single node, called root-node; (b) input from the environment leads to the creation of *test links*, which are used when discriminating between perceptual stimuli and between actions, respectively; (c) with further input, both nets grow additional tests and branches; and (d) nodes within a net can be connected by *similarity links* and nodes between a perceptual and an action net can be connected by *production links*

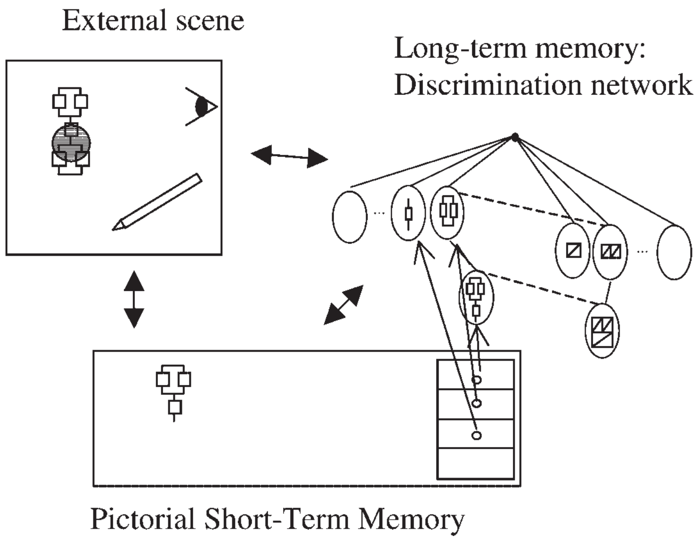

Figure 3. Application of CHREST to learning multiple representations in an electricity curriculum. The system has a simulated eye and pen, for interacting with the external world (top left); the field-of-view of the eye is indicated by the semi-transparent circle. The pictorial short-term memory (bottom) can hold a number of pointers to items of information (bottom right) within its long-term memory, in addition to an iconic representation of that information (bottom left). The long-term memory (top right) uses a network of tests (solid links; for clarity, the individual tests are not shown) to index items of stored information; some of these items are associated with lateral links (dotted links). The bidirectional arrows indicate that information can flow to-and-from each pair of modules in the CHREST architecture

CHREST's first domain of application was chess, where it has simulated data about perception (De Groot & Gobet, 1996), learning (Gobet & Jackson, 2002), and memory (Gobet & Simon, 1996a, 2000; Gobet & Waters, 2003). A variant of CHREST has also been used to explain data about problem solving (Gobet, 1997). Three examples will suffice to illustrate the type of phenomena explained by this model.

When eye movements are recorded in a recall task during the first 5 s of presentation of a chess board, there are striking differences between the eye fixations of experts and novices (De Groot & Gobet, 1996). For example, experts' fixations are rapid (on average 260 ms) with relatively little variability ($sd = 100$ ms), and cover most of the important squares (as defined by the chess semantics of the position). By contrast, novices' fixations are slower (on average 310 ms), show a large variability ($sd = 140$), and miss many important squares. All these results have been replicated by an earlier version of CHREST, as illustrated in Figure 4. Given the importance attached in the CHREST model to eye movements in acquiring knowledge and solving problems, it was critical to show that the model can simulate human data well, as it did.

A criticism often directed towards computational modelling is that simulations 'only' fit the empirical data, but do not make new predictions, as good theories should. A clear illustration that this criticism does not apply to CHREST is offered by the recall of random chess positions. At the time when the simulations were carried out, there was a unanimous belief in the literature that the skill effect with the recall of game positions should disappear with random positions. Yet, the CHREST simulations suggested that masters should perform slightly better, as the large number of chunks they hold in LTM should

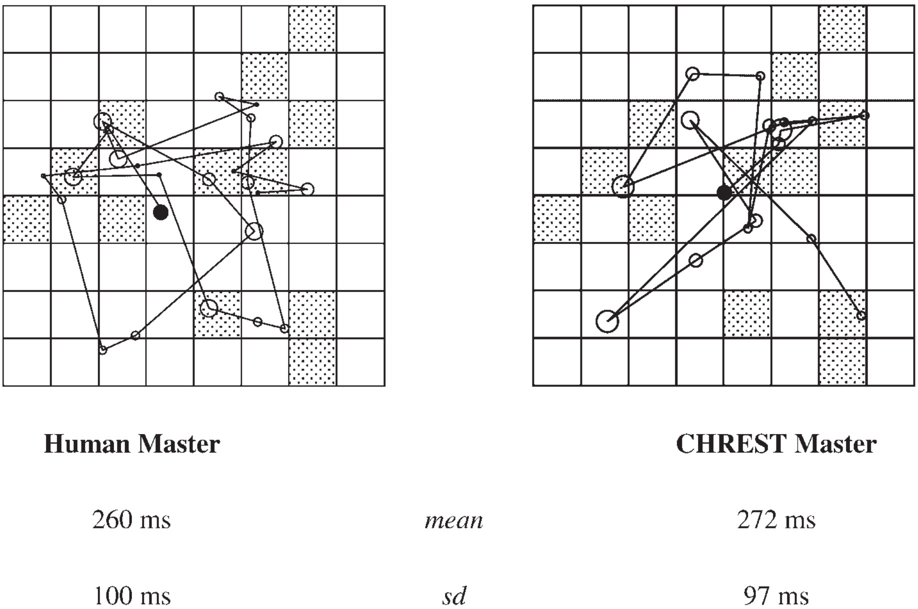

Figure 4. Example of a Master's eye movements for a specific position (left) and its simulations by CHREST (right). The circles are proportional to fixation durations, and the black circle is the first fixation. The shaded squares represent the 12 most important squares in the position. The numbers below indicate the mean and the standard deviation of fixation times across all positions and all subjects. (After De Groot & Gobet, 1996)

enable them to recognize a few patterns by chance, even in random positions. This prediction was supported both by a reanalysis of the available studies and by new experiments (Gobet & Simon, 1996b, 2000). More strikingly, CHREST correctly predicted, not only qualitatively but also quantitatively, that there should be a skill effect with the 'truly random' positions proposed by Vicente and Wang (1998). (See Gobet & Waters, 2003, for details.)

Random positions are also of interest to test the time parameters associated with CHREST's learning mechanisms (see above). These parameters were estimated using data from verbal learning (Simon, 1969), and it was not obvious that they would generalize to visual stimuli. In fact, they do. As recall is poor with random positions (even with masters, although they do better than weaker players), there is plenty of room for improvement through learning. Gobet and Simon (2000) carried out an experiment where the presentation time of game and random positions was systematically varied from 1 s to 60 s. The simulations replicate the human data relatively well with respect to the percentage correct (see Figure 5 for the results with random positions). As discussed by Gobet and Simon, the simulations also account for the size of chunks used and the type and number of errors made.

Beyond chess, CHREST has simulated expertise data about memory for computer programs (Gobet & Oliver, 2002). It has also modelled a variety of results in the acquisition of language, which can be seen as a type of expertise. For example, it gives a good account of phenomena such as children's subject-omission errors, case-marking errors, and use of infinitives (Croker, Pine, & Gobet, 2001; Freudenthal, Pine, & Gobet,

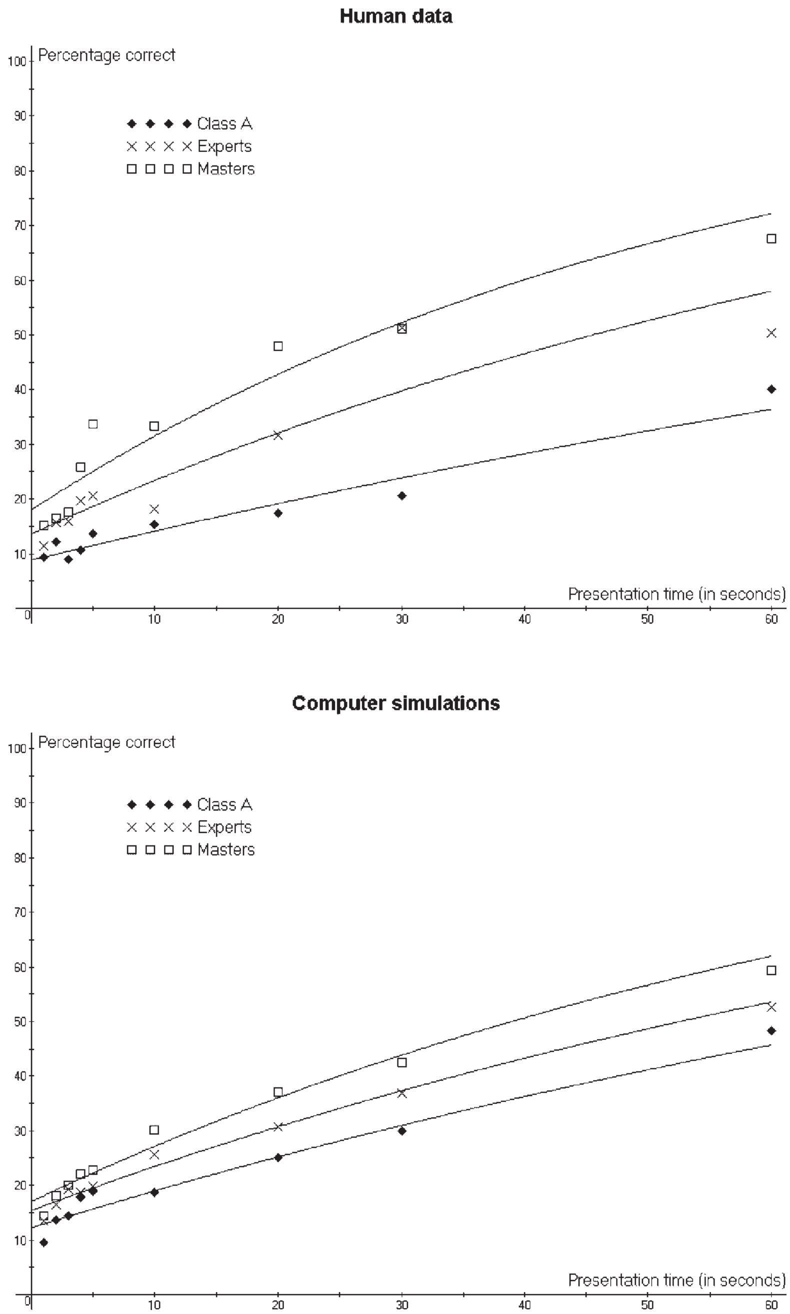

Figure 5. Percentage of pieces correctly replaced in random positions as a function of presentation time and skill level. The top panel shows the human data, and the bottom panel the CHREST simulations. The best fitting exponential growth function is also shown. (After Gobet & Simon, 2000)

2001, 2002). Some of these simulations have been carried out not only in English, but also in Dutch and Spanish.

In several scholastic and scientific domains, students learn how to acquire multiple representations and how to use them (e.g. Ainsworth, 1999). Lane, Cheng, et al. (2000) investigated this question by modelling how novices being taught a curriculum on electric circuits learn to use two diagrammatic representations conjointly—the standard representation typically found in physics textbooks, and a representation where quantitative properties of the domain are encoded in the diagrams (see Figure 3 for an illustration). In this work, visual STM and the simulated eye movements are critical both for learning the chunks that encode stimuli in each of the representations and combining these representations. Perceptual chunks are also essential in learning to perform problem-solving actions to draw the required diagrams. A comparison with human learners solving diagrammatic problems indicated that the match between the chunks drawn by humans and CHREST was satisfactory (Lane, Cheng et al., 2000).

As has been well established by research in learning and instruction, information is often better learned when it is encoded in several modalities and with a fair amount of redundancy. Lane, Sykes, and Gobet (2003) demonstrated how CHREST can encode information both visually and verbally, and how low-level information can be disambiguated by high-level knowledge. Additional models have addressed questions of direct concern for education. Gobet (1999) reports a model of the balance beam task, a classical task in developmental psychology, and draws parallels between the EPAM framework and Piaget's theory of development. Finally, simulations have also been carried out on concept formation (Gobet, Richman, Staszewski, & Simon, 1997) and implicit learning (D. Freudenthal & F. Gobet, in preparation).

IMPLICATIONS FOR EDUCATION

Several articles have discussed the consequences of chunking mechanisms for education. Within the larger framework of adaptive production systems, Herbert Simon has addressed this question either with reference to learning (Langley & Simon, 1981), problem solving (Simon, 1980), methods for learning mathematics from examples and by doing (Zhu & Simon, 1988; Zhu, Lee, Simon, & Zhu, 1996), or as a critique of situated learning and constructivism in education (Anderson, Reder, & Simon, 2000). Gobet and Wood (1999) focused on tutorial systems, comparing the implications of EPAM and CHREST with those of ACT-R (Anderson, Corbett, Koedinger, & Pelletier, 1995) and SuccessMaker (Suppes & Zanotti, 1996). Finally, Gobet and Jansen (in press; see Table 1) derived several practical recommendations for chess training and teaching from the template theory. Here, the aim is to explore what general principles the EPAM and CHREST models (which together we will call *chunk-based models*) offer for education. Most of the time, both theories have similar consequences; however, there are a few instances where recommendations can be derived from CHREST, but not EPAM, for example with respect to learning schemata. These cases will be clearly identified.

The role of practice and the cost of acquiring knowledge

A direct consequence of chunk-based theories is that the time at task is essential to gain knowledge, as each process in the growth of the discrimination network (creating new

Table 1. Some of the educational principles derived from the template theory by Gobet and Jansen (in press) for teaching chess. See the original article for a detailed justification of these recommendations, which are likely to apply to other domains as well

- Teach from the simple to the complex
- Teach from the known to the unknown
- The elements to be learnt should be clearly identified
- Use an 'improving spiral,' where you come back to the same concepts and ideas and add increasingly more complex new information
- Focus on a limited number of types of standard problem situations, and teach the various methods in these positions thoroughly
- Repetition is necessary. Go over the same material several times, using varying points of view and a wide range of examples
- Avoid spending too much time on historical and anecdotal details
- At the beginning, don't encourage students to carry out their own analysis of well-known problem situations, as they do not possess the key concepts yet
- Encourage students to find a balance between rote learning and understanding
- Encourage students to keep information in a central filing system or a database
- The ability to look ahead possible moves is made possible by knowledge
  Training this ability *per se* by exercises aiming at improving short-term memory or visualization (e.g. by playing blindfold chess) is not recommended

nodes, adding information to existing node, and, in CHREST, creating lateral links) is subject to a time cost. Thus, in learning algebra or becoming a violin virtuoso, time must be invested. Consistent with this view, it has been shown that the differences in mathematical ability between Chinese, Japanese, and American children can largely be explained by the amount of time spent studying (Stevenson, Lee, & Stigler, 1986). But, as noted by the proponents of deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993), practice in itself is not sufficient to become an expert—playing the piano for fun will not make one a concert pianist. Practice needs to be tailored to the goal of improving performance. As we have seen in the previous section, CHREST learns what it pays attention to; this explains why deliberate practice, and not just practice, is important. In the latter case, a vast number of irrelevant chunks may be acquired, which not only will not improve performance, but may even hinder it.

The role of perception in acquiring knowledge

The relationship between concrete and abstract knowledge has been the topic of intense research, especially within mathematics education (e.g. Gravemeijer, 1997). A key assumption of chunk-based models is that perceptual skills, anchored in concrete examples, play a central role in the development of expertise, and that conceptual knowledge is later built on such perceptual skills. One of the theoretical reasons behind this claim is that perceptual cues offer an efficient way for LTM knowledge—both declarative and procedural—to be retrieved, and that, without such cues, much LTM knowledge would remain 'inert' (Whitehead, 1929). An additional reason is that CHREST's eye movements provide a close link between perception, learning, and problem solving, as shown by Lane, Cheng, et al.'s (2000) simulations of the acquisition of multiple representations in physics.

The implications for education are clear: means should be found to develop perceptual chunks in a given domain. However, an immediate objection is that emphasizing the acquisition of domain-specific and concrete knowledge will lead to difficulties with transfer. Before addressing this paradox, we need to describe the role that teachers play with respect to the acquisition of perceptual chunks.

The role of teachers and tutorial systems

In order to encourage the acquisition of perceptual chunks, an important role for teachers (both human and artificial) is to direct learners' attention to the key features of the material to learn. One way to do this is to segment the curriculum into natural components (Anderson et al., 2000), with perhaps an optimal ordering of these components (see section on 'Order effects in learning'). Presenting components of the right size and difficulty will help students direct attention to the important features of the material, and in turn help the acquisition of perceptual chunks that are appropriate, given the task at hand. But how, then, identify these components, their adequate level of difficulty and ordering? The answer has three parts (Anderson et al., 2000; Gobet & Wood, 1999). The first part is to carry out a careful task analysis of the material to learn, at the level of component skills; in many domains, there exists extensive literature on which to base such an analysis (e.g. Gagné & Briggs, 1974). The second part is to combine this analysis with the study of how these skills interact with broader tasks and contexts. The final part is to build computer simulations that will verify whether the task analysis in fact leads to curricula that can be learned effectively.

Another important role for teachers is to provide feedback, an obvious way to highlight the important features of a problem, and thus favour the acquisition of correct knowledge. Clearly, this is easier to do with private instruction than in the classroom (Bloom, 1984), although the use of tutorial systems may remedy this situation, at least in procedural domains such as geometry where feedback can be relatively easily automated (Anderson et al., 1995).

The question of transfer

It has been known for more than 100 years (Thorndike & Woodworth, 1901) that there is transfer from one domain to another only when there is an overlap between the components of the skills required in each domain. Indeed, much knowledge acquired in school does not transfer (Travers, 1978), and the same applies to the skills acquired in games and sports. For example, contrary to popular belief, there is little evidence that chess skill transfers to other domains (Gobet & Campitelli, in press). Chunk-based theories clearly indicate that the situation gets worse when high levels of expertise are reached, as the acquired knowledge becomes increasingly specialized. This is because the perceptual chunks, which act as the condition part of productions, becomes more selective. In addition, time spent tuning one specific skill will not be devoted to acquiring other skills.

If this analysis is correct, and keeping in mind that it is difficult to predict what skills will be required two or three decades from now, the best option seems to supplement the teaching of specific knowledge with the teaching of metaheuristics that are transferable (Grotzer & Perkins, 2000; Simon, 1980). These may include strategies about how to learn, how to direct one's attention in novel domains, and how to monitor and regulate one's limited resources, such as small STM capacity and slow learning rates.

Order effects in learning

A question that has received attention in recent years is the impact of the order with which the curriculum is taught (Ritter, Nerb, O'Shea, & Lehtinen, in press). Is there an optimal order, or, on the contrary, is order irrelevant? In chunk-based models, where knowledge is organized hierarchically, the accretion of knowledge is determined by the current state of the system and the incoming information from the environment, with no opportunity for global reorganization of knowledge, as is possible in some learning models (Elio & Scharf, 1990).

Figure 6, based on Gobet and Wood (1999), provides an example illustrating the effect that bad ordering can have on such systems: if irrelevant tests are learned early on, they will affect performance at later stages of learning. Moreover, due to the fact that chunk-based

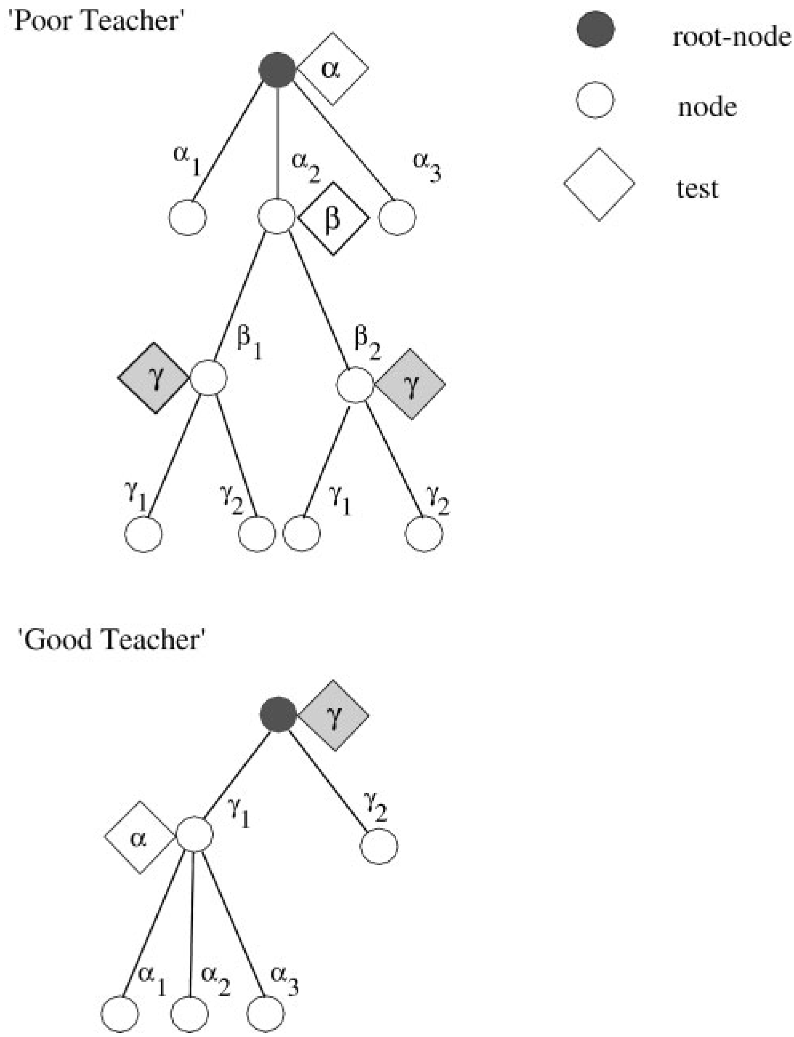

Figure 6. An illustration of the effect of sequencing on curriculum, as predicted by chunk-based models. With a 'poor teacher' (top panel), the learner's attention is not directed to important features of the environment (the γ attribute), which leads to a relatively inefficient network. With a 'good teacher', the discriminative features are learned first, leading to a more efficient network

models construct new chunks using smaller chunks as building blocks, inefficient chunks will recursively propagate to other parts of the network. Considering that pattern recognition happens thousand of times even with simple problems, the effect of suboptimal knowledge becomes evident. A further consequence is that useful connections (including links with potential actions) between pieces of knowledge may be missed due to the inappropriate features encoded in the network.

Gobet and Lane (in press) describe simulations with CHREST showing that changes within the ordering of the learning set have a rather strong impact on the structure of the discrimination network and on the speed of information retrieval. By contrast, differences are not always apparent at the performance level, as measured by the percentage of objects recognized. Interestingly, there are some other cognitive architectures (e.g. ACT-R) which predict that problem sequence should not matter during instruction (Anderson, 1987). While this is clearly a domain where clean new empirical data are necessary, chunk-based models alert us to a potentially serious, but often ignored, element of the way most curricula are designed.

Acquiring productions

Chase and Simon's (1973) idea of using perceptual chunks as conditions to actions is now fully implemented in CHREST (Gobet, 1996; Gobet et al., 2001). As the acquisition of productions is carried out dynamically as a function of the interaction with the environment, CHREST can be classified as an *adaptive production system*, a formalism that has often been used to describe the knowledge that students should ideally acquire (Anderson et al., 1995; Zhu & Simon, 1988). An interesting difference between the production systems used by these authors and CHREST may be mentioned here: while standard systems use a single production for a conceptual condition-action pair, CHREST will create many productions that encode specific perceptual features in the environment. Only then can a more generic and abstract production be acquired, by mechanisms similar to the creation of templates. As a first approximation, what in ACT-R (Anderson & Lebière, 1998) is carried out by the tuning of numerical parameters for a given production corresponds in CHREST to the creation of multiple productions that become increasingly refined.

Regardless of these differences, all production systems alert us to the importance of balancing the acquisition of the *condition* and *action* parts of productions (Zhu et al., 1996). As noted by these authors, this is typically not encouraged in typical textbooks or in traditional classroom instruction, with the unfortunate consequence that students may learn the actions, but not the cues that signal their appropriateness. By contrast, as documented in several mathematics and physics curricula developed in China, students learning from examples and then attempting to solve new problems spend most of their time refining the condition parts of their productions, that is, learning new perceptual chunks (Zhu et al., 1996). One possible reason behind the success shown by some tutoring systems in teaching topics such as algebra or geometry (e.g. Anderson et al., 1995) is that, in addition to providing individual feedback, these systems, by forcing students to solve problems, encourage the creation of more discriminated conditions in the productions they are acquiring.

Acquiring schemata

In spite of the central role of the concept of schema in cognition, there are surprisingly few computational models providing mechanisms explaining how schemata (or templates) are created (Lane, Gobet, & Cheng, 2000). As mentioned earlier, CHREST does provide such mechanisms, where a template is created if a node meets specific criteria relating to its connectivity with other nodes in the discrimination network. Loosely speaking, templates are created in situations where the context presents both constant and variable information. Thus, variability during learning is predicted to foster better long-term performance. In fact, there is substantial empirical evidence supporting this hypothesis (Fazey & Marton, 2002), in domains such as perception (Posner & Keele, 1968), motor skills (Moxley, 1979; Shea & Morgan, 1979), and music (Welker, 1982).

Under the assumption that schemata underpin much of the knowledge acquired in schools—a very common assumption in education research—the way schemata are created in CHREST suggests some important consequences for education: without variation, schemata cannot be created. For example, in the case of elementary mathematics, presenting a narrow range of problems will hamper the acquisition of a sufficient variety of chunks and links connecting them, and, consequently, schemata are not likely to be formed. This outcome is consonant with Fuson's (1992) analysis of the conditions leading to effective teaching of addition and subtraction. Note that the importance of input variability is less apparent in other models of learning and categorization, such as exemplar-based theories (e.g. Estes, 1994) or production systems (e.g. Newell, 1990).

Declarative, procedural, and conceptual knowledge

ACT-R (Anderson et al., 1995) provides a very strong theoretical statement about the respective roles of declarative knowledge (*knowing that*) and procedural knowledge (*knowing how*): knowledge is first acquired declaratively, and only then transformed into a procedural form. According to Gobet and Wood (1999), ACT-R lacks not only mechanisms specifying how declarative knowledge is constructed, but also mechanisms for acquiring knowledge without a declarative stage. In CHREST, there is no strict separation between non-procedural knowledge (which corresponds to chunks and templates) and procedural knowledge (which corresponds to productions), as the learning of both types of knowledge occurs incrementally and implicitly. Verbal knowledge, a subset of non-procedural knowledge, may be instrumental in generating goals and thus directing attention to features of the environment. This characteristic may be used when transmitting verbal instructions to students.

Conceptual knowledge is a tricky term to define in cognitive psychology, as researchers often use it with different meanings. In CHREST, conceptual knowledge is encoded not only in the templates, but also in the web of links (test links, similarity links, procedural links) characterizing the discrimination network. In particular, a node reached by recognition (thus processed by test links) may lead to another node by following lateral links. This organization of knowledge can help us to define 'conceptual understanding'; the possession of some element of knowledge (e.g. a chunk) is not sufficient, but it is also necessary to consider the richness with which this node is indexed, and the density of nodes to which this node is connected (Gobet & Wood, 1999). Thus, to give sufficient basis to conceptual knowledge it is necessary to acquire a richly-connected network of links joining productions and schemata, which are accessible through perceptual chunks.

Acquiring multiple representations

Recently, there has been substantial research on the role of multiple representations in education, although results about their efficacy have been mixed (Ainsworth, 1999). The application of CHREST to modelling the learning of multiple representations in a physics curriculum is instructive in this respect (Lane, Cheng et al., 2000). The simulations suggest that, while multiple representations can support performance in the task, they also take up considerable effort and time from learners. In the worst case, the effort may be spent without any benefit. This is because having efficient representations requires more than the presence of procedural knowledge associated with these representations: it also requires an efficient indexing for accessing them. In fact, learning multiple representations requires duplicating the same information in different formats. Although redundancy is certainly an important aspect of human memory and understanding, CHREST draws our attention to the fact that it also has a cost, in particular with respect to the time spent in learning (Gobet & Wood, 1999).

The use of multiple representations is only one of many learning devices that have flourished with the advent of modern educational technology. It may be appropriate here to comment on the atheoretical way with which much educational research handles the latest technological advances, showing more interest in exploiting them than in understanding the learning mechanisms involved. Examples include the use of LOGO, hypertexts, multimedia, e-learning, and virtual reality in educational settings. Chunk-based models actually warn us against any excess of optimism in the use of new technologies, as long as they do not help circumvent the key limiting constants of human cognition (i.e. attention, STM, and learning rates). In many instances, these new learning environments may present distractions that interfere with what should ideally be learnt. For example, the use of hypertexts turned out to be disappointing educationally (Theng, Jones, & Thimbleby, 1996), a result that should not surprise us as this technology requires students to spend an inordinate amount of time figuring out what link to select next and how to impose some structure on the entangled web of information they are navigating—each imposing a heavy load on STM and attention.

The role of individual differences and talent

There are vast individual differences in people's cognitive abilities (Ackerman, 1987; Sternberg, 2000), both at the novice and expert levels. The thorny question concerns the origin of these differences. Do they reflect previous practice and study, as proponents of deliberate practice (Ericsson et al., 1993) suggest, or do they find their basis in biological and perhaps genetic factors, as champions of the talent approach (Plomin & Petrill, 1997) argue? It is likely that the available data do not allow us to reach a definite answer about this question.[3] For example, in the case of expert behaviour, we have poor measures of the environmental input over years of practice, and even poorer measures of individual differences at the start of this practice. In addition, it is not implausible that the road to expertise sometimes follows a chaotic path, in the mathematical sense that small differences may lead to sizeable qualitative differences years after. A classic illustration of this idea is Simon's (1955) mathematical model showing that the acceptance or non-acceptance of the first paper submitted by a

[3]Based on research into music and chess expertise, proponents of deliberate practice are likely to object to this claim. However, in chess, one of the few domains where expertise levels can be measured quantitatively, deliberate practice accounts only for about half of the variance (Charness, Krampe, & Mayr, 1996), suggesting that other factors are involved. In addition, more recent data on chess practice highlight large individual differences in the amount of deliberate practice that players need to reach master level (G. Campitelli & F. Gobet, submitted; The role of talent, domain-specific practice, and critical period in expert performance).

beginning scientist may to a considerable extent determine their scientific eminence decades later. Given that too many factors are not controlled in current research, such events may remain unnoticed.

Our limited understanding of the complex dynamics underlying expert behaviour does not belittle the fact that individual differences are important in instruction. At the least, it has been shown that feedback tailored to individual students provides better instruction than feedback given in a classroom (Bloom, 1984). What do chunk-based theories have to say about this question?

While emphasizing the role of practice and study, Chase and Simon (1973) were neutral with respect to the origin of individual differences. In fact, this question has rarely been addressed within the chunk-based tradition; for example, most computer simulations were carried out on average data, with fixed time and capacity parameters. However, it is an interesting question as to whether these parameters could reflect individual differences (Charness et al., 1996). This issue was addressed in a practical class taught at the University of Nottingham, where a total of 60 students were given a standard pair-associate task with verbal material. Consistent with the literature on verbal learning (McGeoch, 1946), there were important individual differences in the time required to reach criterion. Students were then requested to simulate their own performance with CHREST (see Gobet & Lane, 2004). The modelling environment enabled them to vary both strategies and time parameters used for creating and elaborating chunks. The outcome of this exercise was that participants' performance could not be explained fully by the strategies used, and that there were important differences in the parameters used by each student to model their own data.

We may mention two important implications for education (see also Mayer, 2000). First, while individual differences tend to be diluted by large amounts of practice, they play a large role in the early stages of studying a domain, which characterizes much of classroom instruction. Second, as seen earlier, taking into account individual differences may lead to better instruction, because instruction can be optimized for each student, including feedback on progress, organization of material, and choice of learning strategies to be taught.

DISCUSSION

This article has first discussed how chunk-based theories can provide a parsimonious account of empirical data in a number of psychological domains, including expert behaviour. It has then attempted to derive some educational principles from these theories. We have focused our attention on EPAM-IV and CHREST, which offer a synthesis combining the concepts of chunking, schema and retrieval structure. These two models highlight the role of perception in learning and the necessity of acquiring a large number of chunks. They incorporate mechanisms interacting with the environment and picking up its statistical structure. They also are able to account for a variety of empirical data, and have made predictions that were later validated by empirical data. These theories can be considered as 'constructivist' in that they provide mechanisms showing how knowledge is constructed and used.

An analysis of the mechanisms used by chunk-based theories enabled us to infer a series of implications for education, which were in some cases contrasted with those made by alternative theories. Among others, we have discussed the difficulty of transferring knowledge from one domain to another, the effect that curriculum sequence may have on learning, the conditions best suited for the acquisition of schemata, and a computational definition for the difficult notion of conceptual understanding. We now address limits of our approach, directions for further research, and the extent to which computational modelling may further help our understanding of learning and instruction.

While we were able to derive several educational principles from chunk-based theories, we also recognize some limits in our approach. Learning has been simulated in several domains within our framework, but so far there has been only one domain where the acquisition of a school curriculum has been modelled (physics of basic electric circuits). We have considered cognitive feedback from teachers but ignored the equally important motivational feedback. Indeed, although factors such as motivation, the social environment offered by peers, and emotions are central in a general theory of learning, our framework has little to say about them, except that they will affect learning through attentional mechanisms and the amount of practice students will spend on the task. Finally, this article can be seen more as a programme of research in need of further empirical validation than as a set of well-validated principles. So far, we have used these principles for suggesting detailed practical prescriptions in only one domain, chess (Gobet & Jansen, in press; see Table 1).

These limits suggest possible directions for future work. Perhaps, besides the need for empirical validation we have just mentioned, the most important addition would be to provide a broader account of motivation. As suggested by Gobet and Wood (1999), another worthwhile alley would be to develop a tutorial system directly based on chunk-based theories, where, contrasting with the stress given by current computer-based tutoring systems on procedural knowledge, the importance of perceptual learning would be emphasized. Such a system could in principle enable the development of curricula *in silico* before they are used in the classroom.

Broadly speaking, computational modelling has been used in two different ways in learning and instruction. Some researchers have chosen to construct computer models that simulate the learners' behaviour. Other investigators have chosen to develop tutoring systems providing instruction to learners. (There has sometimes been an overlap between both approaches, as illustrated by research with ACT-R (Anderson et al., 1995).) A common benefit of both approaches is that they provide means for formulating theories, either of learning or instruction, with a great deal of precision and completeness. Given the complexity of the phenomena to explain and the fact that they evolve as a function of time and of the interaction with the environment, we argue that computational modelling offers the best medium for developing theories of learning and instruction. There is simply no way that verbal, or even mathematical, theories could make predictions about the success of a learner or the impact of a curriculum, as this requires combining factors such as curriculum contents, amount of practice, individual differences, strategies used, social environment in the classroom, teacher's style, to mention but a few. Although we are still far from a general theory of learning and instruction, computer models have been developed that account for substantial subsets of the processes involved.

In this paper, we have attempted to analyse what chunk-based theories tell us about education. One of the advantages offered by these theories is that they provide well-specified learning mechanisms, including mechanisms for acquiring schemata. We have also called for more theory-led research in education, in particular with respect to new technologies. Although further empirical validation is needed, the implications we have derived from chunk-based theories may help improve the way instruction is delivered in our schools.

ACKNOWLEDGEMENT

I am grateful to Guillermo Campitelli, Peter Lane, and two anonymous reviewers for comments on this paper.

REFERENCES

Ackerman, P. L. (1987). Individual-differences in skill learning: an integration of psychometric and information-processing perspectives. *Psychological Bulletin, 102*, 3–27.

Ainsworth, S. E. (1999). A functional taxonomy of multiple representations. *Computers and Education, 33*, 131–152.

Anderson, J. R. (1987). Production systems, learning, and tutoring. In R. Neches (Ed.), *Production systems models of learning and development* (pp. 437–458). Cambridge, MA: MIT Press.

Anderson, J. R., & Lebière, C. (Eds.). (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned. *The Journal of the Learning Sciences, 4*, 167–207.

Anderson, J. R., Reder, L. M., & Simon, H. A. (2000, summer). Applications and misapplications of cognitive psychology to mathematics education. *Texas Education Review.*

Bloom, B. S. (1984). The 2-sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*, 4–16.

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

Charness, N., Krampe, R., & Mayr, U. (1996). The role of practice and coaching in entrepreneurial skill domains: an international comparison of life-span chess skill acquisition. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 51–80). Mahwah, NJ: Lawrence Erlbaum.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. *The Psychology of Learning and Motivation, 16*, 1–58.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Croker, S., Pine, J. M., & Gobet, F. (2001). Modelling children's case-marking errors with MOSAIC. In E. M. Altmann, A. Cleeremans, C. D. Schunn, & W. D. Gray (Eds.), *Proceedings of the 4th International Conference on Cognitive Modeling* (pp. 55–60). Mahwah, NJ: Erlbaum.

De Groot, A. D. (1978). *Thought and choice in chess* (2nd English ed.; first Dutch edition published in 1946). The Hague: Mouton Publishers.

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

Elio, R., & Scharf, P. B. (1990). Modeling novice-to-expert shifts in problem solving strategy and knowledge organization. *Cognitive Science, 14*, 579–640.

Ericsson, K. A. (Ed.). (1996). *The road to excellence*. Mahwah: Erlbaum.

Ericsson, K. A., & Smith, J. (Eds.). (1991). *Studies of expertise: Prospects and limits*. Cambridge: Cambridge University Press.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406.

Estes, W. K. (1994). *Classification and cognition*. Oxford: Oxford University Press.

Fazey, J. A., & Marton, F. (2002). Understanding the space of experiential variation. *Active Learning in Higher Education, 3*, 234–250.

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

Freudenthal, D., Pine, J. M., & Gobet, F. (2001). Modelling the optional infinitive stage in MOSAIC: a generalisation to Dutch. In E. M. Altmann, A. Cleeremans, C. D. Schunn, & W. D. Gray (Eds.), *Proceedings of the 4th International Conference on Cognitive Modeling* (pp. 79–84). Mahwah, NJ: Erlbaum.

Freudenthal, D., Pine, J. M., & Gobet, F. (2002). Subject omission in children's language: the case for performance limitations in learning. In W. D. Gray, & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Meeting of the Cognitive Science Society* (pp. 334–339). Mahwah, NJ: Erlbaum.

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

Fuson, K. C. (1992). Research into whole number addition and subtraction. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 243–275). New York: Macmillan.

Gagné, R. M., & Briggs, L. J. (1974). *Principles of instructional design*. New York, NY: Holt, Rinehart and Winston.

Gobet, F. (1996). Discrimination nets, production systems and semantic networks: elements of a unified framework. In *Proceedings of the 2nd International Conference on the Learning Sciences* (pp. 398–403). Evanston, Il: Northwestern University.

Gobet, F. (1997). A pattern-recognition theory of search in expert problem solving. *Thinking and Reasoning, 3*, 291–313.

Gobet, F. (1999). Simulations of stagewise development with a symbolic architecture. In J. P. Dauwalder, & W. Tschacher (Eds.), *Dynamics, synergetics and autonomous agents* (pp. 143–156). Singapore: World Scientific.

Gobet, F., & Campitelli, G. (In press). Education and chess: a critical review. In T. Redman (Ed.), *Education and chess*.

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

Gobet, F., & Jansen, P. J. (In press). Training in chess: a scientific approach. In T. Redman (Ed.), *Education and chess*.

Gobet, F., & Lane, P. C. R. (2004). CHREST tutorial: simulations of human learning. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (p. 3). Mahwah, NJ: Erlbaum.

Gobet, F., & Lane, P. C. R. (In press). How do order effects arise in a cognitive model? In F. E. Ritter, J. Nerb, T. O'Shea, & E. Lehtinen (Eds.), *In order to learn: How ordering effects in machine learning illuminate human learning and vice versa*. New York, NY: Oxford University Press.

Gobet, F., & Oliver, I. (2002). *A simulation of memory for computer programs* (Technical report No. 74). University of Nottingham (UK): Department of Psychology, ESRC Centre for Research in Development, Instruction and Training.

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

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. (1996c). Templates in chess memory: a mechanism for recalling several boards. *Cognitive Psychology, 31*, 1–40.

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.

Gobet, F., & Waters, A. J. (2003). The role of constraints in expert memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 29*, 1082–1094.

Gobet, F., & Wood, D. J. (1999). Expertise models of learning and computer-based tutoring. *Computers and Education, 33*, 189–207.

Gobet, F., Richman, H. B., Staszewski, J. J., & Simon, H. A. (1997). Goals, representations, and strategies in a concept attainment task: the EPAM model. *The Psychology of Learning and Motivation, 37*, 265–290.

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

Gobet, F., de Voogt, A. J., & Retschitzki, J. (2004). *Moves in mind: The psychology of board games*. Hove, UK: Psychology Press.
Gravemeijer, K. (1997). Mediating between concrete and abstract. In T. Nunes, & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective*. Hove, UK: Psychology Press.
Grotzer, T. A., & Perkins, D. N. (2000). Teaching intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 492–515). Cambridge, MA: Cambridge University Press.
Hoffman, R. R. (1992). *The psychology of expertise: Cognitive research and empirical AI*. New York: Springer-Verlag.
Holding, D. H. (1985). *The psychology of chess skill*. Hillsdale, NJ: Erlbaum.
Klein, G. A. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.
Lane, P. C. R., Cheng, P. C. H., & Gobet, F. (2000). CHREST+: investigating how humans learn to solve problems using diagrams. *AISB Quarterly, 103*, 24–30.
Lane, P. C. R., Gobet, F., & Cheng, P. C. H. (2000). Learning-based constraints on schemata. In *Proceedings of the Twenty Second Annual Meeting of the Cognitive Science Society* (pp. 776–781). Philadelphia, USA: Erlbaum.
Lane, P. C. R., Sykes, A. K., & Gobet, F. (2003). Combining low-level perception with expectations in CHREST. In F. Schmalhofer, R. M. Young, & G. Katz (Eds.), *Proceedings of EuroCogSci 03: The European Cognitive Science Conference 2003* (pp. 205–210). Mahwah, NJ: Erlbaum.
Langley, P., & Simon, H. A. (1981). The central role of learning in cognition. In J. R. Anderson (Ed.), *Cognitive skill and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
Mayer, R. E. (2000). Intelligence and education. In R. J. Sternberg (Ed.), *Handbook of intelligence*. Cambridge: Cambridge University Press.
McGeoch, J. A. (1946). *Psychology of human learning: An introduction*. New York, NY: Longmans.
Moxley, S. E. (1979). Schema—Variability of practice hypothesis. *Journal of Motor Behavior, 11*, 65–70.
Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
Plomin, R., & Petrill, S. A. (1997). Genetics and intelligence: what is new? *Intelligence, 24*, 53–78.
Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353–363.
Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory with EPAM IV. *Psychological Review, 102*, 305–330.
Richman, H. B., Gobet, F., Staszewski, J. J., & Simon, H. A. (1996). Perceptual and memory processes in the acquisition of expert performance: the EPAM model. In K. A. Ericsson (Ed.), *The road to excellence*. Mahwah, NJ: Erlbaum.
Ritter, F. E., Nerb, J., O'Shea, T., & Lehtinen, E. (Eds.). (In press). *In order to learn: How ordering effects in machine learning illuminates human learning and vice versa*. New York, NY: Oxford University Press.
Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology-Human Learning and Memory, 5*, 179–187.
Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika, 52*, 425–440.
Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.
Simon, H. A. (1980). Problem solving and education. In F. Reif (Ed.), *Problem solving and education* (pp. 81–96). Hillsdale, NJ: Erlbaum.
Simon, H. A., & Barenfeld, M. (1969). Information processing analysis of perceptual processes in problem solving. *Psychological Review, 7*, 473–483.
Simon, H. A., & Gilmartin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 5*, 29–46.
Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence*. Cambridge: Cambridge University Press.
Stevenson, H. W., Lee, S., & Stigler, J. W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science, 233*, 696–699.
Suppes, P., & Zanotti, M. (1996). *Foundations of probability with applications*. Cambridge: Cambridge University Press.
Theng, Y. L., Jones, M., & Thimbleby, H. (1996). Lost in hyperspace—Psychological problem or bad design? In *Proceedings of the first Asia Pacific Conference on Computer Human Interaction* (pp. 387–396). Singapore: Information Technology Institute.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 9*, 374–382.
Travers, R. M. W. (1978). *An introduction to educational research*. New York, NY: MacMillan.
Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review, 105*, 33–57.
Welker, R. L. (1982). Abstraction of themes from melodic variations. *Journal of Experimental Psychology-Human Perception and Performance, 8*, 435–447.
Whitehead, A. N. (1929). *The aims of education*. New York, NY: MacMillan.
Zhu, D., & Simon, H. A. (1988). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*, 137–166.
Zhu, X., Lee, Y., Simon, H. A., & Zhu, D. (1996). Cue recognition and cue elaboration in learning from examples. *Proceedings of the National Academy of Sciences, 93*, 1346–1351.
Zsambok, C. E., & Klein, G. A. (Eds.). (1997). *Naturalistic decision making*. Mahwah, NJ: Erlbaum.

