PERGAMON

Computers & Education 33 (1999) 189–207

COMPUTERS & EDUCATION

www.elsevier.com/locate/compedu

Expertise, models of learning and computer-based tutoring

F. Gobet*, D. Wood

ESRC Centre for Research in Development, Instruction and Training (CREDIT), School of Psychology, University of Nottingham, Nottingham NG7 2RD, UK

Abstract

In a wide and diverse range of contexts, from academic disciplines through to games and sports, analyses of what it takes to be judged an expert have established a number of common claims. In this paper, we identify and discuss the theoretical significance of this research in relation to a formal, computational theory of expertise (EPAM). The main thrust of our paper is the argument that the theory both helps to identify and explain theoretical limitations on some influential approaches to computer-based tutoring, and offers a means of overcoming some of these. We argue that, without 'knowledge-based' models of the learning process, attempts to develop effective, computer-based tutoring systems have achieved limited progress towards the goal of helping learners to construct links between their procedural knowledge and conceptual understanding. Current knowledge-based approaches to learner modelling need to be developed in two main directions to reach this goal. First, they will have to integrate a theoretically sound account of the relation between perception and memory (such as that developed within the EPAM approach) in order to build upon what has already been achieved to date in relating processes of learning, memory and problem solving. Second they need an extended theory of declarative (or conceptual) knowledge and its relation to procedural skills. We illustrate how the EPAM model of expertise can be exploited towards these ends, and draw out a number of implications for the design and current limitations of computer-based tutoring systems. © 2000 Elsevier Science Ltd. All rights reserved.

1. The theory of expertise

Research on expertise, which has often focused on top performers in a field, has uncovered a

* Corresponding author. Fax: +115-951-5324.
E-mail address: frg@nott.psyc.ac.uk (F. Gobet).

PII: S0360-1315(99)00032-9

number of general principles and cognitive mechanisms governing the process of becoming an expert (Chase & Simon, 1973; Chi, Feltovitch & Glaser, 1981; Ericsson, 1996). A first principle is that the achievement of expertise requires the acquisition of a large knowledge base, which can be analysed and represented in terms of theoretical constructs such as chunks, rules and schemata. A second holds that it takes a long time and experience to become an expert — Simon's '10-year rule' is most usually taken as the best estimate. A third principle, more contentious, is that experts do not differ from non-experts in basic information-processing powers, nor in their native endowments; differing only in the amount of 'deliberate practice' they have spent on the domain (Ericsson, Krampe & Tesch-Römer, 1993). Finally, there is considerable evidence that experts' skills only minimally transfer from their own domain of expertise into other areas of knowledge (e.g. Chase & Ericsson, 1982): becoming an expert in one field confers little by way of cognitive benefits on performance in others.

Given the amount of commitment, time and experience needed to achieve expert status, coupled with the lack of generalisation of expertise across domains, and the limited time given over to schooling, it follows that any aim of helping all learners to become experts in all or even most areas of knowledge addressed in the school curriculum is unrealistically ambitious. However, the thesis explored here is that an understanding of the processes involved in *becoming* an expert have important educational implications for the design and evaluation of computer-based learning environments. We start to explore this claim with a brief critique of a contemporary, neo-behaviourist approach to the design of computer-based tutoring systems, and then move to a knowledge-based tutoring system. Finally, strengths and limitations of these two approaches are discussed in light of the EPAM theory, a theory of expertise stressing the role of perception.

2. Modelling the learning process

Perhaps the most theoretically motivated and widely applied neo-behaviourist approach to the design of computer-based tutoring systems is the integrated learning system entitled SUCCESSMAKER, which derives from a mathematical approach to modelling the process of mastery learning formulated by Suppes and his colleagues (Suppes & Zanotti, 1996). This system is a focus for classroom-based evaluation studies overviewed in the paper by Wood, Underwood and Avis in this volume. The process of learning in Suppes' approach is modelled by estimating three key parameters: (a) an empirically derived, normative estimate of the difficulty of each element or 'strand' of the curriculum to be taught; (b) a summary estimate of the initial level of each learner's current level of performance (e.g. school grade level equivalent); and (c) an individualised measure of (current) rate of learning which is derived from samples of each learner's recent history of task performance. These parameters are exploited in a mathematical, Bayesian model to assess the rate at which learning is currently proceeding. Provided that the current estimate of learning rate falls within predetermined limits, then the learner is moved on through the curriculum. Where learning falls below the acceptable limits, then the tutor should revisit, and offer further tutoring, in prerequisites skills, in an attempt to enable the learner, subsequently, to make further progress.

Neither the analysis of curriculum 'strands' nor the definition of learning prerequisites within the tutoring domain are derived from the model of learning. Both are based on a priori analyses of the knowledge domain. The learning model itself thus lacks any explicit 'semantics' or model of the knowledge domain. Further, it makes no appeal to cognitive processes such as perception, reasoning or misconception, and involves no attempt to theorise about relations between performance and knowledge. This is why we classify this species of approach as neo-behaviourist, and as a model of mastery learning, rather than a knowledge-based approach.

3. Contingent correction and explanation

Even within the seemingly well structured domain of mathematics, there is ample evidence that the relation between task performance on the one hand (e.g. the ability to solve a problem, or the tendency to produce a particular type of error), and conceptual understanding on the other (i.e. the knowledge which underpins the correct solution or the error), is neither direct nor simple. For example, extensive research into children's learning of small number addition and subtraction problems has revealed many classes of situations where what appears to be the same 'sum' involved in different kinds of word problem demands different conceptual understanding for solution. For instance, where one problem involves the addition or subtraction of elements from a single set, whilst a second involves the comparison or the equalisation of two distinct sets, the levels of difficulty involved, and the stage at which children come to master the two problems, can vary significantly, even where, objectively, the same numbers and a common arithmetical procedure is implicated (see Fuson, 1992, for a thorough and analytical review of this area).

It follows from this that children often master one class of problems several years before another, even in cases where the same procedural knowledge would seem to be called for. As a result, success in the application of a given procedure, or being taught how to use that procedure, in the context of one class of problems will not ensure the recognition of the procedure's relevance nor its application when the construction of different 'mental models' in response to distinct types of problem is called for. Consequently, interpreting both successful answers and errors from the learner is problematic unless one has a psychologically valid theory of the relations between problem type, a given learner's conceptual knowledge and their procedural skills. Thus, the provision of contingent and useful error feedback, and the formulation of an explanation which the learner is likely to understand, demands more from a tutor than a check to see if the learner's action agrees with a particular answer to a problem.

This example is offered as a specific instance to illustrate the general hypothesis that where there are one-to-many mappings between procedural performance and conceptual understanding the theoretical interpretation and explanation for either errors or successful actions is not simple. The complexity of the task of mapping performance measures onto a robust model of conceptual understanding explains, we suggest, why research into the impact of computer-based systems which have tried to help learners to correct their 'buggy procedures' (Sleeman, Kelley, Martinak, Ward & Moore, 1989) and to model learner misconceptions (Anderson, 1987) have proved difficult, and of limited success. If such systems are not capable of interpreting success and error in a robust way, they cannot be expected to provide feedback, or give explanations, or select task experiences, which are likely to be contingent upon the knowledge and mental state of the learner.

In principle, Suppes' approach to mathematical modelling of learning is extensible to any curriculum analysis which meets certain criteria (e.g. analysis into strands and the availability of norms to assess strand difficulty). For example, a curriculum model which was sensitive to the kinds of distinctions between different problem types found in word problems explored above could be constructed, norms established and the mathematical approach applied. In this way, it might be possible to develop learning systems which could enhance some aspects of conceptual understanding.

Although the behavioural approach is potentially extensible in theory, we suggest that it will prove more theoretically satisfying and practically useful in the long run to integrate an explicit and generative account of how a learner's performance relates to their knowledge.

4. Knowledge-based approaches to modelling learning and expertise

Computer-based tutoring systems which have grown out of contemporary cognitive theory are based on explicit models of knowledge and some important aspects of its acquisition. Since Anderson's ACT family of cognitive theories have underpinned the most extensive and researched applications of theory to computer-based tutoring (e.g. Anderson, 1987, 1993; Anderson, Corbett, Koedinger & Pelletier, 1995), we take this as the paradigm case throughout the rest of the paper.

Anderson's theory, like Suppes', does not provide a basis for curriculum design, but involves 'constructing a curriculum under the guidance of a domain expert' (Anderson et al., 1995, p. 167). However, the theory does exert strong constraints on the manner in which the curriculum is analysed and represented. Central to ACT theory is the distinction between declarative and procedural knowledge, and a characterisation of the process of learning as the proceduralisation of declarative knowledge. Declarative knowledge (or 'knowledge that', such as knowing the side–angle–side theorem) comprises facts retrieved from memory, or taken from perception (e.g. as verbal or written utterances, mathematical expressions, programming instructions, elements of diagrams, etc.). Procedural knowledge ('knowing how', such as the ability to use a theorem in a proof) refers to actions taken on the basis of declarative knowledge in the service of goals. Learning involves finding out, through problem solving, what procedures any declarative knowledge calls for when in pursuit of goals. A major epistemological claim of ACT theory is the contention that human knowledge can be decomposed and represented in terms of such 'rules of the mind' (Anderson, 1993).

An analysis of curricula content expressed as a set of such rules provides the model of target knowledge to be taught and learned. By interacting with the learner, setting problems and monitoring performance, it is possible for a tutoring system to construct a model of an individual learner's knowledge of the domain, also expressed as a set of production rules. A comparison of the rule set in the learner model with that defining the target knowledge identifies the extent to which a given learner is or is not likely to have mastered curriculum knowledge. Gaps between the two provide a source of teaching goals for the tutoring system. The selection of problems for instruction is also constrained by the ACT theory which

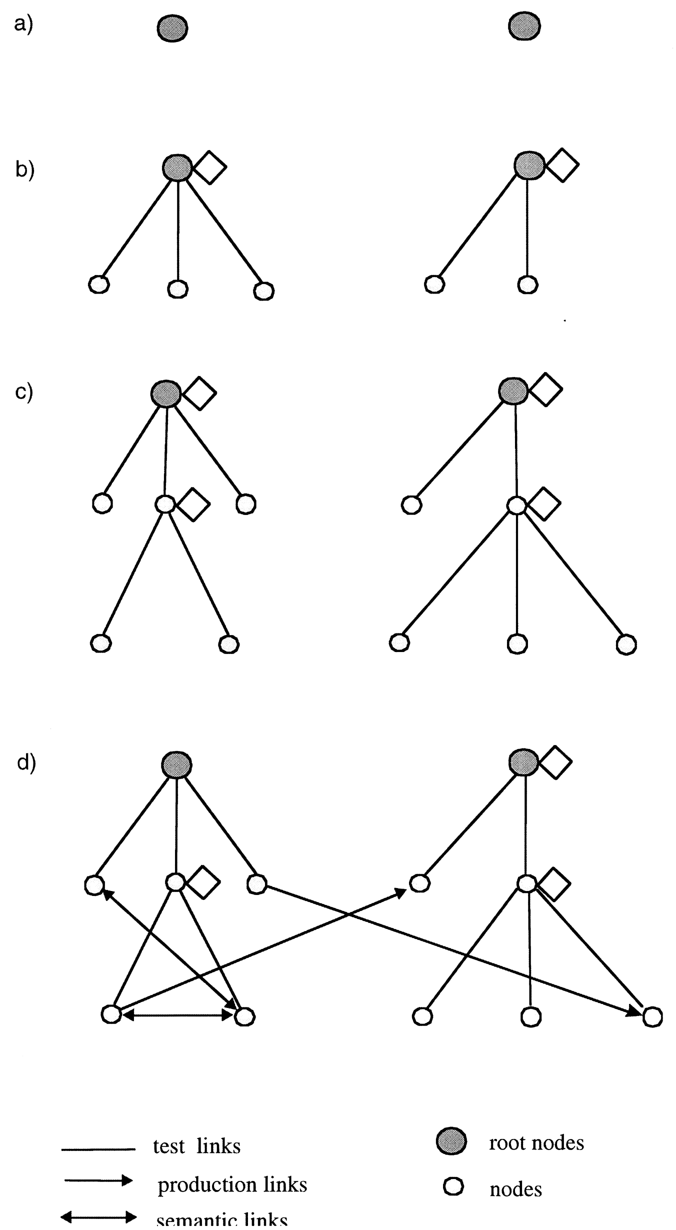

Fig. 1. Diagrammatic representation of learning in EPAM, and of how two discrimination nets can be combined. (a) The two discrimination nets start both with a single node, called root-node; (b) Input from the environment leads to the creation of a test; depending of the outcome of the test, a different branch will be taken; (c) with further input, both nets grow additional tests and branches; and (d) In the extended EPAM framework, nodes within a net can be connected by *semantic links* and nodes between a perceptual and an motor net can be connected by *production links*.

identifies specific limits acting on the cognitive capacities of the learner such that the tutor will "only introduce new rules when the old rules have achieved a sufficient degree of strength that they no longer interfere with the acquisition of new rules" (Anderson, 1987, p. 456). Learner performance in response to each new problem is monitored and the learner model is continually updated on the basis of that performance to identify rules that have been mastered; and so on.

Anderson argues that this conceptualisation of learning provides an explanation for, and a means of circumventing, what Whitehead (1929) refers to as 'inert knowledge'. Thus, while a learner may be able to memorise, recall and recognise factual knowledge, if they do not also have the necessary procedural skills to act on such knowledge in pursuit of relevant goals, then what the learner knows remains inert. Such inertia can be overcome through problem solving; and effective tutoring can support the problem-solving process by adherence to a number of tutoring principles; principles which also converge on the analysis of tutoring derived from research into scaffolding and contingency (see Wood & Wood, 1996, for more details of this analysis). Quasi-experimental evaluations of tutoring systems constructed on the basis of ACT have produced substantial evidence for significant gains over conventional group teaching in domains such as learning to programme and areas of mathematics instruction (Anderson et al., 1995).

Anderson's theory has succeeded in achieving a theoretical synthesis and extension of research into learning processes and memory. In the next section, we argue that to develop more complete theories of conceptual knowledge (and to design tutoring systems which can better support conceptual learning), we will need to further extend the knowledge-based approach to accommodate theoretical insights into work on knowledge and perception. The main source of these insights is research into computational models of expertise.

5. Expertise, the chunking theory and EPAM

Chase and Simon's (1973) chunking theory, one of the most influential theories of expertise, proposes that the acquisition of expertise is made possible by the construction of a large discrimination net[1] of relatively small perceptual chunks, which encode key features of the domain. These chunks act as conditions on productions (see Fig. 1). Since the expert has built up around 10,000 of such chunks, the chances that they will be able to assimilate, recognise and act upon what they perceive in their domain of knowledge is high; explaining both why experts are quick to zoom in to the key aspects of the problem at hand and how, as de Groot (1978) put it, they virtually 'see' possible solutions.

Working within the framework of adaptive production systems and learning by doing and from examples, Zhu, Lee, Simon and Zhu (1996) have found that learning the condition side of productions takes longer than learning the action side. They proposed that students learn

[1] A discrimination net consists of a set of nodes (chunks) connected by branches, which together form a tree-like structure. The nodes possess tests, which can be applied to check features of the external stimuli. The outcome of each test determines which branch will be taken below a node. Once a node has been sorted to, information can be added to it in order to represent the external stimulus in more detail.

the relevant conditions in simple situations first and that, as they move to more complex problems, they acquire productions that include goals and subgoals into their conditions, thus allowing them to control search. This indicates the importance of cues in conditions and the necessity of constructing chunks of growing complexity.

The formal models which have been implemented to test chunking theory are based on the assumption that other aspects of experts' cognitive abilities, such as the capacity of their short-term memory, and the rate with which they can learn and memorise new chunks and productions, are shared with non-experts. The success of these models is thus one measure of the plausibility of the theoretical claim that expertise results from differences in experience rather than talent. Another significant feature of the theory, which we develop later, is the assumption that once nodes have been created in the discrimination net they remain there and are never forgotten. Thus, any failure to make use of past experience results not from a process of forgetting but from the processes by which the network grows with experience. These processes may render certain nodes in the network inaccessible, such that the fruits of experience, though still retained, cannot be retrieved.

The most developed aspect of Simon's chunking theory lies in the workings out and testing of the perceptual learning processes put forward to explain how the acquisition of new chunks takes place (e.g. Simon & Gilmartin, 1973). The mechanisms underpinning these processes have also been tested in computational, EPAM models of perception and memory designed to explain findings from a wide range of empirical investigations into phenomena in verbal learning, memory and concept formation (see Feigenbaum & Simon, 1984; Richman, Staszewski & Simon, 1995; or Gobet, Richman, Staszewski & Simon, 1997, for overviews). Consequently, the proposed processes of perceptual learning and chunking have been tested and evaluated against a wide range of empirical data. As with chunking theory, these EPAM models are based on the assumption that chunks are encoded in a discrimination network. When an external 'object', or perceptual feature, elicits a node, learning can occur in two ways. If the information in the node that is accessed under-represents the object (i.e. when some aspect(s) of what is perceived about the object are novel), then new features are added to that node. The internal representation of the object is thus elaborated and enriched. If the information in the node that is accessed contains features that are not consistent with the external object (e.g. where a perceived object fails to exhibit attributes which define known objects), then a new node to represent this object is created in the network (below the node first accessed). Thus, as the EPAM model is exposed to an increasing amount of information from the environment (i.e. becomes experienced), its discrimination net grows dynamically in direct response to both the content and sequential structure of that environmental information: EPAM is thus a self-organising system.

6. Expanding EPAM: The CHREST framework

Recently, the chunking theory and EPAM have been revised in two main ways (Gobet & Simon, 1996, in press; Richman et al., 1995) in response to evidence showing that the proposed model of expert memory was not sufficiently rich and robust. The changes were mainly motivated by memory research showing that experts can store information in LTM faster than was proposed by the chunking theory. For example, experiments with the digit-span task indicate that, following intensive training, individuals can encode up to about 100 digits rapidly. Two main extensions to the theory are relevant to our current concerns. First, mechanisms have been defined and implemented whereby the 'small' chunks of the original theory can now evolve into larger and more complex data structures; schemas with slots and default values (these schemas are called *templates*). Template slots are created automatically when some information in the tests below a chunk (recall that a chunk is simply a node in the discrimination net) recur often but with some variation in content. For example, in the case of chess, tests may check the type of piece (the variable) placed on the same square (the constant) on the chess board. Once a slot has been created, it is assumed that storing information in it is a fast process. Second, experts can acquire *retrieval structures* deliberately, which facilitate later encoding in the same way mnemonics such as the method of loci facilitates the memorisation of a shopping list (see also Chase & Ericsson, 1982).

After a template has been evoked by recognition, or after a retrieval structure has been instantiated, the slots in either can be filled in rapidly. This follows because the knowledge structure (template) is already in place, and only values in slots have to be added. One consequence of this extension is that encoding into long-term memory can be done faster than in the original chunking theory, and this leads to a better fit between model and data. For example, the revised mechanisms, implemented as computer programs, account for an extensive and detailed body of empirical data from research into expertise in domains such as the digit span task (Richman et al., 1995) and chess (Gobet & Simon, 1996, in press; Gobet, 1998).

However, even these extensions to the original chunking theory fail to account for the richness of experts' knowledge, and ways to further enrich the framework have been proposed by Gobet (1996). The key idea is that, in addition to creating and connecting chunks,

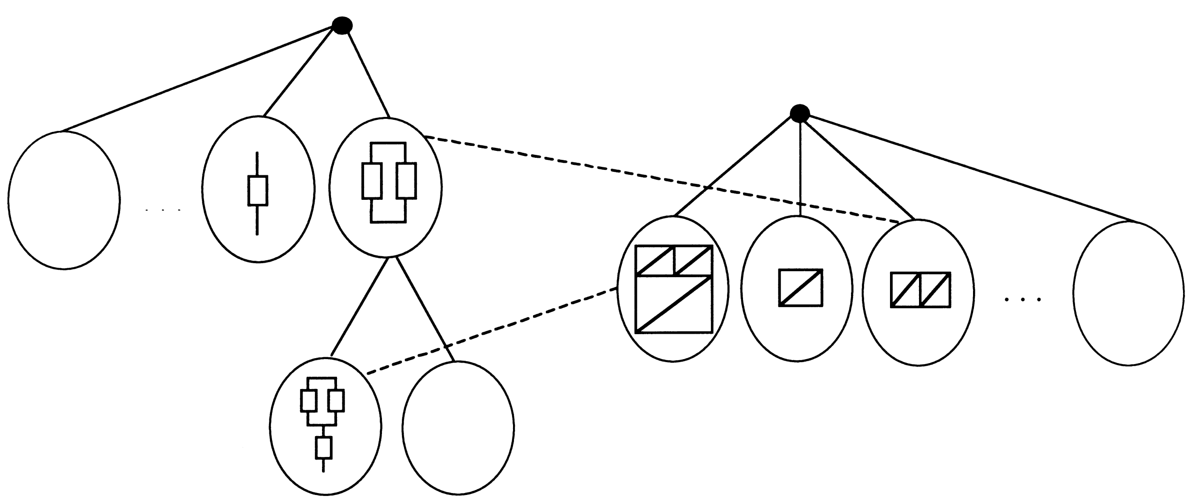

Fig. 2. This figure illustrate an additional way with which networks can be combined (equivalence links). In learning about electrical circuits and AVOW diagrams, EPAM combines representations for standard circuit diagrams (on the left) with AVOW diagrams (on the right).

productions and templates, nodes in the discrimination network can be connected by similarity or by relational (semantic) links, thus providing mechanisms for implementing the growth of associative memory in tandem with perceptual learning (see Fig. 1).

This extended theory — called CHREST, for Chunk Hierarchy and REtrieval STructure — is being tested in our centre in applications to domains such as chess expertise (Gobet & Simon, 1996; in press), young children's acquisition of both syntax (Gobet & Pine, 1997) and vocabulary (Jones, Gobet & Pine, 1999), the balance scale task (Gobet, 1999) and learning with Law Encoding Diagrams (Lane, Cheng & Gobet, 1999). In all cases, computer implementations exist, and the programs carry out learning.

Two examples may illustrate our approach. In the case of language acquisition, where the input to the model is based on empirical studies of parental speech to individual children, the system acts as a distributional analyser of the statistical properties of the environment, picking up patterns that recur often. In the LED case, the training input consists in the type of instructions, diagrams and solved examples given to (human) students. The net growth describes the acquisition and the connection of two different representations in physics, the first coding electric circuits using the standard notation in physics textbooks, the second coding the type of Law Encoding Diagrams developed by Cheng (this volume). These two nets are in turn connected, using production links, to various drawing and problem-solving procedures (see Fig. 2).

With respect to expertise and instruction, the aim of these models is to test rigorously several theoretical claims. First, that linked, multiple representations are necessary to explain the progression from the reliance upon superficial features of the problem to the use of deep principles, a progression often noted in the transition from novice to expert (Chi et al., 1981). Second, that this progression naturally leads to a forward type of search, where experts start from the givens of a problem instead of using a goal-directed mode, a phenomenon that has been observed in several domains of expertise, such as in physics (Simon & Simon, 1978). Within our approach, this means that they have ways to translate superficial, perceptual features of a problem into deep-level features, either through productions or through conceptual knowledge, and then use these features as conditions of productions. This is both a question of having more knowledge (principles of physics richly interconnected both together and with perceptual cues) and a better organised knowledge (links between representations).

In these applications, learning occurs through the application of simple mechanisms which use the information from the environment in a goal-directed manner. While the idea of associative memory is an old idea going back to Aristotle, we believe that these computer implementations, which are continuously growing procedural and declarative knowledge from a dynamically evolving discrimination net, take the idea and extend it in ways that have shed fresh light on the learning process and revealed new explanatory power in such associative mechanisms. In related research, Landauer and Dumais (1997) have demonstrated how the use of powerful computer resources proves that associationist mechanisms can go a long way in explaining semantic aspects of human memory and cognition.

Although the extended EPAM framework does not yet offer the computational intelligence and computing facilities of systems such as Soar (Newell, 1990) and ACT (Anderson, 1993), we will claim that the framework, which emphasises the role of perceptual memory more than these two systems, complements them in the aim of developing a good candidate for a formal model of expertise and of children's acquisition of expertise. It is promising for the study of expertise (Gobet, 1998) because (a) it offers a combination of perception, learning, problem solving and attention; (b) it keeps the strengths of the chunking theory, which accounts for many data in the literature, while removing its weakness with the addition of the concept of template; and (c) its computer implementation is already able to simulate a large number of phenomena in expertise research.

The framework is also promising for the study of children's learning, because (a) again, it offers a combination of perception, learning, problem solving and attention; (b) it embodies the view that learning in school can be seen as a type of acquisition of expertise, a view that has gained support in recent years (Wood, 1998a); (c) empirical data on children's learning are consistent with chunking theory (e.g. Chi, 1978; Opwis, Gold, Gruber & Schneider, 1990). The implications of this framework for the design of tutoring systems is the focus of the next sections.

7. Theories of learning, the design of learning environments and theoretical limits on computer-based tutoring

The earlier papers in this special edition have a shared concern with trying to formulate principles for the design of computer-based learning environments and the development of methodologies for their empirical evaluation. We are also concerned with the nature and evaluation of such principles, but evaluation here rests on conceptual and theoretical analysis rather than empirical evidence. The main aims are to explore theoretical limits on what we can achieve in the task of learner modelling in the context of computer-based tutoring. Since the nature of the learner model defines limits on the pedagogical goals of such systems, this exercise also provides a strategy for trying to identify what we might and might not expect to achieve with such tutors in supporting learning.

8. Procedural skills and conceptual knowledge

Given the arguments we have put forward about the need for a knowledge-based approach to the design of learning environments, it follows that epistemological questions about how we should conceptualise the *nature* of knowledge, and its external and internal representation, are crucial to system design and evaluation. Here, we focus on an epistemological question which divides contemporary cognitive theories of knowledge; the distinction between declarative and procedural knowledge. We then try to demonstrate how the positions taken on this epistemological question lead to significant implications for issues of curriculum development and tutoring.

The distinction between declarative and procedural knowledge has long aroused controversy in educational theory (Hiebart and Carpenter, 1992, p. 77ff). The position taken by the ACT theory on this question is explicitly and succinctly put forward by Anderson et al. (1995, p. 169; our emphasis in italics):

"The ACT theory distinguishes between declarative knowledge...and procedural knowledge...*The assumption of the theory is that goal-independent declarative knowledge initially enters the system in a form that can be encoded more or less directly from observation and instruction*. Cognitive skill depends on converting this knowledge into production rules...which represent procedural knowledge...*the theory assumes that production rules can only be learned by employing declarative knowledge in the context of problem-solving activity*."

The ACT theory is thus clear that knowledge is first acquired declaratively, and then translated into a procedural form, as we outlined above. What is lacking in ACT, we believe, is (a) provision for knowledge acquisition without the declarative stage and (b) a more explicit account of the nature of the mechanisms allowing declarative knowledge to be constructed and structured.

Evidence from research on implicit learning (e.g. Reber, 1993) indicates the possibility of the acquisition of procedural knowledge without a declarative stage. It follows from this that, through such 'implicit' learning, *knowing how* can predate *knowledge that*. The controversy surrounding this issue is confused and confusing, however. Anderson rejects some critics of the ACT position on the grounds that evidence which shows that people can know how to do things without being able to verbalise that knowledge, taken by some as evidence against the view that the declarative predates the procedural, is irrelevant (Anderson & Lebière, 1998, p. 109ff). Conscious awareness of declarative information (such as, for example, that portrayed in a diagram or a mathematical expression) in the course of acquiring productions does not, of necessity, entail the ability to verbalise that declarative knowledge later. However, irrespective of where one stands on this issue, the fact is that the nature of declarative knowledge (e.g. what the learner attends to and perceives as they look at a diagram) is at least under-specified in the current versions of ACT theory.

In our extended EPAM framework, perceptual chunking, the creation of productions and of schemas occurs implicitly. Declarative knowledge may serve to generate goals (i.e. such as learning the procedural implications of declarative knowledge in context) and may direct attention to features of the environment, as it does in ACT, but within EPAM, attention may be directed in other ways as well, as by internal goals. This being the case, unless a computer-based tutor is able to infer these goals, the probability that any attempt they make to scaffold learning will prove contingent on learner knowledge is minimal.

Because, in ACT, both declarative and procedural information are encoded by the modeller, the theory is silent about how declarative knowledge such as schemata (as opposed to rules) is or should be learnt. Our framework, which also emphasises the acquisition of rules, makes it clear that nodes in the discrimination net should also be connected by semantic links. In this framework, a node reached by discrimination may lead to another node by following semantic links. Thus, not only the possession of some element of declarative knowledge, or node, matters, but also (a) the richness with which this node is indexed, and (b) the density of nodes to which this node is connected. These two aspects give some computational meaning to 'conceptual understanding' as a richly connected network of links connecting productions and schemas, that is accessible through perceptual chunks. This definition dovetails well with Baxter and Glaser's (in press) characterisation of experts' knowledge, in which "a well-connected knowledge structure links concepts and processes with conditions under which those concepts and processes should be used''.

9. Multiple representations

As Cheng and Ainsworth argue in this special issue, multiple representations play an important role in the acquisition of knowledge, although the learning contexts under which this holds true are unclear. In ACT, multiple representations are encoded implicitly in rules, and the theory offers no way of exploring and testing hypothesis about the nature of any relations between multiple representations, declarative/procedural knowledge and performance. Within the CHREST model of Law Encoding Diagrams outlined above, multiple representations may sometimes prove useful in supporting the achievement of enhanced, domain-relevant knowledge structures when, at other times, they simply take up learner effort without benefiting subsequent performance.

Having multiple representations has the advantage of increasing the probability that at least one will prove to be computationally efficient (cf. Larkin & Simon, 1987). However, finding the right representation requires two things: (a) the presence of this representation as well as the procedures associated to it; and (b) an efficient indexing of this representation, either through perceptual cues or through semantic links. Learning two types of representation without smooth indexing mechanisms imposes a heavier load on learning without the advantage of being able to use them. Learning efficient links between two types of representation without tuning the procedural utility of these representations is also inefficient — a sort of inert knowledge, to use Whitehead's phrase.

In our framework, learning multiple representations implies both the acquisition of several subnets, partially duplicating information in different format, and the need to create links connecting the nodes from these subnets. While this redundancy may be a key feature of conceptual understanding, and while there is strong evidence that experts do indeed encode information redundantly (e.g. Richman et al., 1995), there is no doubt that the amount of information to acquire increases substantially in comparison to a single type of representation. Thus, learning multiple representations imposes an heavier learning toll on the students, may require a 'critical mass' to be usable, and may therefore turn out to be useful only in the long term.

10. Theories of knowledge and the design of task sequences

For ACT tutors, ''nothing about problem sequence is special, except that it is important to minimise the number of new productions to be learned simultaneously. We have done only a little exploration of problem sequence... but so far have gotten null results, consistent with the ACT theory'' (Anderson, 1987, p. 457). What matters on this account is sufficient exposure to productions for learning to take place, not the order with which the learning sequence is structured. The prediction that sequential order in learning is of such minor significance follows directly from the 'flat', modular way with which productions are encoded in ACT. This view of sequential effects on learning stands in direct contrast to what follows from the EPAM theory, where the potential for hierarchical structures in perceptual and declarative knowledge arise as a natural consequence of the way in which their learning mechanisms work.

To put it boldly, EPAM implies that it is possible to acquire knowledge in a 'wrong' order (relative to a specific set of task demands) and that this has marked, negative effects on both the extent to which learning supports future performance and further learning (see Gobet & Lane, in press, for computational support of this claim). For EPAM, the impact of initial learning on the ease of subsequent learning is considerable. Exposure to a domain-

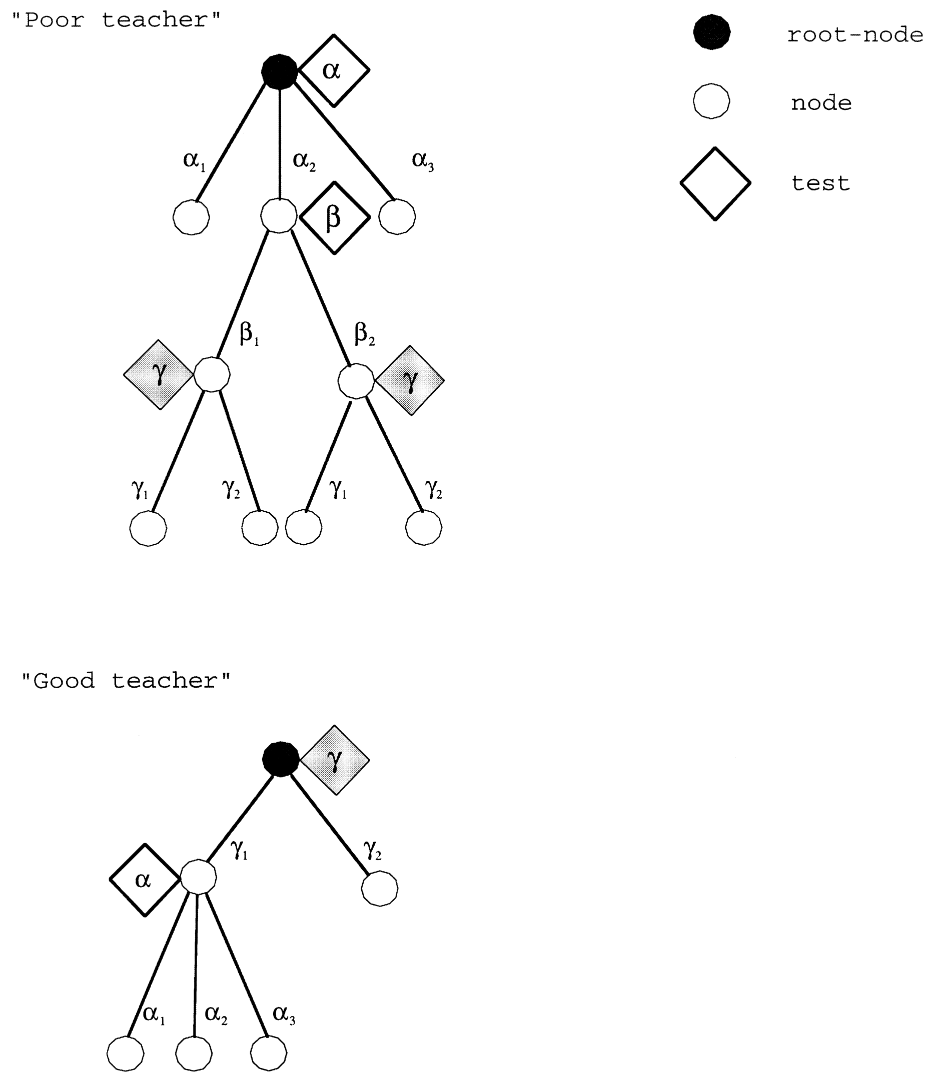

Fig. 3. Effect of sequencing of curriculum. In the case of a 'poor teacher', the learner's attention is not directed to important features of the environment (the γ attribute in the figure), which leads to a relatively inefficient network. In the case of a 'good teacher', the discriminative features are learned first, leading to a more efficient network.

inappropriate learning sequence will lead to the construction of a poorly structured and inefficient discrimination net, hence to inefficient indexing of knowledge, and a consequent high chance of failure to retrieve that knowledge when the demands of the domain require it. In addition, this suboptimal knowledge will propagate through the net, because the chunking mechanisms in EPAM construct knowledge recursively from smaller chunks. Thus, the negative effects of poor task sequencing early on in learning will be cumulative. The recognition of task relevant knowledge will be slower in a poorly structured than in a well structured net. Since the basic EPAM recognition processes recur thousands of times during, say, the attempt to solve an algebra problem, the summative impact of such micro-level differences in the ability to access memory (and, hence, to benefit from past experience) is considerable. Indeed, such differences in powers of recognition lie at the heart of differences between expert and novice performance.

A second consequence of an inappropriate learning sequence is that the non-optimal organisation of the net may lead to certain important connections between nodes (both through production links and similarity links) being missed, and with them efficient solutions to the problem at hand. Repairing the consequences of a suboptimal curriculum order will be costly. First, a complete, more efficient subnet has to be grown, not only a few nodes changed, as is the case in learning with an efficiently structured net. Any further learning is also unlikely to eradicate the effects of suboptimal knowledge entirely, since nodes are never lost and the nodes encoding such 'poor' knowledge may distributed over the whole EPAM net. A third consequence is that erroneous information may be stored in the net, which can lead to a propagation of errors during performance.

Fig. 3 illustrates one possible effect of the sequencing of the curriculum. Let us assume that feature γ is highly diagnostic in the domain, although this may not be obvious for the non-expert. In the upper part of the figure, the network has been trained without being giving information about the value of γ. As can be seen, less useful tests may be learned (in one case, both α and β) before γ is learnt. This has the consequence that (1) the γ-test has to be learnt several times — a waste of learning time; and (2) unimportant tests have to be carried out, with the risk of error increasing.

Issues of task sequencing also emerge in relation to the design of curriculum breadth. In the simulations we are carrying out in chess, language acquisition and acquisition of multiple representations in physics, the formation of templates and of similarity links is a function of the variety of test links below a node. Without enough variety, neither templates nor similarity links can be created. If the theory is correct, this would suggest that learning in such a way that only a narrow range of problems is experienced in the early stages can impede the creation of well connected and integrated knowledge and inhibit future learning. Critical evaluations of approaches to mathematics instruction suggest that the envisaged, negative impact of an overly narrow curriculum on future learning is a real one, at least in mathematics. "If, from the beginning...children were exposed to a rich range of addition and subtraction situations, if understanding and solving these situations in several different ways were emphasized, and if materials to support alternative solutions were provided and discussed, children might be able to understand considerably more than they do in the present, impoverished, narrow classroom environment" (Fuson, 1992, p. 262).

11. Self-explanation and zone of proximal development

Self-explanation has recently been heralded as a key phenomenon in learning (Chi, DeLeeuw, Chiu & LaVancher, 1994). In our approach, self-explanation affects learning in that it encourages the learner to pay attention to important aspects of the domain and to allow enough time for learning them. It also helps because self-explaining forces one to process a topic deeply, allowing the creation of similarity links, which, in turn may later on be used as retrieval cues (e.g. Ericsson & Kintsch, 1995). Finally, self-explanation allows conflicts and failures of comprehension to emerge and, maybe, to be resolved (Renkl, 1997). These conflicts may direct attention to important feature of the domain. But note that self-explanation can be detrimental as well: self-explaining incorrectly will lead to the learning of irrelevant or misleading knowledge, which may interfere with the acquisition of correct knowledge. Thus, the teaching environment should optimise, perhaps through feedback, the moment when students generate self-explanations. (Knowing when to generate explanations and when to simply learn from the textbook may also be a skill that students have to acquire, although our framework has little to say about this question.)

Related to self-explanation is the question of zone of proximal development (or, to use van Geert's (1983) phrase, ''the zone of proximal attention''). In our framework, such zones may be defined as the moment in the learning process where additional knowledge fits 'naturally' into the current knowledge state. By 'naturally', we mean that adding a new piece of knowledge (i.e. a set of nodes and links connecting them) will lead to an elaboration of the net that will be useful in the acquisition of further knowledge and its use. An 'unnatural' addition of knowledge amounts to knowledge that will hamper further learning and possibly impair performance. Again, what human and artificial tutors have to anticipate is when to provide additional information.

12. Using cognitive models to pre-evaluate curricula

The future of cognitive models in education will be in the evaluation and refinement of educational ideas and methods in artificio, with a high probability of being correct, before testing these ideas and methods in vivo, with the expensive methodology that the latter implies. For example, curricula on teaching Law Encoding Diagrams can be tested first using our computational model, improved, and tested again. This cycle can be carried out several times, much more rapidly than when human learners are used. Such an approach, where the goal is to improve a product and test its sensitivity to parameter changes, is common in engineering sciences, but is still rare in education (Suppes & Zanotti, 1996) where hypothesis-testing approaches still dominate research (Grant, 1962).

13. Some implications for principles of tutoring

The implications of this theory for the design of effective tutoring systems are profound. The constraints and demands that will have to be satisfied in order to develop such systems on the basis of the EPAM theory are far more severe than those which constrain tutors based on ACT. An EPAM tutor would have to possess the capacity to monitor and model both the presence or absence of productions (which the ACT tutors do) and the perceptual cues that elicit productions (which ACT tutors cannot do). It would also need to model the way in which, for a given learner, concepts are connected to these perceptual cues and to other concepts. Thus, such a tutor should be able to keep track not only of a student model of procedural knowledge, but also of a student model of conceptual and perceptual knowledge. Although techniques have been developed to solicit and model the structure of conceptual understanding in the field of 'knowledge acquisition' (cf. Shadbolt & Burton, 1990), which could, in principle, be used to assess conceptual understanding in a tutoring context, the task of integrating and exploiting these techniques in the context of computer-based tutors has not been attempted, to the best of our knowledge.

A key assumption in EPAM and in our extended EPAM framework is that the acquisition of knowledge (the augmentation of the discrimination net) depends on the way attention is directed to the environment. We speculate that the role of coaches and of teachers is to direct attention to the right features. Why does attention matter in the creation of the discrimination net? It matters because, being a self-organising system, EPAM develops both as a function of its current state and of the input of the environment. Hence further learning will build on previous learning. If this previous learning is inefficient, perhaps because it rests on attention to irrelevant features of the environment, later performance will be hampered, and relocating learning into a 'correct' trajectory will require a massive and costly restructuring of knowledge. Feedback during learning will help the student to focus upon the relevant features of the environment. Here, our framework makes a rather stronger prediction than ACT that not only productions should be acquired, but also that the conditions and actions of these productions should be organised in an 'efficient' way.

The task of modelling and supporting learning in the EPAM/CHREST framework (or in any other computer tutor designed to promote declarative/conceptual learning) will require a far more powerful technology than we currently possess. Perhaps Anderson is making a wise strategic choice in assuming that all procedural knowledge has to originate from declarative knowledge, as this offers him a way to control and monitor when the tutoring environment transmits new knowledge to the learner. However, if the EPAM view is correct, the ability of such models to support the process of implicit learning and the acquisition of perceptual and conceptual knowledge is, in principle, severely limited. And if this is true, then any pedagogical support for such processes which need to be provided will lie outside the competence of currently foreseeable generations of computer-based tutors.

14. Conclusion

Our attempt to articulate what we have achieved, and may hope to achieve, with computer-based tutors provides both a context for evaluating the scope of the theories about learning which underpin system design and a means of etching out their potential role in the wider educational enterprise. Many competencies shown in scaffolding by human tutors — such as monitoring what the learner is likely to be paying attention to, drawing attention to features that have been overlooked, or whose significance the learner does not appreciate — have yet to be sufficiently well understood, formalised or implemented in computer-based systems. We have explored, albeit briefly, some of the theoretical reasons why we think that such scaffolding functions play an important role in tutoring and learning. To the extent that these reasons are sound, it follows that computer-based tutoring systems are currently incapable of supporting a set of important tutoring functions and the learning which they serve. Similarly, our ability to assess and model conceptual knowledge and, hence, to provide help, advice and task experience which are contingent upon that knowledge, is still extremely limited. Here, too, where pedagogical support is necessary to support learning, then computer-based systems will be of little help until there have been considerable theoretical advances in our ability to formalise such aspects of knowledge and technical advances which enable such formalisms to be built into usable and effective computer-based learning environments.

We have also considered the implications of competing theories for issues of task sequencing which are central to the search for principles of curriculum design. In addition to identifying fruitful areas where we might seek to make empirical evaluations of the theories, this exercise also serves to underline the general case that the adoption and use of such systems involves taking on certain epistemological commitments. Conceptual evaluation provides one means of articulating, and opening up to debate, issues about the value, roles and place of the technology in education.

Although we have been preoccupied with trying to find the limits on what we might hope to achieve with computer-based tutors, it does not follow that we have done so in the belief that such systems have no educational contribution to make or role to play. In terms of guiding learners in practices designed to support the acquisition and extension of procedural skills, for example, systems such as those based on ACT theory have already demonstrated their practical value. Key educational issues, as Anderson and his colleagues acknowledge, now revolve around practices we need to develop in order to embed the use of such systems successfully within the classroom. "Teachers seem to require some time in the classroom before they appreciate the 'tutor as teaching assistant' model and can use it to its maximum potential" (Anderson et al., 1995, p. 201). However, we would also add that one way in which we can seek to realise this potential is by trying to articulate and to understand the limits on what such systems can currently hope to achieve. By trying to "identify any gaps between what the system has to offer and what is needed to support learning generally we can start to outline the possible roles that teachers may have to play in integrating the use of systems alongside their other practices" (Wood, 1998b p. 36).

References

Anderson, J. R. (1987). Production systems, learning, and tutoring. In D. Klahr, P. Langley, & R. Neches, *Production systems models of learning and development* (pp. 437–458). Cambridge, Mass: MIT Press.

Anderson, J. R. (1993). *Rules of the mind*. Erlbaum: Hillsdale, NJ.

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., & Lebière, C. (1998). *The atomic components of thought*. Mahwah, NJ: LEA.

Baxter G.P., Glaser R. (in press). Investigating the cognitive complexity of science assessments. In *Educational measurements: Issues and practice*.
Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower, *The psychology of learning and motivation*, *Vol. 16*. New York: Academic Press.
Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase, *Visual information processing*. New York: Academic Press.
Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler, *Children's thinking: What develops?* Hillsdale, N.J: Erlbaum.
Chi, M. T. H., DeLeeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477.
Chi, M. T. H., Feltovitch, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.
de Groot, A. D. (1978). *Thought and choice in chess*. The Hague: Mouton Publishers (original in Dutch, 1946).
Ericsson, K. A. (1996). *The road to excellence*. Mahwah, NJ: Erlbaum.
Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.
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.
Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science*, *8*, 305–336.
Fuson, K. C. (1992). Research into whole number addition and subraction. In D. A. Grouws, *Handbook of research on mathematics teaching and learning* (pp. 243–275). New York: Macmillan.
Gobet, F. (1999). Simulations of stagewise development with a symbolic architecture. In J. P. Dauwalder, & W. Tschacher, *Dynamics, synergetics, and autonomous agents* (pp. 143–156). Singapore: World Scientific.
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. (1998). Expert memory: A comparison of four theories. *Cognition*, *66*, 115–152.
Gobet F., Lane P.C.R. (in press). Order of learning and EPAM. To appear in T. O'Shea, E. Lehtinen, F. Ritter, P. Langley (Eds.), *Order effects in learning*.
Gobet, F., & Pine, J. (1997). Modelling the acquisition of syntactic categories. In *Proceedings of the 19th Annual Meeting of the Cognitive Science Society* (pp. 265–270). Hillsdale, NJ: Erlbaum.
Gobet, F., Richman, H., Staszewski, 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., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*, 1–40.
Gobet F., Simon H.A. (in press). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*.
Grant, D. A. (1962). Testing the null hypothesis and the strategy and tactic of investigating theoretical models. *Psychological Review*, *69*, 54–61.
Hiebart, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws, *Handbook of research on mathematics teaching and learning* (pp. 65–100). New York: Macmillan.
Jones G., Gobet F., Pine J.M. (1999). *Modelling vocabulary acquisition: An explanation of the link between the phonological loop and long-term memory*, Tech. Rep. No. 61. University of Nottingham (UK): Department of Psychology, ESRC Centre for Research in Development, Instruction and Training .
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.
Lane P.C.R., Cheng P.C.-H., Gobet F. (1999). *Problem solving with diagrams: Modelling the implicit learning of perceptual information*, Tech. Rep. No. 59. University of Nottingham (UK): Department of Psychology, ESRC Centre for Research in Development, Instruction and Training .
Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousands words. *Cognitive Science*, *11*, 65–99.
Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
Opwis, K., Gold, A., Gruber, H., & Schneider, W. (1990). Zum Einfluss von Expertise auf Gedächtnisleistungen

sowie deren Selbsteinschätzung bein Kindern und Erwachsenen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *22*, 207–224.

Reber, A. S. (1993). *Implicit learning and tacit knowledge*. New York, NY: Oxford University Press.

Renkl, A. (1997). Learning from worked out examples: A study of individual differences. *Cognitive Science*, *21*(1), 1–29.

Richman, H. B., Staszewski, J., & Simon, H. A. (1995). Simulation of expert memory with EPAM IV. *Psychological Review*, *102*, 305–330.

Shadbolt, N., & Burton, M. (1990). Knowledge elicitation. In E. N. Wilson, & E. N. Cortlett, *Evaluation of human work*. London: Taylor & Francis.

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

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. S. Siegler, *Children's thinking: What develops?* Hillsdale, N.J: Erlbaum.

Sleeman, D., Kelley, A. E., Martinak, R., Ward, R. D., & Moore, J. L. (1989). Studies of diagnosis and remediation with high school algebra students. *Cognitive Science*, *13*, 551–568.

Suppes, P., & Zanotti, M. (1996). *Foundations of probability with applications. Selected papers, 1974–1995*. Cambridge: Cambridge University Press.

van Geert, P. (1983). *The development of perception, cognition and language: A theoretical approach*. London: Routledge and Kegan Paul.

Whitehead, A. N. (1929). *The aims of education*. New York: Macmillan.

Wood, D. (1998a). *How children think and learn* (2nd ed.). Oxford: Blackwell Publishers.

Wood D. (1998b). *The UK ILS Evaluations: Final Report*. DfEE/BECTa: Coventry.

Wood, D., & Wood, H. (1996). Vygotsky, tutoring and learning. *Oxford Review of Education*, *22*(1), 5–16.

Zhu X., Lee Y., Simon H.A., Zhu D. (1996). Cue recognition and cue elaboration in learning from examples. In *Proceedings of the National Academy of Sciences 93*, (pp. 1346–1351).