ELSEVIER
Cognitive Science 26 (2002) 85–112
COGNITIVE SCIENCE
http://www.elsevier.com/locate/cogsci

Spanning seven orders of magnitude: a challenge for cognitive modeling

John R. Anderson*

Psychology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received 1 May 2001; received in revised form 25 September 2001; accepted 10 November 2001

Abstract

Much of cognitive psychology focuses on effects measured in tens of milliseconds while significant educational outcomes take tens of hours to achieve. The task of bridging this gap is analyzed in terms of Newell's (1990) bands of cognition—the Biological, Cognitive, Rational, and Social Bands. The 10 millisecond effects reside in his Biological Band while the significant learning outcomes reside in his Social Band. The paper assesses three theses: The Decomposition Thesis claims that learning occurring at the Social Band can be reduced to learning occurring at lower bands. The Relevance Thesis claims that instructional outcomes at the Social Band can be improved by paying attention to cognition at the lower bands. The Modeling Thesis claims that cognitive modeling provides a basis for bridging between events on the small scale and desired outcomes on the large scale. The unit-task level, at the boundary of the Cognitive and Rational Bands, is useful for assessing these theses. There is good evidence for all three theses in efforts that bridge from the unit-task level to educational applications. While there is evidence for the Decomposition Thesis all the way down to the 10 millisecond level, more work needs to be done to establish the Relevance Thesis and particularly the Modeling Thesis at the lower levels. © 2002 Cognitive Science Society, Inc. All rights reserved.

Keywords: Cognitive modeling; Cognitive architectures; Education; Intelligent tutoring

1. Introduction

The struggle of Psychology has always been to say things of significance to the human experience that have a rigorous scientific foundation. The enormity of human experience and our strong preconceptions about that experience have made this a very difficult task.

* Tel.: +1-412-268-2781; fax: +1-412-268-2844.
E-mail address: ja+@cmu.edu (J.R. Anderson).

PII: S0364-0213(01)00062-3

Nonetheless, among the things we want a science of cognition to address are education, training, and behavioral change. How can we arrange for experiences to make ourselves, our children, or others more capable people? Certainly, the modern, rapidly changing, technological society has increased the need to acquire new competences. This paper is concerned with the issue of how a science of cognition can address both the learning of formal, academic competences like high-school mathematics and the learning of practical, cognitive competences like air-traffic control.

The goal of effective education illustrates the dilemma of grounding significance in rigorous science. Significant changes in human potential take at least 100 hours to achieve, approximately the time investment required for a semester course. Cognitive psychologists are often concerned with effects as small as ten milliseconds. Ten milliseconds and 100 hours are seven (actually 7.556) orders of magnitude apart. This paper will argue for the **Decomposition Thesis** that the learning which takes place over a 100 hours can be decomposed into learning events involving small units of knowledge and occupying brief periods of cognition. Some might view the Decomposition Thesis as obviously true but there are others (Shepard, 1991) who regard it as obviously false. An alternative hypothesis is that, rather than gradual changes in small units of knowledge, learning depends on abrupt conceptual reorganizations of knowledge. In different theories such conceptual reorganizations can vary in their form from shifts of problem-solving strategy to moments of insight when large bodies of knowledge become transformed. Siegler (1996) has referred to this latter viewpoint as the theory of the “immaculate transition.”

Even if the Decomposition Thesis is correct it leaves open the relevance of the 10 ms level. Is there any reason to believe that learning can be improved by paying attention to events that are measured in tens of milliseconds? Throughout science there are appropriate levels of analysis for applications. One does not design bridges using quantum mechanics (nor would one want to design education at this level). A recent panel of the National Research Council (Bransford, Brown & Cocking, 1999) was put together to address the implications of the science of learning for education. While their report does contain some discussion of biological factors it contains no discussion of events taking place over periods of 10s of milliseconds. The smallest scale events discussed took at least 10 seconds to occur. The neglect did not reflect a conscious decision on the part of the panel that such factors were irrelevant. Rather, it reflected the presupposition of those who constituted the panel that such factors could not be relevant and so were not represented on the panel. It also reflects a frequent practice in cognitive science approaches to education to begin with events on the order of 10 seconds and work upwards. This paper raises the question of whether there is a role for events on the order of 10 milliseconds.

The paper will discuss the evidence for the **Relevance Thesis**, that the microstructure of cognition is relevant to educational issues. Support of a sort for this thesis can be gained by analogy to medicine where molecular biology has become increasingly important and where fine-grained measurements serve important diagnostic functions. I will review examples that show that fine-grained temporal aspects of cognition are important both to learning and to the diagnosis of learning. However, I can only point to a precious few cases where learning has been positively impacted by paying attention to the fine-grained temporal structure of cognition. This is because so little effort has been made to utilize fine-grained temporal information.

One reason for the neglect of fine-grained temporal data are technical. It has been very difficult to gather fine-grained temporal data to guide educational interventions and it has been intractable to trace out the consequences of such fine-grained information for instruction. The technical barriers to gathering fine-grained temporal data are falling as improved methods are becoming available for speech recognition, machine vision, collection of eye-movement data, and other ways of getting high-density measurements of students. However, collecting such fine-grained data are only part of the problem; one also has to find a way to use them. This paper will discuss the **Modeling Thesis**, which is the claim that cognitive modeling can solve the technical barrier of tracing out the consequences of such fine-grained information for instruction.

Cognitive modeling depends on the use of cognitive architectures. Cognitive architectures not only allow us to model the fine-grained temporal structure of cognition, but they also allow us to span parts of cognition that have traditionally been treated as separate in cognitive psychology. Educational applications do not respect the traditional divisions in cognitive psychology. For instance, high-school mathematics involves reading and language processing (for processing of instruction, mathematical expressions, and word problems), spatial processing (for processing of graphs and diagrams), memory (for formula and theorems), problem solving, reasoning, and skill acquisition. To bring all of these aspects together in a cognitive model one needs a theory of the cognitive architecture (Newell, 1990). A cognitive architecture both specifies what is constant across these many domains and how the components from these various domains are integrated. A number of cognitive architectures have been used for education or training purposes including Soar (Newell, 1990), Cognet (Zachary, LeMentec & Ryder, 1996), and our own ACT-R (Anderson & Lebiere, 1998).

In summary, this paper will discuss three propositions of successively greater specificity. The first is the Decomposition Thesis that learning on the large scale is composed out of learning on the small scale. The second is the Relevance Thesis that greater educational effectiveness can be achieved by paying attention to fine-grained temporal detail. The third is the Modeling Thesis that these gains can be achieved by modeling cognition within a cognitive architecture. While it would be convenient rhetorically if one could simply argue that each proposition is true or false (and simpler yet if one could argue for the same conclusion to all three propositions), the current state of knowledge does not allow for such simple conclusions. To appropriately qualify our conclusions requires that we articulate the different time scales at which cognition occurs. Newell (1990) proposed such an articulation that the next section of the paper describes.

2. Newell's bands of cognition

Table 1 reproduces Newell's rendition of the four bands of cognition—the Biological, the Cognitive, the Rational, and the Social. As the table indicates each successive band captures the human experience at roughly 3 orders of magnitude greater than the previous. Ten millisecond effects are at the upper level of Newell's Biological Band while educational effects of consequence are firmly in his Social Band. Newell thought a theory of cognitive architecture was most relevant to the Cognitive Band. Newell thought that issues of cognitive

Table 1
Newell's Time Scales of Human Action

Scale (sec)	Time Units	System	World (theory)
10^7	months		
10^6	weeks		Social Band
10^5	days		
10^4	hours	Task	
10^3	10 min	Task	Rational Band
10^2	minutes	Task	
10^1	10 sec	Unit task	
10^0	1 sec	Operations	Cognitive Band
10^{-1}	100 msec	Deliberate act	
10^{-2}	10 msec	Neural circuit	
10^{-3}	1 msec	Neuron	Biological Band
10^{-4}	100μs	Organelle	

architecture became relatively unimportant at the Rational Band and were completely irrelevant at the Social Band. Nevertheless, this paper will argue that fine-grained temporal factors at the Biological Band do influence higher-level outcomes and that education can be made more effective by paying attention to the lower level.

Bruer (1998) has argued that trying to link biology to education would be a "bridge too far." He suggests that cognitive psychology serves as an "island" to link research on the brain with research on instruction; that is, there should be one bridge from biology to cognitive psychology and another bridge from cognitive psychology to education. His discussion is really concerned with the choice between psychological and biological explanations and not with time scales. Most of what he describes as cognitive psychology would be situated in the upper end of the Newell's Biological Band, the Cognitive Band, and the lower end of the Rational Band. It certainly includes simple laboratory tasks, where 10 millisecond effects are measured. Thus, in part his argument depends on the plausibility of building a bridge from 10 msec effects to complex educational outcomes. Without taking a stance on whether a bridge from the discipline of biology to the discipline of education is a bridge too far, this paper will consider the issues involved in bridging the time scale required for his cognitive psychology-to-education bridge.

The paper will examine the plausibility of three "bridges" of successively longer spans. It will first consider spanning the 4 orders of magnitude in going from what are called unit tasks (the top end of Newell's Cognitive Band, taking on the order of 10 seconds) to educational outcome in courses taking on the order of 100 hours (a "long" bridge). Then the paper will discuss how these unit tasks can be broken down into primitive actions (Newell's Deliberate Acts) taking hundreds of milliseconds and assess whether instruction can be enhanced by paying attention to these primitive actions (a "longer" bridge). Then the paper will show how parallel, subsymbolic processes control 10 millisecond differences in these primitive actions and consider whether differences in these subsymbolic processes have significant impact for educational achievement (the "longest" bridge). The paper will show that the Decomposition Thesis has strong support all the way down. The existing evidence for the Relevance Thesis weakens as we go down to the subsymbolic level. There remain technical barriers to establishing the Modeling Thesis below the unit-task level.

3. The long bridge: from unit tasks to instruction

Card, Moran and Newell (1983) noted that there was a strong tendency for people to decompose the performance of a task into unit tasks that took less than a minute to complete. People focus their attention on solving one of these unit tasks, ignoring for the moment the larger task structure. For instance in text editing, the editing of an entire page would be divided up into a set of tasks like inserting a new line or correcting a word. There is a similar unit-task division in academic tasks. For instance, students solving algebra equations tend to treat each transformation (rewriting) of the equation as a unit task. In most cases like these, there is a near independence among the unit tasks that fosters this division.

There are a number of successful demonstrations that learning can be enhanced by cognitive models that deal with cognition at the unit-task level. Many of these involve what could be called training tasks. For instance, the Soar project (Jones, Laird, Nielsen, Coulter, Kenny & Koss, 1999) has become involved in simulating pilot behavior to provide training experience for Air Force pilots who work with or against the simulated pilots. As another example, Cognet has been deployed to provide team training on shipboard anti-air warfare tasks (Zachary, Cannon-Bowers, Burns, Bilazarian & Krecker, 1998). There have also been successful intelligent tutoring efforts for domains as diverse as physics (VanLehn, Niu, Siler & Gertner, 1998) and introductory computer literacy (Graesser et al., 2000). In all of these domains the fundamental principle has been to provide models of human behavior accurate at the grain size of about 10 seconds. As an example of such an effort that I am familiar with, this paper will consider the Cognitive Tutor Project (e.g., Anderson, Corbett, Koedinger & Pelletier, 1995). This project is useful because it shows what it would mean to support the Decomposition, the Relevance, and the Modeling Theses. We can use this example at the unit-task level as a standard for what would be similar evidence at lower levels.

3.1. Cognitive tutors

The Cognitive Tutor Project was based on an analysis of the unit tasks that were necessary to achieve competence in a number of domains of mathematics and computer programming. These were represented as production rules (but not to be confused with production rules in ACT-R or SOAR, which are much more fine-grained). The following are typical of the tutor production rules that we proposed:

Lisp

IF the goal is to get the nth element of the list
THEN code "car"
and set as a subgoal to get the
n-1st tail of the list

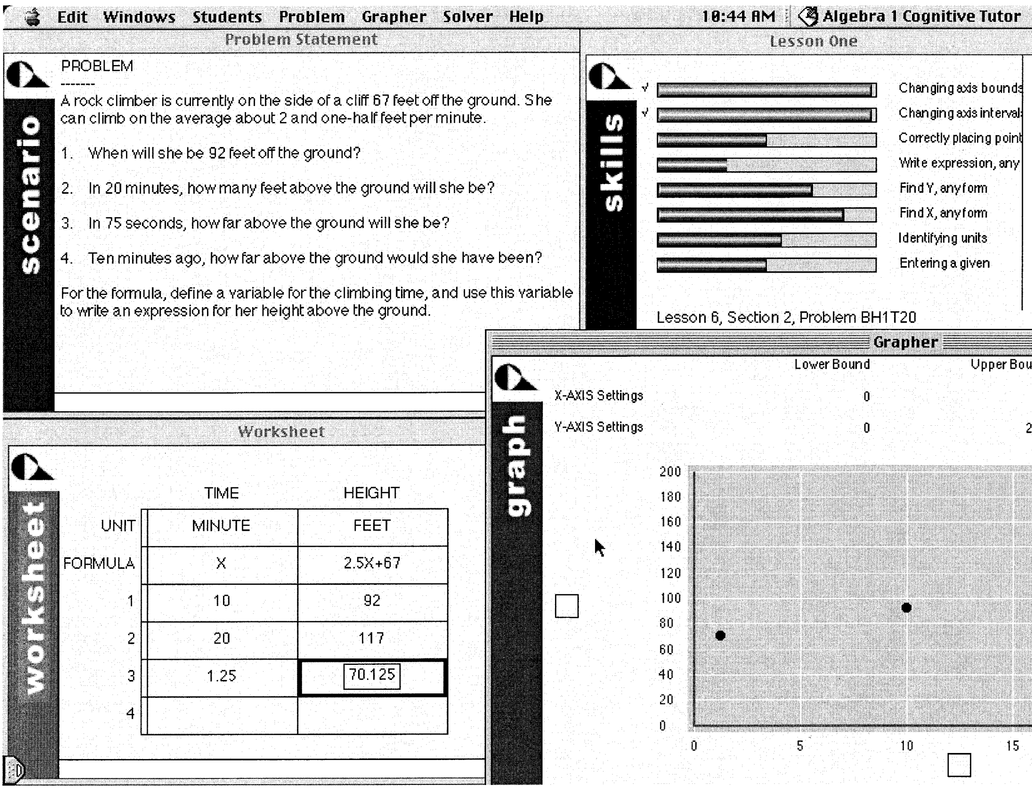

Fig. 1. The tutorial interface for the algebra tutor.

Geometry
IF the goal is to prove two triangles congruent
THEN set as subgoals to prove
corresponding parts congruent

Algebra
IF the goal is to solve an equation in x
THEN set as subgoals to graph the right
and left sides of the equation
and find the intersection point(s)

A typical course involves on the order of 500 such rules. Given the assumption that learning in these domains involves the acquisition of such production rules, it followed that we should try to diagnose whether students have acquired such production rules and provide instruction to remediate any difficulties they have with specific rules. This led to the design of cognitive tutors that ran production-rule models in parallel with the students and attempted to interpret student behavior in terms of the rules. The tutors use computer interfaces in which students can solve problems—for instance, Fig. 1 shows the interface used in the most widely disseminated tutor, the PAT tutor for Algebra 1 (Koedinger, Anderson, Hadley &

Mark, 1997). A process called **model tracing** tries to find some sequence of productions that produces the behavior exhibited by the student. The model-tracing process is key to the success of the tutors because it allows us to interpret the student behavior. The end of the paper will consider some of the technical issues surrounding model tracing.

The interpretation of student behavior, delivered by the model-tracing process, controls the tutorial interactions in 3 principal ways:

(1) If students ask for help it is possible to give help appropriate to where they are in the problem-solving process.
(2) If students correctly perform the behavior corresponding to a rule the student model updates its confidence that they know the rule. This provides cognitively-based mastery criteria for promoting students through the curriculum.
(3) If students make errors that are modeled by "buggy" rules in the student model it is possible to intervene and provide appropriate remedial instruction.

Each of these instructional maneuvers is predicated on the validity of the cognitive model and the attributions the model-tracing process is making about the student at the unit-task level (approximately 10 seconds). There are numerous assessments that establish the effectiveness of these tutors (e.g., Anderson, Corbett, Koedinger & Pelletier, 1995; Koedinger, Anderson, Hadley & Mark, 1997). The tutors are now being used to deliver instruction to about 100,000 students (Corbett, Koedinger & Hadley, in press). Critically for the purpose of this paper, the research on these tutors show how the three theses are supported.

3.2. Decomposition thesis

The typical production rule spans about the interval of a unit task (10 seconds) and the tutoring technology allows us to track their learning. In a number of analyses we have shown that performance on some complex problem (e.g., writing a LISP function) can be predicted from the state of learning of individual components (e.g., Anderson, Conrad & Corbett, 1989; Corbett, Anderson & O'Brien, 1995).

3.3. Relevance thesis

Evidence for the Relevance Thesis involves showing that paying attention to things at the unit-task level enables an improvement of instruction that would not be possible at a higher level. There are two fairly well established examples of this. The first involves feedback. The tutors use their diagnosis of what is happening at the 10-second level to give feedback at this grain size. It has been shown (Corbett & Anderson, 2001) that learning is less rapid if the same feedback is given at a larger grain size (and hence less immediate). It has also been shown that learning is less rapid if the feedback is more generic about right and wrong and does not present the specific information surrounding what is wrong at the unit-task level and what to do at the unit-task level (Anderson, Corbett, Koedinger & Pelletier, 1995). The second example involves knowledge tracing. Here we track how well students are doing at the level of individual rules and present problems that are relevant to dealing with those rules they are having difficulty with. This is in contrast to just giving students more general practice. We have shown (e.g., Corbett & Anderson, 1995; Corbett & Knapp, 1996; Corbett & Trask, 2000) that this practice, guided by diagnosis at the unit task level, leads to better learning. Corbett (2001) has shown that the combined effect of focused feedback and knowledge tracing is to produce achievement gains that are comparable to the high-water mark of 2-sigma improvement reported for human tutoring by Bloom (1984).[1]

3.4. Modeling thesis

The support for the Modeling Thesis is already implied in our discussion of the relevance thesis. It is because of the ability to model trace and knowledge trace that we are able to give targeted instruction and targeted practice. Thus, cognitive modeling is the enabling technology that has allowed events at the 10-second level to be used to direct instruction. Moreover, this is not unique to our cognitive tutors but, as described earlier, cognitive modeling of different varieties has been successfully used in a number of other instructional systems.

3.5. Conclusion

Thus, there is good evidence for the Long Bridge going from 100s of hours to 10s of seconds. However, the very success of these efforts that stop at the 10-second level may make one wonder whether there is anything to be gained by decomposing further. Indeed, if one reads Newell's discussion of the unit-task level, one might wonder whether it would be possible for finer levels to have any effect on instructional objectives. Newell (1990) argued that fine-grained cognitive processes do not have much influence above the unit-task level. Behavior becomes more a function of the characteristics of the task than any properties of the architecture. Or more precisely, behavior becomes a function of what we know about the task:

> ". . . as the system has more time, the solutions it finds will reflect whatever knowledge it has. As that happens, we will have less and less a chance of ascertaining what processing was involved in determining the solution . . . there will be a shift towards characterizing a system as a knowledge-level system, that is, just by the goals and knowledge it contains, without regard to the way in which the internal processing accomplishes the linking of actions to goals" (Newell, 1990, p. 150).

Thus, one might conclude that the unit-task level is the right level for applications because it captures the basic units of knowledge without the irrelevant implementation details. In many domains like those addressed by our cognitive tutors, one does not really care about the details of how students achieve the unit tasks. One only cares that the students can achieve them. However, the microstructure of these unit tasks can still be significant for two reasons. First, it can be used to help diagnose whether in fact students have mastered the unit tasks. Second, these details can be predictive of retention and transfer of this knowledge. Subsequent sections of this paper will give examples of these potentials.

4. The longer bridge: from primitive actions to instruction

Unit tasks are sequences of lower-level actions. When a student performs a unit task such as entering an answer into a cell of the PAT tutor (Fig. l), that student engages in a set of actions such as attending to various parts of the screen, making mouse movements, and key presses. These are events that can be measured in 100s of milliseconds and are basically Newell's deliberate acts (see Table 1). There is reason to believe that getting the detail right at this level can be critical to instructional goals. This is particularly apparent in work on computer-generated forces (Pew & Mavor, 1998) where people train with synthetic partners and against synthetic opponents. It is not enough just to perform the unit tasks right; the synthetic agents need to perform these actions like real people. For instance, it was important in the TacAir Soar effort (Jones et al., 1999) that the synthetic pilots not only make the turns that real pilots do but they make them with the timing of real pilots.

The original GOMS work of Card, Moran, and Newell was concerned with how the individual actions were put together to achieve unit tasks. It was intended and has been proven to be a useful methodology for assessing the design of systems. For instance, Gray, John and Atwood (1993) provided a GOMS analysis of the task of telephone toll and assistance operators. They showed that a consideration of the timing of the primitive actions of the operators relative to system delays led to counterintuitive predictions about the relative merits of different workstations. These predictions were confirmed, leading the telephone company to make a decision that saved millions of dollars.

A number of modeling systems have now arisen for predicting how the individual actions combine to achieve unit tasks. These include Meyer and Kieras's EPIC (Meyer & Kieras, 1997a, b), the merge of Soar and Epic (Chong, 1999), and our own ACT-R/PM (Byrne & Anderson, in press). A critical issue in these systems is understanding how parallel threads of perceptual and motor processing are put together with cognitive processing. While there is some dispute about the degree of seriality of the cognitive component, there is agreement that the human system can run perceptual, motor, and cognitive processes in parallel—for instance, we can read a screen, while we type letters, while we decide the message we want to compose.

Fig. 2 illustrates the parallel threads of processing and dependencies in a simulation of expert performance in an air traffic controller task (Ackerman & Kanfer, 1994) that Lee (2000) developed in ACT-R/PM. This illustrates the detail involved in a 10-sec unit task. While the reader is referred to Lee's thesis for an explanation of the content of the individual boxes, the basic point is there are four parallel lines of processing that need to be coordinated. These involve the execution of motor tasks (in this case key presses), preparation of these motor actions, cognition required for decisions and goal settings, and access of critical perceptual information (in this case visual). There are dependencies between these streams such that an action in one stream cannot execute until an action is taken in another stream. The bold line indicates the critical path in terms of what depends on what else. As can be seen, all four of the processes of motor execution, motor preparation, cognition, and visual attention are on the critical path at different times.

The success of such models provides support for the Decomposition Thesis. Lee and Anderson (2001) have gone on to show that development of expertise involves improvement in the individual components. Thus learning of a complex whole can be decomposed into learning of very small pieces. However, demonstrations like this largely leave the Relevance Thesis untouched. Is there reason to believe that high-level instruction objectives can be improved by paying attention to such low-level details?

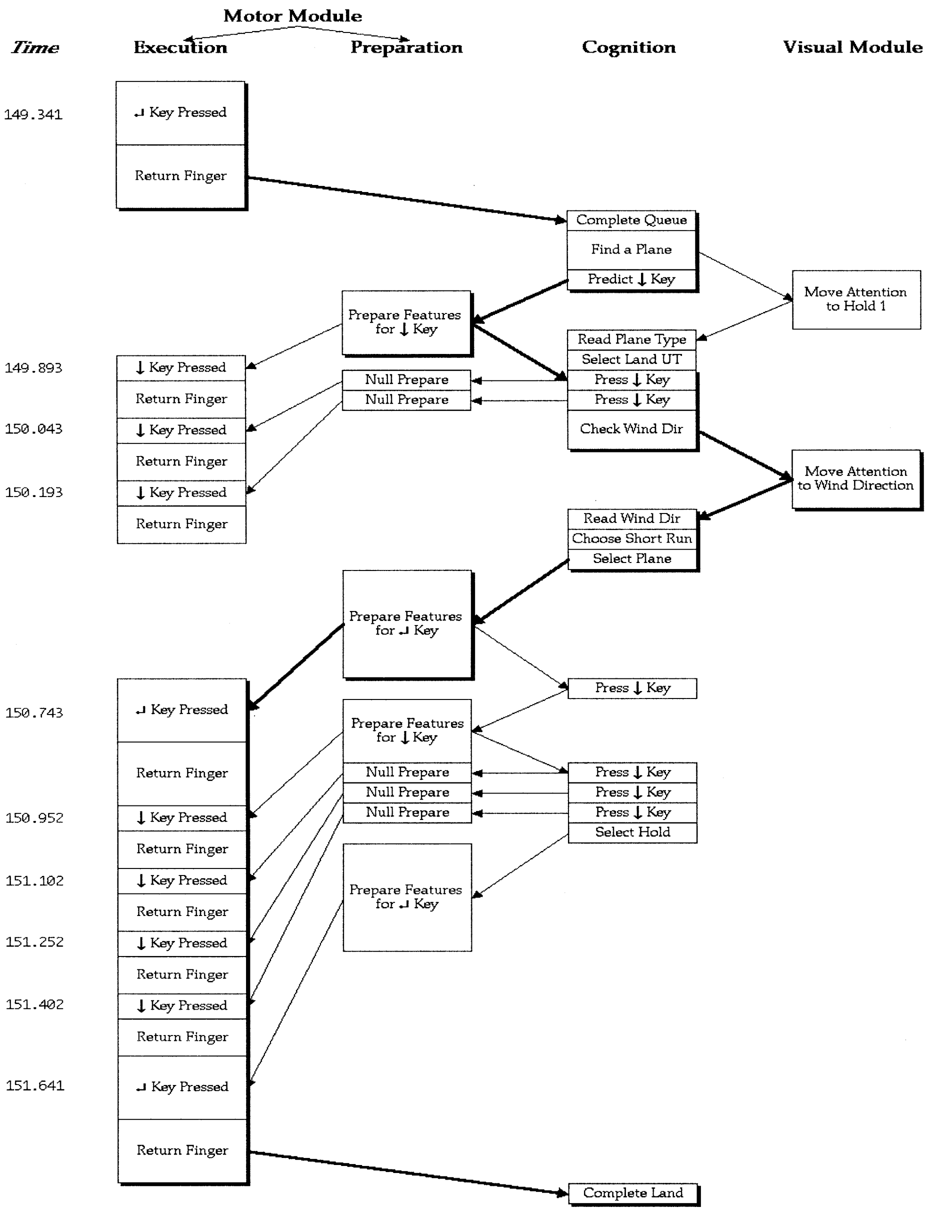

Fig. 2. Critical path of a land unit task in Lee's (2000) air-traffic control simulation.

4.1. Eye movements and tutoring

There is one demonstration of increased instructional effectiveness in our lab that was achieved by tracking primitive actions—in this case eye movements. This is part of a larger effort that we call "high-density sensing" which is concerned with what instructional leverage can be gained by tracking eye movements, facial expressions, and speech information. All of these provide a moment-by-moment read-out of the student and are part of an effort to make the computer tutor as sensitive to the student as is a human tutor. A typical computer tutor only "senses" a student when the student types something or makes a mouse movement. In contrast, a human has constant visual and auditory access to the student and can detect where the student is attending and when the student appears puzzled.

In one investigation of this issue, Gluck (1999, see also Anderson & Gluck, 2001) has looked at the instructional leverage one might gain by monitoring student eye movements while interacting with an algebra tutor. The tutor he developed is a simplified version of the PAT tutor used in schools and its interface is illustrated in Fig. 3 (compare with Fig. 1). Fig. 3a shows a screen display as it appears at the beginning of a problem. The student's task is to fill in the column labels and units, enter a variable and an expression written in terms of that variable, and then answer whatever questions are present. Fig. 3b displays the completed problem. The first question in that problem is an example of what is called result-unknown problem and the second is an example of what is called a start-unknown problem. The key aspect of the problem is the expression 12 + 45x, which the student has filled in. The principal goal of the lesson is to teach the student how to create such expressions and use them to solve problems.

Gluck found numerous points where eye movements indicated opportunities for instructional leverage. For instance, about 40% of all tutor messages to students are not read but current tutors proceed on the assumption that the messages have been read. Often students will quickly self-correct when they do not read the message, but if there is a long latency and the students do not read the message, the odds become high that they will make an error.

As another example, eye movements allow us to disambiguate methods of solving problems. Fig. 4 shows two examples of students solving a result-unknown problem in the tutor. In both cases the student produces the right answer and so there is no bases in the answers they type for suspecting a difference in the solution process. However, the fixation patterns are very different. The critical difference is that in (a) the student is fixating the algebraic expression and using it in the solution of the problem while in (b) the student is rereading the verbal statement and ignoring the expression. This reflects a difference in early mastery of algebra where some students use algebraic expressions while others fall back on pre-algebraic reasoning skills (Koedinger & MacLaren, 1997). Gluck showed that students solved problems more rapidly and successfully when they displayed the eye movements typified in part (a) than when they displayed the eye movements typified in part (b).

Also as part of his dissertation, Gluck was able to develop a cognitive model (in ACT-R/PM) that solved these problems and predicted the distribution of eye movements. Thus, his research provides further evidence that it is possible to decompose complex learning into primitive actions, supporting the Decomposition Thesis. However, what about the Relevance and Modeling Theses?

As a test of these theses, Scott Douglass and I decided to take Gluck's research and see if we could not develop instructional interventions for the tutor that were triggered by eye movements. We identified a number of occasions where the eye movements indicated an instructional opportunities and presented brief auditory messages lasting about 2 seconds. For instance, we noted occasions when students did not read instruction and where it seemed that instruction was critical. One such case, discussed above, occurred when students made

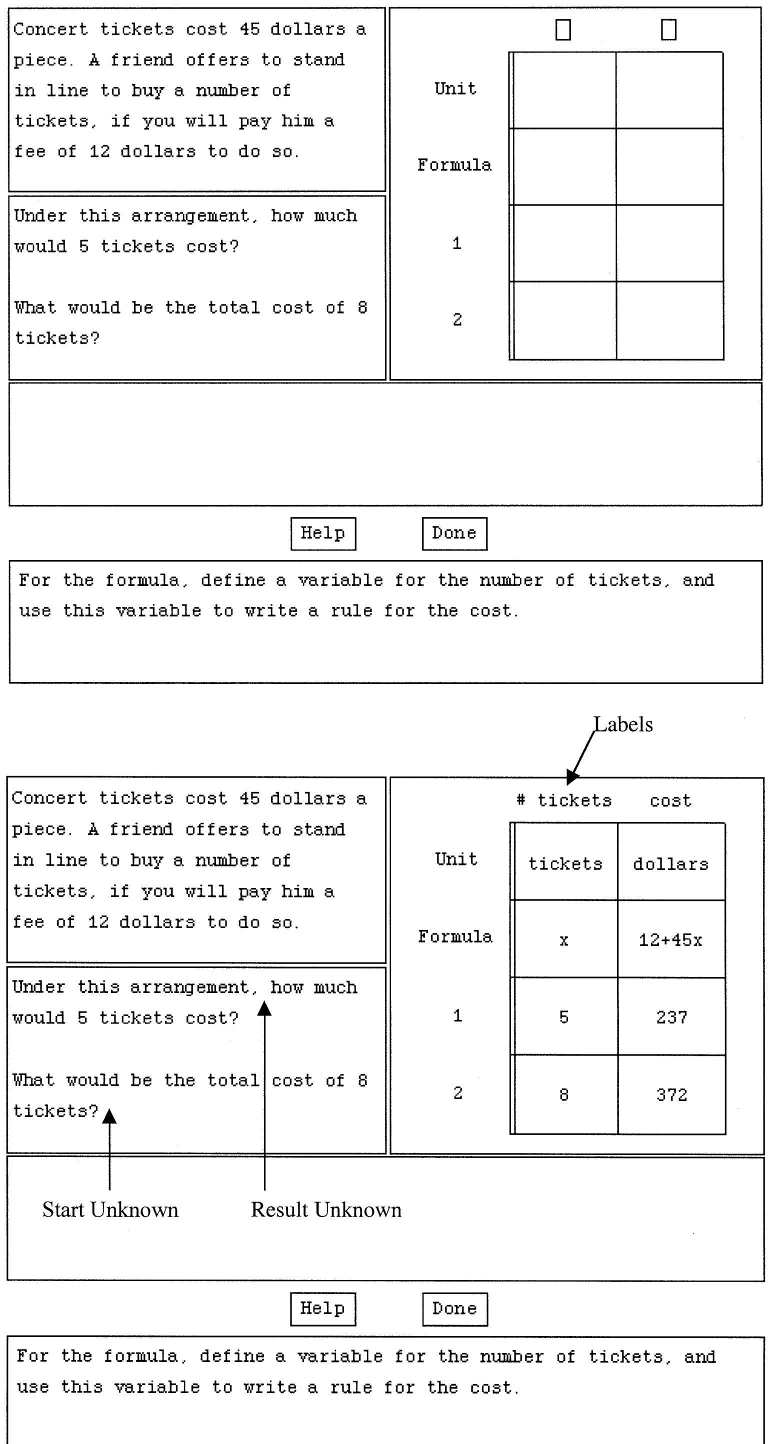

Fig. 3. The tutor screen (a) at the beginning of the problem and (b) at the end of the problem.

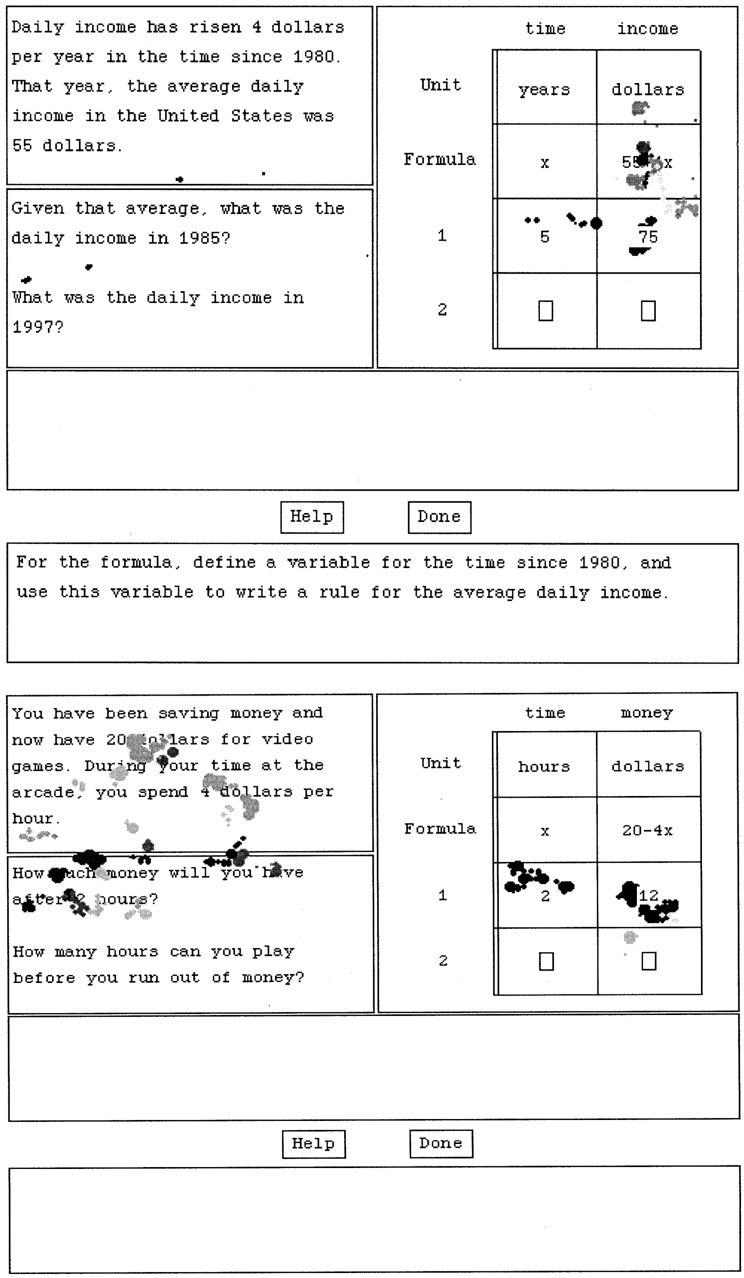

Fig. 4. Eye movements (a) when a student calculates the answer using the expression and (b) when a student calculates the answer using the problem statement. The blotches reflect fixation points that go from dark to light with the passage of time.

an error, failed to self correct, and did not read the error message. When 10 seconds had passed without students fixating the error message for more than 1/4 second, they heard "Read the help message." As another example, we identified occasions where student eye

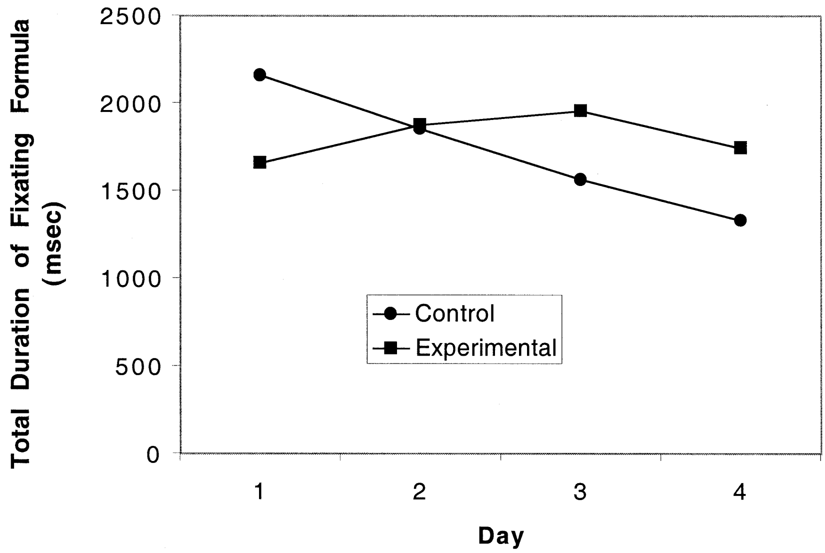

Fig. 5. Amount of time spent fixating the formula in the control and experimental conditions as a function of the amount of practice.

movements indicated a deficiency in their problem-solving strategy. For instance, if as in Fig. 4b, the student failed to solve a start unknown problem and failed to fixate the expression, we would direct the student to use the expression. More precisely, if students made an error and failed to fixate the expression for at least 1/4 s, they heard the message "Try using the formula to compute your answer."

In a test, where we compared this augmented tutor with a tutor that was otherwise identical (including collection of eye movements), we found that students were able to reach the same level of mastery 20% faster with the augmented instruction. One might argue that this result does not really demonstrate that we are improving instructional effectiveness by attending to the fixation-by-fixation behavior of the student. Perhaps all the leverage simply comes from simply getting students to read more of the instructional messages. The Relevance Thesis at the primitive action level requires evidence that we utilize information only detectable at this fine grain size. Moreover, we would like to show that our eye-movement tutor has impact on behavior at this fine grain size.

Fig. 5 provides one example of such a refined effect. This shows how the fixation pattern of students changes over the 4 days with respect to amount of time spent encoding the formula. Fig. 4b illustrated the curious pattern where students would not use the formula in solving the problem. As noted, the tutor would intervene when they displayed this pattern and made an error and it would suggest that they try to use the formula. This figures plots how much time students spent reading the formula during their first attempt to fill in the cells in the start-unknown and result-unknown cells. In the control condition students show decreased use of the formula, probably reflecting the general speed up in problem solution. However, students in the experimental condition maintained a constant rate of use of the formula. The difference in the linear trends in the two functions is significant ($t(38) = 2.12$,

$p < .05$). This indicates that indeed detailed monitoring of students was having an impact on the fine-grain details of their problem-solving behavior.

This is by no means an example of the best that can be achieved by eye movements. This is just a first attempt. It involved an early lesson with minimal student confusions and consequently few opportunities for instructional augmentation. Nonetheless, it is an existence proof that instructional leverage can be gained by paying attention to the fine-grained temporal detail of the student's behavior. Thus, it is evidence for the Relevance Thesis. By taking advantage of technology that allows us to track the student at a subsecond level we were able to improve the effectiveness of the instruction. These instructional interventions were also informed by the analysis provided by ACT-R/PM and Gluck's ACT-R/PM model. However, unlike our cognitive tutors there was not a running cognitive model simulating student eye movements and directing the instructional interventions. One reason for this has to do with the non-determinism of behavior at this fine temporal grain size and the difficulty of predicting the exact trajectory of eye movements. I will return to this issue at the end of the paper.

5. The longest bridge: from the subsymbolic level to instruction

The previous section still leaves open the status of the l0-millisecond effects that have fascinated cognitive psychologists. When we get to effects as small as 10 milliseconds, they can no longer be accounted for by which primitive actions people perform but rather how fast they perform these actions. To understand these latency effects we are beginning to penetrate into Newell's Biological Band. Also, this band is relevant to understanding accuracy effects. When we look to understand whether primitive actions are successful (key to the instructional example that will be developed in this section) we again are looking at effects which have their explanation below Newell's Cognitive Band. The approach in cognitive psychology has largely been not to actually model the biological processes but rather to describe them at some level of abstraction. This level is called the **subsymbolic level** (e.g., in Anderson & Lebiere, 1998).

There seems a general consensus in cognitive science that the underlying neural processes can be abstracted as a set of parallel computations taking tens of milliseconds to complete. Rumelhart and McClelland (1985) explicitly describe their PDP model as such an abstraction. Newell considered the parallel firing of elaboration productions in SOAR in these terms and justified their timing (about 10 ms per production) as an abstraction from knowledge about neural firing. Global memory models (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1993) all seem to involve such a conception. Instance-based models like Logan (1988) or Nosofsky and Palmeri (1998) postulate a parallel retrieval process. The ACT-R architecture assumes a parallel stage of activation processing and pattern matching before the firing of any individual production rule.

Many of these theories postulate that success of higher-level cognition depends on the fluency with which these lower-level processes can progress. They also postulate that practice can have an impact on cognition by improving these subsymbolic processes. As this proves important to an instructional example from a dissertation by Haverty (1999) and

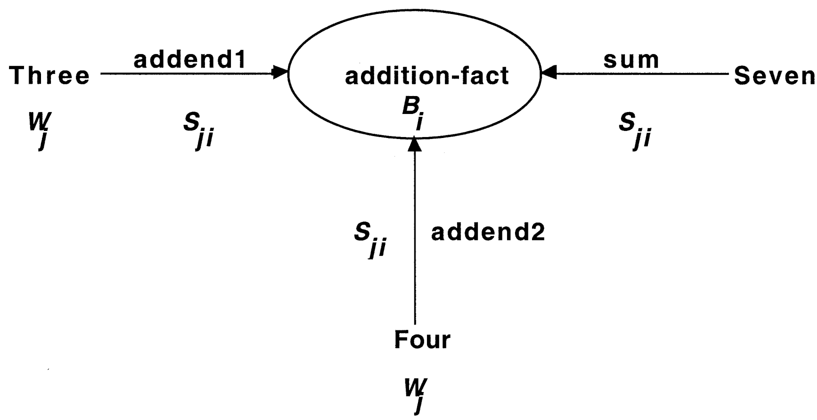

Fig. 6. A representation of the subsymbolic quantities determining the activation of a chunk that encodes 3 + 4 = 7.

because it is relatively easy to explain, I would like to develop in some detail how subsymbolic factors in ACT-R control fluency in the retrieval of declarative facts. Declarative memory in ACT-R consists of units called chunks that encode basic facts like the addition fact $3 + 4 = 7$. Fig. 6 shows a graphic representation of this fact along with various critical subsymbolic quantities. Associated with each chunk is an activation level that determines how successfully and rapidly it can be retrieved. The activation of a chunk i is defined as

$$A_i = B_i + \sum_j W_j S_{ji} \qquad \textbf{Activation Equation}$$

where B_i is the base-level activation of chunk i and reflects how recently and frequently the chunk has been used. The summation $\Sigma\, W_j S_{ji}$ is the associative component and reflects the additional boost to activation that comes from the current context. The W_j's reflect the attentional weightings, called source activations, given to elements in the context that are part of the current goal. In Fig. 6 *Three* and *Four* are the elements of the goal that serve as sources. The S_{ji}'s are the strengths of associations from the elements j to chunk i.

The activation of a chunk will determine the fluency with which that knowledge can be processed. One way of measuring fluency is retrieval time and, according to ACT-R, retrieval time is determined by the level of activation according to the following equation:

$$\text{Time to retrieve chunk } i = Fe^{-A_i} \qquad \textbf{Retrieval Time Equation}$$

A typical estimate of the parameter F is less than a second and this reflects the idea that retrieval will take less than a second. According to ACT-R, variations in activation levels produce many of the effects measured in tens of milliseconds, such as priming effects.

5.1. Fan effects and practice

Historically, the ACT theory of declarative retrieval has focused on tasks that require participants to retrieve facts from declarative memory. The second experiment in Pirolli and

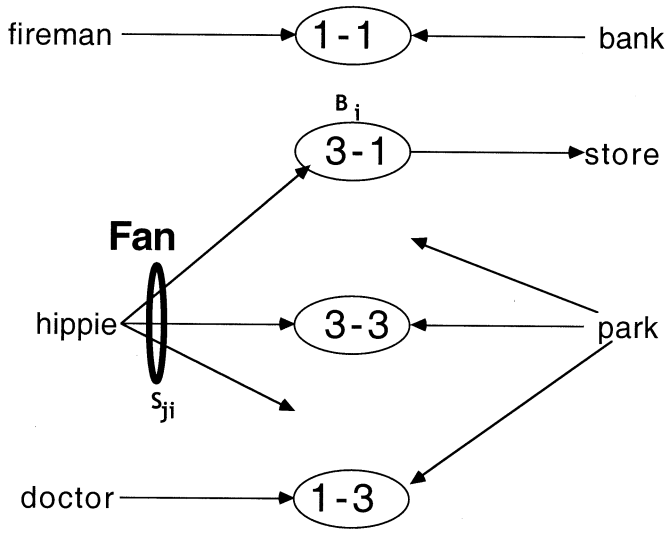

Fig. 7. An illustration of the structure of the material from Pirolli and Anderson's fan experiment.

Anderson (1985) is a good one to illustrate the contributions of both base-level activations (B_i) and associative strengths (S_{ji}) to the retrieval process. This was a fan experiment (Anderson, 1974) in which participants were to try to recognize sentences such as "A hippie was in the park." The number of facts (i.e., fan) associated with the person (e.g., hippie) could be either 1 or 3 and the fan associated with the location could be either 1 or 3. Participants practiced recognizing the same set of sentences for 10 days. Fig. 7 illustrates how to conceive of these facts in terms of their chunk representation and subsymbolic quantities. Each oval in Fig. 7 represents a chunk that encodes a fact in the experiment. As a concept like hippie is associated with more facts, there are more paths emanating from that concept and, according to ACT-R, the strengths of association S_{ji} will decrease.

Fig. 8 illustrates how the activations of these chunks vary as a function of fan and amount of practice. There are separate curves for different fans, which correspond to different associative strengths (S_{ji}). The curves rise with increasing practice because of increasing base-level activation. Fig. 9 illustrates the data from this experiment. Participants are slowed in the presence of greater fan but speed up with practice. The practice in this experiment gets participants to the point where high-fan items are recognized more rapidly than low-fan items were originally recognized. Practice also reduces the absolute size of the effect of fan but it remains substantial even after 10 days practice.

According to the ACT-R theory the effect of fan is to reduce the strength of association, S_{ji}, from a term like hippie to the chunk encoding a fact. As argued in Anderson and Lebiere (1998), the strength of association is approximately $S - ln(fan)$. In Anderson and Reder (1999), we used values of S around 1.5 in fitting the data in that paper and this is the value used for fitting the data in Fig. 9. The effect of practice is to increase the base-level activation of the facts. According to Anderson and Lebiere, an item with *n* units of practice will have an approximate base-level activation of $.5*ln(n)$ and this is what was used in fitting the data. Fig. 8 shows the activation values that are gotten from combining the base-level activation

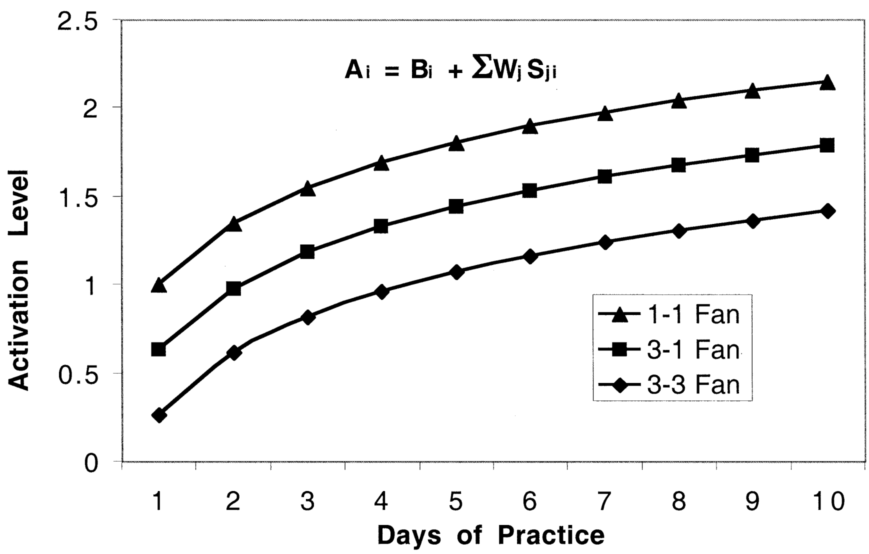

Fig. 8. Growth of activation for different fan conditions of Pirolli and Anderson (1985).

with the source activation according to the Activation Equation, setting the weights, W_j, in this experiment to .333 (as used in Reder and Anderson, because each of the three content terms (hippie, in, park) in the sentence gets an equal 1/3 source activation). These are parameter-free predictions for the activation values. As can be seen, they increase with practice with low-fan items having a constant advantage over high-fan items.

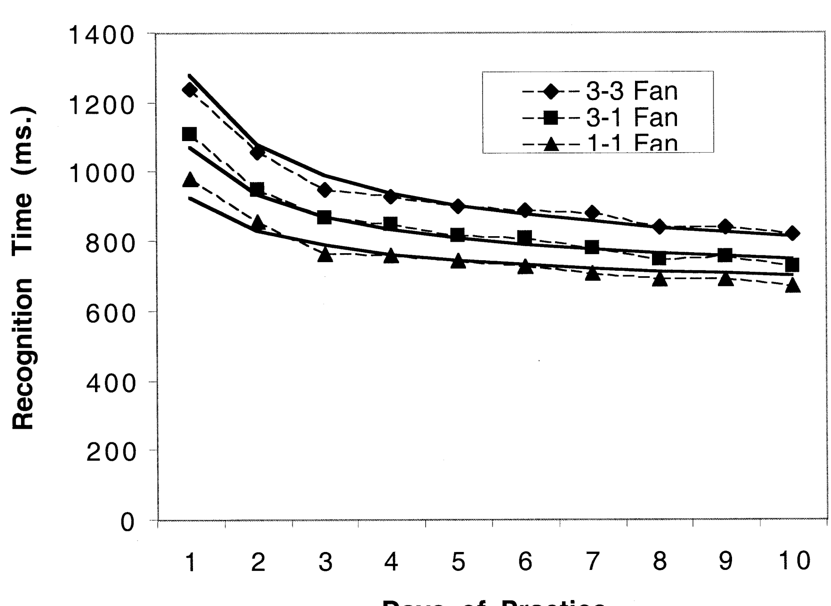

Fig. 9. Recognition latencies, $T = I + Fe^{-A_i}$, determined by the activation levels in Fig. 4.

The Retrieval Time Equation specifies how to convert these activation values into predicted recognition times:

$$\text{Recognition time } = I + Fe^{-A_i}$$

where I is an intercept time reflecting encoding and response time while F is the latency scale factor given in the Retrieval Time Equation. Thus, fitting the model required estimating two parameters and these were $I = 597$ ms. and $F = 890$ ms., which are quite similar to the parameters estimated in Anderson and Reder (1999). The value of I is also quite reasonable as the time to encode the words and emit a response (key press). The overall quality of fit is good with a correlation of .986. Moreover, it is worth noting that this correlation does not depend on the parameter estimates I and F but only on e^{-A_i}, which means that it measures a parameter-free prediction of ACT-R. The effect of I and F is only to scale this critical quantity onto the range of the latencies.

Comparing extreme points, Fig. 9 shows the activation processes having latency effects of almost 600 milliseconds. This is close to the maximum that ACT-R would attribute to subsymbolic activation effects since each production cycle (symbolic level) is postulated to be under a second. As adjacent points on the individual curves illustrate, activation can produce much smaller differences between conditions of the size of 10 msec. or smaller. This is the factor in the ACT-R theory that produces many of the latency effects that have fascinated experimental cognitive psychology. While not wanting to declare ACT-R correct in this paper, I think the theory does show that the Decomposition Thesis can hold all the way down to the level of 10-millisecond effects. That is, it describes how learning at this subsymbolic level can result in changes in the timing of the primitive actions and ACT-R shows how these primitive actions can be combined to produce significant events like the solution of an equation.

5.2. Mastery of arithmetic facts

However, what about the Relevance Thesis? Is there any reason to suppose that such tiny effects have any consequence for our target application, which is learning over 100 hours? These latency effects are only the most sensitive measure one can have of activation effects on fluency in access. Activation also influences the probability that a chunk will be over threshold and so retrieved. Also, in many cases multiple chunks will compete for retrieval and the most active will be selected. Activation-based effects on retrieval probability play a major role in developing mastery of the basic arithmetic facts such as the addition table (e.g., $3 + 4 = 7$) and the multiplication table (e.g., $3 * 4 = 12$). As has been noted (e.g., Campbell, 1995), learning these facts is like a huge fan experiment with particular numbers like 3 associated to many facts. Children spend years mastering these facts. In the ACT-R theory they need this amount of practice to build up the base-level activations to overcome the high levels of interference. Lebiere (1998) produced an ACT-R simulation that successfully modeled the rich data on the development of arithmetic facts from early childhood to adulthood. That model was given the same amount of practice over the same simulated time as people e perience and developed the same level of competence. Lebiere's simulation gives still further support for the Decomposition Thesis.

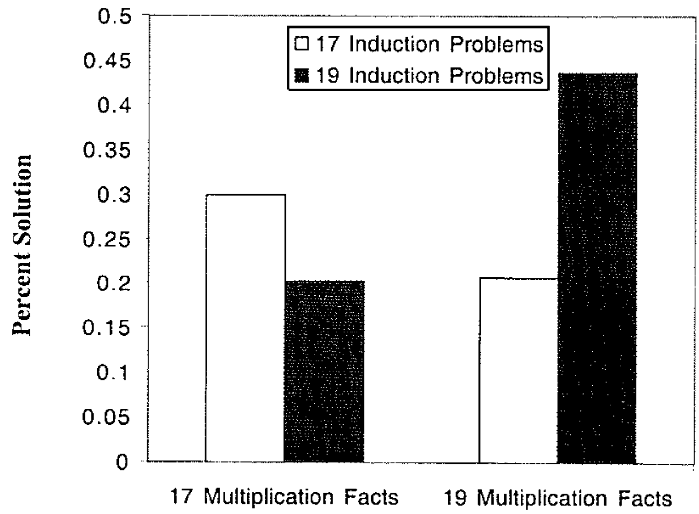

Fig. 10. Effect of practice of multiplication facts on solution of induction problems from Experiment 1, Haverty, 1999.

While these low-level activation processes may be controlling knowledge of the number facts one can question the importance of such knowledge to "real" arithmetic competence. For instance, mathematics educators appear to be uncertain about the importance of practice of the number facts (NCTM, 1989, 2000). Recently, Haverty (1999) has completed a dissertation looking at the relationship between such number knowledge and ability to detect patterns in sequences such as 2, 5, 10, 17, 26, 37 . . . , which is generally thought to reflect a desirable mathematical ability. In a correlational study with seventh graders she showed a fairly strong relationship ($r = .61$) between knowledge of number facts and ability to solve such inductive problems.

Such a correlation is open to a number of interpretations and so she followed this up with a series of experimental studies in which she gave students practice on specific number facts (some students the 17-times table and other students the 19-times table—she chose these high numbers to avoid prior practice). She looked at the solution of induction problems that involved those facts (e.g., representing the sequence 38, 57, 76 as $19*(X + 1)$ involves solving a problem that involves the 19 facts). Fig. 10 shows the results from one experiment in Haverty's thesis, classifying students by which facts they studied and which problems they were tested on. Students achieved 37% correct when the problems matched the facts while 21% when they did not. In a second study, the numbers where 35% and 25%. Haverty controlled for availability of just the numbers "19" and "17" by having the 17-multiplication-fact groups study 19-addition facts and vice versa.

Finally, Haverty, Koedinger, Klahr and Alibali (2000) have published an ACT-R-based model in which availability of such knowledge does play a critical role in solving inductive problems. Fluency in access to declarative information is critical in ACT-R because each production cycle depends on retrieval of needed information. For instance, having highly available the square sequence 1, 4, 9, 16, 25, 36 . . . enables the model to see the pattern in the sequence 2, 5, 10, 17, 26, 37 . . . . This model successfully handled data from college

Table 2
Status of the Theses in the Paper

	Decomposition Thesis	Relevance Thesis	Modeling Thesis
Unit-task Level	Supported	Supported	Supported
Primitive-action Level	Supported	Supported	Open
Subsymbolic Level	Supported	Open	Open

students (in contrast to the 7th graders in the Haverty dissertation). In conclusion, it seems that things happening at a subsymbolic level can be critical to significant mathematical competences that we aspire to train over long periods of time.

The total of the Haverty research effort comes somewhat short of establishing the Relevance and Modeling Theses in their strongest form. Her research shows that relatively gross manipulations of practice can have consequence for high-level outcomes like success on induction problems, and this can be modeled in terms of low-level factors like activation of number facts. An ideal demonstration, from the perspective of the Relevance and Modeling Theses, would be to show that detailed monitoring and modeling of 10 millisecond effects would enhance the effectiveness of the practice manipulation by selecting the optimal problems for practice. That is a demonstration that is still lacking.

The desired demonstration might compare two groups of students—one of whom received just general practice of arithmetic facts and another group of subjects who received selective practice depending on the speed with which they retrieved particular facts. If we could show greater improvement in the speed of retrieval of the group with focused practice we would provide the kind of detailed evidence required at the subsymbolic by the Relevance Thesis—that data at the 10 msec grain size guided instruction that improved performance at this grain size. This is the kind of evidence that we were able to show at the higher grain sizes—in the cognitive tutors, that information at the unit-task level enabled instruction that led to improvement in performance at this level and, in the eye-movement study, that information at the fixation level enabled instruction that led to improvement at this level. However, this is not enough. It would also be necessary to show that such low-level improvements also led to improvement in something significant like mathematical induction. The modeling thesis would be established if one could show that cognitive modeling at this level was critical to defining effective interventions. For instance, one might estimate base-level activations of critical facts and select for practice those items with low base-level activations. One would need to show that item selection determined by these model-based estimates led to better results than a system that used some global statistic like mean latency.

6. Discussion

Table 2 summarizes where we stand on the three theses outlined at the beginning of the paper. The evidence for the Decomposition Thesis seems strong. While there still is plenty of controversy about the details of the right cognitive architecture and there are many details remaining to work out about just how various competences decompose into 10 millisecond effects, all the evidence indicates that such a decomposition is possible and will be achieved. Achieving it in its entirety will be an accomplishment for cognitive science on the order of the Human Genome Project. In the terms of Koedinger (1999), this would be the Human Menome Project.

With respect to the Relevance and Modeling Theses, the evidence is much less complete. We have at least one demonstration that paying attention to primitive actions at the level of 100s of milliseconds can result in a improvement in the educational outcome. Thus, there is evidence for the Relevance Thesis down to the level of primitive actions. Haverty's dissertation is consistent with the proposal that effects at the levels of 10s of milliseconds can result in high-level educational achievement. However, as noted in the previous section, her research falls short of the sharp evidence we have at the higher levels.

With respect to the Modeling Thesis, we are still lacking evidence below the unit-task level. While cognitive models exist that can predict these effects of low-level manipulations, we are still awaiting the technology to turn these models into systems that can optimize learning outcomes. The remaining part of this paper is something of an appendix that discusses some issues that might be involved in such a technology.

6.1. The model tracing below the unit-task level

There are undoubtedly many potential ways to use cognitive modeling at the temporal grain size of 10s of milliseconds to improve instruction. However, here I will focus on the prospect of using cognitive modeling at this level in the same way that it has been successfully used for unit tasks (tens of seconds) in our cognitive tutors. The key idea in this application is model tracing.

In model tracing one has a model that will generate the various sequences of actions that students will generate. Because one has to be able to match many students (ordinary, buggy, or particularly insightful) this student model is often a "super model" in that it represents the union of the tendencies of many students. As a matter of practical fact, one does not really capture the tendencies of the total population. Students do sometimes perform unrecognizable actions and the tutor more or less has to punt, often acknowledging that it is not following what the student is doing with some default comment (e.g., "unrecognizable action"). In point of practice, we probably fail to recognize about 1% of the student actions in a mature cognitive tutor (and about one-fifth of these are correct solutions that the tutor does not "understand"). One of the promises of model tracing at a finer grain is that one may be able to recognize even more of the actions.

A cognitive model in such a tutor is fundamentally nondeterministic. Wherever the student is in the problem solution, the model is usually capable of generating a wide variety of next steps. The average number of next steps reflects the branching factor in the space of possible solution sequences that the model can generate. In the face of any significant branching factor and solutions of any significant length, it would not be possible to generate all of the possible solutions and ask which one matches the student. Model tracing works because we can use a prune-as-you-go strategy at the unit-task level (in part because the relative independence of unit tasks). One can use the student actions to determine which path the student is on. Often there will be a unique action associated with each next step in the tree of possible

(a)

(b)

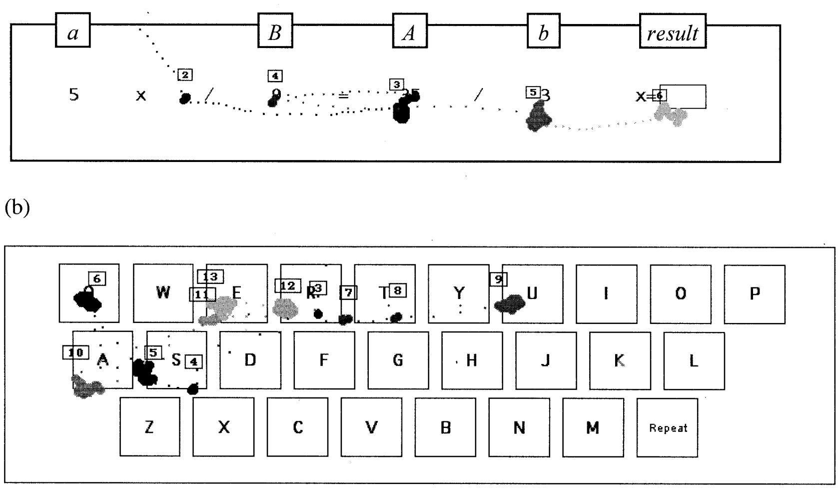

Fig. 11. Sample protocols from (a) the equation-solving task and (b) the eye-typing task from Salvucci (1999).

solutions. When this is not the case, there are usually only a few viable interpretations and often those ambiguities will disappear after a few actions. When this is also not the case, we can often take a high-probability guess or present a query to the students to determine their intentions. Therefore, the nondeterminism seldom becomes serious in our cognitive tutors. If none of these maneuvers work, then we essentially ignore the detail behind this ambiguity. For instance our algebra tutor handles the ambiguity illustrated in the contrast between Figs. 4a and 4b by essentially ignoring the distinction in solution processes and just crediting the student with knowing how to solve the problem in some way. While an immense amount of work goes into engineering the knowledge in the tutors, into designing their interactions, and into supporting them in the classroom, the fundamental reason for their success is because we can tame the exponential explosion in the solution paths at the unit-task level without greatly compromising instructional effectiveness.

6.2. Model tracing and eye movements

The challenge in applying the model-tracing methodology below the unit-task level is to find similarly effective methods of taming the exponential explosion in the model-tracing process. Salvucci's (1999, see also Salvucci & Anderson, in press) approach to tracing eye movements is an illustration of both the difficulties posed by the problem and a potential solution. The type of problem Salvucci's methods addressed can be illustrated by the two examples of fixation sequences in Fig. 11. The fixation sequence in Fig. 11a comes from an experiment where the student is solving the problem $5x/9 = 35/3$. Students in his experiment had solved a large number of problems of the form **ax/B = A/b** where a divided evenly into **A** and **b** divided evenly into **B**. Students knew they could solve them as **(A/a)*(B/b)**. The fixations in Fig. 11a are numbered 2 through 6 according to where they occurred (Fixation 1 is off the problem). There are substantial ambiguities in interpreting the intent of individual fixations. As an example, let us focus on interpreting Fixation 2. The student might be encoding the "/", which the fixation is close to. On the other hand the fixation sequences often began with a general orienting fixation to the equation and so it could be that. It could be a fixation in which the student was encoding the "5" (**a** term) in the periphery. Also, many students would calculate **B/b** first and so it could be a fixation in which the student is encoding the 9 (**B** term). However, in the overall context of the problem, this is probably a fixation on the **a** term, followed by fixation 3 on the **A** term, calculation of **A/a**, followed by fixations 4 and 5 on the **B** and then **b**, calculation of the **B/b**, and a final fixation 6 where the answer was typed.

This illustrates some of the fundamental problems that come at this level. A single action is often quite ambiguous because it could be associated with multiple intentions. This example also illustrates a case where, by the end, there is a best interpretation. In this case the ambiguity is largely because that the location of the eye fixation can be some distance from the item being encoded. However, there are other reasons for ambiguity. Consider the other example in part (b) of Fig. 11, where Salvucci is looking at the task of "eye typing." In eye-typing a person is supposed to look at the keys on the keyboard that they want to type. The person intends to look at the keys that type "square." Here we see multiple fixations on a single letter and spurious fixations (perhaps looking for letters). For instance, fixation 3 on the R is spurious followed by two fixations (4 and 5) on S, the first letter to be typed. Even if one could identify the first intended letter as S, there remains considerable ambiguity as to what word the person is intending to type.

Salvucci's solution to such problems was to use hidden Markov models (HMMs) in which there are states associated with various intentions. HMMs find the most probable interpretation of the sequence of fixation. In the case of equation solving his HMM model was able to find more accurate interpretations than human coders in much less time (participants were instructed on which sequence to use and so "the truth" was known). The time to interpret a sequence of eye fixations in one of his models is proportional to FN^2 where F is the number of fixations and N is the number of states. This proves quite viable in the equation-solving case, but it becomes intractable in the eye-typing case where the number of states is proportional to the number of words in the language. Thus, this solution is not generally feasible and Salvucci has explored various approximations such as recognizing word fragments.

6.3. Conclusions about model tracing below the unit-task level

Salvucci's example illustrates some of the general problems that we face when we go down to the level of primitive actions:

(1) There is high ambiguity that is not quickly resolved. This ambiguity is occurring at too small a grain size for it to be practical to be constantly querying the student as to what their intentions are.
(2) Physical parameters of the action like fixation location are only probabilistically related to the intention for the action.
(3) The are frequent slips and spurious actions.

When one goes down to the subsymbolic level one faces these difficulties and at least one more—which is that states are defined in terms of continuous variables like activations rather than discrete variables. There are Bayesian methods for estimating and updating probability distributions over such continuous variables (e.g., Berger, 1985). Such methods, applied to learning arithmetic facts, might enable one to infer the strength of encoding of such facts and inform selection of which facts to receive greater practice. Such a method could use data such as retrieval time to infer the student's state with great accuracy. However, I know of no such applications. So we do not have an existence proof at the subsymbolic level as we have with Salvucci's algorithm at the level of primitive actions. The computational details of doing this appear formidable.

There is another technical difficulty that we face at the lower levels, which is that such models, being more detailed, take longer to construct and we already invest a lot of time in developing models at the unit-task level. It is noteworthy that we did not employ a method like Salvucci's HMM methodology with a running model in the application using eye movements with the algebra tutor. This was in part because we wanted a quick demonstration without the cost of having to develop a complete model-tracing tutor for the whole task. Having that demonstration in hand it is perhaps worth trying to develop such a tutor and assessing the value added of such an effort. Applying HMM to ACT-R/PM comes down to having each possible goal chunk be a state of the HMM. So the methodology could apply. However, its efficiency and achievement gains are unknown.

Perhaps cognitive modeling of the type illustrated by the cognitive tutors is not the way to capitalize on finer-grain data. Perhaps one can get all the action needed from aggregate measures such as time spent fixating in regions and simple model-free maneuvers like triggers on the amount of time. Interestingly, with respect to the issue of training declarative items, like arithmetic facts, there are successful models like Atkinson (1972) for training of foreign vocabulary items that just use probability of recall and do not use more fine-grained data like 10 msec. latency effects. A challenge here would be to assess what the gain would be of a more detailed-modeling effort.

It is interesting to note that in many applications it is not necessary to know exactly what a person is doing at a fine-grained level. For instance, with respect to Salvucci's eye-movement data in Fig. 11, one does not have to accurately interpret each fixation. In Fig. 11a one only needs to infer the mathematical transformation intended and, in the case of Fig. 11b, the word intended. For many such applications it is enough to know the student's intentions at the unit-task level. However, information below this level can be critical to interpretation of that intention. This suggests spanning the seven orders of magnitude with two bridges—one that models the unit tasks for purposes of tracking instructional objectives and one that decomposes the unit task for purposes of interpreting the unit task. One might use a unit-task model like in the cognitive tutors in conjunction with something like Salvucci's HMM's. In other task domains the microstructure of the unit task can become part of the instructional objective. For instance, in Lee's air-traffic controller model, high levels of skill depended on parallelizing different streams of processing (see Fig. 2). However, in this case too, it may be possible to approach applications with two bridges—one to the unit task and one that decomposes the unit task. Indeed, Lee and Anderson (2001) used this two-pass approach in analyzing the data from their research. So, perhaps the conclusion is that, while it may be possible to go all the way from 10 msec. effects to 100 hr. learning outcomes, it will be necessary to use multiple different sorts of bridges in spanning this distance.

Acknowledgments

This research was supported by the AFOSR Grant F49620-99-1-0086; NSF grant ROLE: REC-0087396, and ONR Grant N00014-96-1-0491. I would like to thank Gwen Campbell, Al Corbett, Christian Lebiere, Marsha Lovett, and Alex Petrov for their comments on an earlier draft of this paper.

Notes

1. Bloom (1984) reported that students who had private human tutors performed 2 standard deviations better than control students from regular classrooms. Other studies have obtained smaller effects (Cohen, Kulik & Kulik, 1982).

References

Ackerman, P. L., & Kanfer, R. (1994). Kanfer-Ackerman air traffic controller task CD-ROM database, data collection program, and playback program. *Office of Naval Research, Cognitive Science Program*.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 5,* 451–474.

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP Tutor. *Cognitive Science, 13,* 467–506.

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

Anderson, J. R., & Gluck, K. (2001). What role do cognitive architectures play in intelligent tutoring systems? In D. Klahr & S. M. Carver (Eds.), *Cognition & instruction: twenty-five years of progress* (pp. 227–262). Erlbaum.

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

Anderson, J. R., & Reder, L. M. (1999). The fan effect: new results and new theories. *Journal of Experimental Psychology: General, 128,* 186–197.

Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology, 96,* 124–129.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analyses.* New York: Springer-Verlag.

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.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learning: brain, mind, experience, and school.* Washington, DC: National Academy Press.

Bruer, J. T. (1998). Education and the brain: a bridge too far. *Educational Researcher, 26,* 4–16.

Byrne, M. D., & Anderson, J. R. (in press). Serial modules in parallel: The psychological refractory period and perfect time-sharing. To appear in *Psychological Review*.

Campbell, J. I. D. (1995). Mechanisms of simple addition and multiplication: a modified network-interference theory and simulation. *Mathematical Cognition, 1,* 121–164.

Card, S., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Erlbaum.

Chong, R. S. (1999). EPIC-Soar and the acquisition of executive process knowledge. Talk presented in a symposium entitled "Integration of Perception, Cognition and Action." In *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society*.

Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Education outcomes of tutoring: a meta-analysis of finds. *American Educational Research Journal, 19,* 237–248.

Corbett, A. T., Anderson, J. R., & O'Brien, A. T. (1993). The predictive validity of student modeling in the ACT Programming Tutor. In P. Brna, S. Ohlsson & H. Pain (Eds.). Artificial Intelligence and Education, 1993: The Proceedings of AI-ED 93. Charlottesville, VA: AACE.

Corbett, A. T., Koedinger, K. R., & Hadley, W. H. (in press). Cognitive Tutors: From the research classroom to all classrooms. In P. S. Goodman (Ed.), *Technology enhanced learning: opportunities for change*. Mahwah, NJ: Lawrence Erlbaum Associates.

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: impact of learning rate, achievement and attitudes. Proceedings of ACM CHI-2001 Conference on Human Factors in Computing Systems, 245–252.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction, 4, 253–278.

Corbett, A. T., & Knapp, S. (1996). Plan scaffolding: impact on the process and product of learning. In C. Frasson, G, Gauthier & A. Lesgold (Eds.), Intelligent tutoring systems: third international conference, ITS 96. New York: Springer.

Corbett, A. T., & Trask, H. (2000). Instructional interventions in computer-based tutoring: differential impact on learning time and accuracy. Proceedings of ACM CHI 2000 Conference on Human Factors in Computing Systems, 97–104.

Corbett, A. T. (2001). Cognitive computer tutors: solving the two-sigma problem. User Modeling: Proceedings of the Eighth International Conference, UM 2001, 137–147.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Gluck, K. A. (1999). *Eye movements and algebra tutoring*. Doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Graesser, A., Wiemer-Hastings, K., Wiemer-Hastings, P., Krenaz, R., & the Tutoring Research Group. (2000). AutoTutor. A simulation of a human tutor. *Journal of Cognitive Systems Research, 1,* 35–51.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: validating a GOMS analysis for predicting and explaining real-world task performance. *Human-Computer Interaction, 8,* 237–309.

Haverty, L. (1999). The importance of basic number knowledge to advanced mathematical problem solving. Doctoral Dissertation, Carnegie Mellon University, Pittsburgh, PA.

Haverty, L. A., Koedinger, K. R., Klahr, D., & Alibali, M. W. (2000). Solving induction problems in mathematics: not-so-trivial pursuit. *Cognitive Science, 24,* 249–298.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Jones, R. M., Laird, J. E., Nielsen, P. E., Coulter, K. J., Kenny, P., & Koss, F. V. (1999). Automated intelligent pilots for combat flight simulation. *AI Magazine, 20,* 27–41.

Koedinger. (1999, September). Coding the human memome: cognitive methods for intelligent tutor design. Invited speaker, Institute for Communicating and Collaborative Systems, and Human Communication Research Centre seminars. University of Edinburgh.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8,* 30–43.

Koedinger, K. R., & MacClaren, B. A. (1997). Implicit strategies and errors in an improved model of early algebra problem solving. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 382–387). Hillsdale, NJ: Erlbaum.

Lebiere, C. (1998). The dynamics of cognition: an ACT-R model of cognitive arithmetic. Ph.D. Dissertation. CMU Computer Science Dept Technical Report CMU-CS-98-186. Pittsburgh, PA.

Lee, F. J. (2000). *Does learning of a complex task have to be complex? A study in learning decomposition*. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA.

Lee, F. J., & Anderson, J. R. (2001). Does learning of a complex task have to be complex? A study in learning decomposition. *Cognitive Psychology, 42,* 267–316.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance. Part 1. Basic mechanisms. *Psychological Review, 104,* 2–65.

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance. Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review, 104,* 749–791.

Murdock, B. B., Jr. (1993). TODAM2: a model for the storage and retrieval of item, associative, and serial order information. *Psychological Review, 100,* 187–203.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics, Inc. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics, Inc.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Cambridge University Press.

Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review, 5,* 345–369.

Pew, R. W., & Mavor, A. S. (1998). *Modeling human and organizational behavior: application to military simulations*. Washington, DC: National Academy Press.

Pirolli, P. L., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 11,* 136–153.

Rumelhart, D. E., & McClelland, J. L. (1985). Levels indeed! A response to Broadbent. *Journal of Experimental Psychology: General, 114,* 193–197.

Salvucci, D. D. (1999). *Mapping eye movements to cognitive processes*. Doctoral Dissertation, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Siegler, R. S. (1996). *Emerging minds: the process of change in children's thinking*. New York: Oxford University Press.

Shepard, L. A. (1991). Psychometricians' beliefs about learning. *Educational Researcher, 20,* 2–16.

VanLehn, K., Niu, Z., Siler, S., & Gertner, A. (1998). Student modeling from conventional test data: a Bayesian approach without priors. In B. P. Goettle, H. M. Halff, C. L. Redfield, & V. J. Shute (Eds.), *Intelligent Tutoring Systems: 4th International Conference, ITS98* (pp. 434–443). Berlin: Springer Verlag.

Zachary, W., Cannon-Bowers, J., Burns, J., Bilazarian, P., & Krecker, D. (1998). An advanced embedded training system (AETS) for tactical team training. In B. P. Goettl, H. M. Halff, C. L. Redfield, & V. J. Shute (Eds.), *Intelligent tutoring systems: Proceedings of the 4th International Conference, ITS '98* (pp. 544–553). Berlin: Springer Verlag.

Zachary, W., LeMentec, J.-C., & Ryder, J. (1996). Interface agents in complex systems. In C. N. Ntuen and E. H. Park (Eds.), *Human interaction with complex systems: conceptual principles and design practice*. Kluwer Academic Publishers.