Cognitive Science 29 (2005) 313–341

Human Symbol Manipulation Within an Integrated Cognitive Architecture

John R. Anderson

Psychology Department, Carnegie Mellon University

Received 28 April 2004; received in revised form 12 September 2004; accepted 20 September 2004

Abstract

This article describes the Adaptive Control of Thought–Rational (ACT–R) cognitive architecture (Anderson et al., 2004; Anderson & Lebiere, 1998) and its detailed application to the learning of algebraic symbol manipulation. The theory is applied to modeling the data from a study by Qin, Anderson, Silk, Stenger, & Carter (2004) in which children learn to solve linear equations and perfect their skills over a 6-day period. Functional MRI data show that: (a) a motor region tracks the output of equation solutions, (b) a prefrontal region tracks the retrieval of declarative information, (c) a parietal region tracks the transformation of mental representations of the equation, (d) an anterior cingulate region tracks the setting of goal information to control the information flow, and (e) a caudate region tracks the firing of productions in the ACT–R model. The article concludes with an architectural comparison of the competence children display in this task and the competence that monkeys have shown in tasks that require manipulations of sequences of elements.

Keywords: mathematics; cognitive architecture; education; learning; problem solving; comparative psychology; brain imaging

1. Introduction

Adaptive Control of Thought–Rational (ACT–R; Anderson et al., in press; Anderson & Lebiere, 1998) is most fundamentally a theory of central cognition. One function of this article is to present an overview of that theory. This article also presents an illustrative application of the theory to algebra equation solving. Algebra equation solving is a uniquely human cognitive activity and provides a relatively well-contained opportunity to address the question of what is unique about human cognition. I compare the requirements of this task with the requirements of other sequential tasks that nonhuman primates have been shown capable of performing. The article emerges with a tentative proposal for what is unique about human cognition, from the

Requests for reprints should be sent to John R. Anderson, Psychology Department, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: ja@cmu.edu

framework of the ACT–R theory. Such comparative analyses relate to issues of brain realization, and this article also describes the preliminary mapping of components of the ACT–R theory onto brain regions. This mapping has enabled the use of functional MRI (fMRI) data to inform theory development.

2. The ACT–R architecture

According to the ACT–R theory, cognition emerges through the interaction of a number of independent modules. Fig. 1 illustrates the modules relevant to algebra equation solving:

1. A visual module that might hold the representation of an equation such as $3x - 5 = 7$.
2. A problem state module (sometimes called an imaginal module) that holds a current mental representation of the problem. For instance, the student might have converted the original equation into $3x = 12$.
3. A control module (sometimes called a goal module) that keeps track of one's current intentions in solving the problem—for instance, one might be trying to apply the unwind strategy described later.
4. A declarative module that retrieves critical information from declarative memory such as that $7 + 5 = 12$.
5. A manual module that programs the output such as $x = 4$.

Each of these modules is capable of massively parallel computation to achieve its objectives. For instance, the visual module is processing the entire visual field and the declarative module searches through large databases. However, each of these modules suffers a serial bottleneck such that only a little information can be put into a buffer associated with the module—a single object is perceived, a single problem state represented, a single control state

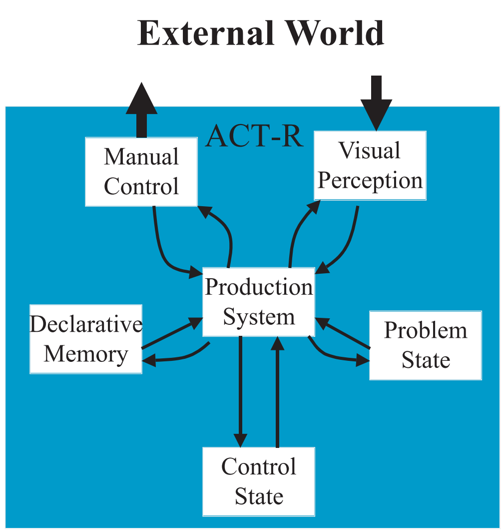

Fig. 1. The interconnections among modules in ACT–R 5.0.

maintained, a single fact retrieved, or a single program for hand movement executed. Formally, each buffer can only hold what is called a *chunk* in ACT–R, which is a structured unit bundling a small amount of information. ACT–R does not have a formal concept of a working memory, but the current state of the buffers constitutes an effective working memory. Indeed, there is considerable similarity between these buffers and Baddeley's (1986) working memory "slave" systems.

Communication among these modules is achieved via a procedural module (production system in Fig. 1). The procedural module can respond to information in the buffers of other modules and put information into these buffers. The response tendencies of the central procedural module are represented in ACT–R by production rules. For instance, the following might be a production rule for transforming an equation:

IF the goal is to solve the equation
  and the equation is of the form Expression – Number1 = Number2
  and Number1 + Number2 is Number3 has been retrieved
THEN transform the equation to Expression = Number3

This production responds when the control chunk encodes the intention to solve an equation, as shown in the first line; when the problem state chunk represents an equation of the appropriate type—second line—for example, $3(x - 2) - 4 = 5$; when a chunk encoding an arithmetic fact has been retrieved from memory—see third line—in this case $4 + 5 = 9$; and appropriately changes the problem representation chunk—see fourth line—in this case to $3(x - 2) = 9$.

The procedural module is also capable of massive parallelism in sorting out which of its many competing rules to fire, but like the other modules it has a serial bottleneck in that it can only fire a single rule at a time. Because it is responsible for communication among the other modules, the production system comprises the central bottleneck (Pashler, 1994) in the ACT–R theory. Therefore, cognition can be slowed when there are simultaneous demands to process information in distinct modules. As already noted, the other modules themselves can also be bottlenecks. All of the bottlenecks are in the communication among modules; within modules things are massively parallel. (Fig. 3, later in the paper, illustrates in some considerable detail how this parallelism and seriality mix.) Documenting the accuracy of this characterization of human cognition has been one of the preoccupations of research on ACT–R (for instance, Anderson, Taatgen, & Byrne, 2004).

In addition to the overall flow of control, major concerns of ACT–R involve the "internal components" (declarative memory, production memory, control state, and problem state). With respect to peripheral modules (and ACT–R has more than just the visual and manual modules represented here) we have been content to implement approximations that capture the major results documented in the literature. Indeed, much of ACT–R's perceptual–motor system is a reimplementation of Executive-Process Interactive Control's (EPIC; Meyer & Kieras, 1997) perceptual–motor system. Following EPIC's lead we have found that we cannot understand central cognition unless we have reasonably accurate models of its interface with the external world. For a substantial fraction of the ACT–R community, particularly those concerned with human–computer interaction issues, this perceptual–motor system is critical.

The central declarative memory and the procedural production system have substantial similarities. Selection among different declarative memories proceeds by means of computations on continuously valued activation quantities, and selection among productions proceeds by means of continuously valued utility computations. These computations give ACT–R many of the graded properties considered virtues of connectionist systems and avoid the sharp edges associated with symbolic systems. Indeed, there was a connectionist implementation of ACT–R (Lebiere & Anderson, 1993). However, the theory also has an important symbolic level that plays a critical role in accounting for such things as the acquisition of competence in algebra. The distinction between the symbolic level of facts and productions and the subsymbolic level of activations and utilities is critical to the ACT–R theory.

Until recently, the problem state and the control state were merged into a single goal system. There have been a number of developments to improve ACT–R's goal system (Altmann & Trafton, 2002; Anderson & Douglass, 2001), and this is another development. There were two reasons for choosing to separate control state (*goal buffer*) and problem state knowledge (*imaginal buffer*). First (and this was the source of the idea to separate the two aspects), our imaging data indicated that the parietal region of the brain reflected changes to problem-state information, but the anterior cingulate reflected control-state changes. Later, this article elaborates on the neural basis for this distinction. Second, the distinction offered a solution to a number of nagging problems we had with the existing system that merged the two types of knowledge. One problem was that our goal chunks often seemed too large, violating the spirit of the claim that chunks were supposed to only contain a little information. This is because they contained both problem-state information and control-state information, which could both involve a number of elements.[1] Also, the control information was getting in the way of storing useful information about the problem solution in declarative memory. For instance, arithmetic facts such as $3 + 4 = 7$ might represent the outcome of a counting process or of an effort to comprehend a sentence. Because the control information was separate and would be different in the case of these two sources for the same arithmetic fact, we effectively were creating parallel memories storing the same essential information. Now, with control and problem state separated, the differences between the counting and comprehension can be represented in different control chunks, whereas the common result would be represented identically in a single problem solution chunk. By factoring control information away (in what we are now calling the *goal module*), one can accumulate abstract memories of the information achieved in the problem state.

3. Algebra equation manipulation

With this brief overview of the ACT–R theory, let us turn to algebra equation solving, which is a domain that offers special opportunities for understanding the nature of human intelligence. Although there are now a number of demonstrations of basic arithmetic competence in other primates, it would be generally conceded that algebra is a uniquely human capability. Algebraic expressions and the operations that can be performed on them represent a domain of substantial cognitive complexity, but unlike many human accomplishments (such as natural language) it is a domain that can be tractably characterized and studied. In the first year of high school algebra with an investment of typically less than 200 hr, suitably prepared students can learn to take a sequence of symbols such as

$$4 = x * [(x + 8) + 7] \tag{1}$$

and rewrite it into the form

$$x^2 + 15x - 4 = 0 \tag{2}$$

in preparation for using the quadratic formula, which they can also apply. Interestingly, students come to prefer writing this out directly without writing any intermediate expressions—despite the urging of some mathematics teachers not to do so. Children have such a facility at manipulating these representations in their heads that it is easier to do the operations mentally than to write out intermediate expressions.

Before continuing this discussion, it is important to make a couple of caveats. First, I am not implying that the ability to engage in such symbol manipulation is the most important part of first-year high school algebra, nor that students spend the majority of their 200 hr mastering this. The goal of the algebra course is to relate multiple representations of mathematical relations (including graphical and verbal) to enable flexible problem solving (Koedinger, Anderson, Hadley, & Mark, 1997). Just being able to engage in such manipulations would be a rather useless skill unless students could relate such manipulations to other things, especially real-world problems. Nonetheless, even though such algebraic manipulations are a small part of the complete picture, they already establish a high level of complexity to human cognition—a level of complexity that is all the more remarkable given that it is mastered in such a brief period of time.

Second, being able to achieve this competence depends critically on what has already been established in earlier grades—in particular, knowledge of arithmetic facts, fractional representations, and how to parse expressions such as the previous example. Students struggle if they arrive in algebra without these prerequisites. Thus, the modeling task in this article is to account for the acquisition and performance of algebraic transformations, assuming the background of such knowledge.

As the previous example illustrates, algebra manipulation is basically a string manipulation task in which one string of symbols is transformed into another. The final section of this article considers what might be uniquely human about this string manipulation task versus other sequential skills that nonhuman primates can do.

The experiment to be modeled in detail looks at the learning of a particularly reduced version of algebra symbol manipulation—solving of simple linear equations—converting expressions such as

$$3x - 5 = 7 \tag{3}$$

$$\text{into } x = 4. \tag{4}$$

These sorts of equations can be solved by what has been called the *unwind* strategy. Such equations have a number on one side and an expression with a single occurrence of the variable on the other. The variable can be isolated by inverting each operation (in the previous example the "– 5" is eliminated by adding 5, and the "3*" by dividing by 3)—peeling away the layers of the expression until the variable is exposed. Many equations do not immediately start out in this form but can be simplified so that they are in the appropriate form. The model in this article assumes that the equations are already in a form to which the unwind strategy can immediately apply. The earliest instruction on equation solving tends to focus on such problems and teaches students the justifications for these transformations as well as providing practice on how to perform them, although students are typically not taught to think of this as a general unwind strategy but rather as a series of more specific operations. We have found, however, that beginning students are quite capable of understanding the general unwind principle and its justification and can use it with its full generality.

In the experiment to be modeled in detail (Qin et al., 2004) 10 students ages 11 to 14 spent 6 days practicing solving such equations. The first day (Day 0) they were given private tutoring on this class of equations, using the unwind principle, and practiced paper and pencil solutions of such problems with a private human tutor. On the remaining 5 days they practiced on a computer the solution of three classes of equations:

$$\text{0-step: e.g., } 1x + 0 = 4 \tag{5}$$

$$\text{1-step: e.g., } 3x + 0 = 12,\ 1x + 8 = 12 \tag{6}$$

$$\text{2-step: e.g., } 7x + 1 = 29 \tag{7}$$

Each day they went through 10 computer-administered blocks of such equations. Each block consisted of 16 trials with four instances of the four possible types of equations (there are two subtypes for the 1-step equations). Fig. 2 presents their latency and the predictions of a model that will now be described.

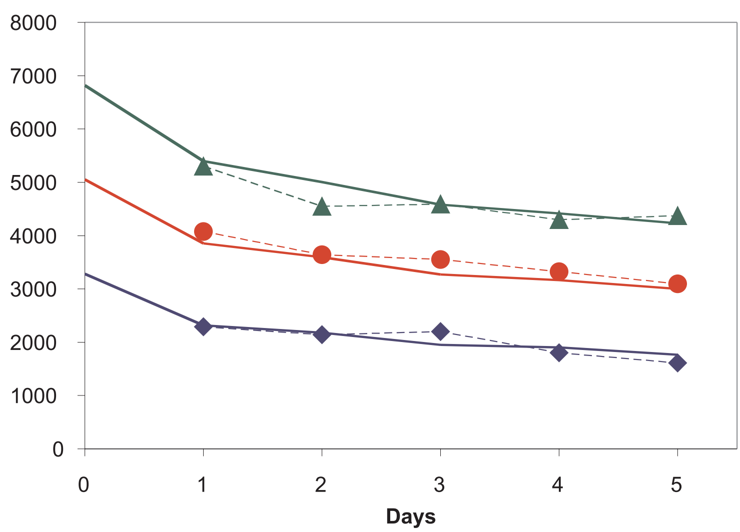

Fig. 2. Mean solution times (and predictions of the ACT–R model) for the three types of equations as a function of delay. Although the data were not collected, the predicted times are presented for the practice session of the experiment (Day 0).

4. The ACT–R model

The ACT–R model begins with a set of declarative instructions, given in Table 1, which encode the unwind strategy. To illustrate how these instructions[2] apply to example equations, first consider a simple 0-step equation such as:

$$1*x + 0 = 2 \tag{8}$$

These instructions imply a sequence of operations that can be summarized:

Instruction 1a: Create image " = 2"
Instruction 2b: Unwind right "1*x + 0"
Instruction 3a: Focus on "1*x" and unwind it
Instruction 2c: Unwind left "1*x"
Instruction 4a: Focus on "x" and unwind it
Instruction 2a: The answer is 2

While for a 2-step equation such as

$$7*x + 3 = 38 \tag{9}$$

they imply a sequence of operations that can be summarized:

Instruction 1a: Create image " = 38"
Instruction 2b: Unwind right "7*x + 3"
Instruction 3b: Change image to " = 38 – 3" and then to " = 35" and focus on "7*x" and unwind it.
Instruction 2c: Unwind left "7*x"
Instruction 4b: Change image to " = 35/7", and then to " = 5" and focus on x and unwind it.
Instruction 2a: The answer is 5

Fig. 2 shows that ACT–R is able to reproduce the speedup seen in the participants. The key to understanding this speedup in the ACT–R model is to understand how the previous instructions were interpreted. These instructions are encoded as declarative structures, and ACT–R has general interpretative productions for converting these instructions to behavior. For instance, there is a production rule that retrieves the next step of an instruction:

IF one has retrieved an instruction for achieving a goal
THEN retrieve the first step of that instruction

There are also productions for retrieving particular arithmetic facts such as

IF one is evaluating the expression "*a* operator *b*"
THEN try to retrieve a fact of the form "*a* operator *b* = ?"

Using such general instruction-following productions is laborious and accounts for the slow initial performance of the task.

Although multiple types of learning are occurring in this experiment, it is mainly production compilation that is accounting for the speedup (see Taatgen, 2005, this issue; Taatgen & Anderson, 2002). This is a process by which new production rules are learned that collapse what was originally done by multiple production rules. In this situation the initial instruction-following productions are compiled over time to produce productions to embody procedures that efficiently solve equations. For instance, the following production rule is acquired:

IF the goal is to unwind an expression
  and the expression is of the form "subexpression + 0"
THEN focus on the subexpression

Fig. 3a illustrates a typical trial at the beginning of Day 1, and Fig. 3b illustrates a typical trial at the end of Day 5. In both cases the model is solving the 2-step equation, $7*x + 3 = 38$. The figure illustrates when the various modules were active during the solution of the equation and what they were doing. The Day 1 trial (Fig. 3a) takes 6.1 sec and the Day 5 trial (Fig. 3b) takes 4.1 sec. However, these do not reflect the extremes of the learning curve according to ACT–R. The very first trial on Day 0 takes 8.4 sec in the model. With an infinite amount of practice, the model would take 1.7 sec during which it would only read the equation and type the answer, having compiled the answer into production rules for that problem. Still, the contrast between parts a and b of Fig. 3 gives a sense for what is happening over the course of learning. It is worth emphasizing a number of general features of the activity in the figure before discussing the detail of what is happening in individual buffers:

Multiple modules can be active simultaneously. For instance, early on in Fig. 3 there is a point where the goal module is noting that it is implementing the unwind strategy, an image of the right-hand side of the equation (" = 38") is being encoded in the imaginal buffer, the next step in the unwind strategy is being retrieved, and the visual system is encoding the left-hand side of the equation. Certain of these activities tend to be on the critical path because they are taking longer than the other processes, and further processing has to wait for them to complete. In these cases, the times of the other operations have no effect on total time. For instance, often the visual encoding of the equation is holding up other operations and the durations of these other operations do not matter.

Table 1
English rendition of instructions given to ACT–R model for equation solving

1. To solve an equation, encode it and
   a. If the right side is a number, then imagine that number as the result, and focus on the left side and unwind it.
   b. If the left side is a number, then imagine that number as the result, and focus on the right side and unwind it.
2. To unwind an expression
   a. If the expression is the variable, then the result is the answer.
   b. If a number is on the right unwind-right.
   c. If a number is on the left unwind-left.
3. To unwind-right, encode the expression (of the form "subexpression operator number") and
   a. If the operator is + or – and the number is 0, then focus on the subexpression and unwind it.
   b. Otherwise invert the operator, imagine it as the operator in the result, imagine the number of the expression as the second argument in the result, evaluate the result, and then focus on the subexpression and unwind it.
4. To unwind-left encode the expression (of the form "number operator subexpression") and
   a. If the operator is * and number 1 then focus on the subexpression and unwind it.
   b. Otherwise check that the operator is symmetric, invert the operator, imagine it as the operator in the result, imagine the number as the second argument in the result, evaluate the result, and then focus on the subexpression and unwind it.

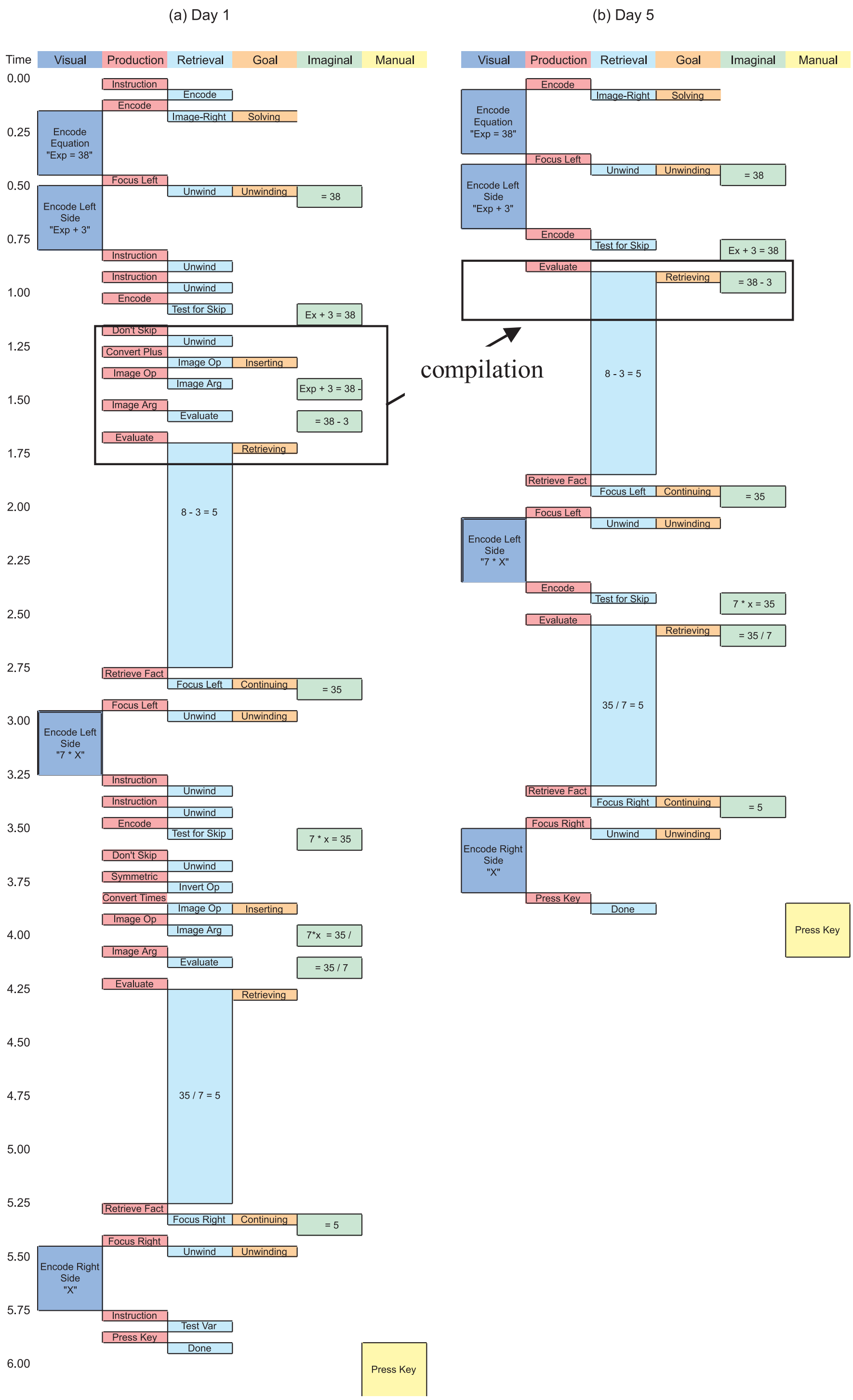

Fig. 3. Comparison of the module activity in ACT–R during the solution of a two-step equation on Day 1 (a) with a two-step equation on Day 5 (b). In both cases the equation being solved is $7{*}x + 3 = 38$.

Much of the speedup in processing is driven by collapsing multiple steps into single steps. A particularly dramatic instance of this is noted in Fig. 3 where five production firings, five retrievals, two control settings of the goal, and two imaginal transformations are compressed into one production, one retrieval, one control setting, and one imaginal transformation.[3] Production compilation can compress these internal operations without limit. What it cannot collapse are the external operations such as visual encodings or manual operations. These external operations define the bounds of the compilation. Although the example in Fig. 3 shows multiple productions being collapsed, the actual learning process proceeds slowly in ACT–R and takes all 5 days to achieve the transformation in Fig. 3. Given enough practice ACT–R would collapse all equation solving simply into a series of visual encodings and manual operations, and there would be no effect of equation complexity (nor any real thought occurring). However, to do so ACT–R would have to essentially build into production rules the capacity to recognize each possible equation and produce its solution. The combinatorics of this are so overwhelming (so many different possible equations) that it would never happen in the normal course of learning to solve equations.

A second, lesser source of speedup is the reduction of retrieval times. This reflects an increase in the base-level activation of the facts used in this experiment and as such it is an example of subsymbolic activation learning. This subsymbolic learning is a relatively minor contributor to the learning in Fig. 2 for two reasons. First, the basic instructions get used over and over again and are already strongly encoded during Day 0, and there is not that much room for further speedup. Second, the arithmetic facts do not repeat very often over the course of the experiment and are getting little practice. In other situations, subsymbolic activation processes can be a major player in performance. However, over the period of time studied in this experiment, the major learning is happening at the symbolic level in terms of creating new production rules.

Now let us consider what is happening in each of the modules as the model goes from Day 1 to Day 5:

1. Visual: On both days four encoding operations take place, which each take 300 msec. Each encoding has the resolution to pick up two terms in the expression. Therefore, the first encodes *Exp* = 38, where *Exp* denotes what cannot be analyzed. The second analyzes this into *Exp* + 3, the third into 5 * *Exp,* and the final encodes the *x*.
2. Procedural: The number of productions fired reduces substantially over the 5 days—from 26 to 11 in the example in Fig. 3. This reflects the compilation of productions into ones that do more work. Many of the productions, even on Day 1, were compiled from the originals that were used on Day 0. Each production takes 50 msec according to the ACT–R theory.[4]
3. Retrieval: Most of the retrievals in Fig. 3 involve retrieval of steps of instruction, and these decrease dramatically from 24 to 9 in the specific example. These instructions are illustrated as taking about 50 msec, but this is only approximate. There are also two long retrievals of arithmetic facts. As noted earlier, all of these retrievals are speeding up with time, but the speedup effect is most apparent for the arithmetic facts.
4. Goal: The goal holds information about control state. Different points in the problem solving can have identical patterns in the other buffers, and it is the responsibility of the goal buffer to keep track of what to do next. As I elaborate later, the major control issue is keeping track of when it is time to retrieve and when it is time to unwind. The number of control settings

changes relatively little over the course of the experiment, decreasing from 10 to 8 in the specific example.

5. Imaginal: The imaginal, or problem state, buffer holds a partial representation of the equation. The number of changes to this buffer shows a small drop in Fig. 3 from nine to seven. The reduction reflects cases when two imaginal operations are collapsed into a single one.

6. Manual: The manual programming does not change over the course of the experiment. A single final finger press is needed that takes 250 msec to program and execute.

5. Use of brain imaging to provide converging data

The complexity of the picture in Fig. 3 provides a striking contrast to the simplicity of the data in Fig. 2. It is hard to justify all those boxes and assumptions on the basis of three simple learning curves. Reflecting this, past theories that we have developed (e.g., Anderson, Reder, & Lebiere, 1996) have often been cast much simpler, not because we thought things were that simple but because we were keenly aware of the assumptions-to-data ratio. However, in honest moments with ourselves we knew more was going on. Indeed, Fig. 3 probably underrepresents the true complexity. Although for many purposes ignoring this complexity is fine, it left us with a picture of mathematics learning that may have failed some of our theory-based efforts to improve mathematics learning. It is better to have a complete theory and determine which details are not relevant and can be ignored, rather than simply never considering the details in the first place.

Brain-imaging data allow us to track these individual components in more detail. Although it still does not provide all of the converging evidence one would want, it goes a long way to justifying the detail in Fig. 3. The study reported in Fig. 2 was actually performed in an fMRI scanner on Days 1 and 5. The trials took 21.6 sec on all days, to facilitate analysis on the scanner days. On the scanner days, an image of much of the brain was taken each 1.2 sec. During the first 1.2 sec, children looked at a fixation point. Then they had up to 10 scans or 12 sec to complete solving the equation, and they pressed a key giving an answer as soon as they had solved the equation. These 10 scans were followed by an additional 7 scans or 8.4 sec to let the hemodynamic response to the equation go down to baseline. Students gave their answer in a data glove in which they pressed one of the 5 fingers on their right hand to indicate an answer of 1 to 5 (all problems had these numbers as answers).

5.1. Regions of interest

We have now completed a large number of fMRI studies of many aspects of higher level cognition (Anderson, Qin, Sohn, Stenger, & Carter, 2003; Anderson, Qin, Stenger, & Carter, 2004; Qin et al., 2003; Sohn, Goode, Stenger, Carter, & Anderson, 2003; Sohn et al., in press), and based on the patterns over these experiments we have made the following associations between a number of brain regions and modules in ACT–R. In this article we are concerned with five brain regions and their ACT–R associations:

1. Caudate (procedural): Centered at Talairach coordinates $x = -5$, $y = 9$, $z = 2$, this is a subcortical structure.

2. Prefrontal (retrieval): Centered at $x = -40$, $y = 21$, $z = 21$, this includes parts of Brodmann Areas 45 and 46 around the inferior frontal sulcus.
3. Anterior cingulate (goal): Centered at $x = -5$, $y = 10$, $z = 38$, this includes parts of Brodmann Areas 24 and 32.
4. Parietal (problem state or imaginal): Centered at $x = -23$, $y = -64$, $z = 34$, this includes parts of Brodmann Areas 39 and 40 at the border of the intraparietal sulcus.
5. Motor (manual): Centered at $x = -37$, $y = -25$, $z = 47$, this includes parts of Brodmann Areas 3 and 4 at the central sulcus.

It is important to emphasize that we had these regions defined and associated with ACT–R modules before performing this experiment. Thus, the research we are reporting here contrasts with the more typical practice of performing exploratory analyses to find what regions give significant changes in activation and trying to interpret their significance after the fact. As such this confirmatory approach is not subject to the issue of trying to correct for false alarms that haunts the exploratory approach.

It is worth briefly noting the past reports in which we identified these regions and associated them with the particular modules in the ACT–R theory. The original publication in the series (Anderson et al., 2003) was focused on looking for the brain correlates of the Anderson, et al. (1996) ACT–R model for algebra equation solving in adults. The two experiments in that article performed exploratory analyses that converged on regions close to the prefrontal, parietal, and manual regions defined previously. Based on that study, we defined these previously mentioned regions and conducted a number of studies focused on verifying their properties. Across these studies we maintained the exact same Talairach definition of these regions. The studies by Anderson et al. (2004), Anderson et al. (in press), & Qin et al. (2003) were focused on better separating of parietal and prefrontal activities (which are often highly correlated) and confirming that the prefrontal was more associated with retrieval and the parietal was more associated with representational changes. That research also showed that these regions responded to the number of retrievals and representational changes and not to the duration of time that the retrieval products or representations were held during the problem solving. Interestingly, we found that the motor region was involved in rehearsal of the results to bridge delay periods, not the parietal or prefrontal regions. The research by Sohn et al. (2003) and Sohn et al. (2005) used the fan effect to confirm that the prefrontal and not the parietal region responded to time to perform an individual retrieval.

Although all of these published articles only reported on the prefrontal, parietal, and motor regions, we did collect data on predefined anterior cingulate cortex (ACC) and caudate regions in all of these studies. Our interest in the ACC was forced by the fact that exploratory studies kept revealing that it showed strong effects from the experimental manipulations. The caudate region of the basal ganglia area only sometimes showed significant effects in the exploratory analysis, and we will see it suffers from a rather poor signal-to-noise ratio. However, our interest in it was driven by our association of the basal ganglia with the production system (Anderson et al., in press) and by other published reports associating the basal ganglia with procedural memory (Ashby & Waldron, 2000; Hikosaka eti al., 1999; Poldrack, Prabakharan, Seger, & Gabrieli, 1999; Saint-Cyr, Taylor, & Lang, 1988).

Although we have our own independent evidence for the associations of these regions with the ascribed ACT–R functions, it certainly is the case that the ascriptions are consistent with other ideas in the literature. As noted previously, our interest in the caudate was basically driven by other research reports. Others (Dehaene, Piazza, Pinel, & Cohen, 2003; Reichle, Carpenter, & Just, 2000) have found a parietal region that reflects imagery and visual representation, and a number of researchers (Buckner, Kelley, & Petersen, 1999; Cabeza, Dolcos, Graham, & Nyberg, 2002; Donaldson, Petersen, Ollinger, & Buckner, 2001; Fletcher & Henson, 2001; Lepage, Ghaffar, Nyberg, & Tulving, 2000; Wagner, Maril, Bjork, & Schacter, 2001; Wagner, Paré Blagoev, Clark, & Poldrack, 2001) have found a strong memory response in the vicinity of our prefrontal region.

The situation in the literature with respect to the ACC is complex. A number of theorists have postulated that it is involved in controlling cognition, much as is being proposed here. For instance, Posner & Dehaene (1994) have described the ACC as "involved in the attentional recruitment and control of brain areas to perform complex tasks" (p. 76). D'Esposito et al. (1995) have identified it with Baddeley's (1986) central executive and Posner & DiGirolamo (1998) have related it to Norman & Shallice's (1986) SAS. However, there are other theories of the ACC. One theory relates it to error detection. This is supported by the error-related negativity in event-related potentials that has been observed when errors are made in speeded response tasks (e.g., Falkenstein, Hohnsbein, & Hoorman, 1995). Dehaene, Posner, and Tucker (1994) were able to localize the error-related negativity as residing within the ACC. However, ACC activity occurs in many more situations than just when there are errors, and another interpretation of its activity is that it is just a reflection of task difficulty as indexed by errors or reaction time (Paus, Koski, Caramanos, & Westbury. 1998). On the other hand, it does not always respond to task-difficulty factors that affect latency (we will see in our experiment that it reflects number of transformations but not practice). Carter, et al. (2000) argued that the real function of the ACC is monitoring for conflict among potential responses and that other regions of the cortex actually respond to the conflict once detected. MacDonald, Cohen, Stenger, and Carter (2000) found that in a Stroop task, when participants are warned that it will be a difficult color trial, there is greater activation in the prefrontal region in preparation for the task. In contrast, when the actual Stroop task is presented the ACC responds to a difficult color trial. Thus, they argue that, unlike the Posner and Dehaene proposal, the prefrontal cortex, and not the ACC, is responsible for control and that the ACC, rather, monitors for conflict, such as that which occurs in the Stroop task. This conflict is often interpreted as conflict among competing responses, and this interpretation is applicable to the Stroop task. However, in our more complex tasks ACC activity reflects transformations and retrievals that do not involve any overt responses or competitions among responses. Sohn, Albert, Jung, Carter, & Anderson (2004) reported a study in which the ACC is clearly involved in controlling attention and not just monitoring conflict. Unlike the MacDonald et al. (2000) experiment, it is highly active in preparation for an upcoming cognitive task, and its activation varies with the anticipated difficulty of that task. At the end of this article we elaborate on the function for the ACC in the ACT–R model.

We should also comment on the restriction of these regions to the left hemisphere. In the case of the motor region this restriction is obvious because participants are responding with their right hand. We and the other researchers we mentioned have found stronger responses in the left parietal and left prefrontal in these kinds of symbolic tasks. The restriction to the left caudate and left ACC is largely done for consistency, but more often than not the response is stronger in the left region. This is somewhat surprising in the case of the ACC because the left and right ACC regions are adjacent to each other.

Given that these regions reflect their ascribed function so well, one is tempted to assume that the function is actually performed in that cortical region. Although this is a plausible inference that many make, it is not necessary to the logic of our approach. We only require that we have a brain region whose activity reliably reflects a particular information processing function. Even if we assume that the function is performed in that region, there is no reason to suppose that its activity will only reflect that function. Nonetheless, we have been fortunate over the series of studies that we have performed that the regions seem to be rather pure indicators of their ascribed functions.

Finally, there is no claim that the ascribed function is restricted to these regions. With respect to retrieval we suspect there was a similar response in the hippocampus, but our scanning parameters in this experiment did not include the hippocampus, and it appears not to give as strong a signal when we do. With respect to the control, it is almost certain that dorsolateral, prefrontal structures play a role in control as well as the ACC. With respect to the caudate, we would expect to find a similar response in other structures connected to the basal ganglia, particularly the dorsal thalamus (and indeed we often do). Elsewhere (Anderson et al., 2004) we have reviewed proposals (Amos, 2000; Frank, Loughry, & O'Reilly, 2000; Houk & Wise, 1995; Wise, Murray, & Gerfen, 1996) that the basal ganglia perform functions similar to those that we ascribe to the production system. Finally, note that our list of five regions does not contain a region that corresponds to the visual module. This is because our scanning parameters also did not include the relevant visual regions. Other studies have found the expected responses in the visual cortex with visual presentation and the auditory cortex with auditory presentation (Sohn et al., in press).

5.2. Predicting the BOLD response

We have developed a methodology for relating the profile of activity in modules such as those in Fig. 3 to blood-oxygen-level-dependent (BOLD) responses from the brain regions that correspond to these modules. Fig. 4 illustrates the proportion of time that a particular module was active at various points during a trial on Day 1 (Part a) and Day 5 (Part b) for the two-step equations. These numbers would be directly obtainable from Fig. 3, except that Fig. 4 reflects the average engagement over the whole day not just at the beginning of Day 1 (Fig. 3a) and the end of Day 5 (Fig. 3b). The basic model we have developed of the BOLD response claims that while a module is engaged, it is producing a hemodynamic response in the corresponding region. We have adopted the standard gamma function that other researchers (e.g., Boyton, Engel, Glover, & Heeger, 1996; Cohen, 1997; Dale & Buckner, 1997; Glover, 1999) have used for the BOLD response. If the module is engaged it will produce a BOLD response t time units later according to the function:

$$B(t) = m\left(\frac{t}{s}\right)^{a} e^{-(t/s)}$$

where m governs the magnitude, s scales the time, and the exponent a determines the shape of the BOLD response such that with larger a the function rises and falls more steeply. The time to

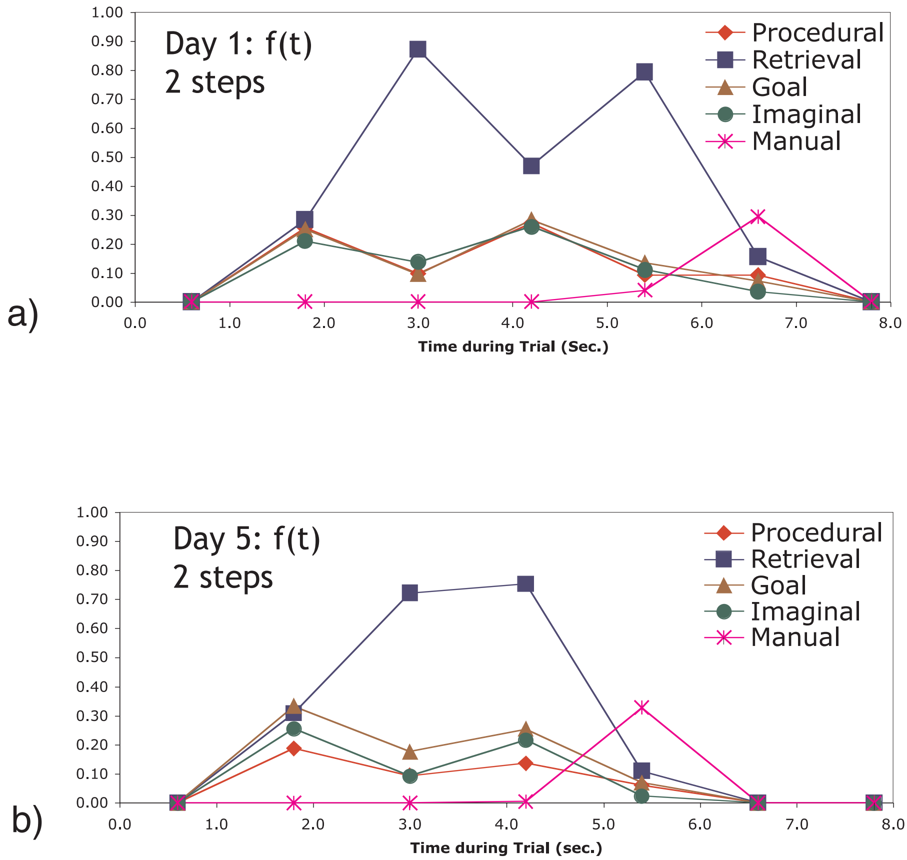

Fig. 4. The degree of engagement of the various modules during a trial on Day 1 (part a) and Day 5 (part b).

peak for the BOLD response is $a*s$, and the magnitude area under the curve is $m*s*\Gamma(a)$ where Γ is the gamma function, [$\Gamma(a) = (a - 1)!$]. Fig. 5 illustrates the effect of different choices of time scale and exponent on the shape of the BOLD response. To facilitate comparison, the magnitude parameters for the curves in this figure have been set so that the maximum response is 1 for all functions. As can be observed, a larger *a* produces a quicker rise and fall, whereas a larger *s* stretches the duration of the BOLD response.

The BOLD response accumulates whenever the region is engaged. Thus, if $f(t)$ is an engagement function giving the probability that the region is engaged at time *t*, then the cumulative BOLD response can be obtained by convolving the two functions:

$$CB(t) = \int_0^t f(x)B(t-x)dx$$

This is the observed BOLD response. Its area is proportional to the total time that the region is engaged. Thus, if a module is active for *T* sec, then the area under the BOLD response is $T*m*s*\Gamma(a)$.

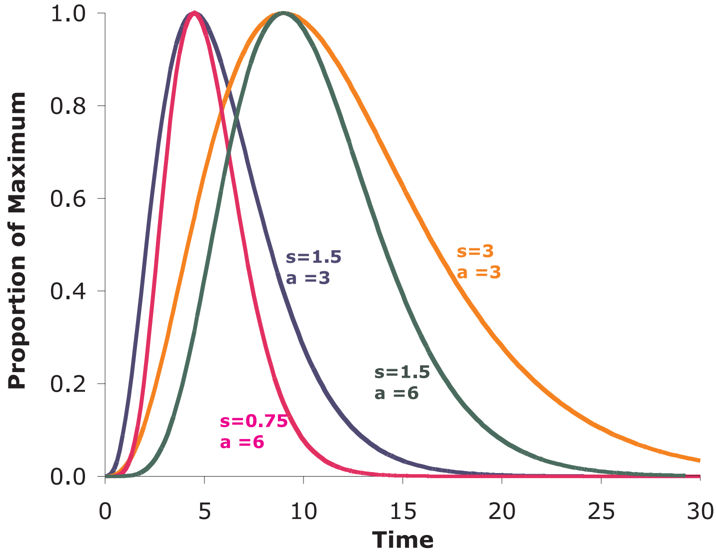

Fig. 5. An illustration of the impact of different choices of the exponent (a) and time scale on the shape of the hemodynamic function. To facilitate comparison, the magnitude parameter (m) has been scaled so that all of these functions have a maximum of 1.0.

In summary, a model for the time course (Fig. 3) of this task yields engagement functions $f(t)$ such as those in Fig. 4. By convolving the engagement functions with the BOLD function one can obtain predictions for the BOLD response in the regions associated with the modules. Most of the parameters of this model are set according to prior values established for ACT–R, but fitting the latency in Fig. 2 did require estimating parameters for the time to encode the equation and the duration of the retrievals. Having now committed to the time course of each module, predictions immediately follow for the time course of the cumulative BOLD response. The exact height and shape of the BOLD response depends on the magnitude (m), the scale (s), and the exponent (a) for the region that corresponds to that module. However, the strong parameter-free prediction is that the relative areas under the BOLD responses in two conditions for a region will reflect the relative amounts of time this region is engaged in these two conditions. Thus, the BOLD response provides a direct check on assumptions about the amount of time various modules are engaged in doing a task.

Table 2
Parameters estimated and fits to the BOLD response

	Motor/ Manual	Prefrontal/ Retrieval	Parietal/ Imaginal	Cingulate/ Goal	Caudate/ Procedural
Magnitude (m)	0.531	0.073	0.231	0.258	0.207
Exponent (a)	3	3	3	3	3
Scale (s)	1.241	1.545	1.645	1.590	1.230
Correlation	.975	.963	.969	.981	.975
Chi-square (105 df)	88.93	82.60	95.21	123.27	81.03

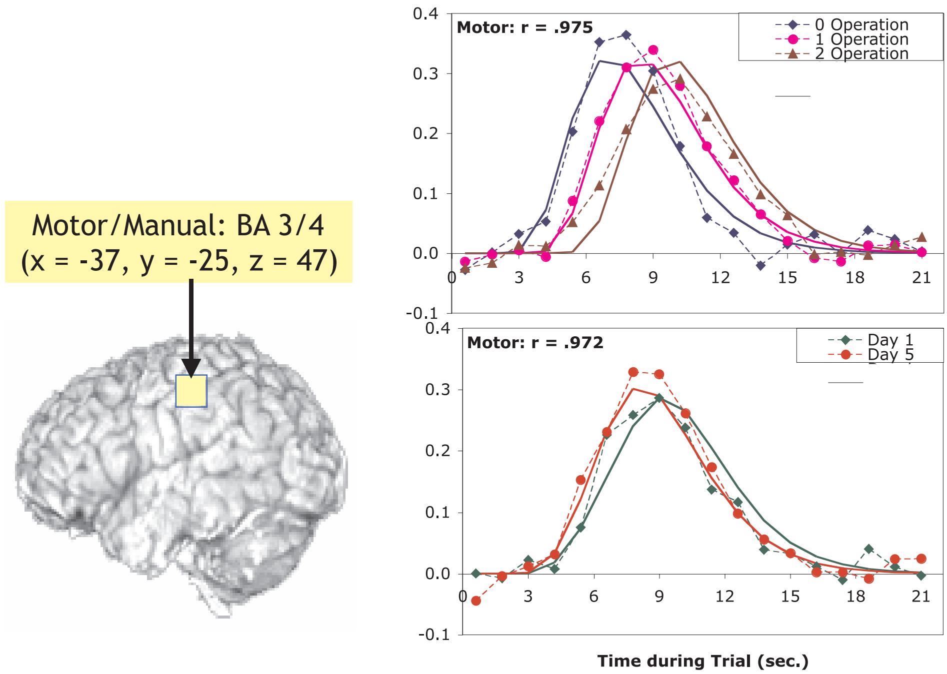

Fig. 6. Use of module behavior to predict percent increase in BOLD response in various regions: (a) Manual module predicts motor region; (b) Retrieval Module predicts prefrontal region; (c) Control/Goal module predicts anterior cingulate region; (d) Imaginal/Problem State module predicts parietal region; (e) Procedural module predicts

Table 2 gives the estimated parameters for the BOLD response and Fig. 6 shows how well this model predicts the BOLD responses in the six conditions achieved by crossing day and number of steps of transformation for each of the five associated regions. To simplify matters and to make the functions more comparable, the exponent of the BOLD response was set to 3 for all regions. To keep the data presentation readable and get better estimates, Fig. 6 either averages over days or over conditions.[5]

5.3. Characterizing the differences among the brain regions

The first impression one probably gets from Fig. 6 is that the BOLD responses for the five regions look a lot alike. All show a characteristic hemodynamic response that goes up and comes down with the trial structure. Furthermore, most regions show a stronger response for more transformations and a stronger response on Day 1. This is quite characteristic of imaging results where disparate regions of the brain give quite similar responses to the material. Without a strong theory to guide one's expectations, one is in danger of missing the differences and concluding that the whole brain (or at least those regions that respond—not all regions in the brain respond to the task structure in this experiment) is reflecting a global response to the task. However, if one knows where to look, there are characteristic differences. Although this one experiment does not reveal all the differences in the behavior of all five regions, it does reflect

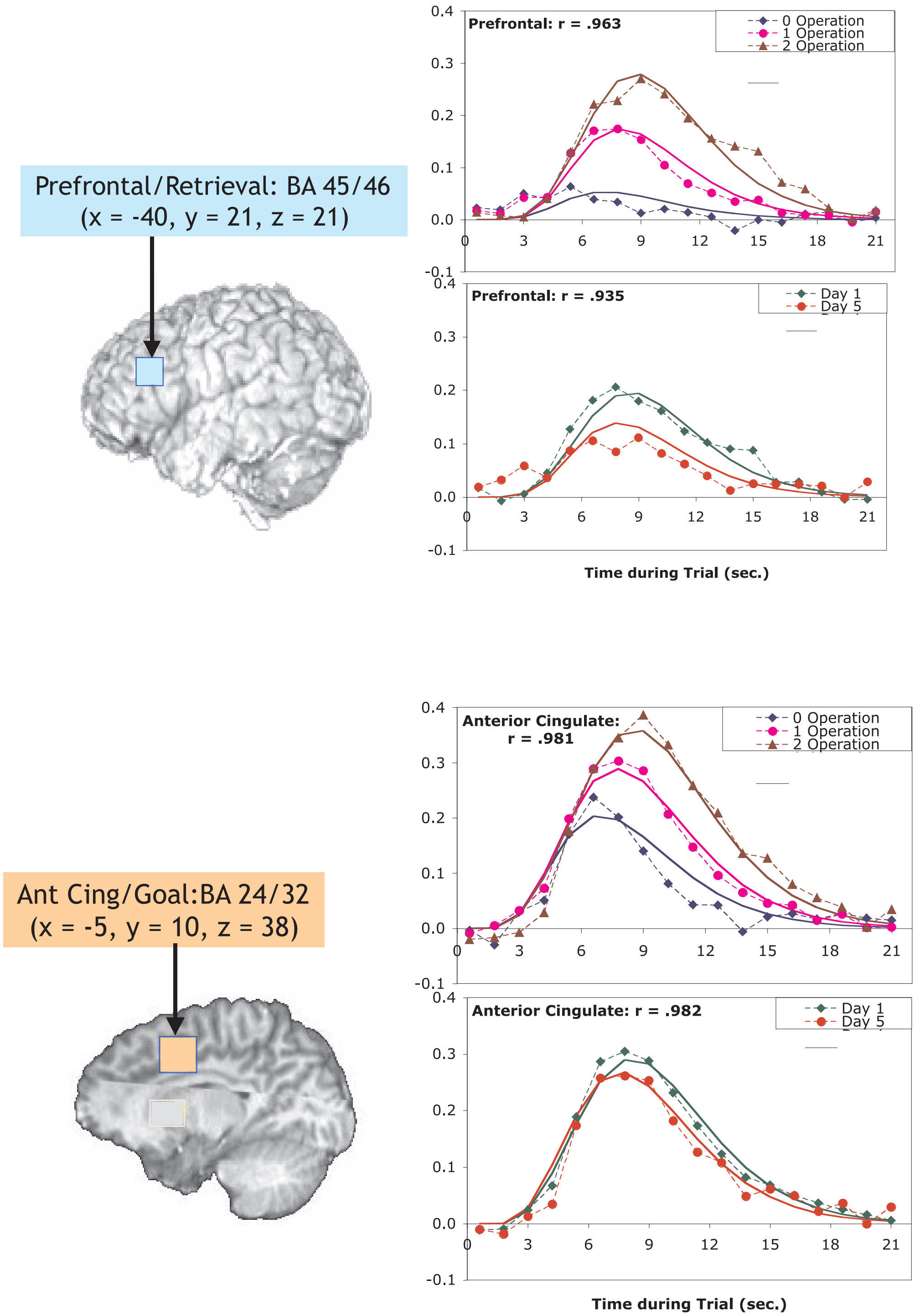

Fig. 6. *(continued)*

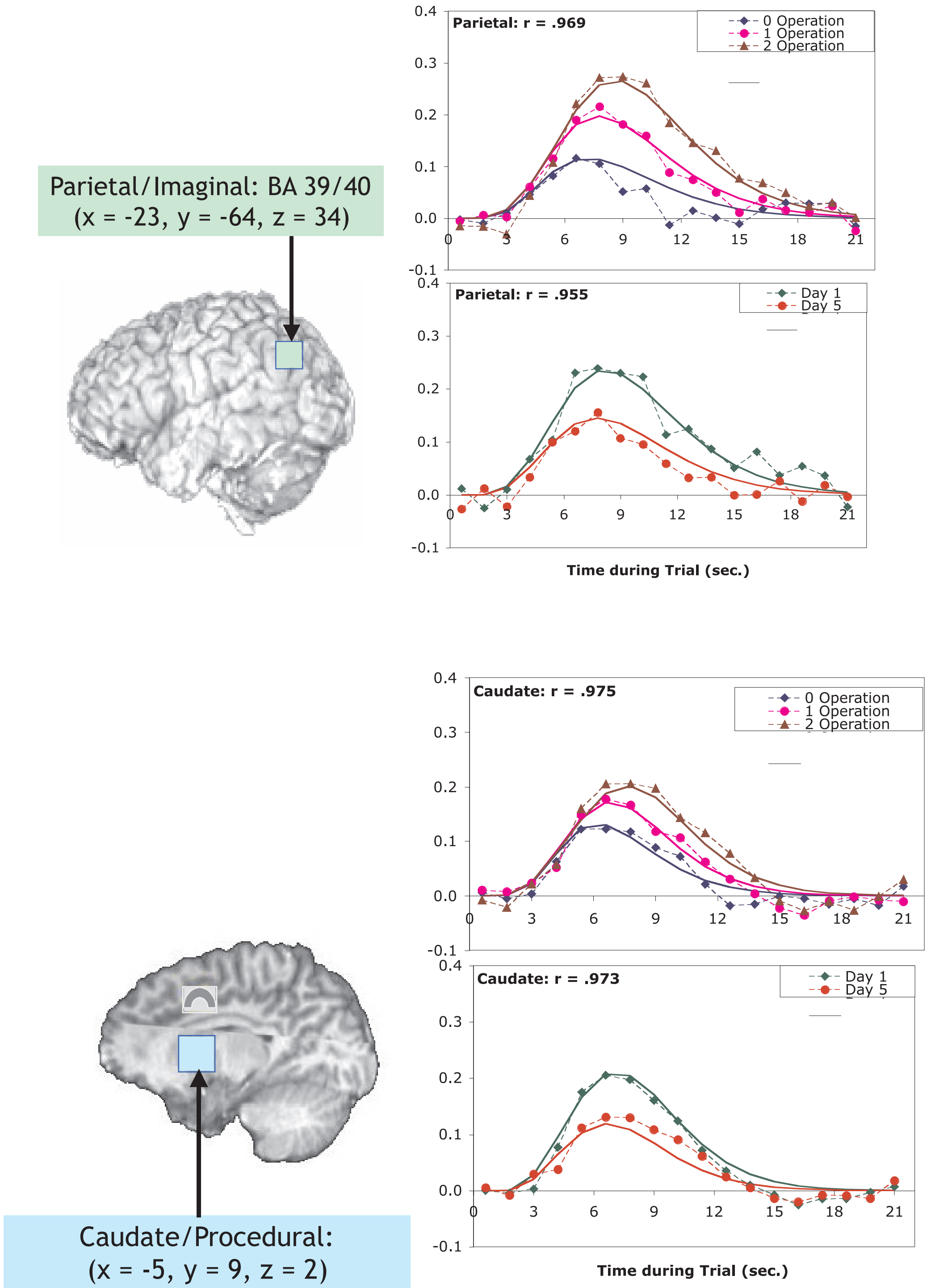

Fig. 6. *(continued)*

many of the important differences that we have identified over our experiments. These are enumerated in the following paragraphs.

First, and most apparent, in Fig. 6a the motor region is giving basically the same hemodynamic response in all conditions. The effect of the slower conditions is to delay when that hemodynamic response occurs. This is what would be expected given a relatively strong understanding of what regions of the brain control the hand. Although the motor region is transparently giving a different response than the other regions (both on theoretical and observational grounds), its correlation with the BOLD responses in other regions averages .66. Thus even it might be confused with the other regions unless one had a theory to tell one where to look to find the relevant differences.

Two of the other regions have distinct signatures. The prefrontal region (Fig. 6b) is distinguished by the very weak response it generates in the case of 0 steps. According to the model this case involves some brief retrievals of instructions but no retrieval of number facts. We have often modeled this condition by assuming no retrieval and predicted a flat function, but a slight rise can be discerned. The striking feature of the anterior cingulate (Fig. 6c) is that there is almost no effect of learning, whereas there is a robust effect of number of steps on magnitude of the response. The goal component in ACT–R is engaged in maintaining the state at points where the system is engaged in a retrieval of an arithmetic fact (this is because the retrieval buffer cannot be used to hold the next step). Every time it engages in retrieval of an arithmetic task it must note this so that it will wait for the fact before going on. Once the fact is retrieved it must reset the state so that it can proceed with unwinding. Thus, the number of retrieval operations is one factor influencing the number of state-setting operations in the goal buffer. The number of arithmetic retrievals changes in this experiment with the number of steps in solving the equation because each step requires retrieval of a fact. However, there is little reduction in these retrievals with practice. In principle, with enough practice they would eventually drop out, but there are so many individual facts that they just do not repeat enough in equation solving. In other research on learning (Qin et al., 2003) where retrievals did not involve a large database, we did find a substantial learning effect over 5 days in the anterior cingulate.

The other two regions (the parietal in Fig. 6d and the caudate in Fig. 6e) can be distinguished from the other three regions because they lack the features that identify the other three. However, there is little difference in the response that we see in these two regions. They approximately reflect the average response of all the areas, showing substantial effects of both number of steps of transformations and delay. There is a subtle difference between the two with the caudate showing a relatively larger effect of days and the parietal showing a relatively larger effect of steps. The caudate is fit according to the number of rules, which naturally increases with steps and decreases with days. These steps often are accompanied by changes in the problem representation, and this is why the two regions are so strongly correlated. We find differences between these two regions in experiments that vary modality of presentation from visual to aural with the parietal responding less to auditory presentation than visual and the caudate responding more (Sohn et al., 2005). Note in the comparisons of Fig. 6d and 6e, that the caudate gives a relatively weak response and has a poorer signal-to-noise ratio. This is unfortunate because according to the theory it should be the one region that is involved in all cognitive tasks, reflecting the number of production rules fired.

5.4. Assessing goodness of fit

The figures contain measures of correlations between the predictions and observed behavior. These are averaged over either days or operations, but Table 2 gives correlations among all 108 points for each region. Although this is a conventional measure of quality of fit, it has a number of problems. For instance, correlation is only sensitive to whether the shapes match up and not to whether the actual predicted numbers match up.

The quantitative correspondence can be assessed by the chi-square statistics in the table, which measure the degree of mismatch against the noise in the data. They are calculated as

$$\chi^2 = \frac{\sum_i (\hat{X}_i - \bar{X}_i)^2}{S^2_{\bar{X}}}$$

where the denominator is estimated from the interaction between conditions and participants. This has 105 *df,* calculated as 108 minus the three parameters estimated for the BOLD function. By this measure all of the areas are being modeled as well as can be expected because they all yield nonsignificant chi-squares (it would have to be 130 or greater to be significant at the .05 level). However, the chi-square statistic is not a perfect measure of fit for a number of reasons. First, it depends on the assumption that errors in individual points are independent, which is unlikely in this case. Second, it depends on the assumption that the gamma function is exactly the correct characterization of the BOLD response. Both of these problems reflect the fact that this measure may weight fitting the exact shape of the curves too much.

Anderson et al. (2003) offered an alternative measure of goodness of fit that avoids this concern with curve shape and parameter estimation. It simply tests whether the areas under the curves are in proportion to the time a module is active. They proposed calculating the following measure of proportionality:

$$\text{Proportionality} = \frac{\left(\sum_i T_i A_i\right)^2}{\sum_i T_i^2 \sum_i A_i^2}$$

where T_i is the amount of time a region is engaged in a condition and A_i is the amount of area under the BOLD response. In this experiment, the summations are over the six conditions defined by crossing the three levels of equation complexity with the 2 days of practice. This is like R^2 in some ways but has some differences. As a simple example of the differences consider the relation between the numbers 10, 11, and 12 and 0, 1, and 2. They have an R^2 of 1 but a proportionality of only .67 because the first set of numbers is almost equal proportionally, but the ratio of the second set is quite different. As another example, the numbers 10, 11, and 12 and the numbers 10, 12, and 10 have an R^2 of 0 but a proportionality of .987. One can calculate a degree of proportional misfit, which we call *misproportionality,* as 1 – Proportionality. These misproportionalities are reported in Table 3a, and one can observe that in all cases the region is best fit by its assigned module.

Table 3
Correlation between various modules and the BOLD response in various brain regions

	Motor	Prefrontal	Cingulate	Parietal	Caudate
a. Misproportionalities					
Manual	0.011	0.308	0.100	0.263	0.119
Retrieval	0.326	0.017	0.059	0.041	0.061
Goal	0.103	0.119	0.006	0.113	0.040
Imaginal	0.212	0.069	0.026	0.035	0.033
Procedural	0.190	0.109	0.049	0.037	0.023
b. Chi Squares					
Manual	88.93	452.05	724.66	426.40	333.89
Retrieval	493.22	82.60	350.32	101.88	133.13
Goal	255.91	194.94	123.27	171.74	111.01
Imaginal	384.66	125.66	210.47	95.21	101.82
Procedural	347.05	163.76	286.28	114.93	81.03

As a different approach, Table 3b reports the outcome of trying to fit each module to each region's activation profile and calculating a chi-square measure of misfit. With 105 *df* the 5-percentile tails for the chi-square distribution are at 82 and 130. As we noted with respect to Table 2, all the modules give acceptable fits (less than 130) to their ascribed regions. A few other modules give acceptable fits to other regions, although not as good. In particular, the modules other than the manual module all give approximately equal fits to the parietal and caudate regions. As noted, these regions approximately show the average response of all the regions. Note in Table 3 that the misproportionalities almost perfectly predict the chi-squares down a particular column (a column holds constant the noise in the data that determines the denominator for the chi-square).

In summary, the good fit of the model to the BOLD responses does not depend on the estimation of the parameters that characterize the BOLD function. Part (a) of Table 3 establishes a parameter-free measure of strength of association between ACT–R component and brain region and part (b) establishes that parameter estimation cannot make other components fit other brain regions. Except for trying to distinguish between the parietal and caudate, whose responses were not well discriminated by this experiment, the proposed associations provide a much better explanation of the data than any alternative set of associations.

6. The capacity for re-representation: A uniquely human trait?

Having now analyzed in some detail the nature of one instance of algebra symbol manipulation, I would like to close by reflecting on the question of what in these processes might be uniquely human. For this purpose it would be useful to describe an example of serial behavior that has been observed in the Rhesus Macaques monkey (Terrace, Son, & Brannon, 2003). In the experiment reported by Terrace et al., monkeys learned four 7-item lists of pictures. On any particular trial the monkeys were shown one set of seven pictures randomly arrayed on the screen, and they had to select them in the correct order. They were able to achieve over 65% correct reproduction of the entire lists—a number that would compare favorably with humans in similar circumstances. Terrace et al. offered a number of varieties of evidence to argue that the monkeys are operating off a declarative representation of the list order and not some type of procedural representation. For instance, monkeys can correctly order two items from different lists that they have never seen paired before. These serial tasks are interesting because they bear certain superficial similarities to algebra symbol manipulation. In algebra symbol manipulation a child is shown one array of symbols (the equation) and must produce another array of symbols (the solution—in our task we reduced this to a single key press, but it is typically more complex as in the example of writing of the quadratic expression at the beginning of this article). Similarly, the monkey is shown one array of symbols and must produce a sequence of symbols or actions.

Although the performance of monkeys in these serial tasks is in many ways remarkable, there is a significant difference in the behavioral capacity involved in transforming algebraic equations and that involved in manipulating serial lists. The most obvious difference is in the generativeness of the child's algebraic capacity. The child is capable of responding to an arbitrary number of new expressions. Even simply using the unwind strategy to solve equations, children are capable of solving an infinite number of equations. Moreover, the generativeness in algebraic symbol manipulation goes beyond this—for instance, a child can factor, expand, and do many other operations to transform one string of symbols into another. This kind of generativeness has much in common with the generativeness of natural language, the most frequently mentioned instance of human intellectual superiority. However, as noted in the introduction, unlike language, the formal properties of algebraic manipulation are more or less completely understood. Also as the reported experiment illustrates, it is much easier to experimentally study the learning of algebra.

Although the differences between the child's algebraic symbol manipulation and the monkey's serial reproduction may seem obvious, the challenge is to identify what in the ACT–R architecture is associated with this difference. Before addressing this question, I should acknowledge an alternative hypothesis that the difference just reflects prior knowledge: Children successful at algebra understand a number system, know their number facts, know how to parse these expressions, and know how to follow instructions. The only factor in this list that can be discounted with certainty is the importance of number knowledge because people are quite capable of learning artificial algebras (e.g., Blessing & Anderson, 1996; Qin et al., 2003) that have no numeric reference. Also, recent research has expanded upward our estimate of primates' understanding of number (see Hauser & Spelke, in press, for a review). The other differences might support the argument that monkeys would be capable of doing algebra if it were practical to teach them this skill and its prerequisite background knowledge. I cannot disprove this possibility, but it seems unlikely. Moreover, it does turn out that there is a significant ACT–R architectural capability required for algebraic manipulation that is not required for the serial reproduction tasks.

ACT–R models for the serial reproduction tasks require visual, manual, and retrieval buffers that work in formally similar ways to the models for the algebraic tasks. These, then, cannot be the source of the differences. However, these buffers in themselves do not allow the mental re-representation that is key to algebraic symbol manipulation. Although children can perform these re-representations on paper, writing out transformation after transformation, and thus saving themselves the need for mental re-representation, they prefer to do it mentally as long as things do not get so complex as to exceed their capacity to hold a representation of the critical algebraic material.

Fig. 7 illustrates the transformations in Fig. 3, somewhat simplified and designed to make salient the key architectural issues rather than to be faithful to all the details of the learning simulation. As in Fig. 3 the equation being solved is $7*x + 3 = 38$. The simplified mental image of the equation just holds the intermediate result, but it is the critical piece of information in that it is what is not supported by external information. For instance, at one point the image in Fig. 7 holds an internal representation of 35 that is intermediate between the original equation and the final answer of 5. Being able to hold onto such an internal representation, detached from either stimulus or action, is critical to the model's algebraic competence. It is tempting to point to the parietal cortex as what is enabling this algebra problem solving, because the parietal cortex is what seems to be holding the image of the intermediate result. The region of parietal cortex that

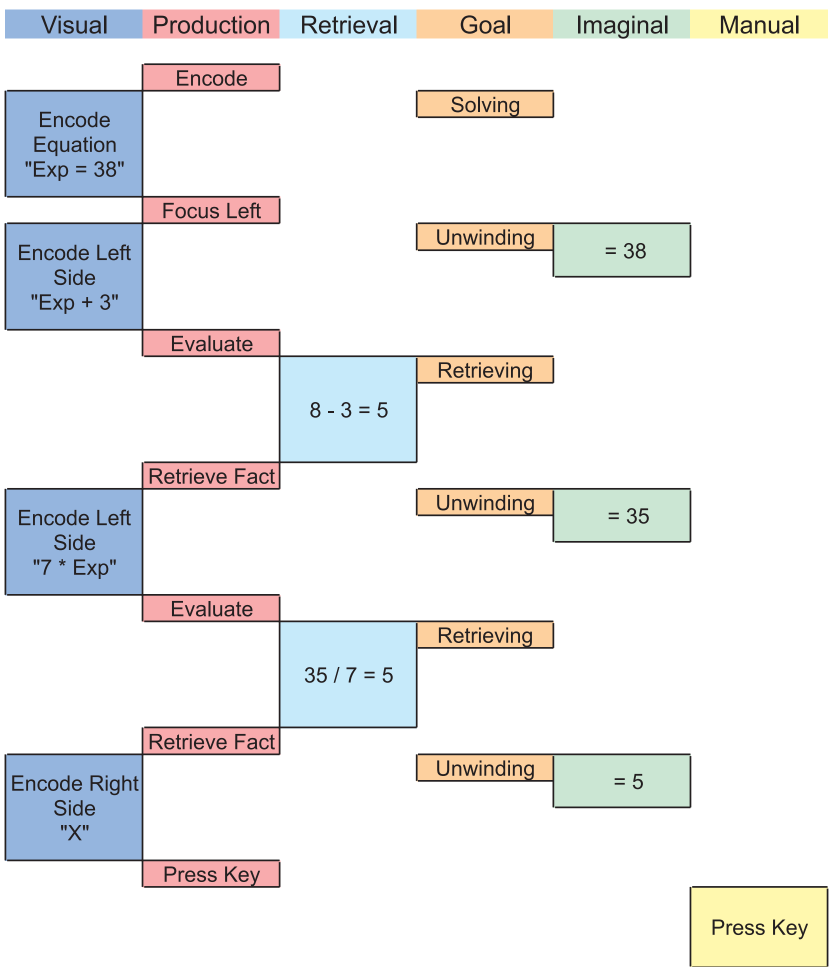

Fig. 7. A representation of the basic buffer operations required to implement the unwind strategy in ACT–R to solve the equation $7*x + 3 = 38$.

corresponds to the imaginal module does not appear to have a homologue in the monkey brain (Zilles & Palomero-Gallagher, 2001).

Although there may be some special properties to the human ability to hold such intermediate results, it is not totally discontinuous from the ability of the monkey. This becomes apparent when one tries to develop an ACT–R model for the monkey task of ordering two items from two lists. Terrace et al. (2003) showed a generative capacity to the monkey's serial knowledge in that it can take a pair of elements from different lists, which it has not seen together, and correctly order it with high accuracy. Fig. 8 is a similar flowchart for a putative ACT–R model that I developed for this ordering task. The model assumes that the monkey retrieves the location of each item in the pair and creates an image that synthesizes the two locations and then picks the item that is first in this image. Although the imaginal ability in this example may not have all of the flexibility of human imagery in equation manipulation, it seems essential to be able to have some internal synthesis of the two objects to make an appropriate decision. I could not figure a way to do this in ACT–R without such an internal representation. A comparison of Figs. 7 and 8 should make clear that the two tasks do not differ in their capacity demands on the imaginal representation. Both require a relatively small amount to be held in this working memory. In

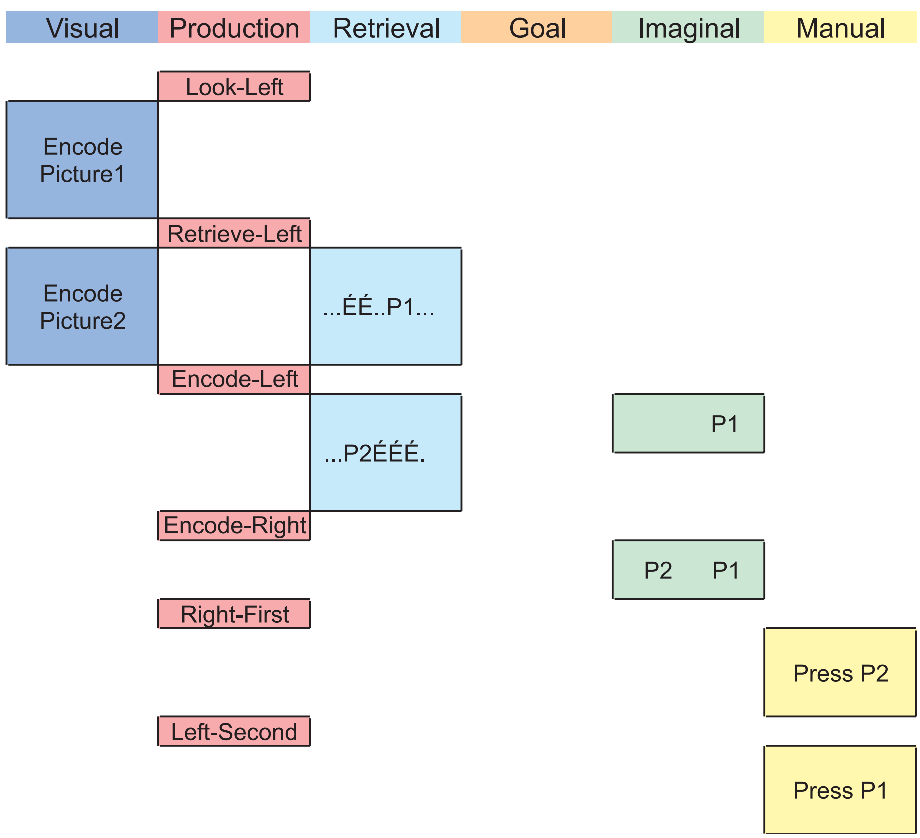

Fig. 8. A representation of the basic buffer operations in ACT–R required to implement the serial ordering task.

fact, the algebra task really requires holding just one number at any time, whereas two items have to be synthesized in the image for the serial task.

Comparing Figs. 7 and 8, however, reveals a striking difference. The model in Fig. 8 does not require any state tests against the goal buffer. I could have built a model that included such state tests (and perhaps such models are appropriate for humans doing this task), but it was unnecessary. The conditions for the firing of the individual productions are determined by the states of the other buffers. Specifically, the presentation of the stimulus is the condition for the look-left production; the encoding of the left element is the condition for the retrieve-left production that requests the position of the left element in its list; the retrieval of an element and an empty image is the condition for the encode-left production that positions the left element in the image and requests retrieval of the right element; the retrieval of an element and an incomplete image is the condition for the encode-right production that similarly places the right element; the completion of the image is the condition for the production that selects the first element; and the manual selection of that element is the condition for the second production that selects the second element.

In contrast, because of the iterative nature of the unwind algorithm, it is not possible to find unique states of the nongoal modules for each production in the algebra model. The model is faced with multiple situations where it has focused on an element in the equation, has retrieved an arithmetic fact, and has an image of an intermediate result. Without the help of the control element in the goal it would not know whether it is time to retrieve another arithmetic fact or perform another transformation of the equation. Therefore, it sometimes skips retrievals or transformations and other times repeats them. In this model repeats are innocuous, but skips mean it fails to solve the problem. For instance, if the model that does not use the control information, faced with the equation $3x + 9 = 15$, it sometimes responds 6 because it omitted retrieving $6/3 = 2$ or because it omitted to use the fact when retrieved.[6]

This goal buffer has been tentatively associated with the anterior cingulate. The anterior cingulate is particularly active in studies where participants have to direct their behavior in a way that violates typical response tendencies. As noted in the introduction, it is particularly active in dealing with conflict in the Stroop task during color naming (Carter et al., 2000). We discussed there the various theories of what is behind this activity. The anterior cingulate has undergone major evolutionary changes that are only found in humans and the closely related great apes (Allman, Hakeem, Erwin, Nimchinsky, & Hof, 2001). These changes, which include a new class of spindle-shaped cells in strongest concentration in the human anterior cingulate, appear to be related to the ability to achieve appropriate behavior in the presence of conflicting stimuli.

So where does this leave the question of what enables the human-unique aspects of algebra problem solving? The critical mental ability seems to be that of re-representation in situations where neither the external situation nor the other internal buffers indicate what to do next. Re-representation in the algebra example involves this alternation between retrieval and transformations of the internal representation, ending finally in a response. One needs some sort of state information to indicate what to do next in a series of steps of re-representation The anterior cingulate is not the only neural structure to have unique properties in the human brain, but it is one of the players in higher level human cognition, and the role it plays seems to be that of maintaining a control state in the presence of ambiguous or conflicting information.

Notes

1. Although the control states in this article will be simple they can contain information about steps, substeps, and other notes such as whether a borrow exists in a subtraction problem.
2. There are a number of comments to make on these instructions. First, these instructions already reflect a distillation of the basic axioms of algebra about doing the same operation to both sides and collapsing results. The basic axioms generate a large search space of possible transformations and students are typically given guidance as to how to select the appropriate transformations (as the students in our experiment were). Second, the instructions require further elaboration to deal with cases such as (6 – subexpression) = 3 or (6 / subexpression) = 3, where there is an asymmetric operator with the subexpression on the right. We did not present any such problems to our students. Third, they treat the special "+0" and "1*" constructions as special cases. Students were given instructions to do so.
3. Although this is a particularly dramatic example of production compilation, there are many other instances in Fig. 3 that I have not noted to avoid overly cluttering the figure.
4. Note that Fig. 3 sometimes gives different productions the same name—the goal in these naming conventions is just to try to indicate the basic function of the rule and different rules are learned with similar functions
5. In fact, none of the regions showed a significant interaction between practice and number of steps or between practice, number of steps, and scan.
6. These problems can be avoided if one puts control information into the image but then this looses the separation of control state and problem state.

Acknowledgments

This research was supported by the NSF grant ROLE: REC-0087396 and ONR Grant N00014–96–1–0491. I would like to thank Herbert Terrace for discussions relevant to preparation of the last section of this article and Jennifer Ferris for her comments on the manuscript.

References

Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., & Hof, P. (2001). Anterior cingulate cortex: The evolution of an interface between emotion and cognition. *Annals of the New York Academy of Sciences, 935,* 107–117.

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26,* 39–83.

Amos, A. (2000). A computational model of information processing in the frontal cortex and basal ganglia. *Journal of Cognitive Neuroscience, 12,* 505–519.

Anderson, J. R., Bothell, D., Byrne,M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review, 111,* 1036–1060.

Anderson, J. R., & Douglass, S. (2001). Tower of Hanoi: Evidence for the cost of goal retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27,* 1331–1346.

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

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30,* 221–256.

Anderson, J. R., Qin, Y., Sohn, M-H., Stenger, V. A., & Carter, C. S. (2003). An information-processing model of the BOLD response in symbol manipulation tasks. *Psychonomic Bulletin & Review, 10,* 241–261.

Anderson, J. R., Qin, Y., Stenger, V. A., & Carter, C. S. (2004). The relationship of three cortical regions to an information-processing model. *Cognitive Neuroscience, 16,* 637–653.

Anderson, J. R., Taatgen, N. A., & Byrne, M. D. (in press). Learning to achieve perfect time sharing: Architectural implications of Hazeltine, Teague, & Ivry (2002). *Journal of Experimental Psychology: Human Perception and Performance.*

Ashby, F. G., & Waldron, E. M. (2000). The neuropsychological bases of category learning. *Current Directions in Psychological Science, 9,* 10–14.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Blessing, S., & Anderson, J. R. (1996). How people learn to skip steps. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 576–598.

Boyton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience, 16,* 4207–4221.

Buckner, R. L., Kelley, W. M., and Petersen, S. E. (1999). Frontal cortex contributes to human memory formation. *Nature Neuroscience, 2,* 311–314.

Cabeza, R., Dolcos, F., Graham, R., & Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *Neuroimage, 16,* 317–330.

Carter, C. S., MacDonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., et al. (2000). Parsing executive processes: Strategic versus evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the U.S.A., 97,* 1944–1948.

Cohen, M. S. (1997). Parametric analysis of fMRI data using linear systems methods. *NeuroImage, 6,* 93–103.

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping, 5,* 329–340.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L., (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20,* 487–506.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science, 5,* 303–305.

D'Esposito, M., Piazza, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., et al. (1995). The neural basis of the central executive of working memory. *Nature, 378,* 279–281.

Donaldson, D. I., Petersen, S. E., Ollinger, J. M., & Buckner, R. L. (2001). Dissociating state and item components of recognition memory using fMRI. *NeuroImage, 13,* 129–142.

Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain, 124,* 849–881.

Falkenstein, M., Hohnbein, J., & Hoorman, J. (1995). Event related potential correlates of errors in reaction tasks. In G. Karmos, M. Molnar, V. Csepe, I. Czigler, & J. E. Desmedt (Eds.), *Perspectives of event-related potentials research* (pp. 287–296). Amsterdam: Elsevier Science.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2000). *Interactions between frontal cortex and basal ganglia in working memory: A computational model* (Tech. Rep. 00-01). Boulder, CO: University Colorado, Institute of Cognitive Science.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage, 9,* 416–429.

Hauser, M. D., & Spelke, E. S. (in press). Evolutionary and developmental foundations of human knowledge: A case study of mathematics. In M. Gazzaniga (Ed.), *The cognitive neurosciences III*. Cambridge, MA: MIT Press.

Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, Z., Nakamura, K., et al. (1999). Parallel neural networks for learning sequential procedures. *Trends in Neuroscience, 22,* 464–471.

Houk, J. C., & Wise, S. P. (1995). Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: Their role in planning and controlling action. *Cerebral Cortex, 2*, 95–110.

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.

Lebiere, C., & Anderson, J. R. (1993). A connectionist implementation of the ACT–R production system. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 635–640). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lepage, M., Ghaffar, O., Nyberg, L., & Tulving, E. (2000). Prefrontal cortex and episodic memory retrieval mode. *Proceedings of National Academy of Sciences U.S.A., 97,* 506–511.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288,* 1835–1838.

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

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Vol. 4. Advances in research and theory* (pp. 1–18). New York: Plenum.

Pashler (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*, 220–244.

Paus, T., Koski, L., Caramanos, Z., & Westbury, C. (1998). Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: A review of 107 PET activation studies. *Neuroreport, 9,* R37–R47.

Poldrack, R. A., Prabakharan, V., Seger, C., &Gabrieli, J. D. E. (1999). Striatal activation during cognitive skill learning. *Neuropsychology, 13,* 564–574.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in Neurosciences, 17,* 75–79.

Posner, M. I., & DiGirolamo, G. J. (1998). Executive attention: Conflict, target detection and cognitive control. In R. Parasuraman (Ed.), *The attentive brain* (pp. 401–423). Cambridge, MA: MIT Press.

Qin, Y., Sohn, M-H., Anderson, J. R., Stenger, V. A., Fissell, K., Goode, A., et al. (2003). Predicting the practice effects on the blood oxygenation level-dependent (BOLD) function of fMRI in a symbolic manipulation task. *Proceedings of the National Academy of Sciences of the U.S.A., 100,* 4951–4956.

Qin, Y., Anderson, J. R., Silk, E., Stenger, V. A., & Carter, C. S. (2004). The change of the brain activation patterns along with the children's practice in algebra equation solving. *Proceedings of the National Academy of Sciences of the U.S.A., 101,* 5686–5691.

Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural basis of strategy and skill in sentence-picture verification. *Cognitive Psychology, 40,* 261–295.

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural learning and neostriatal dysfunction in man. *Brain, 111,* 941–959.

Sohn, M-H., Albert, M. V., Jung, K., Carter, C. S., & Anderson, J. R. (2004). *Pay now or pay later: Preparatory conflict resolution in the anterior cingulate cortex and the prefrontal cortex.* Poster session presented at the 11th Annual Meeting of the Cognitive Neuroscience Society, San Francisco.

Sohn, M-H., Goode, A., Stenger, V. A, Carter, C. S., & Anderson, J. R. (2003). Competition and representation during memory retrieval: Roles of the prefrontal cortex and the posterior parietal cortex. *Proceedings of National Academy of Sciences, 100,* 7412–7417.

Sohn, M-H., Goode, A., Stenger, V. A, Jung, K-J., Carter, C. S., & Anderson, J. R. (2005). An information-processing model of three cortical regions: Evidence in episodic memory retrieval. *NeuroImage, 25,* 21–33.

Taatgen, N. A. (2005). Modeling parallelization and speed improvement in the skill acquisition: From dual tasks to complex dynamic skills. *Cognitive Science, 29,* 421–455.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "Broke"? A model of learning the past tense without feedback. *Cognition, 86,* 123–155.

Terrace, H. S., Son, L. K., & Brannon, E. M. (2003). Serial expertise of rhesus macaques. *Psychological Science, 14*, 66–73.

Wagner, A. D., Maril, A., Bjork, R. A., & Schacter, D. L. (2001). Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral prefrontal cortex. *NeuroImage, 14,* 1337–1347.

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31,* 329–338.

Wise, S. P., Murray, E. A., & Gerfen, C. R. (1996). The frontal cortex-basal ganglia system in primates. *Critical Reviews in Neurobiology, 10,* 317–356.

Zilles, K., & Palomero-Gallagher, N. (2001). Cyto-, myelo-, and receptor architectonics of the human parietal cortex. *Neuroimage, 14,* 8–20.