4

Using Brain Imaging to Guide the Development of a Cognitive Architecture

John R. Anderson

We have begun to use functional magnetic resonance imaging as a way to test and extend the ACT-R theory. In this chapter, we will briefly review where we are in these efforts, describe a new modeling effort that illustrates the potential of our approach, and then end with some general remarks about the potential of such data to guide modeling efforts and the development of a cognitive architecture generally. Brain imaging has grown hand in hand with the movement to a module-based representation of knowledge in the current ACT-R theory. In this chapter, we will first review the ACT-R architecture and its application to brain imaging. ACT-R is a general system, and it is possible to take a model developed for one domain and apply that same model to a second domain. We will describe an instance of this in the second section of the chapter. Then, in the third section, we will try to draw some lessons from this work about the connections between such a modeling framework and brain imaging.

We have begun to use functional magnetic resonance imaging (fMRI) brain imaging as a way to test and extend the adaptive control of thought–rational, or ACT-R theory (Anderson & Lebiere, 1998). In this chapter, I will briefly review where we are in these efforts, describe a new modeling effort that illustrates the potential of our approach, and then end with some general remarks about the potential of such data to guide modeling efforts and the development of a cognitive architecture generally. Brain imaging has grown hand in hand with the movement to a module-based representation of knowledge in the current ACT-R theory (Anderson et al., 2005). In this chapter, we will first review the ACT-R architecture and its application to brain imaging. ACT-R is a general system, and it is possible to take a model developed for one domain and apply that same model to a second domain. We will describe an instance of this in the second section of the chapter. Then, in the third section of the chapter, we will try to draw some lessons from this work about the connections between such a modeling framework and brain imaging.

ACT-R and Brain Imaging

The ACT-R Architecture

According to the ACT-R theory, cognition emerges through the interaction of a number of independent modules. Figure 4.1 illustrates the modules relevant to solving algebraic equations:

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, the model described in Anderson (2005) alternated between unwinding an equation and retrieving arithmetic facts.

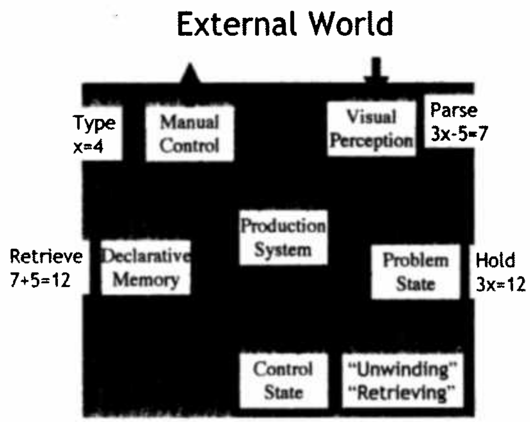


FIGURE 4.1 The interconnections among modules in ACT-R 5.0.

4. A declarative module that retrieves critical information from declarative memory such as that "7 + 5 = 12."
5. A manual module that programs manual responses such as the key presses to give the response "$x = 4$."

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 small amount of 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 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 Figure 4.1). 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 = number3 has been retrieved,
>
> THEN transform the equation to Expression = number3

This production responds when the control chunk encodes the goal to solve an equation (first line), when the problem state chunk represents an equation of the appropriate form (second line, for example, $3[x - 2] - 4 = 5$), when a chunk encoding an arithmetic fact has been retrieved from memory (third line—in this case, $4 + 5 = 9$), and appropriately changes the problem representation chunk (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 as with the other modules, it has a serial bottleneck in that it can only fire a single rule at a time. Since 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 also have bottlenecks. All of the bottlenecks are in the communication among modules; within modules things are massively parallel. (Figure 4.4, later in the chapter, 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 (e.g., Anderson, Taatgen, & Byrne, 2005).

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 the splitting of the goal system into a control module and a problem state module is another development. There were two reasons for choosing to separate control state (goal module) and problem state knowledge (imaginal module). 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, while the anterior cingulate reflected control state changes. Later, the chapter will elaborate 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 both could involve a number of elements. 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 would be different for 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, while the common result would be represented identically in 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.

Use of Brain Imaging to Provide Converging Data

We have associated these modules with specific brain regions, and fMRI allows us to track these modules individually and provide converging evidence for assumptions of the ACT-R theory. 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., 2005) 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 chapter, we will be concerned with five brain regions and their ACT-R associations:

1. Caudate (procedural): Centered at Talairach coordinates $x = -15, y = 9, z = 2$. 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 7, 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 2 and 4 at the central sulcus.

We have defined these regions once and for all and use them over and over again in predicting different experiments. This has many advantages over the typical practice in imaging research of using exploratory analyses to find out what regions are significant in particular experiments. The exploratory approach has substantial problems in avoiding false positives because there are so many experimental tests being done looking for significance in each brain voxel. To the extent that the exploratory approach can cope with this, it winds up setting very conservative criteria and fails to find many effects that occur in experiments. This had lead to the impression (e.g., Uttal, 2001) that results do not replicate over experiments.

Beyond these issues, determining regions by exploratory means is not suitable for model testing.

Being selected to pass a very conservative threshold of significance, these regions give biased estimates of the actual effect size. Also the exploratory analyses typically look for effects that are significant and not whether they are the same. This can lead to merging brain regions that actually display two (or more) different effects that are both significant. For instance, if one region shows a positive effect of a factor and an adjacent region shows a negative effect, they will be merged, and the resulting aggregate region may show no effect.

Predicting the BOLD Response

We have developed a methodology for relating the profile of activity in ACT-R modules to the blood oxygen level dependent (BOLD) responses from the brain regions that correspond to these modules. Figure 4.2 illustrates the general idea about how we map from events in an information-processing model onto the predictions of the BOLD function. Each time an information-processing component is active it will generate a demand on associated brain regions. In this hypothetical case, we assume that an ACT-R module is active for 150 ms from 0.5 to 0.65 s, for 600 ms from 1.5 to 2.1 s, and for 300 ms from 2.5 to 2.8 s. The bars at the bottom of the graph indicate when the module is active.

A number of researchers (e.g., Boyton, Engel, Glover, & Heeger, 1996; Cohen, 1997; Dale & Buckner, 1997) have proposed that the hemodynamic response to an event varies according to the following function of time, t, since the event:

$$h(t) = t^{a}e^{-t}, \qquad (1)$$

where estimates of the exponent have varied between 2 and 10. This is essentially a gamma function that will reach maximum at a time units after the event. As illustrated in Figure 4.2, this function is slow to rise, reflecting the lag in the hemodynamic response to neural activity.

We propose that while a module is active it is constantly producing a change that will result in a BOLD response according the above function. The observed fMRI response is integrated over the time that the module is active. Therefore, the observed BOLD response will vary with time as

$$B(t) = M\int_{0}^{t} d(x)\,h\left(\frac{t-x}{s}\right)dx, \qquad (2)$$

where M is the magnitude scale for response, s is the latency scale, and $d(x)$ is a "demand function" that reflects the probability that the module will be in use at time t. Note because of the scaling factor, the prediction is that the BOLD function will reach maximum at roughly $t = a \times s$ seconds.

As Figure 4.2 illustrates, one can think of the observed BOLD function in a region as reflecting the sum of separate BOLD functions for each period of time the module is active. Each period of activity is going to generate a BOLD function according to a gamma function as illustrated. The peak of the BOLD functions reflects roughly when the module was active but is offset because of the lag in the hemodynamic response. The height of the BOLD function reflects the duration of the event since the integration makes the height of the function proportional to duration over short intervals.

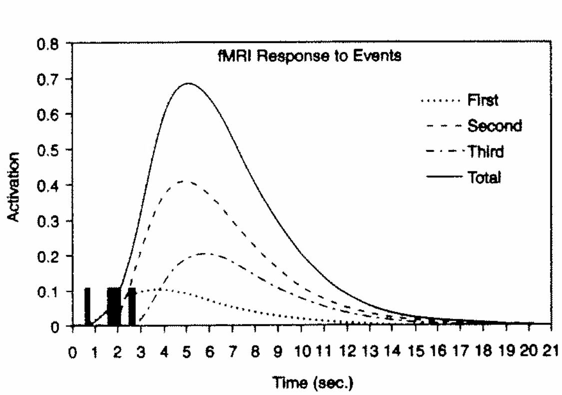


FIGURE 4.2 An illustration of how three BOLD functions from three different events result in an overall BOLD function.

Note that this model does not reflect a frequent assumption in the literature (e.g., Just, Carpenter, & Varma, 1999) that a stronger BOLD signal reflects a higher rate of metabolic expenditure. Rather, our assumption is that it reflects a longer duration of metabolic expenditure. The two assumptions are relatively indistinguishable in the BOLD functions they produce, but the time assumption more naturally maps onto an information-processing model that assumes stages taking different durations of activity. Since these processes are going to take longer, they will generate higher BOLD functions without making any extra assumptions about different rates of metabolic expenditure. The total area under the curve in Figure 4.2 will be directly proportional to the period of time that the module is active. If a module is active for a total period of time T, the area under the BOLD function will be $M\times\Gamma(a+1)\times T$, where Γ is the gamma function (in the case of integer a, note that $\Gamma[a+1] = a!$).

Application of an Existing Model to a New Data Set

The Anderson (2005) Algebra Model

Anderson (2005) described an ACT-R model of how children learned to solve algebra equations in an experiment reported by Qin, Anderson, Silk, Stenger, and Carter (2004). That model successfully predicted how children would speed up in their equation solving over a five-day period. The model used the general instruction-following approach described in Anderson et al. (2004) to model how children learned. Thus, it did not require handcrafting production rules specifically for the task. Rather the model used the same general instruction-following procedures described in Anderson et al. (2004) for learning of anti-air warfare coordinator (AAWC) system. That model was just given a declarative representation of the instructions that children received rather than a declarative representation of the AAWC instructions. The model initially interpreted these declarative instructions, but with practice, it built its own productions to perform the task directly. Only two parameters were estimated in Anderson (2005) to fit the model the model to latency data. One parameter, for the visual module, concerned the time to encode a fragment of instruction from the screen into an internal representation. The other parameter scaled the amount of time it took to perform retrievals in declarative memory as a function of level of activation. All the remaining parameters were default parameters of the ACT-R model as described in Anderson et al. (2004).

Given these time estimates, that model predicted when the various modules of the ACT-R theory would be active and for how long. Moreover, it predicted how these module activities would change over the five-day course of the experiment. Thus, it generated the demand functions we needed to predict the BOLD responses in these brain regions and how these BOLD functions varied with equation complexity and practice. In general, these predictions were confirmed.

Adult Learning of Artificial Algebra

This chapter proposes to go one step further than Anderson (2005). It proposes to take the model in Anderson (2005), including the time estimates and make predictions for another experiment (Qin et al., 2003). This can be seen as a further test of the underlying model of instruction and as a further demonstration of how brain imaging can provide converging data for a theory. Participants in this experiment were adults performing an artificial algebra task (based on Blessing & Anderson, 1996) in which they had to solve "equations."[1] To illustrate, suppose the equation to be solved was

$$②P③4 \leftrightarrow 5, \qquad (3)$$

where the solution means isolating the P before the "↔." In this case, the first step is to move the "③4" over to the right, inverting the "③" operator to a "②"; the equation now looks like

$$②P \leftrightarrow ②5②4. \qquad (4)$$

Then the ② in front of the P is eliminated by converting ②s on the right side into ③s so that the "solved" equation looks like:

$$P \leftrightarrow ③5③4. \qquad (5)$$

Participants were asked to perform these transformations in their heads and then key out the final answer—this involved pressing the thumb key to indicate that they had solved the problem and then keying 3, 5, 3, and 4 in this example (2 was mapped to the index finger, 3 to middle finger, 4 to ring finger, and 5 to little finger). The problems required 0, 1, or 2 (as in this example) transformations to solve. The experiment looked at how participants speed up over five days of practice. Figure 4.3 shows time to hit the first

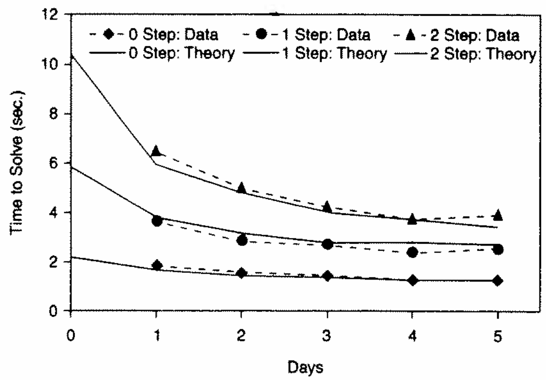


FIGURE 4.3 Mean solution times (and predictions of the ACT-R model) for the three types of equations as a function of delay. Although the data were not collected, the predicted times are presented for the practice session of the experiment (Day 0).

key (thumb press) in various conditions as a function of days.[2] The figure shows a large effect of number of transformations but also a substantial speed up over days. It also presents the predictions from the ACT-R model, which will now be described.

The ACT-R Model

Table 4.1 gives an English rendition of the instructions that were presented to the model. The general strategy of the model was to form an image of the items to the right of the "↔" and then transform that image according to the information to the left of the "↔." In addition to the instructions, we provided the model with the knowledge

TABLE 4.1 English Rendition of Task Instructions Given to ACT-R

1. To solve an equation, first find the "↔," then encode the first pair that follows, then shift attention to the next pair if there is one, then encode the second pair.
2. If this is a simple equation, output it; otherwise process the left side.
3. To process the left side, first find the P.
4. If "↔" immediately follows, then work on the operator that precedes the P; otherwise, first encode the pair that follows, then invert the operator, and then work on the operator that precedes the P.
5. To process the operator that preceded the P, first retrieve the transformation associated with that operator, then apply the transformation, and then output.
6. To output press the thumb, output the first item, output the next, output the next, and then output the next.

1. that ② and ③ were inverses of each other as were the operators ④ and ⑤.
2. the specific rules for getting rid of the ②, ③, ④, and ⑤ operators when they occurred in front of a P

These instructions and other information 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 performing reordering operations such as

IF one's goal is to apply a transformation to an image
and that transformation involves inverting the order of the second and fourth terms
and the image is of the form "a b c d,"
THEN change the image to "a d c b"

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

Production compilation (see Anderson et al., 2004; Taatgen & Anderson, 2002) is one reason the model is speeding up. 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 transform an image
and the prefix is ③
and the image is of the form "a b c d"
THEN change the image to "a d c b"

The model was given the same number of trials of practice as the participants received over the course of the experiment. Thus, we can look at changes in the model's performance on successive days. Figure 4.4a compares the encoding portion of a typical trial at the beginning of the Day 1 and with a typical trial at the end of the Day 5. In both cases, the model is solving the two-step equation:

②P③4↔②5

The figure illustrates when the various modules were active during the solution of the equation and what they were doing. Some general features of the activity in the figure include:

1. Multiple modules can be active simultaneously. For instance, on Day 5 there is a point where the visual module detects nothing beyond the ②5 (encode null right), while an instruction is being retrieved, while the goal module notes that it is in the encoding phase and while an image of the response "2 5" is being built up.
2. Much of the speed up in processing is driven by collapsing multiple steps into single steps. A particularly dramatic instance of this is noted in Figure 4.4 where five production firings and five retrievals on Day 1 (between "encode null right" and "encode equation ②P③④") are collapsed into one each. Production compilation can compress these internal operations without limit.

Figure 4.4b compares the transforming portion of a typical trial at the beginning of the Day 1 and with a typical trial at the end of the Day 5. The reduction in time is even more drammatic here because this portion of the trial involves the retrieval of inverse and transformation rules for getting rid of prefixes. These retrieval times show considerable speed up because of the growth in base-level activation in the declarative representation of these basic facts. Figure 4.4c shows the output portion of a typical trial, which is identical on Days 1 and 5 since production compilation cannot collapse productions that would skip over external actions. Note, however, that the times reported in Figure 4.3 correspond to the time of the thumb press, which is the first key press. Nonetheless, the rest of Figure 4.4c will affect the BOLD response that we will see.

Brain Imaging Data

Participants were scanned on Days 1 and 5. Participants had 18 s for each trial. Figure 4.5 shows how the BOLD signal in different brain regions varies over the 18-s period beginning 3 s before the onset of the stimulus and continuing for 15 s afterward. Activity was measured every 1.5 s. The first two scans provide an estimate of baseline before the stimulus comes on. These figures also display the ACT-R predictions. The BOLD functions displayed are typical in that there is some inertia in the rise of the signal after the critical event and then decay. The BOLD response is delayed so that it reaches a maximum about 4–5 s after the brain activity. In each part of Figure 4.5 we provide a representation of the effect of problem complexity averaging over number of days and a representation of the effect of practice, averaging over problem complexity. None of the regions showed a significant interaction between practice and number of steps or between practice, number of steps, and scan.

Figure 4.5a shows the activity around the left central sulcus in the region that controls the right hand. The effect of complexity is to delay the BOLD function (because the first finger press is delayed in the more complex condition), but there is no effect on the basic shape of the BOLD response because the same response sequence is being generated in all cases. The effect of practice is also just to move the motor BOLD response forward in time.

Figure 4.5b shows the activity around the left inferior frontal sulcus, which we take as reflecting the activity of the retrieval module. It shows very little rise in the zero transformation condition because there are few retrievals (only of a few instructions) in this condition. The lack of response in this condition distinguishes this region from most others. The magnitude of the response decreases after five days, reflecting that the declarative structures have been greatly strengthened and the retrievals are much quicker.

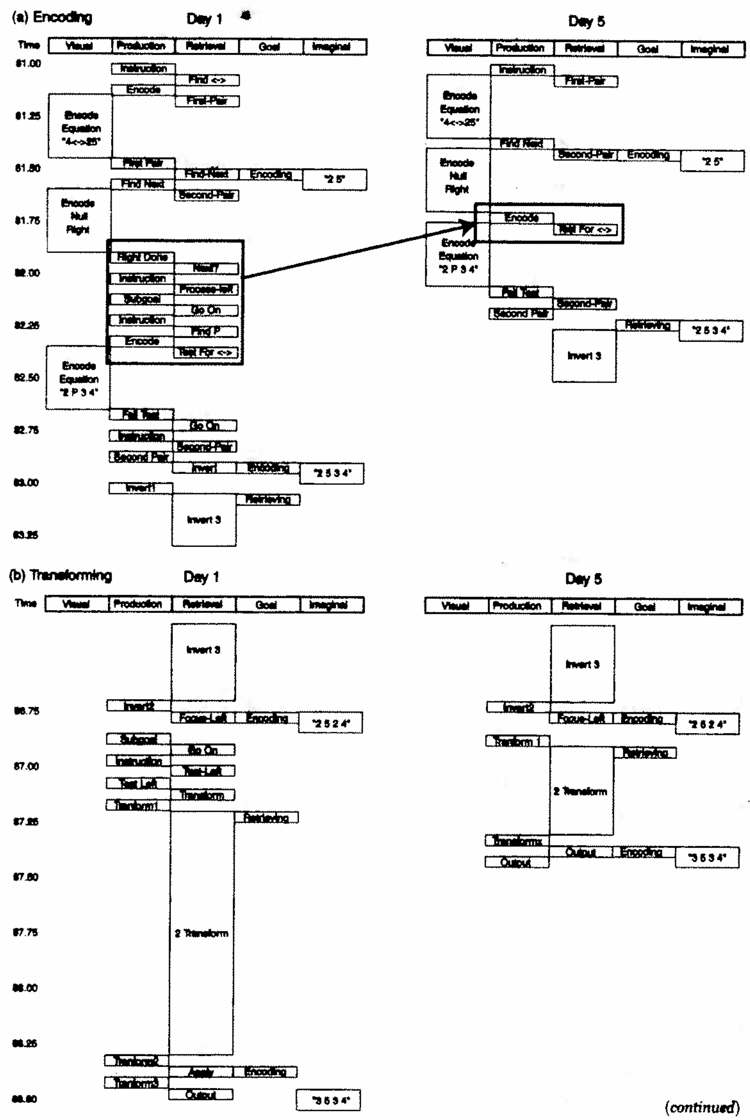


FIGURE 4.4 Module activity during the three phases of a trial: (a) encoding, (b) transforming, and (c) outputting. In the first two phases, the module activity changes from Day 1 to Day 5.

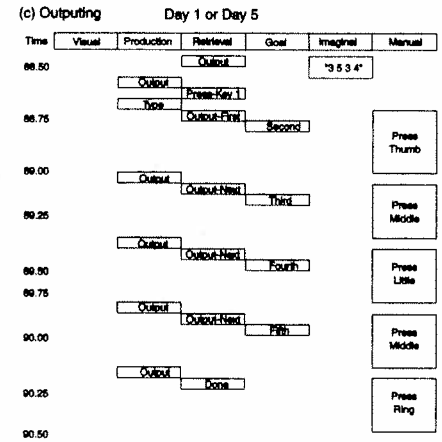


FIGURE 4.4 Continued.

Figure 4.5c shows activity in the left anterior cingulate, which we take as reflecting control activity, and Figure 4.5d shows activity around the left intraparietal sulcus, which we take as reflecting changes to the problem representation. Both of these regions show large effects of problem complexity and little effect of number of days of practice. Unlike the prefrontal region, they show a large response in the condition of zero transformations. There is virtually no effect of practice on the anterior cingulate. According to the ACT-R theory, this is because the model still goes through the same control states, only more rapidly on Day 5. In the case of the parietal region and its association with problem representation, there is a considerable drop out of intermediate problem representations, but most of this happens early in the learning and therefore not much further learning occurs from Day 1 to Day 5.

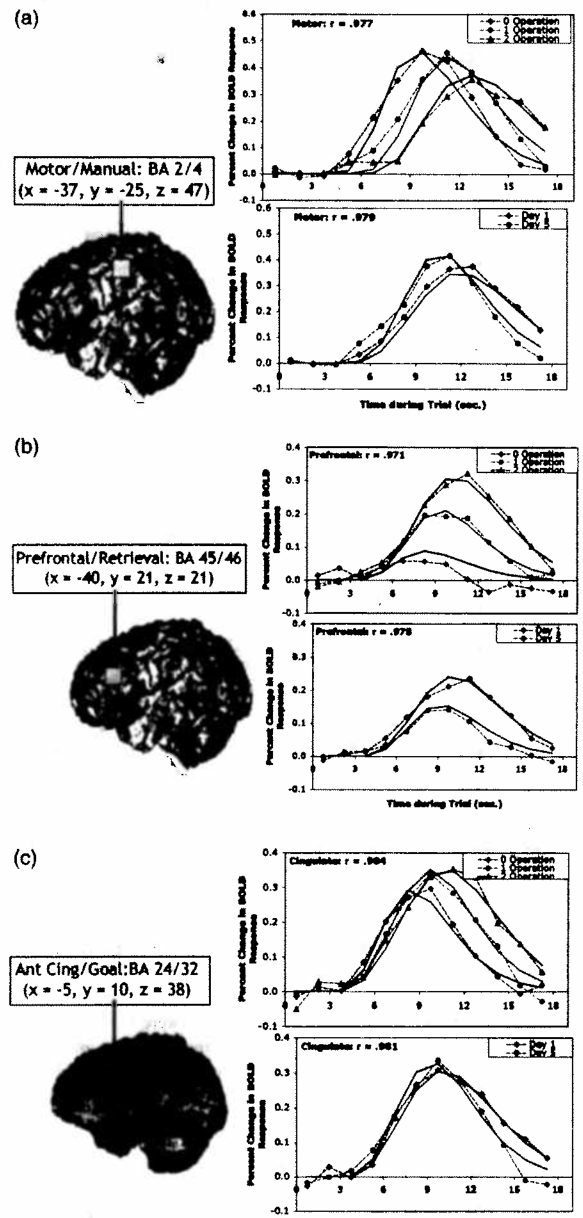
Figure 4.5e shows the activity in the caudate, which is taken to reflect production firing. The signal is rather weak, here but there appears to be little effect of complexity and a substantial effect of practice. The effect of complexity is predicted to be weak by the model because most of the time associated with transformation is taken up in long retrievals and not many additional productions are required. The model underpredicts the effect of learning for much the same reason it predicts a weak effect of practice in the parietal. The effects of practice on number of productions tends to happen early in this experiment and there is not that much reduction after Day 1.

Comments on Model Fitting

The model that yields the fits displayed in these figures was run without estimating any time parameters. This makes the fit to the latency data in Figure 4.3 truly parameter free, and it is remarkable how well that data does fit, given that we estimated parameters with children and now are fitting them to adults. At some level, this indicates that the children were finding learning real algebra as much of a novel experience as these adults were finding learning the artificial algebra and were taking about as long to do the task.

In the case of fitting the BOLD functions, however, we had to allow ourselves to estimate some parameters that describe the underlying BOLD response. To review, there were three parameters—an exponent a that governs the shape of the BOLD response; a timescale parameter s that, along with a, determines the time to peak ($a \times s = \text{peak}$); and a magnitude parameter m that determines just how much increase there is in a region. Table 4.2 summarizes the values of these parameters for this experiment with adults and artificial algebra and the previous experiment with children and real algebra.

We used the same value of a for both experiments and all regions. This value is 3 and it seems to give us



(continued)

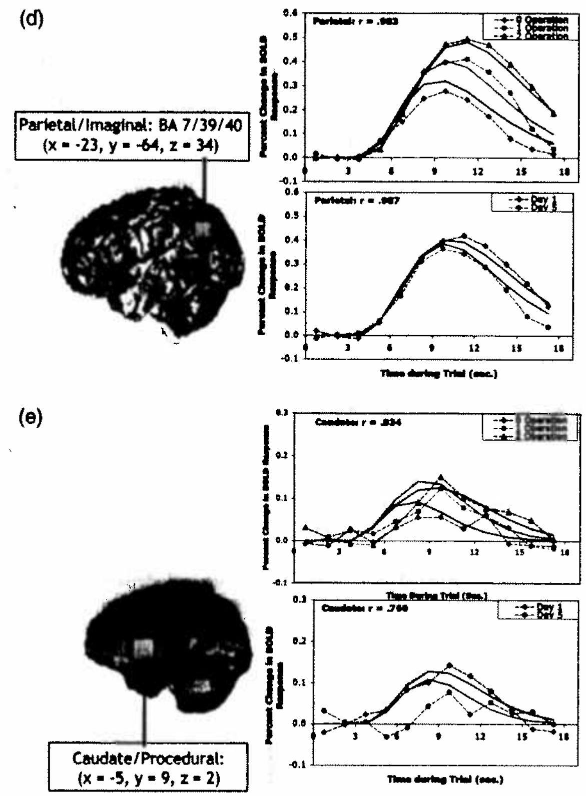


FIGURE 4.5 Use of module behavior to predict BOLD response in various regions: (a) manual module predicts motor region; (b) retrieval module predicts prefrontal region; (c) control/goal module predicts anterior cingulate region; (d) imaginal/problem state module predicts parietal region; (e) procedural module predicts caudate region.

a pretty good fit over a wide range of situations. The value of the latency scale parameter was estimated separately for each region in both experiments. It shows only modest variability and has a value of approximately 1.5 s, which would be consistent with the general observation that it is about 4.5 s for the BOLD response to peak. There is some variability in the BOLD response across subjects and regions (e.g., Huettel & McCarthy, 2000; Kastrup, Krüger, Glover, Neumann-Haefelin, & Moseley, 1999).

The situation with the magnitude parameter, however, does reveal some discrepancies that go beyond naturally expected variation. In particular, our experiment has estimated a motor magnitude that is less than 40% of the magnitude estimated for the children and a parietal magnitude that is almost four times as large. It is possible that these reflect differences in population, perhaps related to age, but such an explanation does not seem very plausible.

In the case of the parietal region, we think that the difference in magnitude may be related to the difficulty in manipulating the expressions. While this is the first time the children were exposed to equations, these expressions had a lot of similarity to other sorts of

TABLE 4.2 Parameters Estimated and Fits to the Bold Response $B(t) = m\left(\frac{t}{s}\right)^{a} e^{-(t/s)}$

		Motor/ Manual	*Prefrontal/ Retrieval*	*Parietal/ Imaginal*	*Cingulate/ Goal*	*Caudate/ Procedural*
Magn(m)	Children	0.531	0.073	0.231	0.258	0.207
	Adults	0.197	0.078	0.906	0.321	0.120
Exponent(a)		3	3	3	3	3
Scale(s)	Children	1.241	1.545	1.645	1.590	1.230
	Adults	1.360	1.299	1.825	1.269	1.153

arithmetic expressions children had seen before in their lives. In contrast, the expressions in the artificial algebra that the adults saw were quite unlike anything experienced before. One might have expected that this would be reflected in different times to parse them but we used the same estimates as with the children—0.1 s for each box in the imaginal columns of Figure 4.4. If we increased this estimate, however, we would have had to decrease some other time estimate to fit the latency data.

In the case of the motor region, we think that the difference in magnitude may be related to the different number of key presses. The adults in this experiment had to press five keys to indicate their answer, while the children had only to press one key. There is some indication (e.g., Glover, 1999) that the BOLD response may be subadditive.

Both discrepancies reflect on fundamental assumptions underlying our modeling effort. In the case of the parietal region, it may be that the same region working for the same time may produce a different magnitude response, depending on how "difficult" the task is. In the case of the motor region, it may be the case that our additivity assumption is flawed.

While acknowledging that there might be some flies in the ointment with respect to parameter estimates, it is still worth asking how well the model does fit the data. We have presented in these figures measures of correlation between data and theory. While these are useful qualitative indicants, they really do not tell us whether the deviations from data are "significant." Addressing this question is both a difficult and questionable enterprise, but I thought it would be useful to report our approach. We obtained from an analysis of variance how much the data varied from subject to subject. This is measured as the subject-by-condition interaction term, where the conditions are the 72 observations obtained by crossing difficulty (3 values) with days (2 values) with scans (12 values). This gives us an error of estimate of the mean numbers going into the figures as data (although in these figures we have averaged over one of the factors). We divided the sum of the squared deviations by this error term and obtained a chi-square quantity:

$$\chi^2 = \frac{\sum_i (\hat{X}_i - \bar{X}_i)^2}{S_{\bar{X}}^2}, \qquad (6)$$

which has degrees of freedom equal to the number of observations being summed (72) minus the number of parameters estimated (2—latency scale and magnitude). With 70 degrees of freedom, this statistic is significant if greater than 90.53. The chi-square values for four of the five regions are not significant (motor, 70.42; prefrontal, 46.91; cingulate, 48.25; parietal, 88.86), but the estimate for the caudate is with a chi-square measure of 99.56. It turns out that a major discrepancy for the caudate is that the BOLD function rises too fast. If we allow an exponent of 5 (and so change the shape of the BOLD response), we get a chi-square deviation of only 79.23 for the caudate.

It is wise not to make too much of these chi-square tests as we are just failing to reject the null hypothesis. There may be real discrepancies in the model's fit that are hidden by noise in the data. The chi-square test is just one other tool available to a modeler and sometimes (as in the case of the caudate) it can alert one to a discrepancy between theory and data.

Conclusions

The use of fMRI brain imaging has both influenced the development of the current ACT-R theory and provided support for the state of that theory. For instance, it was one of the reasons for the separation of the previous goal structure into a structure that just held control information (currently called the *goal*) and a structure that contained information about the problem state (now called an *imaginal module*). Besides giving us a basis for testing a model fit, the data provided some converging evidence for major qualitative claims of the model—such as that there was little retrieval in the zero transformation condition and that there was little effect of learning in this experiment on control information.

While things are encouraging at a general level, our discussion of the details of the model fitting suggested that there are some things that remain to be worked out. We saw uncertainty about a key assumption that magnitude of the BOLD response only reflects time a module is active. Differences in the magnitude of response in the two experiments in the parietal region suggested that there be different magnitude of effort in a fixed time. Again differences in magnitude of response in the motor region suggested that BOLD effects might be subadditive. On another front, problems in fitting the caudate raised the question of whether all the regions are best fit by the same shape parameter. While use of brain imaging data is a promising tool, it is apparent we are still working out how to use that tool.

We should note that there is no reason such data and methodology should be limited to testing the ACT-R theory. Many other information-processing theories could be tested. The basic idea is that the BOLD response reflects the duration for which various cognitive modules are active. The typical additive-factors information-processing methodology has studied how manipulations of various cognitive components affect a single aggregate behavioral measure like total time. If we can assign these different components to different brain regions, we have essentially a separate dependent measure to track each component. Therefore, this methodology promises to offer strong guidance in the development of any information-processing theory. Finally, we want to comment on the surprising match of fMRI methodology to the study of complex tasks. A problem with fMRI is its poor temporal resolution. However, as is particularly apparent in the behavior of our manual module, the typical effect size in a complex mental task is such that one can still make temporal discriminations in fMRI data. One might have thought the outcome of such a complex task would be purely uninterpretable. However, with the guidance of a strong information-processing model and well-trained participants one not only can interpret but also predict the BOLD response in various regions of the brain.

Acknowledgments

This research was supported by the National Science Foundation Grant ROLE: REC-0087396 and ONR Grant N00014-96-1-0491. I would like to thank Jennifer Ferris, Wayne Gray, and Hansjörg Neth for their comments on this chapter. Correspondence concerning this chapter should be addressed to John R. Anderson, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213. Electronic mail may be sent to ja+@cmu.edu.

Notes

1. The reason for using an artificial algebra is that these participants already knew high school algebra, and we wanted to observe learning.
2. Note that there is a Day 0 when subjects practiced the different aspects of the task but were not metered in a regular task set; see Qin et al. (2003) for details.

References

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

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture, *Cognitive Science*, 29, 313–342.

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

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

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

——, 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.

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

——, Taatgen, N. A., & Byrne, M. D. (2005). Learning to achieve perfect time sharing: architectural implications of Hazeltine, Teague, & Ivry (2002). *Journal of Experimental Psychology: Human Perception and Performance, 31*, 749–761.

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.

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.

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

Huettel, S., & McCarthy, G. (2000). Evidence for refractory period in the hemodynamic response to visual stimuli as measured by MRI. *NeuroImage, 11*, 547–553.

Just, M. A., Carpenter, P. A., & Varma, S. (1999). Computational modeling of high-level cognition and brain function. *Human Brain Mapping, 8*, 128–136.

Kastrup, A., Krüger, G., Glover, G. H., Neumann-Haefelin, T., & Moseley, M. E. (1999). Regional variability of cerebral blood oxygenation response to hypercapnia. *NeuroImage, 10*, 675–681.

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

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 United States of America, 100*, 4951–4956.

——, 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 National Academy of Sciences, 101*, 5686–5691.

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.

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

Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain.* Cambridge, MA: MIT Press.

5

The Motivational and Metacognitive Control in CLARION

Ron Sun

This chapter presents an overview of a relatively recent cognitive architecture and its internal control structures, that is, its motivational and metacognitive mechanisms. The chapter starts with a look at some general ideas underlying this cognitive architecture and the relevance of these ideas to cognitive modeling of agents. It then presents a sketch of some details of the architecture and their uses in cognitive modeling of specific tasks.

This chapter presents an overview of a relatively recent cognitive architecture and its internal control structures (i.e., motivational and metacognitive mechanisms) in particular. We will start with a look at some general ideas underlying this cognitive architecture and the relevance of these ideas to cognitive modeling.

In the attempt to tackle a host of issues arising from computational cognitive modeling that are not adequately addressed by many other existent cognitive architectures, CLARION, a modularly structured cognitive architecture, has been developed (Sun, 2002; Sun, Merrill, & Peterson, 2001). Overall, CLARION consists of a number of functional subsystems (e.g., the action-centered subsystem, the metacognitive subsystem, and the motivational subsystem). It also has a dual representational structure—implicit and explicit representations in two separate components in each subsystem. Thus far, CLARION has been successful in capturing a variety of cognitive processes in a variety of task domains based on this division of modules (Sun, 2002; Sun, Slusarz, & Terry, 2005).

A key assumption of CLARION, which has been argued for amply before (see Sun, 2002; Sun et al., 2001; Sun et al., 2005), is the dichotomy of implicit and explicit cognition. In general, implicit processes are less accessible and more "holistic," while explicit processes are more accessible and crisper (Reber, 1989; Sun, 2002). This dichotomy is closely related to some other well-known dichotomies in cognitive science: the dichotomy of symbolic versus subsymbolic processing, the dichotomy of conceptual versus subconceptual processing, and so on (Sun, 1994). The dichotomy can be justified psychologically, by the voluminous empirical studies of implicit and explicit learning, implicit and explicit memory, implicit and explicit perception, and so on (Cleeremans, Destrebecqz, & Boyer, 1998; Reber 1989; Seger, 1994; Sun, 2002). In social psychology, there are similar dual-process models, for describing socially relevant cognitive processes (Chaiken & Trope, 1999). Denoting more or less the same distinction, these dichotomies serve as justifications for the more general notions of