

# ACT-R: A Theory of Higher Level Cognition and Its Relation to Visual Attention

**John R. Anderson, Michael Matessa, and  
Christian Lebiere**  
*Carnegie Mellon University*

---

## ABSTRACT

The ACT-R system is a general system for modeling a wide range of higher level cognitive processes. Recently, it has been embellished with a theory of how its higher level processes interact with a visual interface. This includes a theory of how visual attention can move across the screen, encoding information into a form that can be processed by ACT-R. This system is applied to modeling several classic phenomena in the literature that depend on the speed and selectivity with which visual attention can move across a visual display. ACT-R is capable of interacting with the same computer screens that subjects do and, as such, is well suited to provide a model for tasks involving human-computer interaction. In this article, we discuss a demonstration of ACT-R's application to menu selection and show that the ACT-R theory makes unique predictions, without estimating any parameters, about the time to search a menu. These predictions are confirmed.

---

**John R. Anderson** is a cognitive scientist with an interest in cognitive architectures and intelligent tutoring systems; he is a Professor of Psychology and Computer Science at Carnegie Mellon University. **Michael Matessa** is a graduate student studying cognitive psychology at Carnegie Mellon University; his interests include cognitive architectures and modeling the acquisition of information from the environment. **Christian Lebiere** is a computer scientist with an interest in intelligent architectures; he is a Research Programmer in the Department of Psychology and a graduate student in the School of Computer Science at Carnegie Mellon University.

---

## CONTENTS

1. INTRODUCTION
  2. A THEORY OF THE VISUAL INTERFACE
  3. VISUAL ATTENTION
    - 3.1. Sperling Task
    - 3.2. Subitizing Task
    - 3.3. Visual-Search Task
    - 3.4. Conclusions
  4. APPLICATION TO MENU-SELECTION DATA
  5. CONCLUSIONS
- 

## 1. INTRODUCTION

ACT-R, as originally developed by Anderson (1993), was a model of higher level cognition. That model has been applied to modeling domains like the Tower of Hanoi, mathematical problem solving in the classroom, navigation in a computer maze, computer programming, human memory, and other tasks. By the standards of these fields, it has provided good models of human cognition. However, by the standards of human-computer interaction (HCI), it has had a serious failing. It has ignored many of the details by which the subject interacted with the external environment. All of the applications, either in the laboratory (e.g., Anderson, Reder, & Lebiere, 1996) or in the classroom (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995), have involved people reading a computer screen and using a mouse and a keyboard, but there was no theory of how this “input” and “output” took place. In Kieras and Meyer’s (1994) terms, we had a theory of “disembodied cognition.” We have recently addressed these lacunae and have developed a theory of how ACT-R interacts with computer applications. Most of these embellishments are concerned with visual attention, although they also address issues of visual perception and motor action. We describe this added theory of visual attention and how it relates to the ACT-R theory of higher level cognition. We describe its application to some classic paradigms in visual attention to establish its credibility. Then we describe its extension to a menu-selection task and its ability to make some novel predictions about that task. First, though, we set forth the basic ACT-R theory of cognition. We are brief in our description of the basic ACT-R theory because descriptions exist elsewhere (Anderson, 1993). Here we just describe enough detail to establish the context in which our theory of visual attention has been developed.

ACT-R assumes that there are two types of knowledge—declarative and procedural. *Declarative knowledge* corresponds to things that we are aware we know and can usually describe to others. Examples of declara-

tive knowledge include "George Washington was the first president of the United States" and "Three plus four is seven." *Procedural knowledge* is knowledge that we display in our behavior but of which we are not conscious. Procedural knowledge basically specifies how to bring declarative knowledge to bear in solving problems.

Declarative knowledge in ACT-R is represented in terms of chunks (Miller, 1956; Servan-Schreiber, 1991), which are schema-like structures consisting of an "isa" pointer specifying their category and some number of additional pointers encoding their contents. Below is a chunk encoding the addition fact that  $3 + 4 = 7$ :

```
fact3 + 4
  isa      addition-fact
  addend1  three
  addend2  four
  sum      seven
```

Production rules specify how to retrieve such declarative knowledge to solve problems. For instance, consider a child working on the 10s column in the following multicolumn addition problem:

$$\begin{array}{r} 234 \\ + 746 \\ \hline 0 \end{array}$$

At this point in time, the following production rule might apply:

```
IF the goal is to add n1 and n2 in a column,
   and n1 + n2 = n3
THEN set as a subgoal to write n3 in the column
```

Applied to the preceding problem, this production rule would retrieve the addition fact  $3 + 4 = 7$  and set the subgoal to write out 7 in the 10s column. At this point, other productions would apply, which would deal with operations like carrying into or out of the column or writing out the answer.<sup>1</sup> All productions in ACT-R have this basic character of responding to some goal, retrieving information from declarative memory, and possibly taking some action or setting a subgoal. In ACT-R, cognition proceeds step by step by the firing of such production rules.

Other aspects of ACT-R involve a theory of subsymbolic, neural-like computations that determine the availability of declarative chunks and

---

1. For a complete model of multicolumn addition, see Anderson (1993, chaps. 1 & 2), where the formal syntax of such production rules is specified.

choice among production rules. These aspects of the theory are important in determining timing when retrieval from memory becomes important and in predicting which paths subjects will explore in complex problems. There is also a learning theory that specifies the acquisition of symbolic chunks and productions as well as subsymbolic continuous-valued quantities associated with the chunks and productions. This learning theory is critical to modeling skill acquisition. We do not elaborate on the subsymbolic or learning aspects of ACT-R because they are not critical to the tasks described here. What is critical is ACT-R's new theory of visual attention.

## 2. A THEORY OF THE VISUAL INTERFACE

Theories of higher level cognition typically ignore lower level processes such as visual attention and perception. They simply assume that lower level processes deliver into working memory some relatively high-level description of the stimulus situation upon which the higher level processes operate. This certainly is an accurate characterization of our past work on the ACT-R theory (e.g., Anderson, 1993). The typical task to which ACT-R has been applied is one in which the subject must process some visual array (the array may contain a sentence to be recognized, a puzzle to be solved, or a computer program to be written). We have always assumed that some processed representation of this visual array is placed into working memory in some highly encoded form, and we modeled processing given that representation.

The strategy of focusing on higher level processes might seem eminently reasonable for a theory of higher level cognition. However, the strategy creates two stresses for the plausibility of the resulting models. One stress is that, by assuming a processed representation of the input, the theorists are granting themselves unanalyzed degrees of freedom in terms of choice of representation. It is not always clear whether the success of the model depends on the theory of the higher level processes or on the choice of the processed representation. Another stress is that theorists may be ignoring significant problems in access to that information that may be contributing to dependent variables such as accuracy and latency. For instance, the visual input often contains more information than can be held in a single attentional fixation, and shifts of attention (with or without accompanying eye movements) may become a significant but ignored part of the processing. For these reasons, we have been encouraged to join several other efforts (e.g., Kieras & Meyer, 1994; Wiesmeyer, 1992) to embed a theory of visual processing within a higher level theory of cognition. The choice to focus on vision is largely strategic, reflecting the fact that most of the tasks that ACT-R has modeled have involved input from the visual modality. To be more exact, most tasks have involved

processing input from a computer screen, and so we developed a theory of the processing of a computer screen. As a fortunate consequence, we have situated ACT-R to be appropriate for HCI applications.

We wanted to remove anything implicit about how our theory related to the behavior we saw from our subjects. To do this, we have our simulation interact with the same software that presents the experiment to the subject. Basically, our ACT-R simulation can operate the computer application just as a subject can: As we describe, the simulation has access to the same computer screens to which the subject has access, must scan these screens as a subject must, and must enter keystrokes and mouse motions as a subject must. The software does not distinguish whether the keystrokes and mouse motions come from ACT-R or a human. The data from the simulation are collected by the same software that analyzes the human's data and are subject to the same analyses. The one difference between our simulation and a human is that ACT-R's whole world is the computer screen, the mouse, and the keyboard, whereas this is only a small part of the human's world.

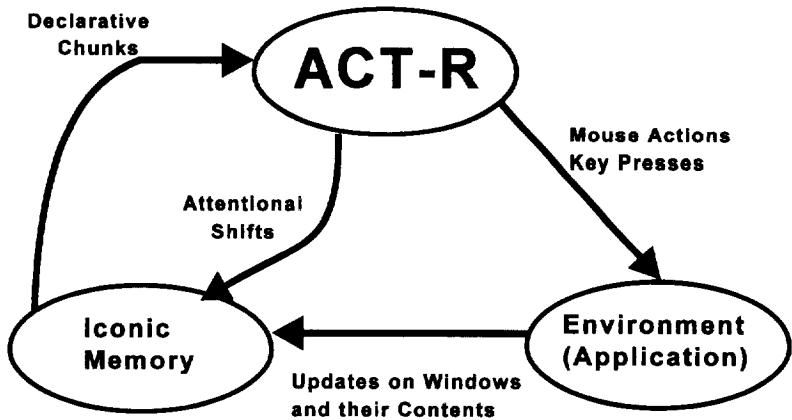
It is important to define our approach from the outset: We require a theory of visual attention and perception that is psychologically plausible, but it is not our intention to propose a new theory of visual attention or perception. Therefore, we have embedded within ACT-R a theory that might be seen as a synthesis of Posner's (1980) spotlight metaphor, Treisman's feature-synthesis model (Treisman & Sato, 1990), and Wolfe's (1994) attentional model. These seem to define the current consensus model of visual attention and perception. What this model does is provide us with a set of constraints that we can then embed within the ACT-R theory of higher level cognition.

Figure 1 provides a basic overview of the system. There are three basic entities to be related. There is the ACT-R system, which we have been describing; there is the environment with which the system is interacting (in our case, the computer application); and there is an iconic memory, which is a feature representation of the information on the screen.<sup>2</sup> As can be seen, there is a limited number of actions that ACT-R can take—it can issue keystrokes and mouse presses to the computer, and it can move its attention around the iconic memory. Wherever it moves its attention, it can synthesize the features located there into declarative chunks that can then be processed by the ACT-R system. The computer program with which it is interacting can issue updates to the screen (and thence to the iconic memory) either spontaneously or in response to actions of ACT-R.

---

2. Thus, what ACT-R "sees" when a word is presented is neither a pixel representation nor a word but a feature description of the word. This corresponds to what the visual cortex receives from lower level routines. ACT-R must recognize the patterns (i.e., words) formed by these feature descriptions.

Figure 1. Relation among ACT-R, the environment, and iconic memory.



We flesh out this basic description in the subsequent sections of this article. In the next section, we focus on the theory of visual attention, which serves to guide attention in its movement about the screen. These are the kinds of cognitive operations that must be addressed in many HCI applications. After describing the theory of visual attention, we address some of the “classic” studies of visual attention. In the final section, we show how these ideas can model, with no additional assumptions, a canonical HCI task—selecting an item from a menu.

### 3. VISUAL ATTENTION

In processing information from a computer screen, we do not have constant access to everything and often have to search for information. On the other hand, we rarely have to do an exhaustive search of the entire screen to find what we are seeking. ACT-R’s theory of visual attention is concerned with how ACT-R finds and extracts information from the iconic memory in Figure 1. The information in the visual icon consists of features, but ACT-R cannot process visual features directly. It can only process chunks representing the objects that these features compose. We have implemented the spotlight metaphor of visual attention, in which a variable-size spotlight of attention can be moved across the visual field. When the spotlight fixates on an object, its features can be recognized. Once recognized, the objects are then available as chunks in ACT-R’s working memory and can receive higher level processing. The following is a potential chunk encoding of the letter *H*:

object	
isa H	
left-vertical	bar1
right-vertical	bar2
horizontal	bar3

We assume that, upon the appearance of an object in the visual field, the features comprising the object (e.g., the bars) are available but that the object itself is not immediately recognized. The system can respond to the appearance of a feature anywhere in the visual field. Only when it has moved its attention to that location can it recognize the conjunction of features that correspond to the object. Thus, for instance, it can respond immediately to a vertical bar but can recognize an *H* only after moving attention to that object. Thus, in order for the ACT-R theory of higher level processing to “know” what is in its environment, it must move its attentional focus over the visual field. In ACT-R, the calls for shift of attention are controlled by explicit firings of production rules. Consequently, it will take time to encode visual information, and we are forced to honor the limited capacity of visual attention.

What information can ACT-R use to guide where it looks on a screen? There are three basic types of information ACT-R can use to guide where attention goes: ACT-R can (a) look in particular locations and directions, (b) look for particular features, and (c) request to scan for objects that have not yet been attended. ACT-R can conjoin these in scanning requests, asking for things like, “Find the next unattended pink vertical bar to the left of the current location.” This kind of search deserves several comments. First, note that ACT-R can search for a conjunction of visual features (pink and vertical). At one time, it had been argued that attention could be drawn only by single features (e.g., Treisman & Gelade, 1980). However, a more current view is that attention can be guided by conjunctions of features but that such conjunction searches are more noisy (Wolfe, 1994).<sup>3</sup> Second, ACT-R can specifically restrict itself to unattended objects. There is evidence that people have difficulty returning attention to attended objects even if they want to (Klein, 1988; Tipper, Driver, & Weaver, 1991). Although ACT-R can restrict itself to unattended objects, it has no more difficulty attending to previously attended objects than to previously unattended objects. Thus, this “inhibition of return” is not modeled in the ACT-R visual component. At some point in time, we might extend ACT-R’s attentional module to incorporate these details, but right now it should be viewed as a system that is consistent at a general level with what we know about visual attention but that does not model the

---

3. However, ACT-R does not yet model this noise.

*Figure 2.* ACT-R can see either the *H* or the *Xs* comprising the *H*, depending on how ACT-R sets its feature scale.

```

X      X
X      X
X      X
XXXXXX
X      X
X      X
X      X

```

microstructure of these attentional processes. As stated earlier, our goal is to focus on how visual attention is used by the cognitive system.

A final general comment is that ACT-R can select the scale of the features for which it searches and the size of the object it is recognizing. Thus, it can recognize either letters or words as objects. Also, depending on how ACT-R sets its feature scale, we would want it to recognize (in Figure 2) either the *H* or the *Xs* comprising the *H*. In fact, subjects can adjust the scale or the spatial frequency at which they are attending to a visual display (Navon, 1977).

The best way to understand how this theory works is to see it applied to various tasks involving visual attention. A constant problem in processing a visual array, such as a computer screen, is finding objects on that screen—whether one is looking for an icon, searching a menu, or scanning a text for a key word. Such visual-search tasks have been a bread-and-butter domain of experimental research on visual attention. There is a rich literature surrounding such tasks, and we want to establish that the ACT-R theory of visual attention is consistent with this literature. We look at three classic paradigms—the Sperling paradigm, subitizing, and speeded search. At one level, the ACT-R theory just implements the existing theoretical understandings of these paradigms. However, the ACT-R implementation does so in precise mechanistic terms and so banishes the homunculus that tends to haunt theories of visual attention. Also, because it places all of these tasks into a single framework, ACT-R allows us to establish that there is an apparent universal of visual attention, which is that it takes about 185 msec to move visual attention. With this parameter established, we are then able to make novel a priori predictions, free of additional parameters, about menu search.

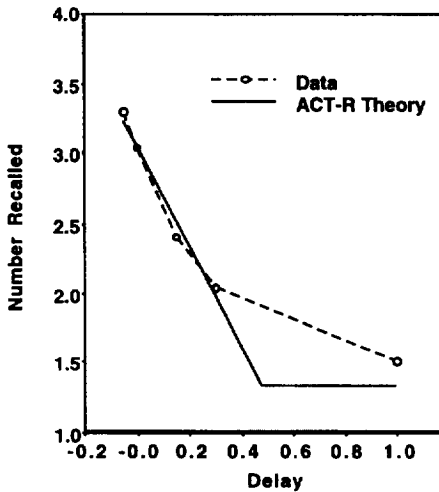
### 3.1. Sperling Task

Sperling (1960) reported a now-classic study of visual attention. Figure 3 illustrates the material Sperling used in one of his experiments. In the whole-report condition, he presented subjects with brief presentations (50

*Figure 3.* An example of the kind of display used in a visual-report experiment. This display is presented briefly to subjects, who are then asked to report the letters it contains.

X	M	R	J
C	N	K	P
V	F	L	B

*Figure 4.* Data from Sperling (1960). Number of items reported from a row of four items as a function of the delay of the cue identifying the row.



msec) of visual arrays of letters (three rows, four columns) and found that, on average, subjects could report back 4.4 letters. In the partial-report condition, Sperling gave subjects an auditory cue to identify which row they would have to report. He found that they were able to report 3.3 letters in that row. When he delayed the presentation of the auditory cue to 1 sec after the visual presentation, he found that subjects' recall fell to about 1.5 letters. Figure 4 shows Sperling's results as a function of the delay in the tone. Because subjects' recall at 1 sec of delay fell to about one third of the whole-report level, the obvious interpretation was that subjects were able to report as many items from the cued row as they happened to encode without the cue. This research has been interpreted as indicating that subjects have access to all of the letters in a visual buffer but that they have difficulty reporting them before the letters decay.

This experiment and other subsequent research have two dimensions of significance. The first dimension is information about the limitations of visual sensory memory. The general importance of this limitation has been questioned (e.g., Haber, 1983), and it certainly does not seem significant to

HCI issues for which we do not receive 50-msec screen presentations. The second dimension of significance is how fast visual attention can move over an array, which is quite relevant to many domains, including the processing of computer screens. For this reason, we believe it is important to show that the ACT-R theory of visual attention can model this result.

We developed a simulation of this task in which the letters in the visual array were encoded by the visual interface as sets of features grouped into unidentified objects. When a report row is not identified, the following production would apply:

**Encode-Screen**

IF one is encoding digits without a tone  
 and there is an unattended object on the screen  
 THEN move attention to that object

After a row has been identified, different productions would fire depending on the tone. For instance, the following production is responsible for reporting the top row:

**Encode-Top-Row**

IF one is encoding digits and there is a high tone  
 and there is an unattended object in the top row  
 THEN move attention to that object

These productions call for attention to be moved to unattended objects. When the production moves attention to the location of that object, the letter would be recognized and a chunk created to encode it. If no tone is presented, **Encode-Screen** will encode any letter in the array, whereas, if a tone is present, productions like **Encode-Top-Row** will encode letters in the cued row. After the visual array disappears, the following production is responsible for report:

**Do-Report**

IF the goal is to report the digits  
 and there is a chunk encoding an item  
 THEN report the item

This production will report only those letters that had been encoded because only these have chunk representations in working memory.

The number of letters encoded in the whole-report procedure is essentially equal to the number of **Encode-Screen** productions that can fire before the iconic memory of the letters disappears. Physically, the stimulus is presented for only 50 msec, but the critical issue is the duration of the stimulus in the system—a parameter we estimate to be 4.4 times the firing

time per production (as 4.4 items are recalled on average). In fitting the data in Figure 4, we estimated the duration of the image to be 810 msec and the time per production to be 185 msec, which is a reasonable estimate for the time for attention to move. Note that  $810 / 185 = 4.4$ .<sup>4</sup>

To understand the fit to the data in Figure 4, we needed to think through how the advantage of the partial report worked. In our analysis, subjects had a one-in-three chance of guessing the right row, in which case they would be able to report the four letters. They had a two-in-three chance of guessing wrong, in which case they would only start encoding the row after switching to that row. We assumed that there was some delay in time for the tone to be perceived and for attention to switch to the correct row (note in Figure 4 that subjects never reported all four items and were doing better given a .05-sec headstart on the tone than a simultaneous presentation). We estimated this switch-over delay to be 335 msec. This can be seen as 150 msec to register the signal (the time for auditory signal to get from the ear to being registered in the goal chunk) and 185 msec for an attention-changing production to fire (same time as value of all other attention-switching productions). Thus, the effective time spent encoding an array if the tone is presented  $t$  msec after the array will be  $810 - t - 335$  msec. Thus, our predicted number of digits reported is:

$$\frac{1}{3} \times 4 + \frac{2}{3} \times \left[ \frac{(810 - t - 335)}{185} \right] = 3.04 - .0036t \quad \text{if } 810 - t - 335 > 0 \text{ (or } t < 475)$$

or

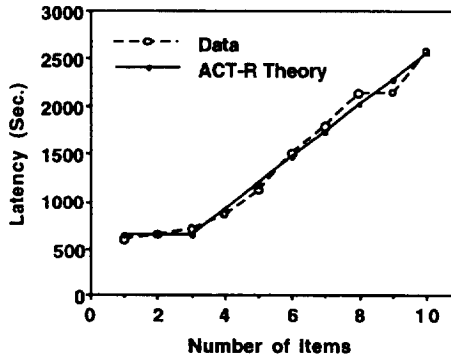
$$\frac{1}{3} \times 4 = 1.33 \quad \text{if } 810 - t - 335 < 0 \text{ (or } t > 475)$$

Figure 4 presents the predictions of this model. As can be seen, the model does a nice job of simulating the data. The ACT-R model of this task is very simple and consists of the production rules given plus a rule to switch from attending to reporting. In part, it is implementing the standard understanding of the data, but it makes clear both the control structure of the task (which is vague in the standard understanding) and the need to postulate the switching time (335 msec) to consistently account for the data. For purposes of comparison with later modeling efforts, the critical number is 185 msec for switching attention. This number comes directly from the slope in Figure 4. Every 185 msec, the memory report is dropping by two thirds of an item.

---

4. An Excel file giving the fit can be found by following the path given in the Background note at the end of this article.

Figure 5. Data from Jensen, Reese, and Reese (1950). Amount of time to name the number of objects in a presentation as a function of the number of objects.



### 3.2. Subitizing Task

In the Sperling task, time is controlled by the duration of the iconic memory, and the goal is to see how many things can be attended to in that time. Another way to measure switching time for attention is to see how long it takes to attend to several objects on a screen. One way to get people to attend to all of the objects on a screen is to ask them to say how many objects there are. This is precisely what is done in a subitizing task (see the recent discussion by Simon, Cabrera, & Kliegl, 1994), in which several objects are presented to a subject, and the subject must identify as quickly as possible how many objects there are on the screen. Figure 5 illustrates the classic result obtained (Jensen, Reese, & Reese, 1950) in this task in which there is an increase in latency with number of digits to be identified. There is an apparent discontinuity in the increase, with the slope being much shallower until three or four items and then getting much steeper. There is about a 50-msec slope until three or four items and approximately a 275-msec slope afterward. Figure 5 also shows the results from the ACT-R simulation that we describe.

The basic organization of the model is to assume that there are special productions that recognize one object, two objects (e.g., lines), three objects (e.g., triangles), and familiar configurations of larger numbers of objects (e.g., the five on a die face) and that there is a production that can count single objects. This is the basic model of the subitizing task that has been proposed by researchers such as Mandler and Shebo (1982). Again, what ACT-R adds to this standard model is an explicit theory of the control structure. The following is some of the productions used in modeling the task:

**Start**

IF the goal is to count the objects starting from a count of 0  
THEN move attention to some object on the screen

**See-One**

IF the goal is to count the objects starting from a count of 0  
and a single object has been seen  
THEN initialize the count to 1

**See-Two**

IF the goal is to count the objects starting from a count of 0  
and a line of two objects has been seen  
THEN initialize the count to 2

**See-Three**

IF the goal is to count the objects starting from a count of 0  
and a triangle of three objects has been seen  
THEN initialize the count to 3

**Attend-Another**

IF the goal is to count the objects and the count is not 0  
and there is another unattended object  
THEN move attention to that object

**Add-One**

IF the goal is to count the objects and the count is not 0  
and another object has been attended  
and X is one more than count  
THEN reset the count to X

**Stop**

IF the goal is to count the objects  
and there are no more unattended objects  
THEN respond with the count

Faced with an array of objects, **Start** will move attention to some part of the screen, and the largest pattern will be recognized. In this model, we have introduced the capacity to see patterns of one, two, and three objects. Depending on which pattern is attended to, one of the productions (**See-One**, **See-Two**, or **See-Three**) will apply to initialize the count. After that point, **Attend-Another** will move attention to other unattended objects, and **Add-One** will add one to the count. When there are no more unattended objects, **Stop** will report the count.

There are several noteworthy aspects of this model. First, it makes clear that successful performance of subitizing depends on ACT-R's ability to tag items in the visual array as attended so that double counts are avoided. Second, beyond three, subitizing depends on retrieval of counting facts. One could have an alternate model that aggregated additional items in units larger than one. Thus, six objects might be achieved by twice attending three objects and adding  $3 + 3 = 6$ . However, retrieval of such addition facts would be much slower than retrieval of counting facts. The model predicts a flat function from one to three and an equal rise from three to four as from four to five—neither of which is quite true. This may reflect the fact that there is some probability of counting in the “sub-three” range and some probability of pattern matching for four elements. Although the model could be complicated to incorporate these ideas, it did not seem worth it. The correlation between prediction and data was already .995.<sup>5</sup>

The most important issue with respect to coordinating this account with our model of the Sperling task is accounting for the 275 msec slope that holds beyond four digits. In fitting this data, we assumed a 185-msec time to switch attention, as in the Sperling model. However, ACT-R does predict the 275-msec slope because it takes approximately an additional 90 msec to retrieve the counting fact in production **Add-One**— $x$  is one more than the count. Although this 90-msec period is estimated for this experiment, it is consistent with estimates in our model of cognitive arithmetic (Lebiere, 1997).

### 3.3. Visual-Search Task

Another way to investigate the time to shift attention is to display an array and ask subjects to search among objects for a specific object. If one can manipulate the number of objects through which a subject must search, one can manipulate search time. The slope of the function relating number of objects attended to search time gives an estimate of time to move attention. This straightforward logic is complicated by the fact that subjects can select which objects to attend to on the basis of the features of the objects. Thus, for instance, in looking for a red object, subjects will not be affected by the number of green objects in the array.

An example reflecting such a paradigm and its complexities is Shiffrin and Schneider's (1977) study of visual search. In their Experiment 2, subjects had to detect a target item when it was presented in a visual display of one to four items (frame size). The target letter was in a memory set of one to four (memory-set size). For instance, subjects might hold a memory set of B and K and be asked if either element occurred in a visual

---

5. An Excel file giving the fit can be found by following the path given in the Background note at the end of this article.

array that contained G, K, M, and F (in which case, they would respond yes). Subjects were either in what was called the varied-mapping condition or what was called the consistent-mapping condition. In the varied-mapping condition, both distractors and the memory-set items were letters (drawn from the same pool on each trial); in the consistent-mapping condition, the memory set was composed of numbers, and the distractors were letters (therefore, they were always drawn from different pools). In general, judgment times increase with memory-set size and frame size, but the effects are much stronger for the varied-mapping condition. Figure 6 (top panel) shows Shiffrin and Schneider's results.

We developed what is a fairly straightforward ACT-R model of this task. It involved the following stages:

1. *Preparation.* Upon receipt of the memory set, an effort was made to find a feature common to all members of the memory set. If there was more than one such feature, then the feature was selected that was least frequent among the distractors. If no one feature characterized all of the items in the memory set, two or more features could be selected. This defined the target feature set. The feature set we used was the features proposed by McClelland and Rumelhart (1981) plus one global feature to encode whether the character was left-facing, right-facing, or symmetric.
2. *Search.* ACT-R directed attention to a location on the basis of the target set of features. Upon presentation of the display, the system examined all positions that had a target feature. If no position had a target feature, it randomly selected one position to view. It would look at more only if more than one position had a target feature. This search was self-terminating in the case of positive trials, but all positions with target features had to be examined in the case of negative trials. The first production that applies to start the scanning is:

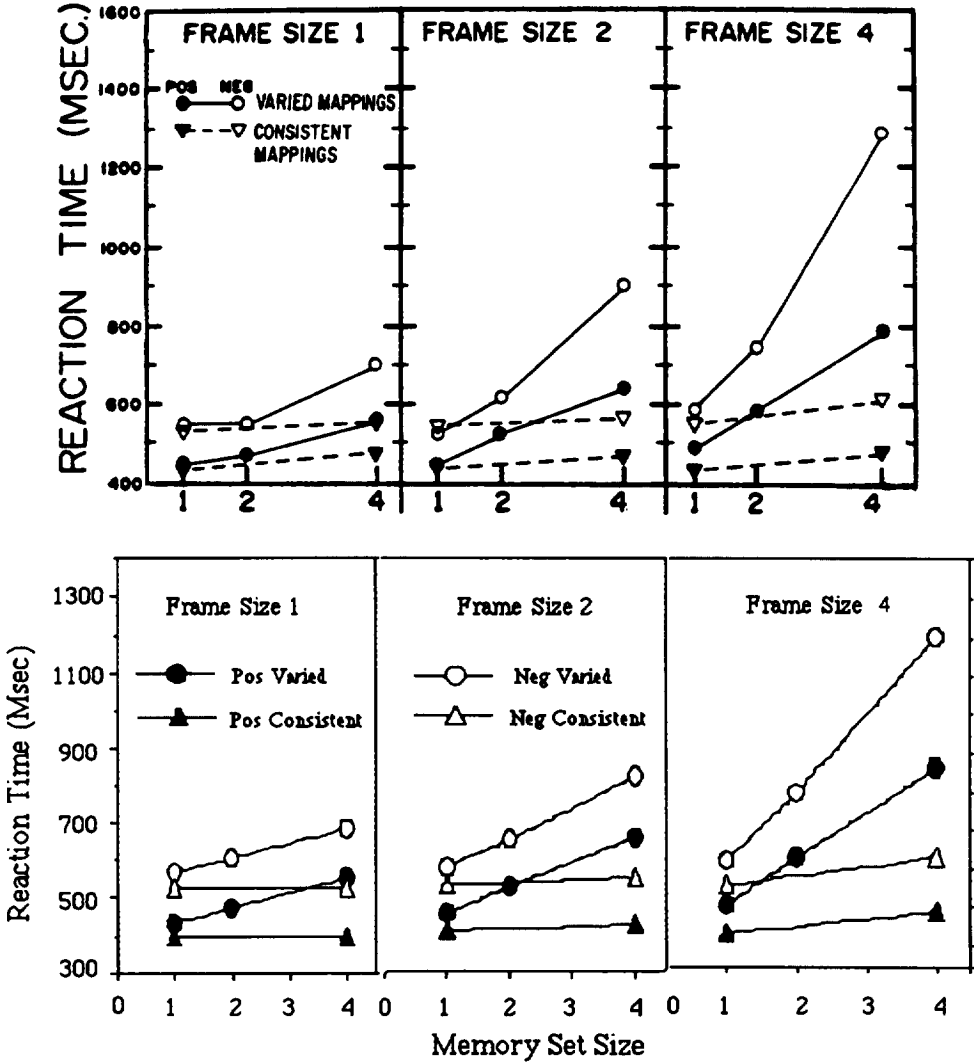
**Encode-First-Object**

**(SHIFT)**

If the goal is to search for an object with feature F  
 and an unattended object with feature F occurs in location L  
 THEN move attention to that object

3. *Judgment.* For each position examined, it decided whether that item was in the memory set. In the consistent-mapping condition, this could be done by simply judging whether the item was a number, which could be done by direct retrieval of a category label. In the varied-mapping condition, it was necessary to determine if the item was in the memory set. This did not require a sequential search but was done by a production pattern-match test whose time increased with the size of the set. This is analogous to the existing ACT model

Figure 6. Results of Shiffrin and Schneider's (1977) simulation (top panel) and result of the ACT-R simulation (bottom panel). Note. Top panel from "Controlled and automatic human information processing: I. Detection, search and attention," by W. Schneider and R. M. Shiffrin, 1977, *Psychological Review*, 84(1), p. 19. Copyright © 1977 by the American Psychological Association.



for fan experiments and the Sternberg task (see Jones & Anderson, 1987). The basic productions for the two conditions are:

<b>Judge-Consistent-Positive</b>	<b>(BASE)</b>
IF the goal is to search for an object with feature F and object attended is a number	
THEN pop the goal and respond yes	
<b>Judge-Consistent-Negative</b>	<b>(SHIFT)</b>
IF the goal is to search for an object with feature F and object attended is a letter	
THEN search for another object with feature F	
<b>Judge-Variied-Positive</b>	<b>(BASE + i * FAN)</b>
IF the goal is to search for an object with feature F and object attended is in the memory set	
THEN pop the goal and respond yes	
<b>Judge-Variied-Negative</b>	<b>(SHIFT + i * FAN)</b>
IF the goal is to search for an object with feature F and object attended is not in the memory set	
THEN search for another object with feature F	
<b>Terminate-No</b>	<b>(BASE + NEG)</b>
IF the goal is to search for an object with feature F and there are no unattended objects with feature F	
THEN respond no	

According to this model, the consistent-mapping condition enjoys two advantages over the varied-mapping condition. First, fewer positions will have to be examined because numbers have fewer features in common with letters than do letters. In the extreme condition (frame size = 4, set size = 4, negative trial), an average of 1.56 item positions had to be examined in the consistent condition and 2.49 item positions in the varied condition. The second advantage is that the target set did not have to be examined during judgment in the consistent condition, and so the condition did not suffer a fan effect. However, there still is a small effect of memory-set size in the consistent condition because there will be more target features to discriminate the targets from the letters as the memory-set size increases.

We developed a mathematical model of this ACT-R theory and fit it against Shiffrin and Schneider's (1977) data. This model required four parameters—a base reaction time (BASE, estimated to be 208 msec), an additional waiting time associated with **Terminate-No** for a negative response (NEG, estimated to be 133 msec), a time to attend to a position (SHIFT, estimated

to be 186 msec), and a fan time per element (*FAN*, estimated to be 40 msec). The parameters associated with each production were given along with the productions. The predictions of the model are displayed in Figure 6 (bottom panel). The  $R^2$  between the data and the predictions is .951, and the average mean deviation in prediction is 38 msec. The parameter values are also quite reasonable. Note the fan parameter is the slope in the typical Sternberg (1969) task. The time to attend to a position is reassuringly close to the estimate (185 msec) we obtained in fitting the Sperling task. Both the fan costs and the negative costs reflect the kinds of times we estimated previously for the effect of these factors on production matching.<sup>6</sup>

### 3.4. Conclusions

We have shown that the ACT-R model is consistent with some of the classic results from visual attention. In each of three tasks, we were able to fit the data assuming just about 185 msec to switch attention. In the Sperling task, attention switching was the only activity. In the subitizing task, there was also time required to set up and increment a count. In the Shiffrin and Schneider (1977) task, judgment time played a significant role. When we go to more cognitively loaded tasks, other processes will play still more significant roles. However, every time visual attention switches, approximately another 185 msec will be added to the processing time.

## 4. APPLICATION TO MENU-SELECTION DATA

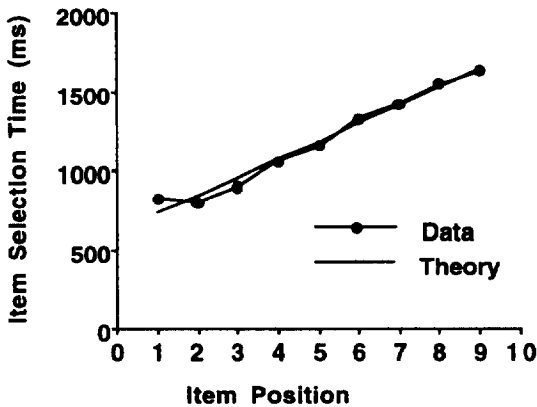
We thought it would serve as a useful illustration to apply this ACT-R system to the same Nilsen (1991) data described by Kieras and Meyer (1994) in their report of the EPIC model. These data are concerned with time to scan a menu as a function of the target position of the item in the menu. The menu consists of a set of digits (1 to 9) randomly ordered vertically. The data to be modeled are the times for subjects to move a mouse from home position above the menu to the target item. Figure 7 shows the time for this action as a function of the serial position of the item in the menu. A linear function is obtained with a slope of 103 msec per position.

It is critical for this study that the items in the menu are ordered randomly. Because the subject does not know where the target item is, a critical component to latency has to be a serial search of the list looking for the target item. Subjects tend to move the mouse down as they scan for the target. Thus, after they identify the target, the distance to move the mouse tends not to vary much with serial position. Thus, our view is that, when the

---

6. An Excel file giving the fit can be found by following the path given in the Background note at the end of this article.

Figure 7 Observed and predicted menu-selection times. Observed data are from Nilsen (1991).



target position is unknown, time is dominated by visual search. In contrast, if the position of the item was known (as in a fixed order menu), the critical latency component might be a Fitts's-law description of the motion.

Our model for this task was essentially the same model as we proposed for Shiffrin and Schneider's (1977) data (Figure 6). We assume that, given a target, subjects selected one of its features and scanned down the menu for the first item with that feature. If this was the target, they stopped. If not, they scanned for the next item that contained the target feature.

The two critical productions are:

#### Hunt-Feature

IF the goal is to find a target that has feature F  
 and there is an unattended object below the current location with feature F  
 THEN move attention to closest such object

#### Found-Target

IF the goal is to find a target  
 and the target is at location L  
 THEN move the mouse to L and click

The first production **Hunt-Feature** moves attention down looking at objects that have the target feature. The movement of attention to an object will cause its identity to be encoded. If it is an instance of the target letter, **Found-Target** can apply. The production **Found-Target** will retrieve the location of the target and move the mouse to that location.

The time to reach a target will be a function of the number of digits that precede it that have the selected feature. It turns out, given the McClelland

**Figure 8. Interaction between target and background.**

Target	Background	
	Number <sup>a</sup>	Letter <sup>a</sup>
Number	1,324	1,293
Letter	1,253	1,366

<sup>a</sup>In milliseconds.

and Rumelhart (1981) feature set, there is a .53 probability that a randomly selected feature of one number will overlap with the feature set of another number.<sup>7</sup> Using the estimate (from Shiffrin & Schneider, 1977) of 186 msec for a shift of attention, we predict  $186 \times .53 = 99$  msec per menu item, which is close to the slope (103 msec) in the Nilsen (1991) data. The fit of our model to the data is illustrated in Figure 7. This is a striking demonstration of how the ACT-R theory can be used to predict new data sets using old parameters.

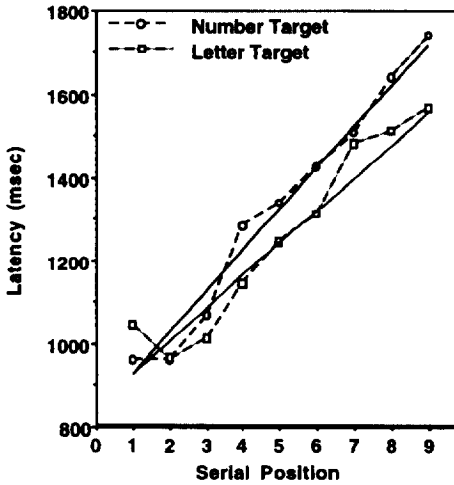
Kieras and Meyer's (1994) EPIC model is able to do an equally good job assuming a pipeline model whereby there are eye movements every 103 msec that will overrun the target. This strikes us as a very improbable speed of eye movement, which is conventionally set at about 200 msec. Kieras and Meyer suggested an alternative model in which as many as three items are processed in each gaze. Either of these models would predict no effect of distractor similarity on search time. In light of studies like Shiffrin and Schneider's (1977), this prediction seems unlikely. In contrast, the ACT-R model would predict, for instance, that it would be easier to find a number in a menu of letters than in a menu of numbers.

To test this prediction, we performed a within-subject menu-search task in which subjects had to select either a capital letter or a digit in a background of letters or digits. Figure 8 presents the results from subjects for menus of nine elements, as in Nilsen (1991). As predicted by ACT-R, subjects are significantly,  $F(1, 20) = 104.77$ ,  $p < .01$ , faster when the distractors are different than the target. This is a confirmation of ACT-R's conception of visual attention and a token of its potential for modeling HCI tasks.

Although the interaction is to be expected, there is one unexpected result in the data—that there is a significant effect of background, with subjects slower (41 msec) in the presence of a letter background,  $F(1, 20) = 29.96$ ,  $p < .001$ . We have no explanation of this effect.

7. Compared to Shiffrin and Schneider (1977; consistent-mapping condition), we did not assume that subjects had enough practice to select the most discriminating feature.

Figure 9. Observed and predicted selection times for numbers versus letters against a letter background. The predictions of the ACT-R theory are given in the solid lines.



The strongest prediction of the ACT-R theory is that there should be a significant Serial Position  $\times$  Target  $\times$  Background interaction. In fact, there are significant Target  $\times$  Position,  $F(8, 160) = 6.49$ ,  $p < .001$ , Background  $\times$  Position,  $F(8, 160) = 4.30$ ,  $p < .001$ , and Target  $\times$  Background  $\times$  Position,  $F(8, 168) = 2.18$ ,  $p < .05$ , interactions. There are significant differences among slopes, with 103 msec in the number-on-number condition, 84 msec in the number-on-letter condition, 80 msec in the letter-on-number condition, and 82 msec in the letter-on-letter condition. The basic effect is a steeper slope in the number-on-number condition. We calculated the mean probability of feature overlap in the conditions. There is a 53% probability overlap of the number-on-number condition, 39% in the number-on-letter condition, 42% in the letter-on-number condition, and 43% in the letter-on-letter condition. Thus, these overlap scores predict that there will be less ability to use features to guide search in the number-on-number condition. This prediction is confirmed.

Figure 9 plots the predictions of the ACT-R theory for number and letter targets holding constant the background as numbers. ACT-R is already committed as to the slopes in these cases. For number targets, it is  $186 \times .53 = 99$  msec (actual slope = 103); for letter targets, it is  $186 \times .42 = 78$  msec (actual slope = 80 msec). The only degree of freedom in estimating this is the "intercept" when the serial position is one. This was estimated as 927 msec. This is a striking confirmation of the ACT-R analysis of menu scanning in comparison to the EPIC model, which fails to predict these effects of Target  $\times$  Background interaction. It also serves more generally to indicate the relevance of research on visual attention to HCI.

## 5. CONCLUSIONS

Our goal in this article has been to describe how we have given eyes to ACT-R. Although the model has a theory of visual perception, we have not concentrated on this but rather have focused on the important role of visual attention in accounting for data patterns. Elsewhere (Anderson & Douglass, 1998), we focused on the important role of visual attention in classic problem-solving tasks such as equation solving. However, here our goal has been to show that we properly model the basic processes of visual attention and that they matter in a traditional HCI task such as menu scanning. A critical value was the approximately 185 msec involved in shifting attention to an item in a visual array. However, if this value is simplistically applied to a task, one can overestimate the time to shift attention because attention has the capacity to focus on items with specific features, and one needs to consider the implication of this focus for search time. For instance, Figure 9 shows that differential focus will result in differential search speed. ACT-R provides an architecture in which to work out these complex interactions with visual attention for both simple and complex tasks.

---

## NOTES

**Background.** ACT-R and its visual interface can be accessed at its website: <http://act.psy.cmu.edu/act/>. This contains a pointer to the ACT-R visual interface page, where the simulations and Excel parameter estimation files can be found.

**Acknowledgments.** We thank Chris Schunn for his comments on the research.

**Support.** We acknowledge Office of Naval Research Grant N00014-96-I-041 in supporting this research.

**Authors' Present Addresses.** John R. Anderson, Michael Matessa, and Christian Lebiere, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: ja+@cmu.edu; mm4b@andrew.cmu.edu; cl+@cmu.edu.

**HCI Editorial Record.** First manuscript received January 20, 1996. Revisions received June 18, 1996, and December 1996. Accepted by Wayne D. Gray. Final manuscript received February 17, 1997. — *Editor*

---

## REFERENCES

- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4, 167-207.
- Anderson, J. R., & Douglass, S. (1998). *Visual attention and problem solving*. Manuscript in preparation.

- Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, *30*, 221-256.
- Haber, R. N. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information processing. *Behavioral and Brain Sciences*, *6*, 1-11.
- Jensen, E. M., Reese, E. P., & Reese, T. W. (1950). The subitizing and counting of visually presenting fields of dots. *Journal of Psychology*, *30*, 363-392.
- Jones, W. P., & Anderson, J. R. (1987). Short- and long-term memory retrieval: A comparison of the effects of information load and relatedness. *Journal of Experimental Psychology: General*, *116*, 137-153.
- Kieras, D. E., & Meyer, D. E. (1994). *The EPIC architecture for modeling human information-processing and performance: A brief introduction* (Report 1, TR-94/ONR-EPIC-1). Ann Arbor: University of Michigan, Department of Electrical Engineering.
- Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature*, *334*, 430-431.
- Lebiere, C. (1997). *An ACT-R model of cognitive arithmetic*. Dissertation proposal, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*, 1-22.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*, 375-407.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353-383.
- Nilsen, E. L. (1991). *Perceptual-motor control in human-computer interaction* (Technical Report 37). Ann Arbor: University of Michigan, Cognitive Science and Machine Intelligence Laboratory.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3-25.
- Servan-Schreiber, E. (1991). *The competitive chunking theory: Models of perception, learning, and memory*. PhD dissertation, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA.
- Shiffrin, W., & Schneider, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1-66.
- Simon, T., Cabrera, A., & Kliegl, R. (1994). A new approach to the study of subitizing as distinct enumeration processing. *Proceedings of the sixteenth annual conference of the Cognitive Science Society*, 929-934. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Sperling, G. A. (1960). The information available in brief visual presentation. *Psychological Monographs*, *74*(Whole No. 498).
- Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist*, *57*, 421-457.
- Tipper, S. P., Driver, J., & Weaver, B. (1991). Object centered inhibition of return of visual attention. *Quarterly Journal of Experimental Psychology*, *43A*, 289-298.

- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.
- Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 459-478.
- Wiesmeyer, M. D. (1992). *An operator-based model of covert visual attention*. Unpublished PhD thesis, Department of Computer Science, University of Michigan, Ann Arbor.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review, 1*, 202-238.