Automaticity and the ACT* theory

JOHN R. ANDERSON
Carnegie Mellon University

An explanation of automaticity within the framework of the Adaptive Control of Thought (ACT*) production system theory (Anderson, 1983, 1987) is presented. There is no automaticity mechanism per se in ACT*. This is as we would expect it to be. It would be the exception rather than the rule that we would find in a scientific theory mechanisms that directly correspond to natural language concepts. The critical question is whether ACT* can give an account of the phenomena associated with the term automaticity. This article is structured as follows: First, I will try to identify the phenomena of automaticity to be explained, then give a brief overview of the ACT* theory, and finally explain how these phenomena of automaticity are to be understood in terms of the theory.

Automaticity and Related Phenomena

In the context of this special issue on automaticity, no detailed descriptions of the phenomena of automaticity are required. It is sufficient to identify automaticity phenomena I will be addressing. I will restrict my attention to automaticity phenomena that accrue with practice of a particular skill. Another use of the term automaticity is to refer to effortless extraction of features in perception but, as discussed in this issue by Treisman, Vieira, and Hayes (1992), there is evidence that these are not the same thing. These perceptual phenomena are beyond the bounds of the ACT* theory. Still another use of automaticity refers to some cognitive process whose operation is not subject to conscious control. Thus, we may say that spread of activation is automatic or encoding of frequency information is automatic. Such phenomena do have interpretation in terms of basic ACT* cognitive mechanisms, but I do not think these are the same as the automaticity properties that accrue as a function of practice of a skill. Logan (in press) has argued that we may need to treat innate automatisms and acquired automatisms separately. I will be concerned with acquired automatisms. Indeed, in what follows (particularly points 2 and 3) I will be concerned really with the effects of practice on skilled performance:

1. Skill speeds up with practice and reduces in error rate. Both

AMERICAN JOURNAL OF PSYCHOLOGY
Summer 1992, Vol 105, No. 2, pp. 165–180
© 1992 by the Board of Trustees of the University of Illinois

performance improvements show up as power law functions (Newell & Rosenbloom, 1981). A power function relates a performance measure, P, to a number of trials of practice, N, by a function of the form $P = AN^{-b}$, where b is an exponent usually less than 1 (which may be interpreted as the learning rate) and A is a measure of performance on the first trial. The fact that it is a power function is not a trivial result. As Mazur and Hastie (1975) note, many have predicted that the learning function should be exponential.

Much recent attention has been paid to the power law improvement in time, but not in error rates. Perhaps one of the reasons for neglect of error rate measures is that in the domain of skilled performance they tend to be relatively noisy and unstable compared with reaction time measures. Logan (1988) has noted that not only do reaction times decrease as a power function of practice, but the standard deviation of these times also decreases as a power function.

2. A frequent belief is that well-practiced skills do not decay with disuse. However, this is only because the amount of forgetting appears small relative to the amount of improvement with practice. For instance, in Kolers's (1976) classic study of learning to read text in different orientations, subjects improved from 15 min per page to 2 min per page over 200 pages and slipped back to 4 min per page after a year's delay. Furthermore, the rate of forgetting with time appears to be a power function also (Anderson & Schooler, 1991; Wickelgren, 1974). Again, a power function is a nontrivial outcome, and others have expressed the opinion that the forgetting function should be exponential (e.g., Loftus, 1985).

3. All practice is not equivalent. In particular, spaced practice is much more effective than massed practice (Bahrick, 1979; Gay, 1973). Note that this spacing effect, along with the practice and retention effects in (1) and (2), creates a much more complex picture of the underlying practice effects that produce automaticity than one typically gets from the literature that simply reports number of trials of practice and reaction time measures.

4. As a skill becomes more practiced, it interferes less with a concurrent task and is less interfered with by a concurrent task.

5. It is relatively difficult to inhibit an automatic process, and thus an automatic process can be more interfering to another ongoing task. Thus, a process like reading a word may run off even though the experimental task is to name the color of the word. Note that this phenomenon pulls in the opposite direction of the previous one, where we noted that automatic processes are less interfering.

6. Automatic processes are less slowed down by the number of alternatives. The classic domain for showing this is the Shiffrin and

Schneider (1977) search task. Often the results are described as showing no effect of number of alternatives. However, inspection of most results shows diminished effects of number of alternatives rather than no effect. If the effect is not significant in some experiment, this says more about the statistical power of that experiment than about whether there is really no effect of number of alternatives.

7. It is easier for a task to become "automatic" if there is a consistent stimulus-to-response mapping. In the context of this statement, the term *automatic* connotes fast processes, little interference by concurrent processes, and little effect of number of alternatives.

The ACT* Theory

According to the ACT* theory, all cognitive behavior is controlled by production rules. Production rules specify the steps of cognition. A typical rule might take the form of the following production which adds two numbers:

IF the goals is to find the sum of $n1$ and $n2$
and $n1 + n2 = n3$
THEN say $n3$.

There are a number of noteworthy features of this production rule. First, it is evoked in response to a specific goal. Second, it is variablized with $n1$, $n2$, and $n3$ serving as slots for specific numbers. Third, this production rule requires retrieving from long-term memory a specific sum (to match to $n1 + n2 = n3$). Thus, if the goal were to find out what 6 and 5 are, in the second line the production rule would have to retrieve "6 + 5 = 11." The long-term memory that holds this fact is referred to as declarative memory to contrast it with production memory. In addition to this general rule, one can have a specific production rule of the form:

IF the goal is to find the sum of 6 and 5
THEN say 11.

This rule does not require a retrieval from long-term memory. One of the learning processes of interest is how one transits from the first, general rule to the second, specific rule.

Knowledge in production rule form is distinguished from knowledge in declarative form in that the production is committed to a specific use. The production rule above is for specifically saying the sum. Different rules would be required to express the 6 + 5 fact as an intermediate step in performing multiple-column addition, a different rule for subtraction, and yet another rule to answer a question such as "What is the relation between 6, 5, and 11?" Consult Anderson, Conrad, and Corbett (1989) and Singley and Anderson (1989) for evidence that knowledge develops in this use-specific way.

There are three relevant learning processes in the ACT* theory. The first concerns encoding knowledge directly from experience. This knowledge is encoded in a declarative form. The second learning process is associated with converting declarative knowledge into a production rule form. Suppose subjects are told, "You are to say *yes* if the digit that appears on the screen is in the set 2,7,4." Suppose the subjects categorize their goal as classify the elements on the screen and that elements in the set 2,7,4 are associated with the category label *yes*. Assuming that they have never done the Sternberg task before, they would not have productions to perform that specific task, but it might not be unreasonable to credit them with more general production rules such as the following:

P1 IF the goal is to categorize an element in location *X*
and *n*1 is in location *X*
THEN set as subgoals to determine the category of *n*1
and to say the category name of *n*1.

P2 IF the goal is to determine the category of *n*1
and *n*1 is in set *X*
THEN *n*1 is in category *X*.

P3 IF the goal is to say the category name of *n*1
and *n*1 is in category *X*
and *R* is the response for *X*
THEN say *R*.

The first production rule would recognize when the digit came on the screen and set subgoals to classify it and say the appropriate response. The second one would recognize category membership based on an extensional definition of the category. The third production rule would retrieve *yes* as the appropriate response.

Except for the goal elements, the condition parts of these production rules require declarative task-specific knowledge to be available: The second clause of P1 requires something like (a) "4 is on the screen" to be available. The second clause of P2 requires something like (b) "4 is in the memory set." The second and third clauses of P3 require things of the order (c) "4 is in the category of the memory set" and (d) "*yes* is the response for the memory set." Fact (a) is an encoding of the perceptual array and (c) is created by P2. However, (b) and (d) represent declarative encodings in memory of the instructions.

Performing the task this way is slow and laborious. With practice, subjects quickly convert to production rules of the form

P4 IF the goal is to categorize an element on the screen
and 4 is on the screen
THEN say *yes*.

This production rule eliminates the need to retrieve the declarative instructions and performs in a single step what took three steps previously to put together. The process of producing task-specific productions like the above is called *knowledge compilation* in the ACT* theory. It can serve to eliminate multiple production firings and to eliminate the need for retrieval from declarative memory. Knowledge compilation does not eliminate the original production rules or declarative knowledge. Rather those remain and are alternative bases for performing the task. Which path the subjects take depends on conflict resolution principles to be discussed shortly.

The third learning process in ACT* involves strengthening the production rules and declarative facts. Each time one of these production rules or declarative facts is used, its strength is increased one increment. The strength of a declarative fact determines how active it is. The actual selection of a production rule is determined by a competition among production rules in which they compete for the activation of the data elements to which they match. Stronger productions do better in this competition.

Conflict resolution refers to the principles that determine which production rule is executed. According to the ACT* theory, the five factors (*A*, *G*, *S*, *I*, *N*) determine the speed and probability of a production rule applying:

A. The level of activation of the data elements to which it matches.
G. The degree to which the condition of the production rule is matched. In the ACT* theory, it is possible for a production rule to fire even if there is only a partial match.
S. The strength of the production rule.
I. The strength of production rules that match to the same declarative elements as the production of interest.
N. The degree of overlap in the data elements to which the competing productions match.

Anderson (1983) noted that these five factors were approximately multiplicative and proposed that speed and performance of a production would vary approximately with the quantity:[1]

$$\text{Performance} \sim \frac{SAG}{NI} \qquad \text{(Proportion 1)}$$

Explanation of Automaticity Phenomena

Although the ACT* theory does not have the concept of automaticity per se, an explanation of most of the phenomena listed for automaticity can be found in its principles of learning and conflict resolution. The most important construct is the buildup of strength of a production rule. To an approximation, we may say that a production is automatic to the degree that it is strong. Of course, a particular skill may involve many productions at different levels of strengthening.

Strengthening phenomena

The actual principles for buildup of strength in ACT* require some discussion. The basic assumption is that every time a production is practiced it receives the same increment in strength. As we shall see, the consequences of this obvious assumption may not be so obvious, but before unpacking these complexities, let us consider the simplified assumption that if an item has had P exposures it has strength P. One might also assume, as was the case in Anderson (1983), that reaction time was inversely related to strength such that the time, T, to retrieve an item with P practices would be given by

$$T = B + a/P \tag{1}$$

where B reflects some intercept unaffected by practice. This is a considerable simplification because it does not reflect factors of data activation, degree of match, and competing productions contained in Proportion 1. However, for now we will assume that those are constant and all summarized in the constant a. This leads to a prediction that reaction times decrease as a hyperbolic function of practice which is a special case of a power function with exponent of -1. These are not what are typically fit to learning curves, but as Newell and Rosenbloom (1981) and Mazur and Hastie (1975) note, in slightly generalized form they actually do a very good job of fitting both reaction time and error rate data.[2]

The error rate function that Anderson (1983) proposed to relate strength to errors was

$$E = e^{-PK}. \tag{2}$$

Such an exponential function is provably wrong as the learning function (Mazur & Hastie, 1975). The decrease in errors is better fit by a power function or by a generalized hyperbolic function. It would not much change the ACT* theory if we proposed such a power function:

$$E = CP^{-d} \tag{3}$$

where C and d are parameters to be estimated.

Forgetting and spacing

Unfortunately for the simplicity of the story, there is good reason to doubt a simple linear relationship between strength and amount of practice. Thus, if strength and not P is the controlling variable, then the reaction time and error functions above are not universally accurate. The fundamental reason for doubting a linear relation is the evidence that there is a decay in strength with delay. Because forgetting follows a power function, Anderson (1982) proposed that the total strength of a trace would be the summation of individual strengthenings of the traces, each decaying according to a power function:

$$S = \sum_{i}^{P} t_i^{-d} \tag{4}$$

where the summation is over the P occurrences of the item, t_i is the time since the ith occurrence and t_i^{-d} is the strength of the ith trace. Anderson (1982) showed that under the special case in which the individual practices are uniformly spaced, this leads to prediction of a power function for practice with exponent $1-d$. That is, the strength of the item is approximately

$$S = aP^{1-d}. \tag{5}$$

To accommodate spacing data, Anderson and Schooler (1991) proposed a further complication to this formula. We proposed that the rate of decay of a particular strengthening itself decays with the time since the last strengthening. Thus, we need to replace d above by a different d_i for each presentation. The formula we proposed for d_i had the following form:

$$d_i = (t_i - t_{i-1})^{-r}. \tag{6}$$

This we were able to show fits an extraordinarily wide range of data about the effects of the number of practices, their timing, and retention delay. Certain assumptions in this formulation are basically unmotivated—in particular, the proposal of power law decay in strengthenings and in the decay in the rate of decay. Anderson and Schooler showed that these assumptions allow the system to be optimized to the environment and suggested that the ultimate explanation of these assumptions may lie in the way that the need to use information may vary over time in the environment. For current purposes, we regard these as primitive properties of the human system, perhaps reflecting how neural connections decay with disuse.

We think it is critical that other theories of power-law learning account for the effects of forgetting and spacing. These effects can be extremely potent and overwhelm any effect of practice per se. One does not have a theory of how practice improves performance if one cannot account for the role of these temporal factors. It may be possible for other theories to incorporate the kind of model I have sketched out here and then proceed forward as they did before. However, they need to show that this is possible or that it is possible in some other way to account for temporal factors.

Other practice considerations

Logan (1988) has made much of the fact that his theory also predicts a power law decrease in the standard deviation of the reaction times. It turns out that this is also true of ACT* as described in Anderson (1983). In ACT*, reaction times had an exponential distribution. The standard deviation of an exponential distribution is equal to its mean, and any process that reduces the mean will reduce the standard deviation. More generally, one might assume a gamma distribution for reaction times in which standard deviation will be proportional to mean if not equal.[3] A gamma distribution is a very common reaction time distribution.

It is often thought (e.g., Carlson, Sullivan, & Schneider, 1989; but see Anderson, 1989) that the knowledge compilation processes in ACT* imply that performance is some mixture of precompiled processing and compiled processing and that the course of learning should be determined by the movement from precompiled to compiled form. However, as developed in Anderson (1982), the course of learning is really controlled by the buildup of strength which is the precondition for complex compilations. The productions and declarative knowledge involved in a complex compilation must be sufficiently strong so that all the relevant information remains active in working memory.

The initial conversion of declarative knowledge into procedural form takes place quite rapidly. Figure 1 illustrates some data collected from our LISP Tutor where we looked at the learning of production rules over trials. The data is plotted on log-log scales to bring out a power function as a straight-line relationship. Clearly, the first trial is discontinuous from the others which may be in a slow power-law learning function. We have frequently found this first-trial discontinuity in our studies of skill acquisition. In most research a peculiarity of the first trial gets lost in practice trials. However, in our tutoring

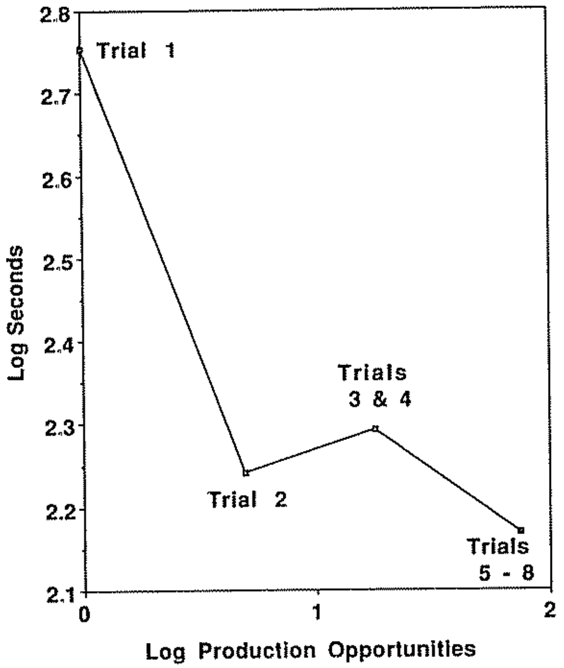

Figure 1. Relationship between number of opportunities to practice a production rule and speed in the LISP Tutor. Data are plotted on a log-log scale.

work we are motivated to focus on the first few trials because that is all the practice many productions get.

Interference phenomena

To understand interference relationships involving automaticity, let us return to Proportion 1 which asserts performance of a production is proportional to the quantity *SAG/NI*. Clearly, as we increase the strength, *S*, of the rule we make it less vulnerable to the factors *N* and *I* which reflect competing rules. Thus, it is easier to perform the task in the presence of a concurrent task as practice increases.

However, let us switch perspectives and consider the impact of practicing another task on performance of a concurrent task. Now the effect of practice will show up in *I* in our production rule formulation. What effect *I* has is determined by *N*. If the competing production rules do not overlap in the data to which they match, there will be no effect. If they do overlap, there will be an interfering effect.

It is useful to consider the application of the ACT* theory to patterns of interference and facilitation in the Stroop task. The typical Stroop task looks at the interaction between naming the color of a word and naming the word. Typically, it is shown that there is an effect on the color-naming task of the word task, but MacLeod and Dunbar (1988) showed that it is really a matter of which task is stronger and that the stronger task will impact the weaker task. Let us consider the proposal in Anderson (1983) for what color-naming and word-naming productions would look like:

IF the goal is to name the color of a word
and the word is in color *X*
and *Y* is the articulatory code for *X*
THEN say *Y*.

IF the goal is to name a word
and the word is spelled as *X*
and *Y* is the articulatory code for *X*
THEN say *Y*.

These two production rules overlap in the third elements of their conditions where they access the articulatory codes. There may be other overlaps among the productions depending upon representational assumptions that I do not want to get into here. For current purposes, it is sufficient to note that in the *SAG/NI* ratio, *N* will not be negligible and there will be a definite interference of *I*. Indeed, it is possible for the wrong production, which partially matches, to fire occasionally. Whether this misfiring is interfering or not depends on whether the word is printed in the same color as its name. If it is, we will get facilitation due to being able to use the fastest of two productions. If it is not, it will lead to an error.

Facilitation phenomena

Now we come to the observation that with practice some skills can become less interfering to concurrent processes. It is an implication of Proportion 1 that this cannot happen if productions stay constant. Increasing the practice of a production will increase its strength and so its tendency to interfere with concurrent productions. The explanation of such facilitation phenomena in the ACT* theory comes from the knowledge compilation process which can produce new productions that overlap less with concurrent productions and also require less maintenance of information in working memory.

Consider the experiment by Spelke, Hirst, and Neisser (1976) who found that by practicing taking dictation subjects were able to reach the point where it no longer interfered with reading comprehension.

Dictation and comprehension productions would overlap to the extent that they made reference to the common verbal and semantic information about words. If dictation productions were compiled so that they went from the phonetic representation directly to the transcription, this overlap and basis for interference would be eliminated. Also, to the extent that subjects needed to rehearse the dictation information to maintain it for transcription, this would take away from rehearsing and keeping active the text material being comprehended. Again, a transformation of dictation productions to something like stimulus-response associations would eliminate this potential for interference.

Number of alternatives

As a skill becomes more automatic, the effect of number of alternatives decreases. Figure 2 illustrates some data of mine (Anderson, 1983) looking at the effect of practice of fact recognition. Subjects practiced recognizing the same set of sentences for 25 days. In the no-fan condition, there was only one fact associated with each word in the sentence to be recognized. In the fan condition, there were

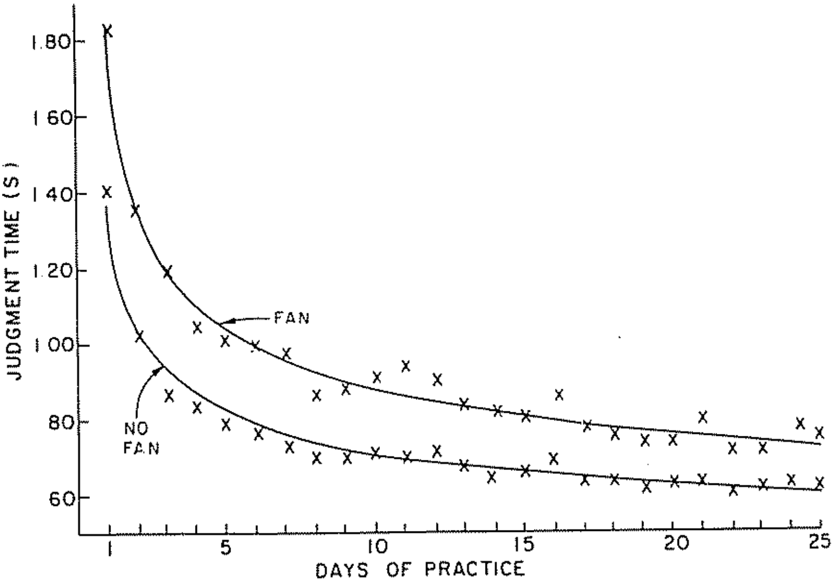

Figure 2. Recognition times for fan and no-fan sentences as a function of practice. The solid lines represent the predictions of the model described in the text. *Note.* Figure from Anderson (1983). Copyright 1983 by Harvard University Press. Reprinted by permission.

two alternative facts. This means that to retrieve the sentence from a word the subject must consider two alternatives in the fan condition. This is a small manipulation of number of alternatives (1 vs. 2 facts), but it is enough for a large effect in a fact retrieval experiment.

As can be seen, the effect of number of alternatives started out at more than 450 ms (fan vs. no-fan conditions), and after 25 days of practice it had reduced to less than 150 ms. This decrease in effect of number of alternatives is often described as if it were a qualitative shift, but as the figure makes clear, it is a gradual, continuous change. Moreover, power functions with the same exponents and intercepts can be fit to the data. The equation for the no-fan condition is

$$T = .36 + .77\left(P - \frac{1}{2}\right)^{-.36} \tag{7}$$

and for the fan condition it is

$$T = .36 + 1.15\left(P - \frac{1}{2}\right)^{-.36} \tag{8}$$

where P is the number of days of practice. Thus, the effect of number of alternatives decreases proportionately with reaction time.

The effect of number of alternatives can be understood in terms of our familiar expression SAG/NI. The more alternatives there are, the lower the activation A of any one of them. Note also that according to this expression, the effects of data activation and production strength are multiplicative. This produces the proportionate relationship between the decrease in reaction time and the effect of number of alternatives.

Effect of consistency

In a consistent mapping condition, the same stimulus is always associated with the same response, or as Logan (in press) has argued, with the same interpretation. To consider the effects of consistency, let us imagine a Sternberg-type task in which some digits are in a positive set and other digits are not. In a consistent condition, the subject has many trials with the same positive and negative digits. We have described earlier how repeated use of the same memory set would cause there to be the compilation of production rules like

P4 IF the goal is to categorize an element on the screen
and 4 is on the screen
THEN say *yes*

and

P5 IF the goal is to categorize an element on the screen
and 7 is on the screen
THEN say *no*.

In this case, specific production rules are learned to classify each digit. In contrast, when the positive digit set is changed from trial to trail, the subject would have to fall back on more interpretive productions[4] like

P6 IF the goal is to categorize an element on the screen
and *n*1 is on the screen
and *n*1 is in the positive set
THEN say *yes*

and

P7 IF the goal is to categorize an element on the screen
and *n*1 is on the screen
and *n*1 is not in the positive set
THEN say *no*.

The first set of productions, acquired in a consistent condition, should be faster because they are simpler and should show no effect of numbers of alternatives. In contrast, the second set, acquired in the inconsistent condition, should be slower. They should also show an effect of number of alternatives because matching "*n*1 is in the positive set" should be affected by the fan out of the positive set. (For a more thorough discussion of the application of the ACT* theory to the Sternberg task, see Anderson, 1983; Jones & Anderson, 1987.)

Usually some effect of number of alternatives is obtained in the consistent condition, and the effect decreases with practice. There may be a residual effect of number of alternatives because a race would exist between the interpretive productions like P6 and P7 and the more direct productions like P4 and P5. Thus, there will be a residual effect of the properties of the interpretive productions in a task that has come to be dominated by the more direct productions.

SUMMARY

To summarize the application of this ACT* model to automaticity, let us consider how it addresses the empirical phenomena we identified earlier:

1. Power law improvement: This is due to accumulation of strength modulated by factors 2 and 3 below.

2. Power law forgetting: This is basically handled in ACT* by

assumption. This assumption can be motivated as an adaptation of memory to the environment.

3. Spacing effects: This is also handled in the ACT* theory by an assumption which can also be shown to be adapted to the structure of the environment.

4. (a) An automatic process is interfered with less: This is a direct consequence of the appearance of the strength factor, *S*, in basic conflict resolution quantity *SAG/NI*.

(b) An automatic process can become less interfering if the process of compilation has created production rules that overlap less with the concurrent task. This is produced by a decrease in the value of *N*.

5. An automatic process can be harder to inhibit and is more interfering: This is produced by an increase in the value of *I*.

6. Automatic processes are less slowed down by number of alternatives: This is predicted by the interaction between production strength and data activation—the multiplicative relationship between *A* and *S*.

7. Consistency promotes automaticity: Consistency can result in special-case productions that eliminate the need for long-term memory access. This can also be a reason for diminishing the effect of number of alternatives.

Notes

Preparation of this article was supported by Grant 87-05811 from the National Science Foundation and Contract N00014-90-J-1489 from the Office of Naval Research.

Correspondence concerning this article should be addressed to John R. Anderson, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213. Received for publication April 18, 1991; revision received August 7, 1991.

1. Actually, Equation 4.1 in Anderson (1983) also involves a quantity, *C*, concerned with the complexity of the condition. This will not be of concern in this article. Equation 4.1 conflated *N* and *I* from Proportion 1 into a single factor.

2. The generalization involves allowing for some prior learning such that *P* in Equation 1 becomes $N + X$ where *N* is the number of trials and *X* is a fixed amount of prior learning.

3. A gamma distribution has a scale parameter, *a*, and an index, *n*, which is often interpreted as number of stages. An exponential is a gamma with $n = 1$. The comment above about proportionality is true if we assume the speed-up is reflected in the scale parameter but the index stays constant.

4. Note that these reflect an intermediate degree of compilation from P1–P3 given earlier.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94*, 192–210.

Anderson, J. R. (1989). A rational analysis of human memory. In H. L. Roedinger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness* (pp. 195–210). Hillsdale, NJ: Erlbaum.

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

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396–408.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forget to ask. *Journal of Experimental Psychology: General, 108*, 296–308.

Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 517–526.

Gay, I. R. (1973). Temporal position of reviews and its effect on the retention of mathematical rules. *Journal of Educational Psychology, 64*, 171–182.

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.

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 554–565.

Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 397–406.

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

Logan, G. D. (in press). Automaticity and memory. In W. E. Hockley & S. Lewandowski (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdoch*. Hillsdale, NJ: Erlbaum.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interferences: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 126–135.

Mazur, J. E., & Hastie, R. (1975). Learning as accumulation: A reexamination of the learning curve. *Psychological Bulletin, 85*, 1256–1274.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Singley, M. K., & Anderson, J. R. (1989). *Transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition, 4,* 215–230.

Treisman, A., Vieira, A., & Hayes, A. (1992). Automaticity and preattentive processing. *American Journal of Psychology, 105,* 341–362.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition, 2,* 775–780.