34

The Adaptive Nature of Memory

JOHN R. ANDERSON & LAEL J. SCHOOLER

Most research on human memory has mainly focused on the question of understanding what memory does and not on why memory does what it does. There have been requests that the field focus more on the function of memory and perhaps as a consequence applied research on memory has been a growing field. Neisser (1976, 1978, 1982) argued that the field should adopt an "ecological approach" to memory and study how memory is used in the real world. He suggests that the principles of memory in the real world might be different from the ones uncovered in the laboratory. Bruce (1985) made a serious effort to define what such an ecological approach would amount to. He noted that research under the ecological banner focused on everyday memory phenomena and he argued that more attention needed to be given to evolutionary explanations for why memory worked as it did.

The major difficulty in achieving insight from a functional approach to human memory is that its function is at once so obvious but also so apparently flawed. It is obviously valuable to have access to one's past and make one's current behavior contingent on the past. On the other hand, human memory with its many failures seems quite flawed. How much insight is there to be gained by noting that memory has an obvious function that it achieves poorly? As Bruce noted, the ecological approach is left with anecdotes about when memory seems to achieve its function well and when it does not.

However, researchers have long wondered whether human memory really was so flawed. One finds occasional arguments (e.g., Bjork & Bjork, 1992) that memory's most apparent deficit, forgetting, may be an adaptive response to the need to focus on currently relevant information. In general, arguments for the adaptiveness of human memory have taken the perspective that the memory system faces constraints and that its behavior is an optimal solution to these constraints. This theme appears in a couple of recent formal models of recognition memory (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997) that argue that recognition memory takes an optimal Bayesian solution to discriminating between the traces of foils and targets. Such optimal theories are shown capable of explaining such phenomena as the list-strength effect (Ratcliff, Clark, & Shiffrin, 1990), which had been problematical for past theories.

This chapter is organized around a proposal for understanding the adaptiveness of the memory system called *rational analysis* (J. R. Anderson & Milson, 1989; Anderson & Schooler, 1991; Schooler, 1993; Schooler & Anderson, 1997). This framework assumes that there is some cost, C, associated with retrieving a memory. This cost may reflect metabolic expenditure in maintaining and retrieving the memory and also the time to search and consider the memory. If the memory proves to be useful to the current purposes, there is some gain, G, in accessing the memory. The problem facing the memory system is to come up with some scheme that minimizes the costs in retrieval while maximizing the gains. Rational analysis also proposes that the memory system can, in effect, assign some probability P to a memory being relevant in advance of retrieving it. Given these three quantities, an adaptive memory system would search memories in order of their expected utilities, $PG - C$, and stop considering memories when a probability P is retrieved such that:

$$PG < C \tag{1}$$

This predicts that people will be able to retrieve most rapidly memories that are most likely to be relevant to their current needs and not recall memories that are unlikely to be relevant. This framework can be elaborated into a theory that makes quantitative predictions about latency and probability of recall.

As Anderson and Milson (1989) discuss, this basic framework applies to many artificial memory systems such as information retrieval systems (Salton & McGill, 1983), libraries (Burrell, 1980, 1985), and file management systems (Stritter, 1977). For instance, libraries try to assess the probability that books will be needed, making those most likely to be needed available in special collections, others in the stacks, others in various off-site storage areas, and giving away still other books. Similarly, it is argued that the human memory system makes the most likely memories available in various sorts of working memories, makes others more or less available in long-term memory, and forgets still others.

There are three quantities in the above analysis—P, G, and C. The one that has been subjected to the most analysis is P, the probability of a memory being needed. The next section will give a formal analysis of this probability which is called *need probability* and its relationship to probability of recall and latency of recall. The subsequent two sections will discuss how this need probability is sensitive to the history of past usage of the memory as well as the current context. The last section of this chapter will discuss how G, the value of the memory, and C, the cost, might reflect effects of the content of the memory.

Part of the effort to give a rational analysis to human cognition (Anderson, 1990) was the claim that the memory system functioned optimally under this analysis. Such claims about optimality are always controversial. This chapter will not push this issue of optimality but rather simply note that this adaptive perspective does offer insight about memory. Sometimes, it yields surprising, quantitative predictions.

The Analysis of Need Probability

Need probability is the probability that a particular memory is needed in a particular context. Letting H stand for the hypothesis that the memory is needed and E the evidence of the elements in the context, the need probability can be denoted as a conditional probability $P(H \mid E)$. This conditional probability could be profitably analyzed in a Bayesian framework. Bayes theorem in odds form is:

$$\frac{P(H|E)}{P(\overline{H}|E)} = \frac{P(H)}{P(\overline{H})} * \frac{P(E|H)}{P(E|\overline{H})} \tag{2}$$

or

Posterior-odds = prior-odds * likelihood-ratio

This offered a useful separation of factors that did involve the current context and that did not. The prior odds, $P(H)/P(\overline{H})$, reflected factors associated with the general history of use of the memory. It is called the *history factor* and will be analyzed in the next section. The likelihood ratio, $P(E \mid H)/P(E \mid \overline{H})$, reflects how the current context determines the probability of being relevant. This is called the *context factor* and will be discussed in the subsequent section. The remainder of this section will do three things. The first is to discuss the issue of possible mechanistic realizations of this mathematical formula. The other two are to articulate how need probability might map onto the dependent measures of latency and probability of correct recall.

A frequent criticism of this analysis is that the Bayesian mathematics is not intuitive and that people are poor at estimating Bayesian probabilities. It should perhaps be obvious, but there is no claim here that people are explicitly calculating these probabilities. Rather, the claim is that the human memory system is behaving as if it were calculating such probabilities. This leaves open the question of how it achieves this "as if" behavior. J. R. Anderson (1993) showed that standard connectionist activation formulas (e.g., Rumelhart & McClelland, 1986) serve to calculate a quantity related to need probability. There are two keys to this insight. The first is to note, as had J. R. Anderson and Milson (1989), that the likelihood ratio $P(E \mid H)/P(E \mid \overline{H})$ can be decomposed into the product of a number of likelihood ratios, one for each element in the context. Then the Bayesian equation above becomes:

$$\frac{P(H|E)}{P(\overline{H})|E)} = \frac{P(H)}{P(\overline{H})} \prod_{j \in E} \frac{P(j|H)}{P(j|\overline{H})} \tag{3}$$

where the product is over the elements j in the context and the $P(j \mid H)/P(j \mid \overline{H})$ are the likelihood ratios of elements j appearing in the context if the memory is needed or not.

The quantity above reflects multiplicative operations, but neural activation is typically thought of as adding influences from various sources. One can get an additive formula by taking logs of both sides of the above equation and having a log-odds formula:

$$\log\left(\frac{P(H|E)}{P(\overline{H}|E)}\right) = \log\left(\frac{P(H)}{P(\overline{H})}\right) + \sum_{j \in E} \log\left(\frac{P(j|H)}{P(j|\overline{H})}\right) \tag{4}$$

This is basically equivalent to the following connectionist activation formula where A_i is activation of unit i, B_i is the base level activation, W_j is the input from element j, and S_{ji} is the strength of association between j and i:

$$A_i = B_i + \sum_j W_j S_{ji} \tag{5}$$

If the unit i reflects a memory, then its activation reflects log posterior odds of the memory being needed, its base-level activation reflects prior odds, and strength of association reflects the log of the likelihood ratio. The ACT-R theory, which uses these activation processes to reflect these Bayesian quantities, has been quite successful in accounting for human memory (J. R. Anderson, Bothell, Lebiere, & Matessa, 1998).

That ACT-R theory also incorporates the rational analysis of the relationship between need probability and the dependent measures of probability of recall and latency of recall. J. R. Anderson (1990) derived what the relationship should be between need odds, n, and the observed behavioral measures. Need odds is $P(H \mid E)/P(\overline{H} \mid E)$ from the prior mathematics.

The analysis given in the introduction implied a step function for probability of recall such that all items with n over a threshold would be recalled and none below would be recalled. However, J. R. Anderson (1990) showed that, if there was some noise in the estimation of need probability, then the odds of a memory, with need odds n, being above threshold is:

$$\text{Odds of recall} = Cn^d \tag{6}$$

In words, the recall odds should be a power function of need odds. A power function means that some term (in this case n) is raised to an exponent (in this case d).[1]

The introduction described a search process that terminates when one of the following two conditions is met: (1) the needed memory is found; or (2) the need odds of the next memory falls below threshold. To predict latency of recall what is important is the first condition: the time it takes to find a needed memory that has been correctly recalled. The analysis assumes a best-first serial search of memory. Therefore, the time to retrieve a particular memory will be proportional to the rank of its need odds among the need odds of all memories. J. R. Anderson and Schooler (1991) show that, assuming need odds are distributed according to the ubiquitous Zipf's law (Ijiri & Simon, 1977), the time to recall a memory with need odds n will be:

$$\text{Time for recall} = An^{-b} \tag{7}$$

That is, the time to retrieve a particular memory should be a power function of need odds. Thus, the prediction is that two dependent measures, odds of recall and latency of recall, should both be power functions of need odds.

The History Factor

J. R. Anderson and Schooler (1991) focused on the effects of amount of practice and retention interval on recall. Somewhat obviously, it is adaptive to make more available memories that are used more often (the practice effect) and less available memories that have not been used for a while (the retention effect). If this is all that an adaptive analysis predicted it would offer only a little insight. Fortunately, more is known about the effects of practice and retention interval than these simple ordinal relationships. It has been shown that performance improves as a power function of practice (Newell & Rosenbloom, 1981). This is usually measured in terms of latency, but the relationship also holds for odds of recall. It is also known that performance deteriorates as a power function of the retention interval (Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991). This is usually displayed in terms of probability of recall, but power function deterioration also describes odds of recall and latency. One of the questions addressed in Anderson and Schooler was whether this adaptive analysis could predict the behavioral power functions.

As noted, this analysis implies that the behavioral measures are related to need odds by a power function. If it turned out that need odds were related to amount of practice by a power function and to retention interval by a power function, the power laws of practice and retention interval would be predicted. That is, if the need odds, n, were related to an independent variable, X, as a power function:

$$n = Fx^g \quad (8)$$

then odds of recall would be related to X by a power function (combining equations 6 and 8):

$$\text{Recall odds} = C(FX^g)^d = CFX^{dg} \quad (9)$$

and latency would also be related to X by a power function (combining equations 7 and 8):

$$\text{Time for recall} = A(FX^g)^{-b} = AFn^{-bg} \quad (10)$$

Thus, the goal became to identify what relationships held between need odds and practice and retention interval. More generally the goal became to study the statistics of the information-retrieval demands that people face in their environment.

Gathering statistics about these informational demands requires detailed records of people's experience in the world. Ideally, researchers would follow people around, tallying their informational needs. Clearly, it is impractical to study the complete history of the informational demands that the environment places on an individual. Instead, J. R. Anderson and Schooler (1991) have studied three environmental databases that capture coherent "slices" of the environment. Two of these databases, word usage in speech to children and in the *New York Times* headlines, are linguistic in nature, but differ in their time scales. A third involves the daily distribution of people who sent the first author (*JA*) electronic mail messages. This database captured aspects of his social environment.

- *Speech to children.* Each word a child hears is another demand to retrieve the meaning of that particular word. MacWhinney and Snow's (1990) CHILDES database is a collection of transcripts of children's speech interactions. Anderson and Schooler analyzed 25 hours of preschool children's verbal interactions donated by Hall and Tirre (1979) that were collected by attaching wireless microphones to the children's clothing. For the analyses based on the Hall and Tirre corpus, a word was defined to be needed each time it was mentioned in an utterance.
- *New York Times.* Reading newspaper headlines requires retrieving the meaning of words that make up the headlines. Two years' (1986 & 1987) worth of New York Times front-page headlines were studied. For the analyses based on the *New York Times* headlines, a word was defined to be needed each time it was mentioned in a headline.
- *Authors of electronic mail.* Each time someone receives mail, demands are made on the memory system to retrieve information about the person who sent it. Three years' worth of JA's mail messages were studied. For these analyses, information about a sender was defined to be needed each time JA received a message written by that sender.

Studying the effect of practice involved looking at the relationship between the probability that an item would occur in a particular interval in the past and the probability that it would occur in the next interval. It was found, as has already been documented (Ijiri & Simon, 1977), that there is a direct linear relationship between the two probabilities of the form:

$$\text{Probability-in-the-future} = a * \text{probability-in-the-past} \quad (11)$$

where a is a fraction typically less that 1 (in these three domains it varied from .76 to 1). Since the actual probabilities are much less than 1, this also implies the relationship between odds will be approximately of the same form:

$$\text{Odds-in-the-future} = a' * \text{odds-in-the-past} \quad (12)$$

This linear function is a special case of a power function and therefore the rational analysis does predict the power functions of practice. However, it is a somewhat degenerate and obvious power function and some of the surprise value in the result was diminished.

The situation is more interesting with respect to the retention function. When an item's environmental need odds are plotted as function of how long it has been since it was last encountered, is the resulting curve a power function? Figure 34.1 shows the results of the analysis of the *New York Times*. Figure 34.1a plots the odds of a word being included in the front page headlines as a function of the number of days since the word was last included. Figure 34.1b shows these same data in log-log coordinates. Here the curve is straight, suggesting that the environmental recency function, like its behavioral counterpart, can be described by a power function ($R^2 = .99$). Similar results held for both the analyses of speech to children and the daily distribution of the sources of electronic mail messages. Recker and Pitkow (1996) have shown that similar statistics are also found in WEB accesses.

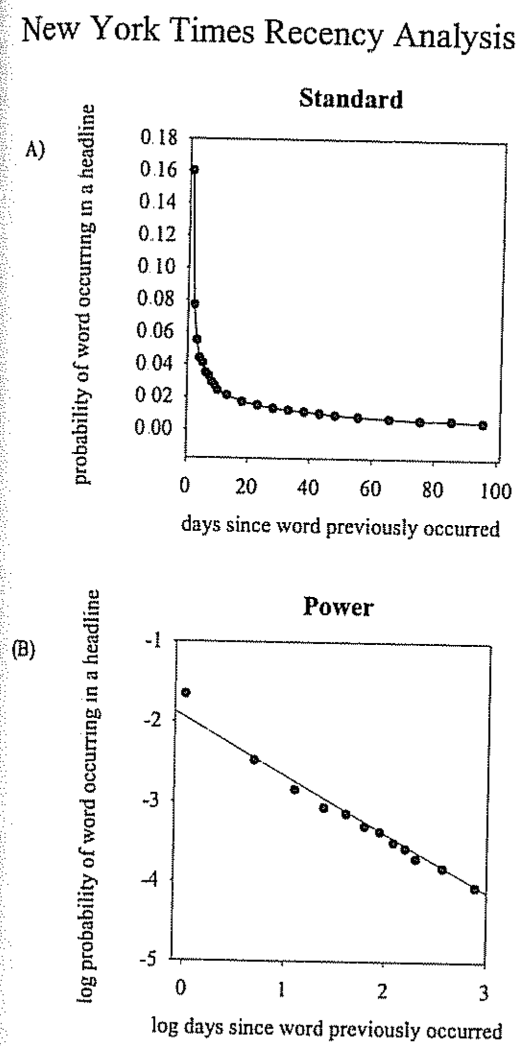

Figure 34.1 (a) Probability of a word occurring in a headline in the *New York Times* on day 101 as a function of how long it has been since the word previously occurred; (b) the same data in log-log form.

In conclusion, need odds is both a power function of amount of practice and retention interval. In one case, this was a novel finding. Thus, in both cases, the behavioral power functions represent adaptive responses to the statistical structure in the environment. This offers some insight into the nature of memory.

Can rational analysis go beyond predicting the parametric forms of the learning and retention curves? One complication in the memory literature is the spacing effect, which involves an interaction between study lag and the retention interval. For short retention intervals memory is often best with short study intervals while the result reverses for long retention intervals. J. R. Anderson and Schooler (1991) looked at their three sources, considering situations where an item had occurred just twice in the last 100 time units. They examined how the interval between these two occurrences (analogous to study lag) interacted with the retention interval (the interval between the last occurrence and the current time—analogous to retention interval). Figure 34.2 shows the results for the three empirical domains. In each case, increasing study lag decreases the probability of encountering a stimulus for short retention intervals and increases this probability for long retention intervals. Thus, the spacing function in memory behavior seems to reflect a similar spacing function in the statistics of the environment.

Finally, J. R. Anderson and Schooler looked at the relationship between amount of practice and retention interval. There has been a history of some controversy about what the retention functions are like for different levels of practice (e.g., Bogartz, 1990; Loftus, 1985; Slamecka & McElree, 1983). When one considers odds of recall one tends to get parallel retention functions for different amounts of practice. Figure 34.3a presents some data from Hellyer (1962) looking at short retention inter-

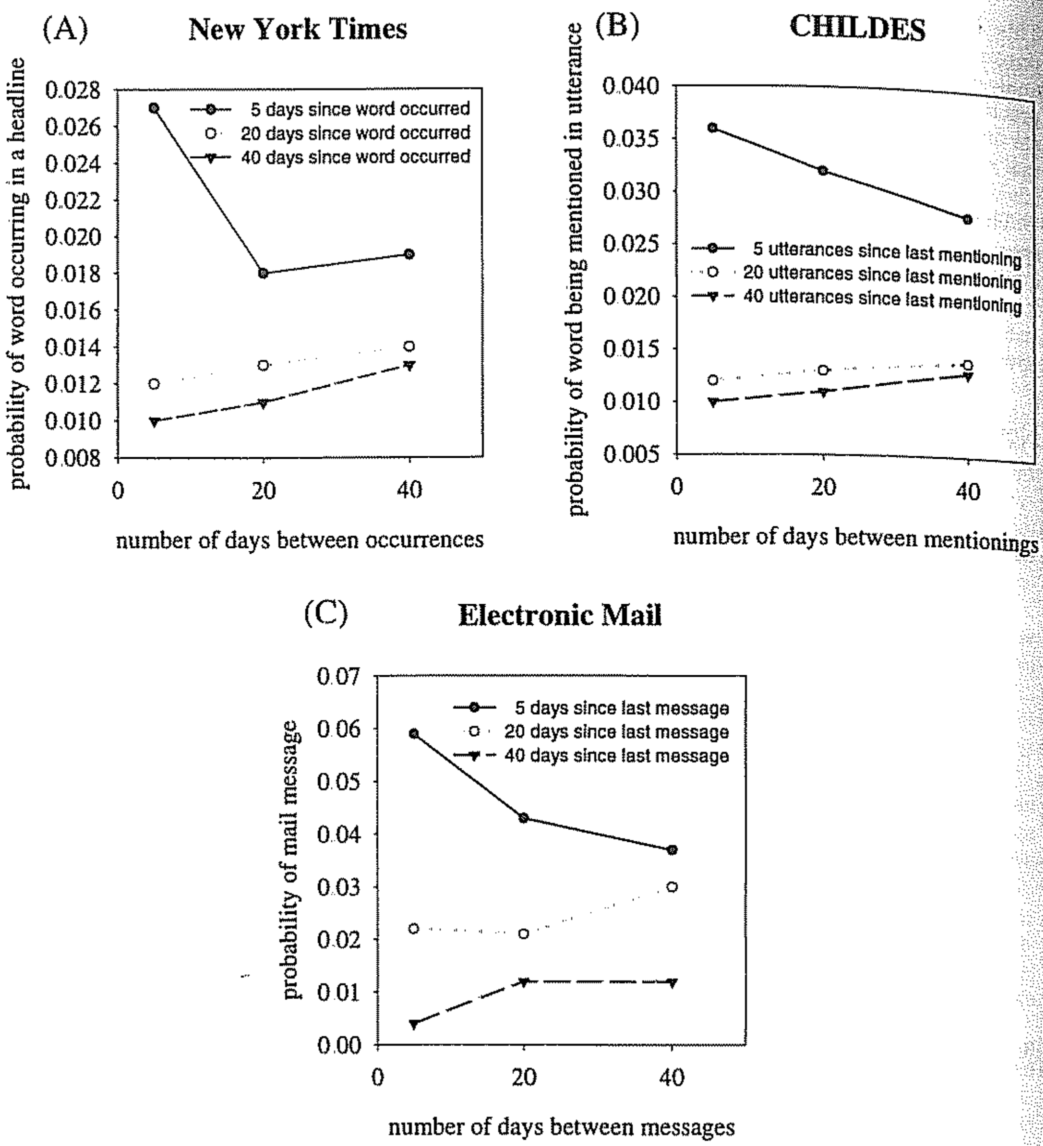

Figure 34.2 The interaction between study lag and retention lag in the (a) *New York Times*; (b) CHILDES database, and (c) electronic mail source.

vals and figure 34.3b presents data from Krueger (1929) looking at relatively longer intervals. In both of these cases, when log odds of recall is plotted as a function of log interval, there are approximately parallel curves for different levels of practice. J. R. Anderson and Schooler looked at environment retention functions in the cases where there had been many occurrences in the last 100 intervals and cases where there had been few. The results for the three data bases are shown in figure 34.4. Again there are parallel retention functions when log odds is plotted as a function of log retention interval. Figures 34.3 and 34.4 imply that when one goes back to the original scales from log scales one will find a multiplicative relationship such that the dependent measure (odds, latency) is a joint power function of practice and retention interval:

$$A \times \text{practice}^{b} \times \text{delay}^{-c}$$

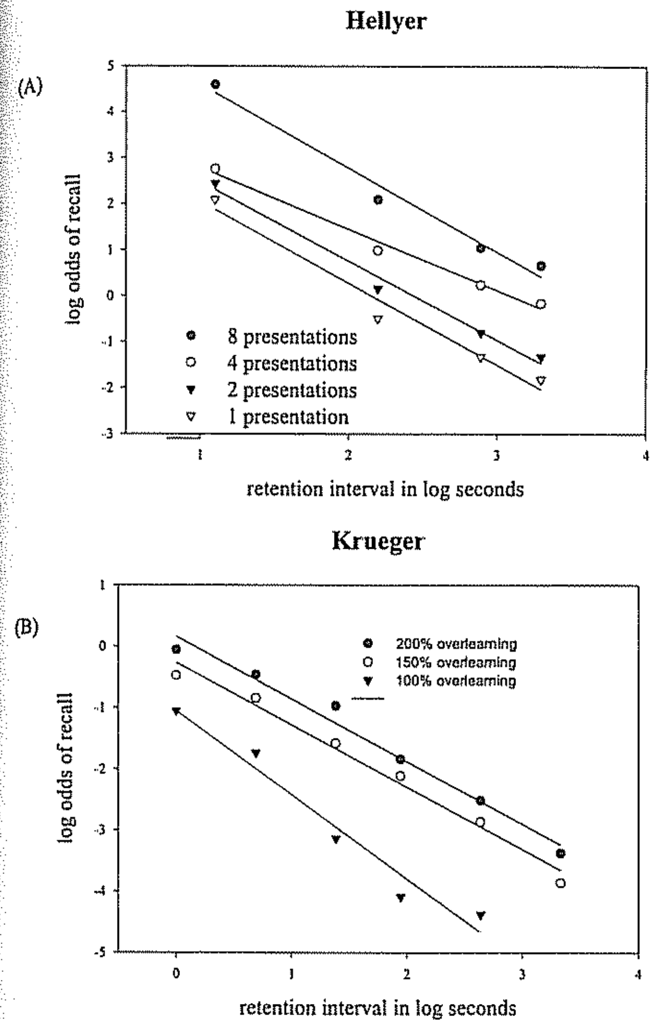

Figure 34.3 (a) Forgetting curves at four practice levels, (from Hellyer, 1962); (b) forgetting curves at three practice levels (from Krueger, 1929).

Thus, sitting behind some of the most robust regularities in human memory there are equally robust regularities in the environment, many of which had not been suspected. This offers insight into the nature of memory and that the history factor does seem to be an adaptive response to the statistics of the environment.

One of the criticisms that has been made of that research is that the three domains investigated by J. R. Anderson and Schooler (*New York Times*, e-mail messages, and caregiver speech to children) all involve human communication. One might think that human communication is determined by human memory. Thus, it might seem circular that properties of human memory can be predicted from properties of such databases. There are problems in making such criticisms go through in detail. For instance, one of the terms in the *New York Times* database is *Challenger* and reflects the *Challenger* explosion. It is a bit bizarre to think human memory caused the *Challenger* explosion and hence its appearance in the *New York Times*. Still, it would be nice to have databases that were free of the influence of human memory.

Human-communication databases were chosen because these tend to be represented as computer records and therefore are subject to computer-based analyses. As humans did not evolve in a world of e-mail and newspapers, one may wonder about the informational demands that were placed on early hominids during critical periods in evolution. It is, of course, impossible to study these environments di-

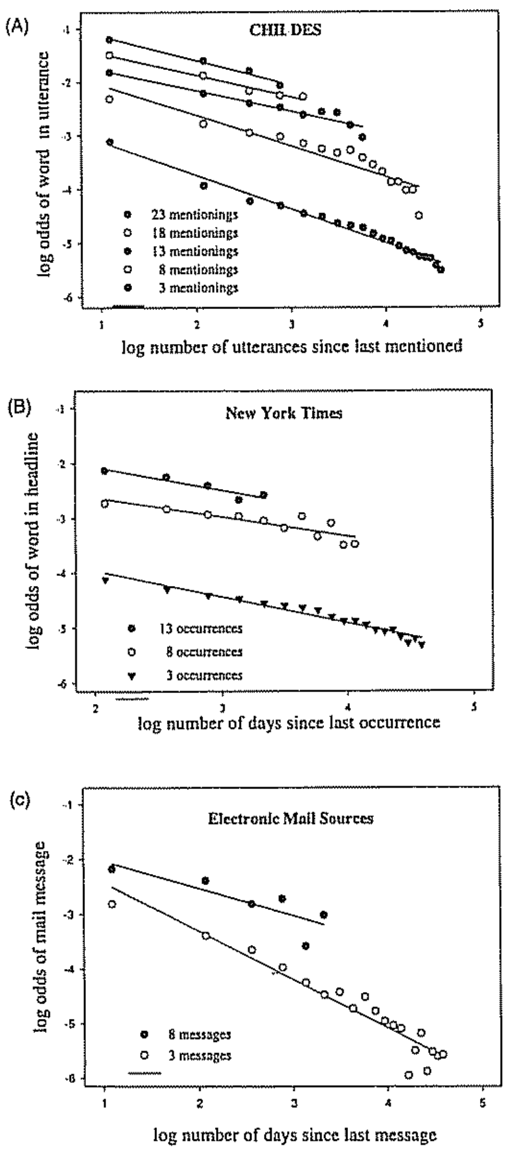

Figure 34.4 Combined frequency and recency effects in (a) the *New York Times*; (b) CHILDES database; (c) electronic mail source.

rectly. However, one can study the informational demands placed on animals whose current ecological niches share something in common with those of early hominids.

The question, then, is which animals fill the appropriate ecological niches. Milton (1981) notes that "field studies are often concerned with savanna-living primates since these are thought to offer possible analogues for hominid evolution" (p. 535). She points out, however, that hominids evolved first in tropical forests, where "the extreme diversity of plant foods in tropical forests, and the manner in which they are distributed in space and time, have been a major selective force in the development of advanced cerebral complexity in higher primates" (p. 535). Thus, "to understand the origins of mental complexity, one must look not only at life in the savannas but also life in tropical forests." (p. 535) Therefore, studying how primates move through forests and savannas represents good starting points for understanding the informational demands that shaped early hominid evolution. The second author (L.S.), in collaboration with Juan Carlos Serio Silva[2] and Ramon Rhine,[3] is analyzing the ranging behavior of howler monkeys through forests (Serio Silva's data), and baboons through savanna (Rhine's data). While these analyses are still preliminary, it does appear that the visitation patterns of howlers and baboons match the statistics of the earlier human communication databases. For example, there appear to be similar decay functions.

Another issue is whether human memory is responsive to changes in the statistical pattern of appearance of information in the environment. In the real world at large, the probability of something appearing decreases with how long it has been since the item has been encountered. However, what if an experimenter changed this statistic and made it more likely that something would appear the longer it had not been encountered? Would human memory respond to the change in the statistics of the environment? It is just this question that has been investigated in recent research by R. B. Anderson, Tweney, Rivardo, and Duncan (1997). In fact, the retention function did change with experience and showed less decay in the case where the passage of time made the reappearance more probable. Memory did not show an increase with retention interval but the decay rate did decrease. The failure to get the retention function to rise with delay may reflect the fact that the local experience did not overwhelm the massive past experience to the contrary. Nonetheless, this is an impressive experimental demonstration that the memory system will respond to the statistics of its experience.

The Contextual Factor

It is well established that the memory system is sensitive to the match between the context

in which a piece of information was studied and the context in which it was tested. This is the basic encoding specificity demonstration of Tulving (1975), where memory for a word is higher if it is tested in the context of the same word as it was studied. In analyses of the *New York Times* and the CHILDES databases, Schooler (1993) showed that a particular word was more likely to occur when other words that had occurred with it in the past were present. For instance, a headline one day mentioned both Qaddafi and Libya, and sure enough a headline the next day that mentioned Qaddafi also mentioned Libya. Stated in the context of such an example the result is rather obvious, but the example makes clear the basic adaptiveness of encoding specificity: if two items have occurred together in the past, they are more likely to occur together in the future. Therefore, an adaptive memory system should show an encoding specificity.

Schooler (1993) tried to explore the extent to which the effects of context went beyond the obvious in the *New York Times* and CHILDES databases. He collected likelihood-ratio measures of association between various words. This was measured as the associative ratio, $P(S \mid q)/P(S)$, that approximates the likelihood ratio common in Bayesian statistics. The denominator of this ratio, $P(S)$, is the base rate probability of needing a memory. The numerator is the conditional probability, $P(S \mid q)$, of needing a memory in the presence of some cue, q. The overall strength of the context is taken to be the product of the associative ratios of each of the individual cues in the context, as per equation 3.

In the same way that the environmental databases were used to investigate the history factor, they can be used to explore the context factor. Calculating the associative ratios requires estimating the base rate frequencies of the items (i.e., words) as well as the many conditional probabilities of finding one item in the presence of another. In these environmental analyses, the base rate probabilities were taken to be the proportion of all the headlines or utterances in which a word appeared. Estimating the conditional probability of finding a word in the presence of another in the same context requires a definition of context. A context was defined to be a headline or utterance; a word's context, then, was the other words that compose the headline or utterance. Table 34.1 shows some words from the headlines along with associates that had particularly high associative ratios. For example, *AIDS*

Table 34.1 Associative rates in the *New York Times* database

Associates	p(AIDS/associate)	p(AIDS/associate)/p(AIDS)
	p(AIDS) = .018	
virus	.75	41.0
spread	.54	29.4
patients	.40	21.8
health	.27	14.6
	p(trade) = .015	
	p(trade/associate)	p(trade/associate)/p(trade)
imports	.77	49.6
gap	.60	38.7
exports	.44	28.7
deficit	.35	22.7
	p(senate) = .015	
	p(senate/associate)	p(senate/associate)/p(senate)
measure	.48	19.4
veto	.46	18.6
bill	.38	15.3
votes	.35	14.3

was included in 1.8% of all headlines, and in 75% of the headlines that included *virus*. The associative ratio for the pair is 41 (i.e., .75/.018), or equivalently, *AIDS* is 41 times more likely to occur in a headline that includes *virus* than its base rate.

By definition a word is more likely to occur in a headline if a strong associate of it occurs. However, what if two strong associates occur in one headline? The earlier Bayesian statistics (see equation 3) implied that there should be a proportionate increase in the odds of the target word occurring. Indeed, this is what Schooler found. The corresponding behavioral prediction, then, is that the probability of retrieving a target word should increase if two strong associates are present rather than one. In fact, there does appear to be an effect of accumulating associates in retrieving a memory (e.g., Bowers, Regehr, Balthazard, & Parker, 1990).

Schooler (1993) and Schooler and Anderson (1997) looked at how this contextual factor combined with the historical factor. Schooler examined how these two factors combined in the *New York Times* database and the child-language database. The results are displayed in figure 34.5. Parts (a) and (b) show that both the presence of a high associate ("strong context" in figure 34.5) and the time since last appearance ("delay" in figure 34.5) affect the probability of occurrence in the critical unit of time. It might appear from parts (a) and (b) of figure 34.5 that the time factor is more important in the presence of a high associate. However, if one converts the odds scale to log odds and the time scale to log time (see J. R. Anderson & Schooler, 1991, for the justification), one gets the functions in parts (c) and (d) of figure 34.5. Those figures show parallel linear functions, which is the additivity predicted by the Bayesian log formula given earlier (equation 4).

The interesting question is whether human memory is similarly sensitive to these factors. Schooler conducted an experiment where he asked subjects to complete word fragments and manipulated whether the fragments were in the presence of a high associate or not and the time since the target word had been seen. The data are displayed in log-log form in figure 34.6. They once again show parallel linear functions, implying that human memory is combining information about prior odds and likelihood ratio in the appropriate Bayesian fashion and making items available as a function of their posterior probability. Schooler and Anderson (1997) note that similar parallel functions have been found in other studies of human memory (e.g., Mäntylä & Nilsson, 1988; Thompson, 1972) but such data had not

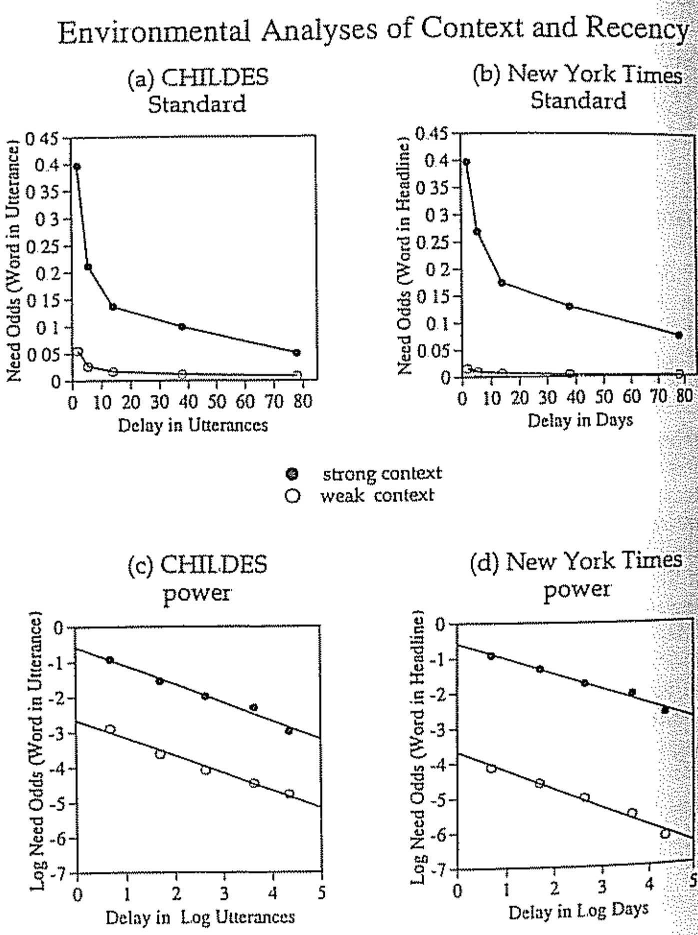

Figure 34.5 Environmental recency curves from the analysis of the CHILDES and *New York Times* database (panels a and b). Log-log transformations are given in panels c and d.

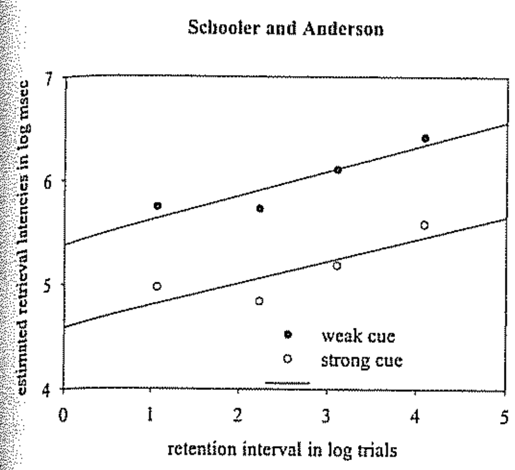

Figure 34.6 Time to retrieve memories (Schooler and Anderson, 1997). The results are plotted in log-log coordinates.

previously been analyzed in these terms. So once again there is a rather unexpected result predicted by an adaptive analysis.

The research on context discussed so far looks at situations where the context and the item are explicitly related—often because both are presented together, as in the Tulving encoding-specificity research. However, contextual effects have been shown of a more general sort where memories are associated to general cues of the environment, of drug state, or emotional mood (for a review, see J. R. Anderson, 1995). One of the interesting aspects of this literature is that the strength of these contextual effects can vary dramatically—sometimes very strong effects are obtained and sometimes null effects are obtained. This makes the point that mere statistical co-occurrence of items is not enough to ensure an association. This might seem to contradict the adaptive analysis that relates memory to the likelihood ratio measuring the statistical co-occurrence of cue and memory. However, this analysis has a significant degree of freedom, which is that it leaves open exactly what are selected as cues for a memory. Clearly, subjects cannot include everything in their current environment as potential cues. Eich and Metcalfe (1989) argue that a critical factor in getting context effects is whether the subject encodes the memory with the context at study. This shows the important role of subject encoding strategy, which is something that will be discussed further in the conclusion.

Effects of Content

Except for the issue of selection of contextual elements, the analyses so far treat memory as if it were totally a function of the statistical patterns with which people encountered events. This runs counter to all of the research that indicates that how subjects study an event and the content of an event have substantial impact on memory for the event. As J. R. Anderson and Milson (1989) show, some of the effects of study strategy can be conceived of as the subject's manipulating his experience with the target memory. For instance, different rehearsal strategies will create an environment in which the statistics will favor some items and not others. J. R. Anderson and Milson show that some of the effects associated with free recall, such as the serial position effect, reflect the statistics of such self-made environments. However, other effects do not seem to have such an explanation. These effects include the difference between shallow and deep processing (Craik & Lockhart, 1972), the effects of the concreteness and imageability of the material (e.g., Paivio, 1971), and the self-reference effect whereby memories involving the self are better (e.g., Rogers, Kuiper, & Kirker, 1977).

However, the emphasis on the statistical properties of the environment reflects only the variable *P* in the original *P*, *G*, and *C* characterization of a rational memory system. It might be less costly to process certain memories than others—that is, they have a lower value of *C* associated with them. Certainly, this would seem to be the implication of Paivio's analysis of the advantage of pictorial material. The claim is that the visual system is just more capable than the verbal system. Unfortunately, claims about such differences in processing costs tend to be basically circular: one winds up claiming memories are better for certain types of material because memory system is better for that material. Without some convergent evidence about processing costs, it does not seem profitable to pursue the *C* variable.

On the other hand, it seems possible to make objective assessments of the relative importance of various memories. One result that seems to fall out directly from an adaptive analysis is that, all other things being equal, memory should be better for more important material (i.e., higher *G*). For instance, people

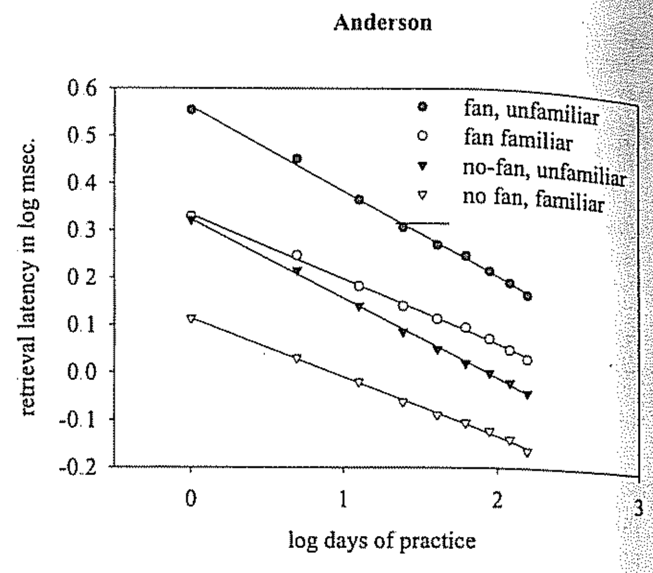

Figure 34.7 Recognition of fan and no-fan sentences about familiar and unfamiliar concepts as a function of practice. The results are plotted in log-log coordinates.

remember things about themselves and other people close to them better than about strangers (Rogers et al., 1977). One can similarly interpret the apparent superiority for flashbulb memories for details about dramatic events (e.g., Palmer, Schreiber, & Foy, 1991). Again, the framework can explain the better retention levels associated with memories that produce high arousal (Levonian, 1972). As another example of the same principle, people tend to show better memory for the meaningful aspects of an event than its unimportant superficial details (e.g., Wanner, 1968).

While it is reassuring to an adaptive analysis that memory is better for more important things, this does not offer very much insight into the nature of memory. As in the case of the history and context factor, one wants to look for detailed, nonobvious predictions. One such prediction involves the effect of practice and retention interval for memories of differential value. Rational analysis would predict that retention or practice curves for more important things would parallel those for less important on log-log scales—just as were parallel curves for different contexts in figure 34.6. This is because P and G multiply in the utility analysis and a multiplicative relationship implies an additive relationship in log-log scale. Unfortunately, in the case of at least one variable plausibly affecting importance—namely arousal—there seems to be a cross-over interaction with retention interval (Levonian, 1972). So this is evidence that the predictions of the adaptive analysis are not always confirmed.

On the other hand, J. R. Anderson (1983) contrasted familiar (Ted Kennedy) and unfamiliar people (Bill Jones) in terms of the fan effects and practice functions for new facts learned about such people. One might assume memories about the familiar people are more important. Typical of other research, J. R. Anderson found that subjects show better memory for the more familiar material (in adaptive terms because such memories are more important) but that they show near identical practice functions (history factor) and fan effects (context factor) for the two types of material. Figure 34.7 displays the results. So this is at least one instance of the nonobvious prediction that the importance of the material elevates memory but does not change the basic memory functions.

Conclusions

Despite the generally successful tone of this review, it is not the case that all factors affecting memory can be understood in terms of an adaptive analysis. As one example, in adaptive framework it is hard to understand the effects of intention to learn on memory. Instructing a subject to learn or paying him to learn can result in differential study patterns (see J. R. Anderson, 1995, for a review) but it appears that, controlling for study pattern, there is no effect of motivation or intention to learn. An adaptive system should be able to give more resources to things that are more important. However, the only way people seem to achieve this is by changing how they study the material.

The effect of study strategy on memory also points to an Achilles' heel of the adaptive analysis. The adaptive analysis works best if one can conceive of memory as responding passively to external statistics in the environment. However, by different rehearsal patterns the subject can actively create his own unobserved environments with its own statistics. In the current terms, subjects can "trick" their memories into treating something as highly probable by giving it the statistics associated with a highly probable memory. This indicates that there are layers of adaptive consideration that go beyond just looking at environmental statistics.

Thus, the adaptive analysis here does not paint a complete picture of memory and it needs to be supplemented by other considerations. Nonetheless, some insight has been obtained about the nature of memory by taking an adaptive analysis. Nonobvious results were found that indicated human memory is adaptive in ways that had not been suspected.

Acknowledgments We would like to thank Mike Byrne for his comments on this paper. Preparation of this paper has been supported by grant SBR-94-21332 from the National Science Foundation.

Notes

1. Note that this implies that the more lawful mathematical relationships should appear if one looks at odds of recall rather than probability of recall. If P is probability of recall, $P/(1 - P)$ is the odds of recall.
2. Juan Carlos Serio Silva is at the Departamento de Ecologia Vegetal Instituto de Ecologia A.C.
3. Ramon Rhine is in the psychology department at the University of California, Irvine.

References

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

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1993). *The rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1995). *Learning and memory: An integrated approach*. New York: John Wiley.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38*, 341–380.

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

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*, 703–719.

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

Anderson, R. B., Tweney, R. D., Rivardo, M., & Duncan, S. (1997). Need probability affects retention: A direct demonstration. *Memory & Cognition, 25*, 867–872.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes*, Vol. 1: *From learning theory to connectionist theory;* Vol. 2: *From learning processes to cognitive processes* (pp. 35–67). Hillsdale, NJ: Erlbaum.

Bogartz, R. S. (1990). Evaluating forgetting curves psychologically. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 138–148.

Bowers, K. S., Regehr, G., Balthazard, C. G., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology, 22*, 72–110.

Bruce, D. (1985). The how and why of ecological memory. *Journal of Experimental Psychology: General, 114*, 78–90.

Burrell, Q. L. (1980). A simple stochastic model for library loans. *Journal of Documentation, 36*, 115–132.

Burrell, Q. L. (1985). A note on aging on a library circulation model. *Journal of Documentation, 41*, 100–115.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Eich, E., & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 443–455.

Hall, W. S., & Tirre, W. C. (1979). *The communicative environment of young children: Social class, ethnic and situation differ-*

ences University of Illinois, Center for the Study of Reading.
Hellyer, S. (1962). Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental Psychology, 64*, 650.
Ijiri, Y., & Simon, H. A. (1977). *Skew distributions and the sizes of business firms*. Amsterdam: North Holland.
Krueger, W. C. F. (1929). The effects of overlearning on retention. *Journal of Experimental Psychology, 12*, 71–78.
Levonian, E. (1972). Retention over time in relation to arousal during learning: An explanation of discrepant results. *Acta Psychologica, 36*, 290–321.
Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 397–406.
MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language, 17*, 457–472.
Mäntylä, T., & Nilsson, L. G. (1988). Cue distinctiveness and forgetting: effectiveness of self-generated retrieval cues in delayed recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 502–509.
McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*, 724–760.
Milton, K. (1981). Distribution patterns of tropical plant foods as an evolutionary stimulus to primate mental development. *American Anthropologist, 83*, 534–548.
Neisser, U. (1976). *Cognition and reality*. San Francisco: Freeman.
Neisser, U. (1978). Memory: What are the important questions? In M. M. Gruneberg, P. Morris, & R. H. Sykes (Eds.), *Practical aspects of memory* (pp. 3–24). New York: Academic Press.
Neisser, U. (1982). *Memory observed: Remembering in natural contexts*. San Francisco: Freeman.
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.
Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.
Palmer, S., Schreiber, G., & Fox, C. (1991). Remembering the earthquake: "Flashbulb" memory of experienced versus reported events. Paper presented at the 32nd annual meeting of the Psychonomic Society, San Francisco.
Ratcliff, R., Clark, S., & Shiffrin, R. M. (1990). The list strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 163–178.
Recker, M., & Pitkow, J. (1996). Predicting document access in large, multimedia repositories. *ACM Transactions on Computer-Human Interaction* (ToCHI), *3*, 352–375.
Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*, 677–688.
Rubin, R. C., & Wenzel, A. E. (1996). 100 years of forgetting: A quantitative description of retention. *Psychological Review, 103*, 734–760.
Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing* (Vol. 1). Cambridge, MA: MIT Press.
Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York: MacGraw-Hill.
Schooler, L. J. (1993). *Memory and the statistical structure of the environment*. Ph.D. Dissertation, Department of Psychology, Carnegie Mellon University.
Schooler, L. J., & Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology, 32*, 219–250.
Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145–166.
Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 384–397.
Stritter, E. P. (1977). *File migration*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
Thomson, D. (1972). Context effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior, 11*, 497–511.
Tulving, E. (1975). Ecphoric processing in recall and recognition. In J. Brown (Ed.), *Recall and recognition*. London: John Wiley.
Wanner, H. E. (1968). *On remembering, forgetting, and understanding sentences: A study of the deep structure hypothesis*. Unpublished doctoral dissertation, Harvard University.
Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science, 2*, 409–415.

35

Memory Models

ROGER RATCLIFF & GAIL MCKOON

The study of models of memory often seems like a backwater in the overall study of memory. Models do not have a prominent place in experimental studies of memory and they are not used or examined by most researchers in the field. This review examines the various questions that models can address, discusses why theory is not as prominent in the memory domain as in other domains of science, and presents an overview of current models. The aim is to show why models should have greater prominence and wider use.

Mainstream models for "long-term" memory take as their database the results of experiments in which subjects are asked to study and learn lists of items (words, nonsense syllables, letters, numbers, sentences, or pictures). Memory is tested in one of a number of ways: asking subjects whether or not an item occurred on the study list (recognition), asking for recall of the items on the study list, asking what item on the study list was associated with a cue, asking for the recency or frequency of appearance of an item on the study list, and so on. The dependent measures are usually accuracy, confidence ratings, or a combination of reaction time and accuracy. The eventual aim is to account for the effects on all the dependent variables for a range of experimental tasks and for a range of experimental manipulations, including the length of the study list, the strengths of the items in memory, the type of material, the similarity among study and test items, levels of processing, rehearsal methods, and so on.

Recent development of models of long-term memory has proceeded relatively independently of other areas of memory research. For example, there has been little contact between the long-term memory models and the findings of implicit memory experiments and there has been little explicit theoretical work in the domain of implicit memory. Over the last 20 years, the domains of reaction time research and memory have not interacted in strongly productive ways, although there has been a recent resurgence of interest in random walk and diffusion reaction time models and so there may soon be more fruitful interactions. The one domain of research with which there is some sharing of representation and process assumptions is categorization. In this domain, subjects are presented with exemplars and through feedback learn how to assign the exemplars to categories. Some models of categorization are essentially long-term memory models in that they assume a representation of a category is built up with learning and the category decision process depends on retrieval from this memory representation.

Short Historical Background

The attempt to produce models of memory that can account for data both qualitatively