2.15 The Adaptive Nature of Memory

Lael J Schooler, Syracuse University, Syracuse, NY, United States
John R Anderson, Carnegie Mellon University, Pittsburgh, PA, United States

Adapted from Anderson, John R., Schooler, Lael J., 2000. Tulving, Endel, Craik, Fergus I. M. (Eds.), The Oxford Handbook of Memory, (pp. 557–570). New York, NY, USA: Oxford University Press, vol. xiv, 700 pp.

2.15.1 Introduction

Most research on the human memory has focused on the question of understanding what memory does and not on why memory does what it does. However, we may be able to gain insights into human memory by considering the functions of memory and how memory achieves those functions. 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 suggested that the principles of memory in the real world might be different than 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 should be given to evolutionary explanations for why memory worked as it did.

The major challenge in achieving insight from a functional approach to human memory is that its function is at once so obvious as well as 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, many of us routinely fail to recall the information from memory that we need, such as the name of an acquaintance at a conference or that Aleppo is a city in war-torn Syria. 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 was really so flawed. One finds occasional arguments (e.g., Bjork and 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 formal models of recognition memory (McClelland and Chappell, 1998; Shiffrin and Steyvers, 1997), which argue that recognition memory takes an optimal Bayesian solution to discriminate between the traces of foils and targets. Such optimal theories are shown to be capable of explaining such phenomena as the list strength effect (Ratcliff et al., 1990), which have been problematical for past theories.

This chapter is organized on a proposal for understanding the adaptiveness of the memory system called *rational analysis* (Anderson and Milson, 1989; Anderson and Schooler, 1991; Schooler, 1993; Schooler and Anderson, 1997). This framework assumes that there is some cost, C, associated with retrieving a memory. This cost may reflect not only the metabolic expenditure for maintaining and retrieving the memory but also the time required 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 cost of 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 before retrieving it. Given these three quantities, an adaptive memory system would search memories in the order of their expected utilities, $PG-C$, and stop considering memories when the probability P is retrieved such that

$$PG < C \quad (1)$$

This predicts that people will be able to most rapidly retrieve memories that are most likely to be relevant to their current needs and avoid recalling memories that are unlikely to be relevant. This framework can be elaborated into a theory that makes quantitative predictions about latency and the probability of recall.

As Anderson and Milson (1989) discussed, this basic framework applies to many artificial memory systems such as information retrieval systems (Salton and 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 and maintaining others in the stacks, 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 aforementioned analysis: P, G, and C. The quantity that has been subjected to the most analysis is P, the probability of a memory being needed. The following section will give a formal analysis of this probability, which is called *need probability*, and its relationship to the probability of recall and latency of recall. The subsequent two sections will discuss how 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. As Simon (1991) pointed out, the predictions of the rational analysis depend little on the optimality assumption, so 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.

2.15.2 The Analysis of Need Probability

Human memory solves a problem much like the Internet search engine Google accomplishes today. The essential analogy is that Google searches for websites germane to the search query and the memory system searches for memories relevant to the current context. Need probability is the probability that a particular memory (or website) is needed in a particular context. In the case of Google, context would be the collection of search terms used to initiate a search, along with the documents, emails, and other materials you have stored with them. For the cognitive system, context is the salient element of the environment or active memories. Letting H stand for the hypothesis that the memory is needed and E the evidence of the elements in the context, need probability can be denoted as a conditional probability $P(H|E)$ that could be profitably analyzed in a Bayesian framework. The 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 offers a useful separation of factors that involve the current context and that do not. The prior odds, $P(H)/P(\overline{H})$, are factors associated with the general history of how frequently and recently a memory has been relevant in the past. It is called the *history factor* and will be analyzed in the following section. The likelihood ratio, $P(E|H)/P(E|\overline{H})$, reflects how the current context determines the probability of being needed. This is called the *context factor* and will be discussed in the next section. The remainder of this section will discuss the issue of possible mechanistic realizations of this mathematical formula and 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. To clarify, 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. Anderson (1993) showed that the standard connectionist activation formulas (e.g., Rumelhart and McClelland, 1986) could serve to calculate a quantity related to need probability. To see how this would work, we note that the likelihood ratio, $P(E|H)/P(E|\overline{H})$, can be decomposed into the product of several likelihood ratios, one for each element in the context. Then the Bayesian equation becomes

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

where the product is over the elements j in the context, C, and $P(j|H)/P(j|\overline{H})$ is the likelihood ratio of element 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 Eq. (3), yielding a log-odds formula:

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

This is basically equivalent to the following connectionist activation formula, where A_i is the 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:

$$A_i = B_i + \sum_{j \in C} W_j S_{ji} \tag{5}$$

If the unit i reflects a memory, then its activation reflects log posterior odds of the memory being needed, base level activation reflects prior odds, and strength of association reflects the log of the likelihood ratio. The adaptive control of thought-rational (ACT-R) theory, which uses these activation processes to reflect these Bayesian quantities, has been quite successful in accounting for human memory (Anderson et al., 1998; Anderson, 2007). Indeed, the declarative module in ACT-R, based on this rational analysis, is frequently used by other cognitive architectures (e.g., the Soar cognitive architecture—Laird, 2012).

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. Anderson (1990) derived the relationship between need odds, n, and the observed behavioral measures, where need odds is $P(H|E)/P(\overline{H}|E)$ from the previous formula.

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

The Introduction section 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, the first condition is important: 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. Anderson and Schooler (1991) showed that, assuming need odds are distributed according to the ubiquitous Zipf's law (Ijiri and 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. Anderson (2007) showed that Eq. (7) can also be derived based on neural processes. Thus the prediction is that the two dependent measures, odds of recall and latency of recall, should both be power functions of need odds.

2.15.2.1 The History Factor

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, then it would offer 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 and 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 and Wenzel, 1996; Wixted and 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 by Anderson and Schooler was whether this adaptive analysis could explain 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,

$$n = FX^{g} \tag{8}$$

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

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

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

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

Thus to determine whether this analysis was consistent with the power laws of forgetting and practice, we needed to determine if need odds bore a power function relationship to practice and retention interval. This led us to a general investigation of 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, Anderson and Schooler (1991) 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. The third database involves the daily distribution of people who sent electronic mail (email) messages to the second author (J.A.). This database captured aspects of his social environment.

2.15.2.1.1 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 (Child Language Data Exchange System) database is a collection of transcripts of children's speech interactions. Anderson and Schooler analyzed 25 h 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 & Tirre corpus, a word was defined as being needed each time it was mentioned in an utterance.

2.15.2.1.2 New York Times

Reading newspaper headlines requires retrieving the meaning of words that make up the headlines. Two years (1986 and 1987) worth of front page headlines were studied from the New York Times. For the analyses based on the New York Times headlines, a word was defined as being needed each time it was mentioned in a headline.

2.15.2.1.3 Authors of Electronic Mail

Each time someone receives a mail, demands are made on the memory system to retrieve information about the person who sent it. Three years' worth of J.A.'s mail messages (collected in the prespam era) were studied. For these analyses, information about a sender was defined as being needed each time J.A. 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 and Simon, 1977), that there is a direct linear relationship between the two probabilities of the form:

$$\text{Probability in the future} = a * \text{Probability in the past} \tag{11}$$

where a is a fraction typically less than 1 (in these three domains, it varied from 0.76 to 1). Since the actual probabilities are much less than 1, this also implies that the relationship between odds will be approximately of the same form:

$$\text{Odds in the future} = a\prime * \text{Odds in the past} \tag{12}$$

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, diminishing some of the surprise value in the result.

The situation is more interesting with respect to the retention function. When an item's environmental need odds are plotted as a function of how long it has been since it was last encountered, is the resulting curve a power function? Fig. 1 shows the results of the analysis of the New York Times. Fig. 1A plots the probability of a word being included in the front page headlines as a function of the number of days since the word was last included. Fig. 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 ($R^2 = 0.99$). Similar results held for both the analyses of speech to children and the daily distribution of the sources of email messages. Recker and Pitkow (1996) have shown that similar statistics are also found in website accesses.

The analyses of J.A.'s email contacts captured aspects of his social environment but were limited in that they were based on a single person and one measure of social contact. Pachur et al. (2014) extended the analysis of social environments to more people and more kinds of contacts. To this end, they asked 40 participants to keep a diary in which they recorded their social contacts for 100 consecutive days. Here social contact was defined as face-to-face interactions and phone calls lasting for more than 5 min, and all electronic communications of at least 100 words. They found that the regularities Anderson and Schooler observed were found in this larger sample of people and for the different kinds of contacts. For example, Fig. 2A shows the recency function for the face-to-face contact. The smooth curves illustrate that a power function does a good job of fitting the overall trends in the data. An interesting deviation from the power function in the face-to-face recency curve is the bumps that occur every seventh day, reflecting the cyclic nature of our weekly schedules.

In conclusion, need odds is both a power function of amount of practice and retention interval. In the case of the recency function, this was a novel finding. Thus, in both cases, the behavioral power functions represent adaptive responses to the statistical structure in the environment.

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 retention interval. For short retention intervals, memory is often best with short study intervals, whereas the result reverses for long retention intervals. Anderson and Schooler (1991) looked at their three sources, considering situations in which 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). Fig. 3 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.

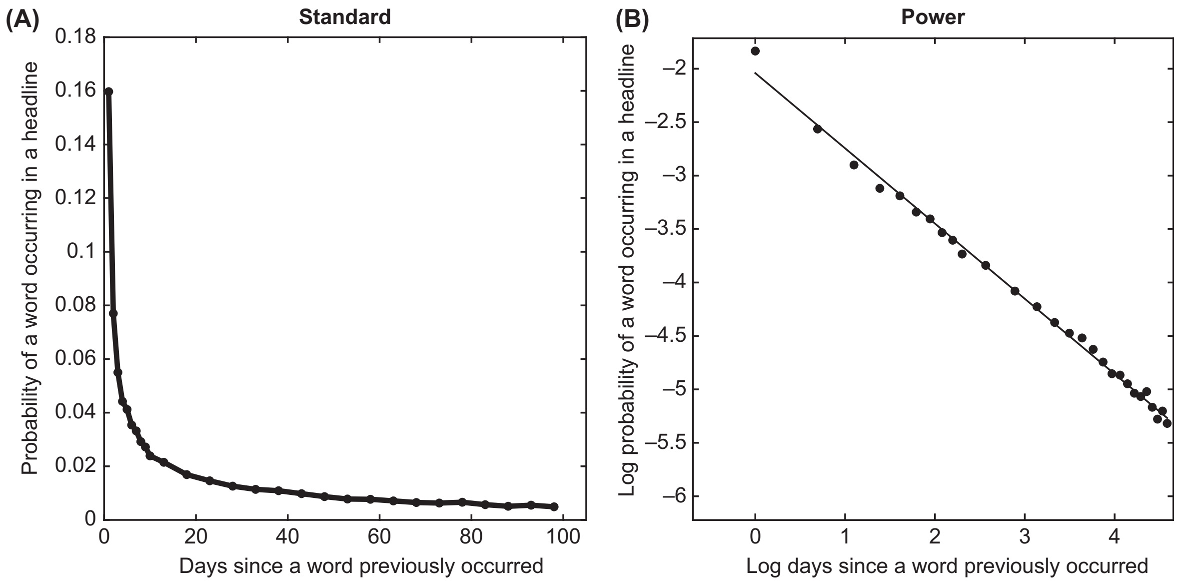

Figure 1 (A) Probability of a word occurring in a headline in the New York Times on day 101 as a function of how long it has been since the word previously occurred. (B) The same data in log–log form. Adapted from Schooler, L.J., 1993. Memory and the Statistical Structure of the Environment. Department of Psychology, Carnegie Mellon University (Ph.D. dissertation).

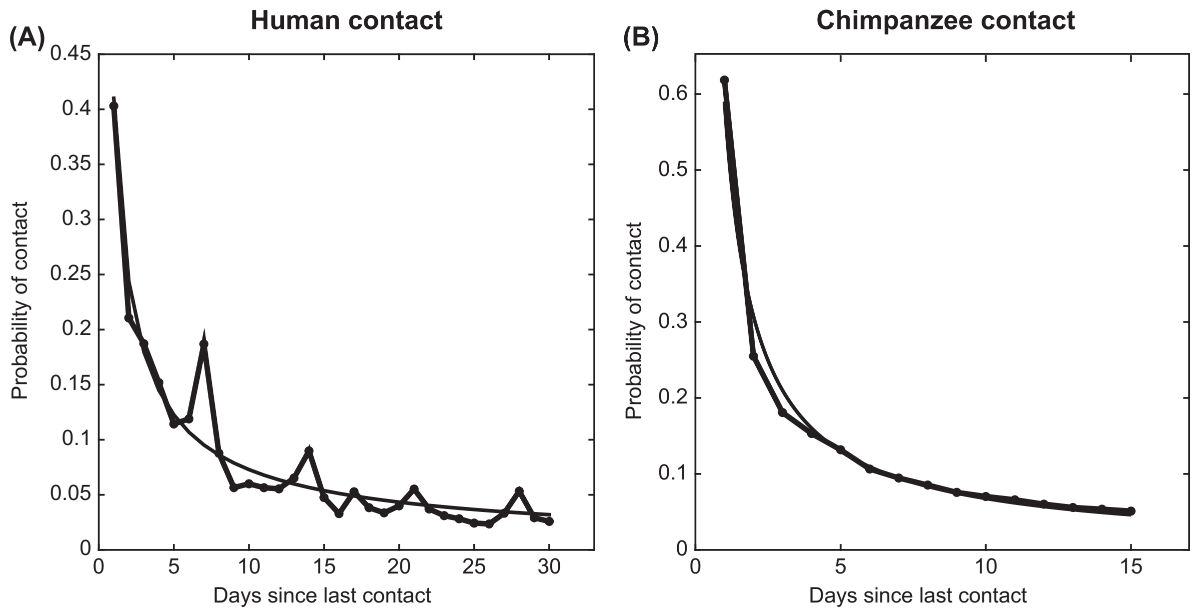

Figure 2 (A) Probability of a person meeting someone as a function of how long it has been since they last saw them face to face. (B) The probability that two chimpanzees will be in contact with each other has a function of how long it has been since they last had contact. Adapted from (A) Pachur, T., Schooler, L.J., Stevens, J.R., 2014. We'll meet again: revealing distributional and temporal patterns of social contact. PLoS One, 9 (1), e86081. http://dx.doi.org/10.1371/journal.pone.0086081; (B) Stevens, J.R., Marewski, J.N., Schooler, L.J., Gilby, I.C., 2016. Reflections of the social environment in chimpanzee memory: applying rational analysis beyond humans. R. Soc. Open Sci. 3, 160293.

Finally, 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 and McElree, 1983). When one considers odds of recall, one tends to get parallel retention functions for different amounts of practice. Fig. 4A presents some data from Hellyer (1962) looking at short retention intervals, and Fig. 4B presents data from Krueger (1929) looking at relatively longer intervals. In both 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. Anderson and Schooler looked at environmental

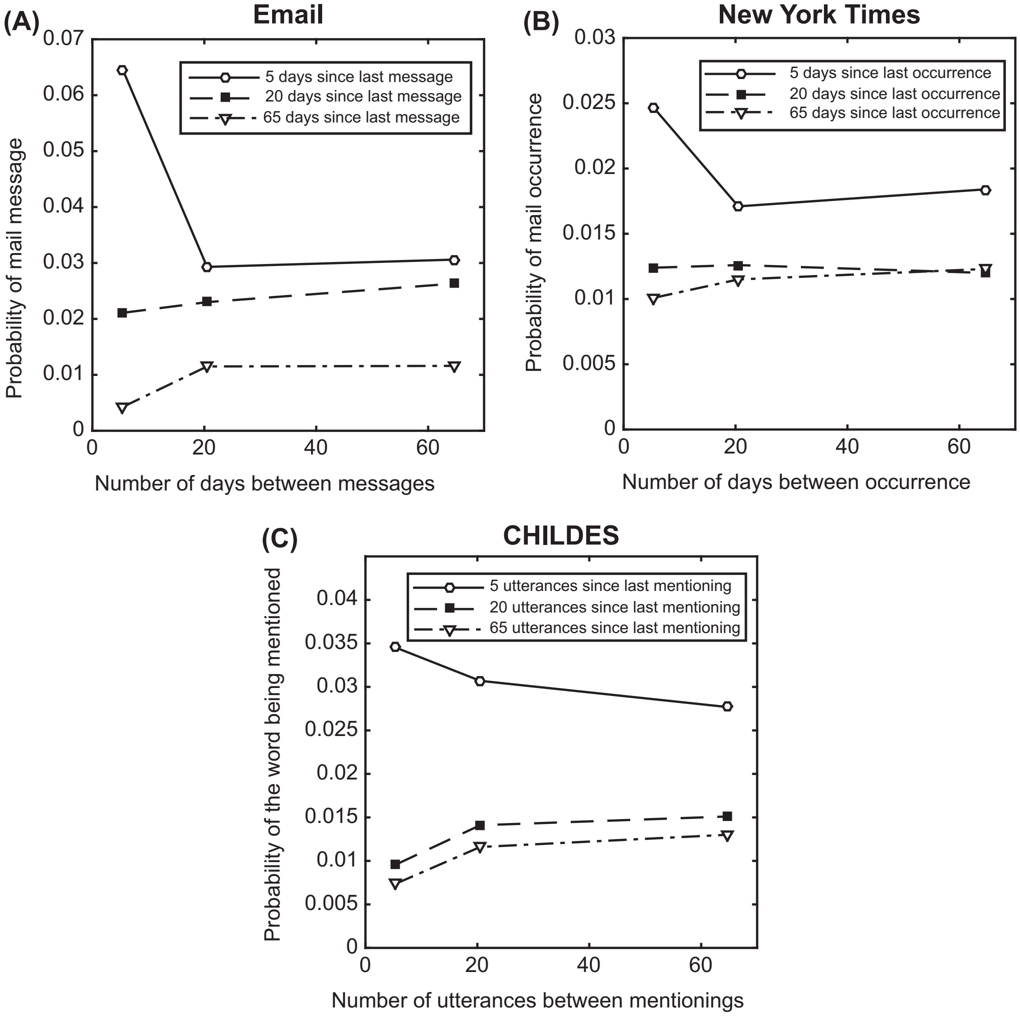

Figure 3 The interaction between the spacing between encountering something in the environment and the lag since it was last encountered. The interaction is found in (A) email authors, (B) words in the New York Times headlines, and (C) the CHILDES database. CHILDES, Child Language Data Exchange System. Adapted from Schooler, L.J., 1993. Memory and the Statistical Structure of the Environment. Department of Psychology, Carnegie Mellon University (Ph.D. dissertation).

recency 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 empirical environments are shown in Fig. 5. Again there are parallel retention functions when log odds is plotted as a function of log retention interval. Figs. 4 and 5 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:

$$A \times \text{Practice}^{b} \times \text{Delay}^{-c}$$

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.

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 typically 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 been since it was last encountered? Would human memory respond to the change in the statistics of the environment? It is just this question that was investigated by Anderson et al. (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. Although memory did not show an increase with retention interval, 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.

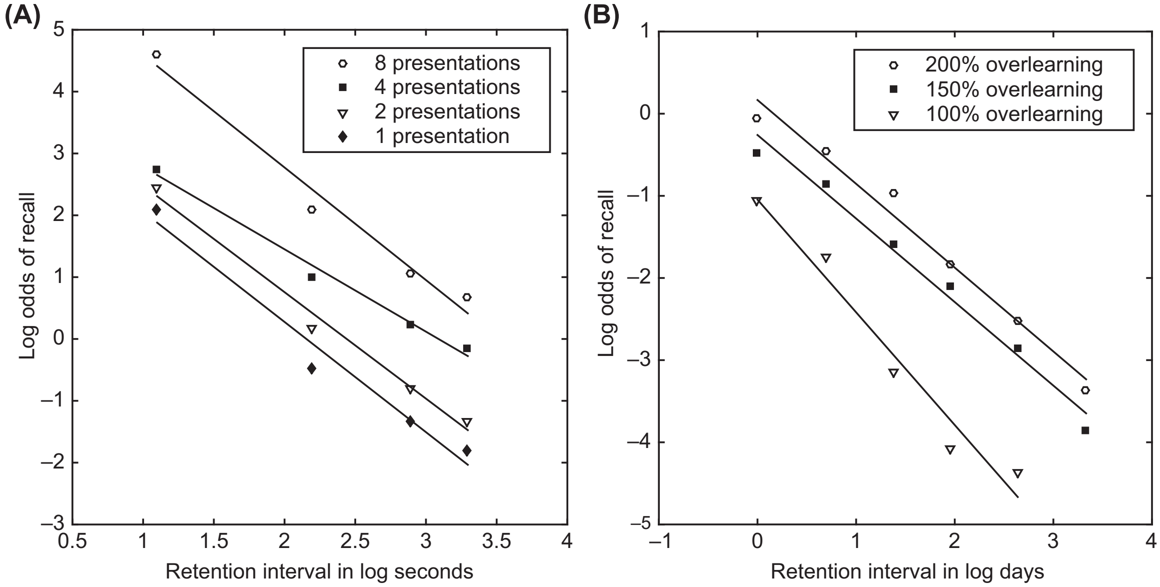

Figure 4 (A) Forgetting curves at four levels. (B) Forgetting curves at three practice levels. Reproduced from (A) Hellyer, S., 1962. Frequency of stimulus presentation and short-term decrement in recall. J. Exp. Psychol. 64, 650; (B) Krueger, W.C.F., 1929. The effects of overlearning on retention. J. Exp. Psychol. 12, 71–78.

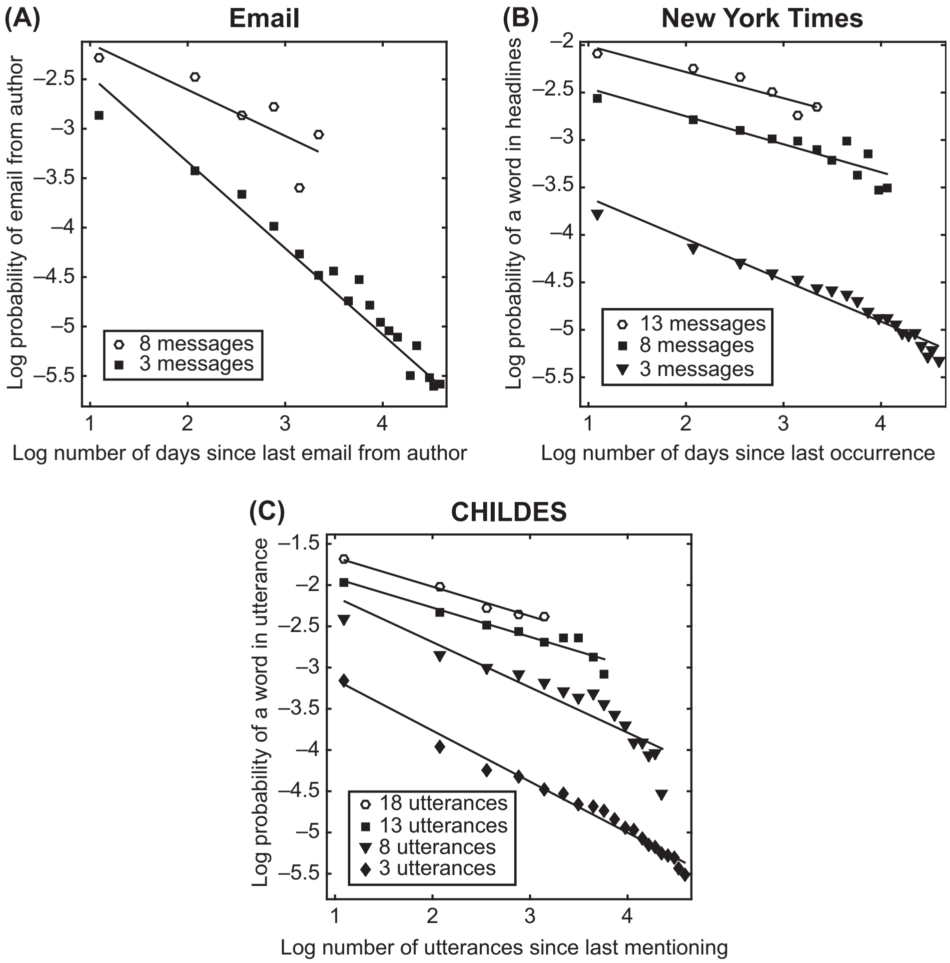

Figure 5 Combined frequency and recency effects in (A) email authors, (B) the New York Times headlines, and (C) the CHILDES database. CHILDES, Child Language Data Exchange System. Adapted from Schooler, L.J., 1993. Memory and the Statistical Structure of the Environment. Department of Psychology, Carnegie Mellon University (Ph.D. dissertation).

2.15.2.2 The Rational Analysis of Primate 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 email and newspapers, one may wonder about the informational demands that were placed on early hominids during critical periods in evolution (e.g., Shettleworth, 2010). It is, of course, impossible to study these environments directly. 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) argues "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. Serio-Silva et al. (2016) analyzed the behavior of howler monkeys through forests and baboons through savanna. It does appear that the visitation patterns of howlers and baboons display frequency and recency functions reminiscent of the human communication databases.

Stevens et al. (2016) extended the rational analysis of memory from humans to chimpanzees, one of our closest relatives. The aim of the work was much like the diary study of Pachur et al., but instead of using data from 100 days of diary entries, it involved 19 years of interactions between 143 chimpanzees living in Kibale National Park in Uganda. The chimpanzees live in fluid groups of 40–50 individuals, and the composition of the groups can change over the course of days or even hours. Those individuals who were observed in the same group on a particular day were counted as being in contact for that day and those who were not observed together were counted as not having contact. Despite the complications inherent in putting together a natural data set on this scale, Stevens et al. analyzed 4.9 million observations about whether or not two individuals had contact on a particular day. Fig. 2B shows the probability that an individual chimpanzee will be in contact with the other as a function of how long it has been since they were last in contact. The power function fit is quite remarkable, with the evidence favoring the power function over the exponential by 11 million times. For comparison, Fig. 2A shows the contact data from Pachur et al. (2014).

Stevens et al. (2016) predicted that chimpanzee retention functions should be described well by a power function. In a delayed matching to sample experiment, a chimpanzee is shown an object that is then hidden. Following a delay the object is presented along with a lure that had previously not been seen. The chimpanzee is rewarded when it chooses the object it had seen before. A retention function can be mapped out by varying the time between when the object was shown and when it was tested. Fig. 6 shows data combined from three delayed matching to sample experiments. As predicted, the power function fits the chimpanzee retention function better than the exponential.

2.15.2.3 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), in which 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

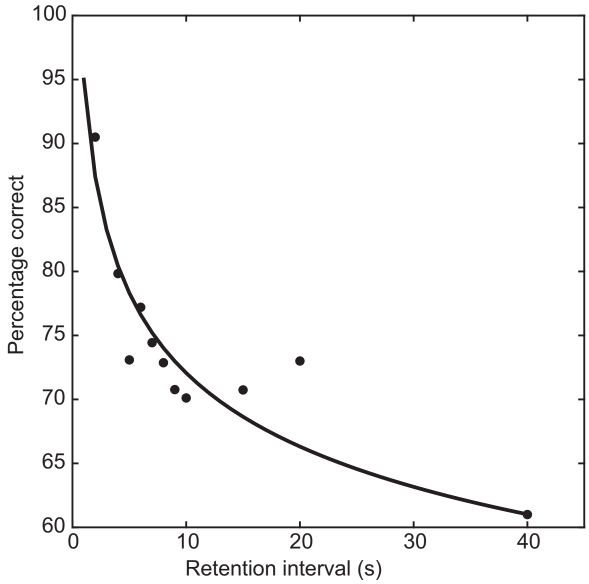

Figure 6 Chimpanzee memory performance on a delayed match-to-sample task as a function of retention interval. Adapted from Stevens, J.R., Marewski, J.N., Schooler, L.J., Gilby, I.C., 2016. Reflections of the social environment in chimpanzee memory: applying rational analysis beyond humans. R. Soc. Open Sci. 3, 160293.

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 most likely a headline that mentioned Qaddafi the next day would also mention 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 encoding specificity.

Schooler and Anderson (1997) explored the extent to which the effects of context went beyond the obvious in the New York Times and CHILDES databases. They collected likelihood ratio measures of association between various words. This was measured as the associative ratio, $P(S|q)/P(S)$, 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, $P(S|q)$, 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 Eq. (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. For example, "AIDS" 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., 0.75/0.018), 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 Eq. 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 et al., 1990).

Schooler and Anderson (1997) looked at how this contextual factor combined with the historical factor. They examined how these two factors combined in the New York Times and CHILDES databases. The results are displayed in **Fig. 7** that shows that both the presence of a high associate ("strong context" in **Fig. 8**) and the time since last appearance ("retention" in **Fig. 7**) affect the probability of occurrence in the critical unit of time. It might appear from **Fig. 8** 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 timescale to log time (see Anderson and Schooler, 1991 for the justification), one gets the functions in **Fig. 7B and D**. These figures show parallel linear functions, which is the additivity predicted by the Bayesian log formula mentioned earlier (Eq. 4).

The interesting question is whether human memory is similarly sensitive to these factors. Schooler and Anderson (1997) conducted an experiment in which they 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 **Fig. 8**. 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) noted that similar parallel functions have been found in other studies of human memory (e.g., Mantyla and Nilsson, 1988; Thomson, 1972), but such data had not 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 in which 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 the general cues of the environment, drug state, or emotional mood (for review, see Anderson, 2000). 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 cooccurrence 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 cooccurrence 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 Conclusions section.

2.15.3 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 the research that indicates that how subjects study an event and the content of an event have substantial impact on the memory of the event. As Anderson and Milson (1989) showed, some of the effects of study strategy can be conceived of as the subject manipulating their 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.

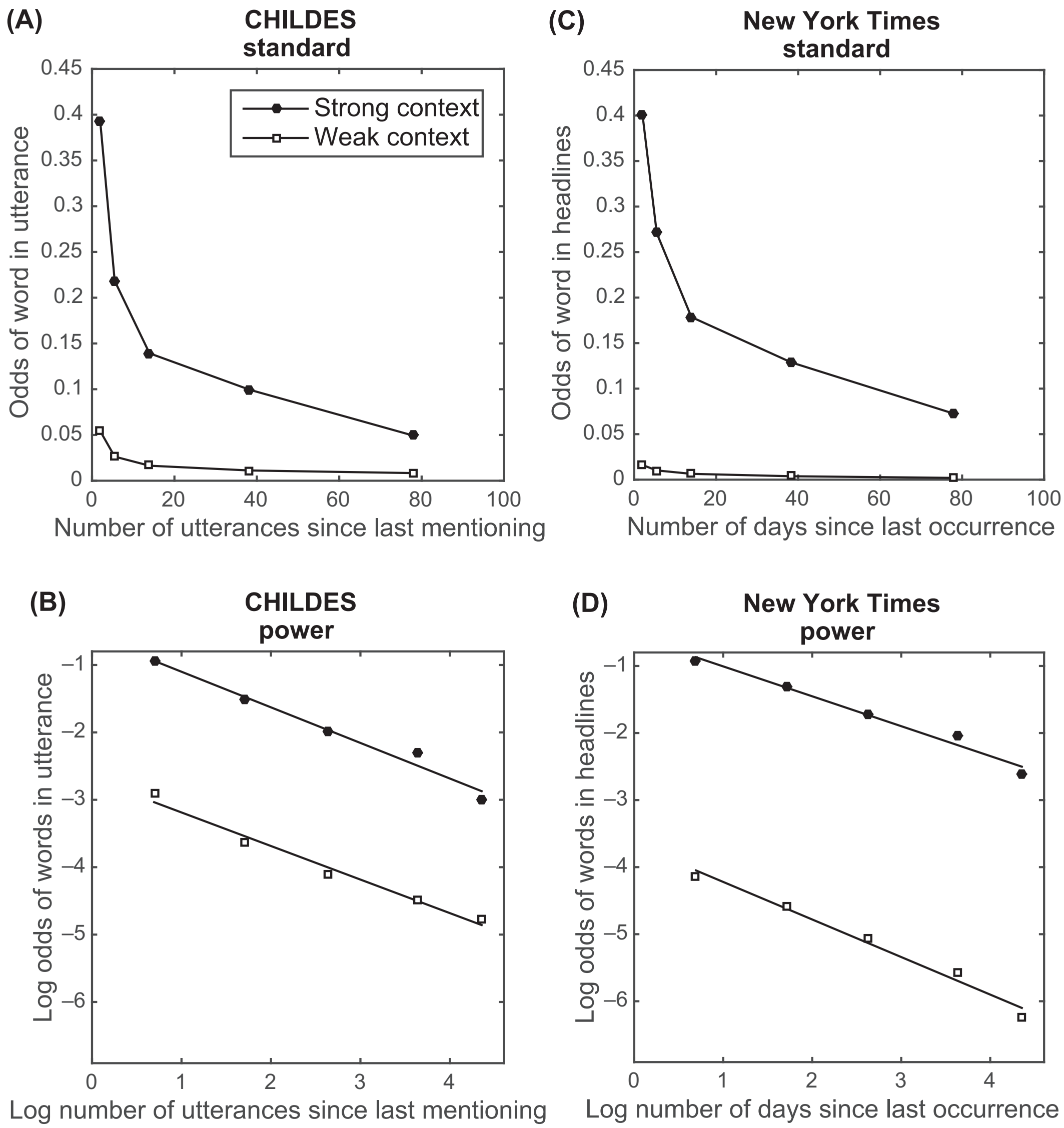

Figure 7 Environmental recency curves as a function of context from the analysis of the (A, B) CHILDES database and (C, D) the New York Times headlines. CHILDES, Child Language Data Exchange System. Adapted from Schooler, L.J., 1993. Memory and the Statistical Structure of the Environment. Department of Psychology, Carnegie Mellon University (Ph.D. dissertation).

Anderson and Milson showed 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 and 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 et al., 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 analyses 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 the 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 remember things about themselves and other people close to them better than about strangers (Rogers et al., 1977). One might be similarly tempted to interpret the apparent superiority for flashbulb memories for details about dramatic events (e.g., Palmer et al., 1991). However, work by Talarico and Rubin (2007) on flashbulb

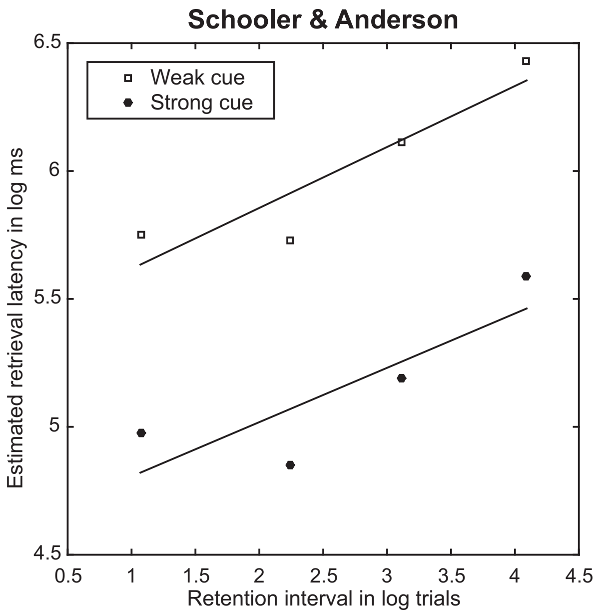

Figure 8 Time to retrieve memories as a function of context and retention interval. Adapted from Schooler, L.J., 1993. Memory and the Statistical Structure of the Environment. Department of Psychology, Carnegie Mellon University (Ph.D. dissertation).

memories following the events of 9/11 calls into question the resilience of flashbulb memories. They found that the consistency of Duke students' recollections of hearing about the events of 9/11 declined at the same rate as their recollections of everyday events. However, the students' confidence in the 9/11 recollections remained high for the flashbulb memories, whereas confidence in their recollections of everyday events fell sharply. Talarico and Rubin conclude that flashbulb memories are not particularly accurate, although they do differ in phenomenology, such as confidence and vividness. As Talarico and Rubin point out, these flashbulb memories result not just from the initial encoding of the events surrounding when the students heard about the events of 9/11, but by the numerous retelling of the events. The social nature of these retellings would make the 9/11 memories particularly susceptible to postevent elaboration. As we discuss in the Conclusions section, such postevent elaborations point to limitations of the rational analysis of memory.

As another example of the principle that recall should be better for more important material, people tend to show better memory for the meaningful aspects of an event than its unimportant superficial details (e.g., Wanner, 1968). Nairne, Pandeirada, and Ferandes (this volume) argue that survival-relevant memories have a mnemonic advantage. In all these cases, there is an alternative explanation to the advantage of such memories, which is that their importance causes us to rehearse or elaborate on them more. In effect, we can warp the statistics of the environment to favor the more important memories.

Whether because of a special advantage or because of special processing, it is reassuring that memory is better for more important things, but it would be more informative if we understood how this interacted with other memory factors. 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. The rational analysis would predict that, although more important memories will display overall better memory, there would be parallel retention functions on log–log scale for memories of different importance just as there were parallel retention functions for memory of different amounts of study (**Fig. 4**). This is because P and G multiply in the utility analysis and a multiplicative relationship implies an additive relationship in log–log scale. In the case of at least one variable plausibly affecting importance, namely, arousal, there seems to be contradictory evidence. Levonian (1972) found a crossover interaction with retention interval. More recently, Mickley Steinmetz et al. (2012) looked at the interaction between arousal by testing performance at retention intervals of 30 min and 24 h. **Fig. 9** plots their data in log–log coordinates, in which the parallel lines indicate that the prediction of the rational analysis does appear to hold. A thorough investigation of this prediction would require testing performance at additional retention intervals and levels of arousal.

The additive utility prediction can also be tested by manipulating the amount of practice along with the importance of information. 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, 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. **Fig. 10** displays the results. So this is an instance of the nonobvious prediction that the importance of the material elevates memory but does not change the basic memory functions.

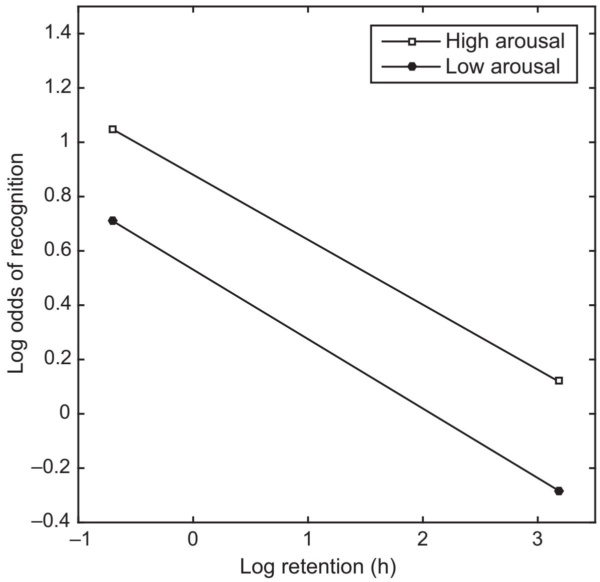

Figure 9 Recognition performance as a function of retention interval and arousal level at the time of study. Adapted from Mickley Steinmetz, K.R., Schmidt, K., Zucker, H.R., Kensinger, E.A., 2012. The effect of emotional arousal and retention delay on subsequent-memory effects. Cogn. Neurosci. 3 (3–4), 150–159. http://dx.doi.org/10.1080/17588928.2012.677421.

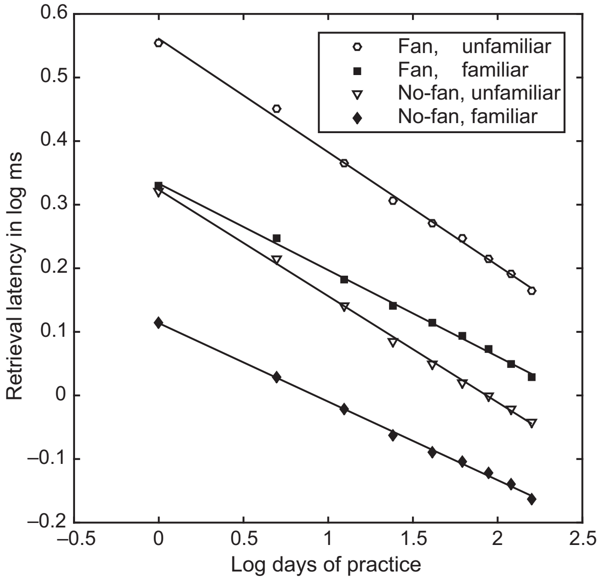

Figure 10 Recognition of fan and no-fan sentences about familiar and unfamiliar concepts as a function of practice. The results are plotted in log–log coordinates. Adapted from Anderson, J.R., Schooler, L.J., 2000. The adaptive nature of memory. In: Tulving, E., Craik, F.I.M. (Eds.), The Oxford Handbook of Memory, pp. 557–570. Oxford University Press.

2.15.4 Conclusions

Despite the generally successful tone of this chapter, it is not the case that all factors affecting memory can be understood in terms of a straightforward adaptive analysis. As one example, in an adaptive framework, it is hard to understand the effects of intention to learn on memory. Instructing a subject to learn or paying a subject to learn can result in differential study patterns (see Anderson, 2000 for a review), but it appears that, controlling for study pattern, there is no effect of motivation or intention to learn. On the face of it, an adaptive system should be able to give more resources to information judged to be 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, people can actively create their own unobserved environment with its own statistics. In the current terms, people 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.

The rational analysis of memory treats the contents of memories as immutable, focusing solely on the issue as to whether or not the memory can be retrieved. Going beyond such a simple characterization of forgetting, Schacter (1999) developed a taxonomy of memory failures, he likens to the seven deadly sins. For example, the sin of *transience* refers to the general erosion of memories over time, essentially the kind of forgetting analyzed in the rational analysis of memory. The sin of *absentmindedness* refers to failing to pay attention with the result that details about the original event cannot be recalled because they were never properly stored in the first place. The sins of transience and absentmindedness describe memory errors of omission. The event is simply forgotten. Other sins characterize the mutability of memories. It is not that a specific event is entirely forgotten, but rather is misremembered in some way. For example, the sin of *suggestibility* refers to a memory that was properly encoded but has been distorted by suggestions at the time of retrieval, as is likely the case for flashbulb memories that are repeatedly recounted over time. Memory demonstrates *biases* generalized from past experience; we may clearly remember books in a professor's office where there had been none. The sins of suggestibility and bias exemplify errors of commission. Schacter makes the point that these apparent sins of commission can serve valuable, adaptive functions. Most professors' offices do indeed have books, so it is a good bet to remember books being there, even when we may not have noticed them in the bookcase. The sin of bias can correct the sin of absentmindedness.

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

See also: 2.16 Adaptive Memory. 2.26 Spacing Effects on Learning and Memory.

References

Anderson, J.R., 1983. The Architecture of Cognition. Harvard University Press, Cambridge, MA.
Anderson, J.R., 1990. The Adaptive Character of Thought. Lawrence Erlbaum Associates, Hillsdale, NJ.
Anderson, J.R., 1993. The Rules of the Mind. Lawrence Erlbaum Associates, Hillsdale, NJ.
Anderson, J.R., 2000. Learning and Memory, second ed. Wiley, New York.
Anderson, J.R., 2007. How Can the Human Mind Occur in the Physical Universe? Oxford University Press, New York.
Anderson, J.R., Bothell, D., Lebiere, C., Matessa, M., 1998. An integrated theory of list memory. J. Mem. Lang. 38, 341–380.
Anderson, J.R., Lebiere, C., 1998. The Atomic Components of Thought. Lawrence Erlbaum Associates, Mahwah, NJ.
Anderson, J.R., Milson, R., 1989. Human memory: an adaptive perspective. Psychol. Rev. 96, 703–719.
Anderson, J.R., Schooler, L.J., 1991. Reflections of the environment in memory. Psychol. Sci. 2, 396–408.
Anderson, R.B., Tweney, R.D., Rivardo, M., Duncan, S., 1997. Need probability affects retention: a direct demonstration. Mem. Cogn. 25, 867–872.
Anderson, J.R., Schooler, L.J., 2000. The adaptive nature of memory. In: Tulving, E., Craik, F.I.M. (Eds.), The Oxford Handbook of Memory. Oxford University Press, pp. 557–570.
Bjork, R.A., Bjork, E.L., 1992. A new theory of disuse and an old theory of stimulus fluctuation. In: Healy, A.F., Kosslyn, S.M., Shiffrin, R.M. (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.
Bogartz, R.S., 1990. Evaluating forgetting curves psychologically. J. Exp. Psychol. Learn. Mem. Cogn. 16, 138–148.
Bowers, K.S., Regehr, G., Balthazard, C.G., Parker, K., 1990. Intuition in the context of discovery. Cogn. Psychol. 22, 72–110.
Bruce, D., 1985. The how and why of ecological memory. J. Exp. Psychol. Gen. 114, 78–90.
Burrell, Q.L., 1980. A simple stochastic model for library loans. J. Doc. 36, 115–132.
Burrell, Q.L., 1985. A note on aging on a library circulation model. J. Doc. 41, 100–115.
Copher, J., Kanfer, A., Walker, M.B., 2002. Everyday communication patterns of heavy and light email users. In: Wellman, B., Haythornthwaite, C. (Eds.), The Internet in Everyday Life. Blackwell, Oxford, UK, pp. 263–288.
Craik, F.I.M., Lockhart, R.S., 1972. Levels of processing: a framework for memory research. J. Verbal Learn. Verbal Behav. 11, 671–684.
Eich, E., Metcalfe, J., 1989. Mood dependent memory for internal versus external events. J. Exp. Psychol. Learn. Mem. Cogn. 15, 443–455.
Hall, W.S., Tirre, W.C., 1979. The Communicative Environment of Young Children: Social Class, Ethnic and Situation Differences. University of Illinois, Center for the Study of Reading.
Hellyer, S., 1962. Frequency of stimulus presentation and short-term decrement in recall. J. Exp. Psychol. 64, 650.
Ijiri, Y., Simon, H.A., 1977. Skew Distributions and the Sizes of Business Firms. North Holland, Amsterdam.
Krueger, W.C.F., 1929. The effects of overlearning on retention. J. Exp. Psychol. 12, 71–78.
Laird, J.E., 2012. The Soar Cognitive Architecture. MIT Press.
Levonian, E., 1972. Retention over time in relation to arousal during learning: an explanation of discrepant results. Acta Psychol. 36, 290–321.
Loftus, G.R., 1985. Evaluating forgetting curves. J. Exp. Psychol. Learn. Mem. Cogn. 11, 397–406.
MacWhinney, B., Snow, C., 1990. The child language data exchange system: an update. J. Child Lang. 17, 457–472.
Mantyla, T., Nilsson, L.G., 1988. Cue distinctiveness and forgetting: effectiveness of self-generated retrieval cues in delayed recall. J. Exp. Psychol. Learn. Mem. Cogn. 14, 502–509.
McClelland, J.L., Chappell, M., 1998. Familiarity breeds differentiation: a subjective-likelihood approach to the effects of experience in recognition memory. Psychol. Rev. 105 (4), 724.
Mickley Steinmetz, K.R., Schmidt, K., Zucker, H.R., Kensinger, E.A., 2012. The effect of emotional arousal and retention delay on subsequent-memory effects. Cogn. Neurosci. 3 (3–4), 150–159. http://dx.doi.org/10.1080/17588928.2012.677421.
Milton, K., 1981. Distribution patterns of tropical plant foods as an evolutionary stimulus to primate mental development. Am. Anthropol. 83, 534–548.

Neisser, U., 1976. Cognition and Reality. Freeman, San Francisco.

Neisser, U., 1978. Memory: what are the important questions? In: Gruneberg, M.M., Morris, P., Sykes, R.H. (Eds.), Practical Aspects of Memory. Academic Press, New York, pp. 3–24.

Neisser, U., 1982. Memory Observed: Remembering in Natural Contexts. Freeman, San Francisco.

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

Pachur, T., Schooler, L.J., Stevens, J.R., 2014. We'll meet again: revealing distributional and temporal patterns of social contact. PLoS One 9 (1), e86081. http://dx.doi.org/10.1371/journal.pone.0086081.

Paivio, A., 1971. Imagery and Verbal Processes. Holt, Rinehart, & Winston, New York.

Palmer, S.E., Schreiber, C., Fox, C.R., November 1991. Remembering the Earthquake: 'Flashbulb' Memories for Experienced Versus Reported Events. Paper presented by Palmer at meeting of the Psychonomics Society, San Francisco.

Ratcliff, R., Clark, S., Shiffrin, R.M., 1990. The list strength effect: I. Data and discussion. J. Exp. Psychol. Learn. Mem. Cogn. 16, 163–178.

Recker, M., Pitkow, J., 1996. Predicting document access in large, multimedia repositories. ACM Trans. Comput. Hum. Interact. (ToCHI) 3, 352–375.

Rogers, T.B., Kuiper, N.A., Kirker, W.S., 1977. Self-reference and the encoding of personal information. J. Pers. Soc. Psychol. 35, 677–688.

Rubin, R.C., Wenzel, A.E., 1996. 100 years of forgetting: a quantitative description of retention. Psychol. Rev. 103, 734–760.

Rumelhart, D.E., McClelland, J.L., 1986. Parallel Distributed Processing, vol. 1. MIT Press, Cambridge, MA.

Salton, G., McGill, M.J., 1983. Introduction to Modern Information Retrieval. MacGraw-Hill, New York.

Schacter, D.L., 1999. The seven sins of memory: insights from psychology and cognitive neuroscience. Am. Psychol 54, 182–203.

Schooler, L.J., 1993. Memory and the Statistical Structure of the Environment. Department of Psychology, Carnegie Mellon University (Ph.D. dissertation).

Schooler, L.J., Anderson, J.R., 1997. The role of process in the rational analysis of memory. Cogn. Psychol. 32, 219–250.

Serio-Silva, J.C., Schooler, L.J., Rhine, R., Chapman, C.A., 2016. Does contrasting ecological niches provide insights into human memory and selective pressures on early hominins? Manuscript in preparation.

Shettleworth, S.J., 2010. Cognition, Evolution, and Behavior. Oxford University Press, Oxford.

Shiffrin, R.M., Steyvers, M., 1997. A model for recognition memory: REM–Retrieving effectively from memory. Psychon. Bull. Rev. 4, 145–166.

Simon, H.A., 1991. Cognitive architectures and rational analysis: comment. In: Van Lehn, K. (Ed.), Architectures for Intelligence. Erlbaum, Hillsdale, NJ, pp. 15–39.

Slamecka, N.J., McElree, B., 1983. Normal forgetting of verbal lists as a function of their degree of learning. J. Exp. Psychol. Learn. Mem. Cogn. 9, 384–397.

Stevens, J.R., Marewski, J.N., Schooler, L.J., Gilby, I.C., 2016. Reflections of the social environment in chimpanzee memory: applying rational analysis beyond humans. R. Soc. Open Sci. 3, 160293.

Stritter, E.P., 1977. File Migration. Stanford University, Stanford, CA (Unpublished doctoral dissertation).

Talarico, J.M., Rubin, D.C., 2007. Flashbulb memories are special after all; in phenomenology, not accuracy. Appl. Cogn. Psychol. 21, 557–578.

Thomson, D., 1972. Context effects in recognition memory. J. Verbal Learn. Verbal Behav. 11, 497–511.

Tulving, E., 1975. Ecphoric processing in recall and recognition. In: Brown, J. (Ed.), Recall and Recognition. Wiley, London.

Wanner, H.E., 1968. On Remembering, Forgetting, and Understanding Sentences: A Study of the Deep Structure Hypothesis. Harvard University (Unpublished doctoral dissertation).

Wixted, J.T., Ebbesen, E.B., 1991. On the form of forgetting. Psychol. Sci. 2, 409–415.