Carnegie Mellon University
Research Showcase @ CMU

Department of Psychology | Dietrich College of Humanities and Social Sciences

November 2008

Recency and Context: An Environmental Analysis of Memory

Lael J. Schooler
Carnegie Mellon University

John R. Anderson
Carnegie Mellon University

Follow this and additional works at: http://repository.cmu.edu/psychology

This Conference Proceeding is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Psychology by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Recency and Context: An Environmental Analysis of Memory

Lael J. Schooler **John R. Anderson**
Department of Psychology
Carnegie Mellon
Pittsburgh PA 15213

Abstract

Central to the rational analysis of memory is the proposal that memory's sensitivity to statistical structure in the environment enables it to optimally estimate the odds that a memory trace will be needed now (Anderson, 1990). These odds are based on (1) the pattern of prior use of the memory (e.g., how recently it has been needed) and (2) the similarity of the current context to the previous contexts in which it has been needed. We have analyzed three sources of informational demand in the environment: (1) speech to children; (2) word usage in front page headlines; and (3) the daily distribution of authors of electronic mail. We found that the factors that govern memory performance, including recency, also predict the odds that an item (e.g., a word or author) will be encountered now. Here we tested a basic prediction the theory makes about the independence between context and recency in the environment by extending our previous analysis of recency in the New York Times. Though the results of four behavioral experiment were inconsistent with this independence assumption, the combination of the rational and environmental analyses were able to account for 94% of the variance in these experiments.

Introduction

A central question of cognitive psychology is: What mechanisms and representations underlie human memory? A less frequently asked question is: Why does human memory display the behavior that it does? Anderson (1990) has proposed an answer to this latter question: memory functions to optimally recall relevant information. Few would disagree with the claim that memory retrieves relevant information. It is the assertion that memory performs this task optimally that runs against intuition. Anderson does not suggest that memory does a perfect job of retrieval; no system could do this given the dual constraints of limited resources and a non-deterministic environment. Rather memory does as good a job as could be done, given these constraints.

A rational system should retrieve a memory when the expected gain of retrieving it, E(R), exceeds the expected gain of not retrieving it, E(~R). Schooler (1993) showed that E(R) > E(~R) will be true when the odds that it will be needed, O_n, exceed some cost benefit ratio:

$$O_n > \frac{\text{costs}}{\text{benefits}} \qquad \text{Eq. 1,}$$

where P_n is the probability that a memory will be needed and $O_n = \frac{P_n}{1-P_n}$.

Within this framework, the mapping between need odds and performance measures, such as odds of recall, are somewhat more complex than they might initially appear. One might expect that need odds and performance would simply be proportional to each other, but this is not the case. Take for example the relationship between need odds and probability of recall. Equation 1 implies that probability of recall should be a step function with a discontinuity at *costs/benefits*. Memories with need odds below this point should never be recalled and those over should always be recalled.

Anderson (1990) notes that noise in the estimate of the distance between need odds and this cut off implies that odds of recall will be a power function of need odds:

$$O_r = \left(\frac{\text{benefits}}{\text{costs}}\right)^{1/s} O_n^{\ 1/s} \qquad \text{Eq. 2,}$$

where s is related to the variance in the estimate.

Estimating Need Odds

Anderson (1990) offers two factors that could reasonably be expected to predict need odds: (1) the pattern of prior use of the memory (*i.e.*, its history) and (2) the similarity of the current context to the previous contexts in which it has been needed. An item's past level of frequency and recency of use, and to a lesser degree spacing between these uses, constitute its history. The context factor measures the strength of the association between a particular memory trace and the current context. Here the context is taken to be particularly salient elements of the environment. Within this framework, memory's estimate of the odds that a particular memory trace will be needed is the product of the history and context factors:

$$O_n = O_n(\text{History})\ \text{Context} \qquad \text{Eq. 3}$$

Anderson (1990) accounted for many empirical findings by applying this equation to a model of the statistical structure of the environment. This model of the environment, however, was based on plausible assumptions about the informational demands that the environment places on memory. These assumptions needed to be tested.

History Factor

We have studied constrained "environments" that place demands on human memory. We have found that the historical factors that govern memory performance also predict the odds with which words are spoken in children's linguistic environments[1] (Schooler, 1993; Anderson & Schooler, 1991; Schooler & Anderson, 1991). These factors include frequency, recency, and spacing between exposures. The results were replicated in the patterns of word use in newspaper headlines and the daily distribution of authors who sent electronic mail to JA.

Recency in Memory

Anderson & Schooler (1991) focused exclusively on the history factor. Here we will describe our analysis of the relationship between context and history. But first, as an example, we will review our analysis of retention--the effect of the interval since the last exposure to a memory item on performance.

The retention function is studied typically by giving subjects a constant experience with one or more exposures to an item, and then testing for recall at various retention intervals. Squire (1989) studied a retention function that was closely tied to subjects' day to day experience. He presented subjects with the potential names of TV shows. They had to decide whether the show had aired. Figure 1 plots subject performance as a function of the number years since the show's cancellation.

The negatively accelerated curve is typical of memory retention functions. Wickelgren (1975) proposed that such curves could best be described as power functions, which will be straight when plotted in log-log coordinates. Figure 2 plots Squire's data in this scale and it does appear relatively straight ($r^2 = .97$).

[1] The speech was from the Hall & Tirre (1979) corpus in MacWhinney and Snow's (1990) CHILDES database.

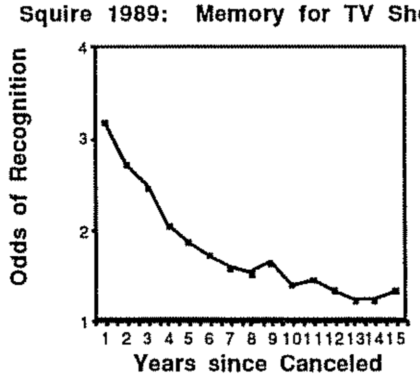

Figure 1. Retention function from Squire (1989) in standard coordinates.

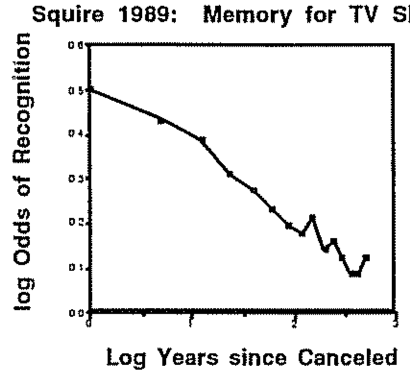

Figure 2. Retention function from Squire (1989) in log log coordinates. Straight line indicates power function.

Recency in the Environment

We show the results of our analysis of the New York Times plotted in log-log coordinates (Figure 3). There we have plotted the log odds of a word being included in a particular front page headline as a function of the log number of days since the word was last included. The curve is straight when plotted in log-log, suggesting that the environmental recency function, like its behavioral counterpart, can be described by a power function ($r^2 = .99$). Similar results hold for the analyses of speech to children and the authors of mail messages.

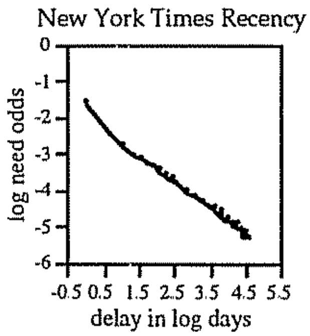

Figure 3. Environmental recency function.

In sum, there is a power function relating need odds (e.g., the odds that a word will be included in the front page headlines on a particular day) and recency. This corresponds to the behavioral relationship between memory performance and recency.

Context Factor

In Equation 3 memory's estimate of a particular item's need odds was taken to be the product of its need odds given its history (*e.g.*, how recently you have seen it) multiplied by the context factor. The history factor captures the predictive power of using an item's past use to predict future use. The context factor gauges the strength of association between the context and a particular item. This information adjusts the estimates of need odds coming from the history factor. Contextual strength is measured in terms of associative ratios that approximate the likelihood ratios common in Bayesian statistics. The denominator of this ratio is the base rate need probability of the item. The numerator is the conditional probability of finding the item in the presence of some cue. The larger this ratio the better the cue, the greater the associative strength. For most items and cues this ratio should hover around one: an item is no more or less likely to appear in the presence of a particular cue. 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:

$$\text{Context} = \prod_{q \text{ in Context}} \frac{P(i/q)}{P(i)}$$

Independence Assumption

As a matter of mathematical convenience, the rational analysis assumes independence between the history and context factors. A strong assumption of independence such as this should be tested. Within the rational analysis framework this can be done both in the environment and in behavior. First, we will go through the environmental predictions and tests, before returning to their behavioral counterparts.

Environmental Tests of Independence

The implications of the independence assumption between the history and context factor can be seen most clearly by taking the natural logs of each side of Equation 3. The environmental analysis of recency revealed that need odds in the environment was in a power relationship with recency (Figure 3). Evidence for this was the nearly perfect linear relationship between log need odds and log recency. By factoring context into the equation, the presumption is that the recency function for progressively stronger degrees of association should be linear and approximately parallel when plotted in log-log. First, we describe the procedure for estimating the associative ratios, followed by the procedure and test of the independence of recency and context in the environment.

Environmental Analysis of Associative Ratios

Calculating the associative ratios requires estimates of the base rate frequencies of the items as well as the many conditional probabilities of finding one word in the presence of another. 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 occurring in the presence of another 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 1 shows an example of a word from the New York Times with particularly strong associations.

p(AIDS) = .018		
Associates	p(AIDS/associate)	p(AIDS/associate) / p(AIDS)
virus	.75	41.0
spread	.54	29.4
patients	.40	21.8
health	.27	14.6

Table 1. Strong associates of AIDS.

Context and Recency in the New York Times and Speech to Children

Method (New York Times)

The aim of this analysis was to estimate recency curves contingent on whether a headline contained a strong associate. A word's strong associates were those words with associative ratios that exceeded 10. The rest were classified as weak associates. The analysis of the New York Times first selected all the words that had not been mentioned for, say, 10 days. For each of these words the analysis checked each of the headlines on the critical day. For a particular word, all the other words in a headline were classified as either strong or weak associates based on their associative ratios. Estimates were made of the odds of a word occurring in a particular headline as function of the number days since it was last seen, and whether the headline contained at least one strong associate.

Method (CHILDES)

Here need odds was taken to be the odds of a word occurring in a particular utterance and was estimated as a function of the number utterances since a word was last mentioned, and whether the utterance contained at least one strong associate.

Results

Figure 4 (on last page) shows the results in standard (a & b) and power (c & d) coordinates. Table 2 (on last page) summarizes the regression fits to the strong and weak context curves in Figures 4c and 4d. For both the CHILDES and New York Times analysis the decay parameters (slopes) for the strong and weak associate curves are approximately equal. In good agreement with the theory, it appears that the resulting curves are linear and approximately parallel when plotted in log-log coordinates. The effects of association and recency in the environment (or at least in the New York Times and in speech to children) are indeed independent of each other.

In sum, the environmental analyses of the New York Times headlines and CHILDES database support the assumption the rational analysis makes about the independence between the history and context factors. Next we explore whether this independence is reflected in behavioral data.

Behavioral Tests of Independence

Predictions

The previous section established that both a word's past history of use and the statistical association of the current context to the word contribute to an estimate of the word's need odds. Further, it was found that the contribution of the history and context factors were in an independent multiplicative relationship. The corresponding behavioral prediction is that when performance is measured in terms of log need odds of recall the retention curves, corresponding to progressively stronger contexts, should be approximately parallel. This prediction can be mapped onto a cued recall task. In such an experiment the cues can be thought of as setting the context for recall. Thus, varying the degree of association between the cue and the target amounts to varying contextual strength.

Procedure

Figure 5 illustrates a typical trial. A trial started with the subject reading a headline and categorizing it as pertaining most to international, national, or local events. After classifying the headline, a plus sign appeared in the middle of the screen. This was followed by a prime. This prime was followed by a word-stem, consisting of the first two letters of a word and underscores to hold the places of the missing letters. The answer to the word stem (space) was drawn from one of the previous headlines that the subject had read.

Experiment					
1 (n = 32)		2 (n = 19)		3 (n = 20)	
short	long	short	long	short	long
1.5	13.3	.22	10.3	1.3	10.3

Table 3. Average retentions for experiments 1, 2 & 3.

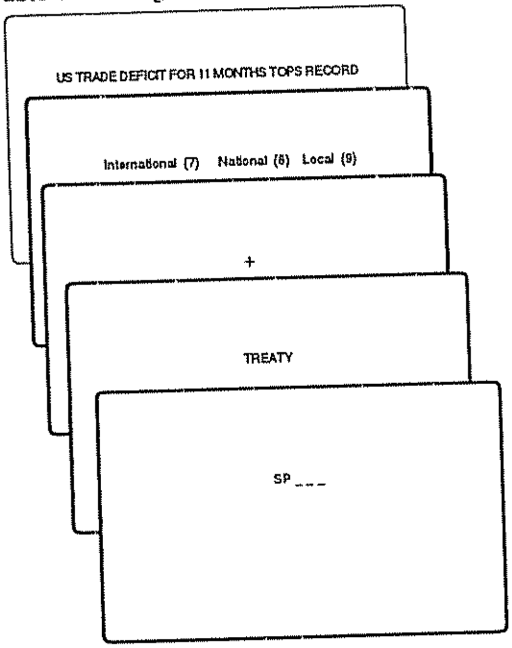

Figure 5. Screens in a typical trial.

Design

To test the predictions of the analysis retention and association were manipulated. The strength of the association between a word stem and a prime was classified as strong or weak, based on their associative ratios derived from the analysis of the New York Times headlines. Retention was measured by the number of intervening trials between when a word was read in a headline and it was the answer to a word stem. Three experiments were run that differed in the average length of their retention intervals. Table 3 lists the average retention intervals for each, along with the number of subjects.

Results

Figure 6 plots the cell means for log odds of recall from experiments 1, 2 and 3 in log-log scale. Based on the results of the power function fits, the slope, or decay rate, for the strong associate curve (-.12) is shallower than the that of the weak associate curve (-.22). The question of the independence of context and recency amounts to asking whether the curves have the same slopes, or decay rates. The difference in slopes is marginally significant at the .1 level; suggesting that context and recency are not independent. Though space precludes going into the details, a fourth experiment, using slightly different methods and materials support these results (Figure 7).

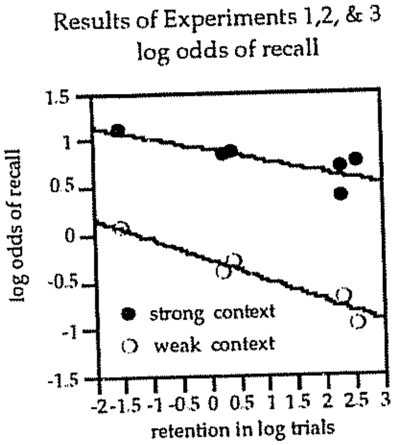

Figure 5. Performance in Experiments 1, 2, & 3 plotted as a function of retention and cue (context) strength.

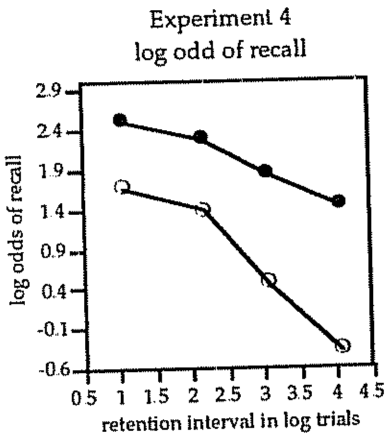

Figure 7. Performance in Experiment 4 (n=33).

In this experiment associative strength was based on word association norms, and the task involved recalling words read in isolation. The decay rate for the strong associate curve (-.45) was significantly smaller than that of the weak associate curve (-.65), $p < .03$.

The prediction the rational and environmental analyses make about the independence between context and recency in memory did not hold up. Nevertheless, we can ask to what extent the variance in the experimental results can be accounted for by a combination of the rational and the environmental analyses.

Performance in each of the experimental conditions corresponds to the results of the environmental analysis of context and recency. For instance, the strong association and long retention interval conditions in Experiments 2 and 3 can be mapped onto a situation where a headline includes a strong associate of the word and the word has not been mentioned for 10 days.

Provided it is plausible to map estimates of need odds based on the environmental analysis directly onto the experiments, and recalling Equation 2, which showed that need odds, O_r, is plausibly in a power function relationship with odds of recall, O_r, then this implies that performance in the various conditions of experiments 1, 2 and 3 should be in a power function relationship with the corresponding results of the analysis of the New York Times.

This in turn implies that there should be a linear relationship between log-need odds and log odds of recall. Figure 8 plots log odds of recall in the three experiments as a function of the corresponding log need odds resulting from the analysis of the New York Times headlines. The relationship does appear to be approximately linear with an R^2 of .94.

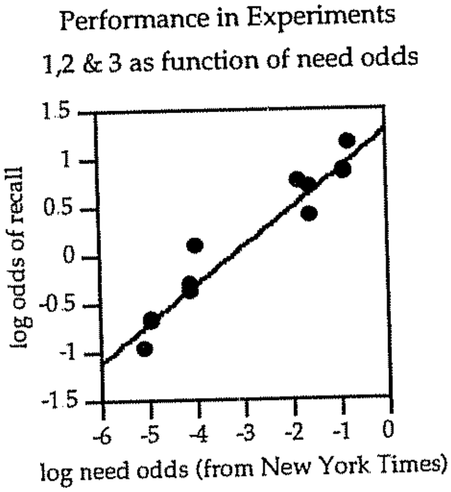

Figure 8. Performance in Experiments 1, 2 and 3 as a function of log need odds.

Conclusion

Within the rational analysis framework the odds that a particular item will be encountered now can be estimated based on the item's past history, and on the statistical association between the item and the current context. Moreover, these estimates based on history and context are argued to be approximately additive in a log-log scale. The environmental analyses reported here were consistent with this prediction. This implies that when memory performance is measured in terms of log odds of recall and retention intervals are measured in a log scale, then the transformed retention functions corresponding to strong and weak contexts should be parallel. In contrast to this prediction, the combined results of four experiments demonstrated that shallower slopes were obtained for the strong context retention curves than for the retention curves associated with weaker contexts. Though this relatively subtle prediction of the rational analysis did not hold up, the results of the environmental analysis and the rational analysis were able to account for 94% of the variance across the results of Experiments 1, 2, and 3.

Environmental Analyses of Context and Recency

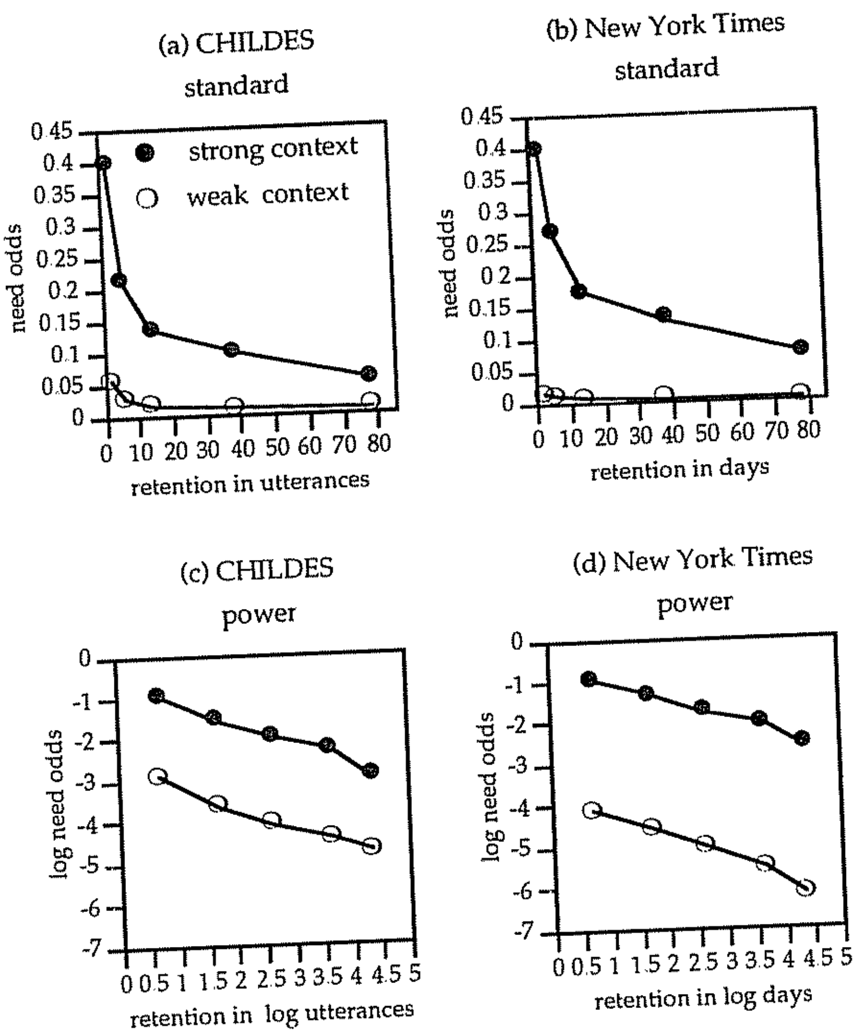

Figure 4. Environmental recency curves from the analysis of the New York Times and CHILDES database. The left panels show the odds of a word being mentioned in a particular utterance as a function of the number of intervening utterances since it was last mentioned and whether the utterance included a strong associate (strong context) or did not (weak context). The right panels show the odds of a word being included in a particular headline as a function of the number of days since the word was last included and whether the headline included a strong associate.

R^2's From Power Fits to the Environmental Analyses of Context and Recency

	CHILDES		New York Times	
	R^2	exponent	R^2	exponent
strong	.98	-.50	.98.	-.44
weak	.98.	-.52	.98.	-.54

Table 2. The degree to which the strong and weak recency curves can be fit by power functions. The estimated exponents are listed as well

References

Anderson, J. R. (1990). The Adaptive Character of Thought. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. , & Schooler, L.J. (1991). Reflections of the Environment in Memory. *Psychological Science*. 2 (6), 396-408.

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.

MacWhinney, B., & Snow, C. (1990). The child language data exchange system: an update. *Journal of Child Language*, 17, 457-472.

Schooler, L.J., & Anderson, J. R. (1991). Does Memory Reflect Statistical Regularity in the Environment? *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* Hillsdale: Erlbaum.

Schooler, L. J. (1993) Memory and the statistical structure of the environment. Ph.D. Diss., Dept of Psychology, Carnegie Mellon.

Wickelgren, W.A. (1975). Single-trace fragility theory of memory dynamics. *Memory and Cognition*, 2, 775-780.