Cognitive Modeling
Lecture 3
Memory

"the memory system […] makes most available those memories most likely to be needed"

(Anderson, 2007, p. 109)

How does it know which memories are most likely to be needed?

The rational basis for declarative memory – **The importance of past history**

Suppose you are a librarian who has to move his library to a smaller building and toss away 50% of the books

The rational basis for declarative memory – **The importance of past history**

Which of these two books would you keep?

Bring this book back before:

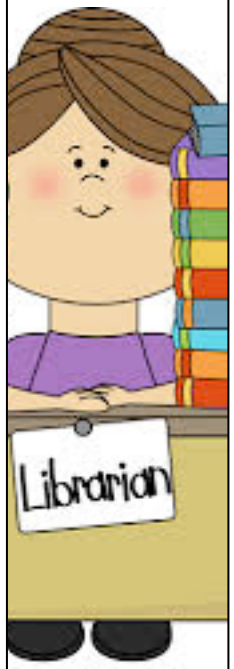
14 Jun 83

21 Feb 84

01 Mar 84

25 Sep 02

05 Jan 14



Bring this book back before:

04 Mar 13

10 Apr 13

05 May 13

21 Jun 13

05 Sep 13

21 Nov 13

Anderson & Schooler, 1991

PSYCHOLOGICAL SCIENCE

Research Article

REFLECTIONS OF THE ENVIRONMENT IN MEMORY

John R. Anderson and Lael J. Schooler

Department of Psychology, Carnegie Mellon University

Abstract—*Availability of human memories for specific items shows reliable relationships to frequency, recency, and pattern of prior exposures to the item. These relationships have defied a systematic theoretical treatment. A number of environmental sources (*New York Times, *parental speech, electronic mail) are examined to show that the probability that a memory will be needed also shows reliable relationships to frequency, recency, and pattern of prior exposures. Moreover, the environmental relationships are the same as the memory relationships. It is argued that human memory has the form it does because it is adapted to these environmental relationships. Models for both the environment and human memory are described. Among the memory phenomena addressed are the practice function, the retention function, the effect of spacing of practice, and the relationship between degree of practice and retention.*

The title of our paper is inspired by the following remark in Shepard (1990): "We may look into that window on the mind as through a glass darkly, but what we are beginning to discern there looks very much like a reflection of the world" (p. 213). [illegible] on how the principles of perception are

current day? Memory would be behaving optimally if it made this memory less available than memories that were more likely to be used but made it more available than less likely memories.

In this paper we examine a number of environmental sources to determine how probability of a memory being needed varies with pattern of past use. However, we first review how availability in human memory varies with pattern of past use. Some aspects of this problem have been extensively studied in empirical studies of human memory.

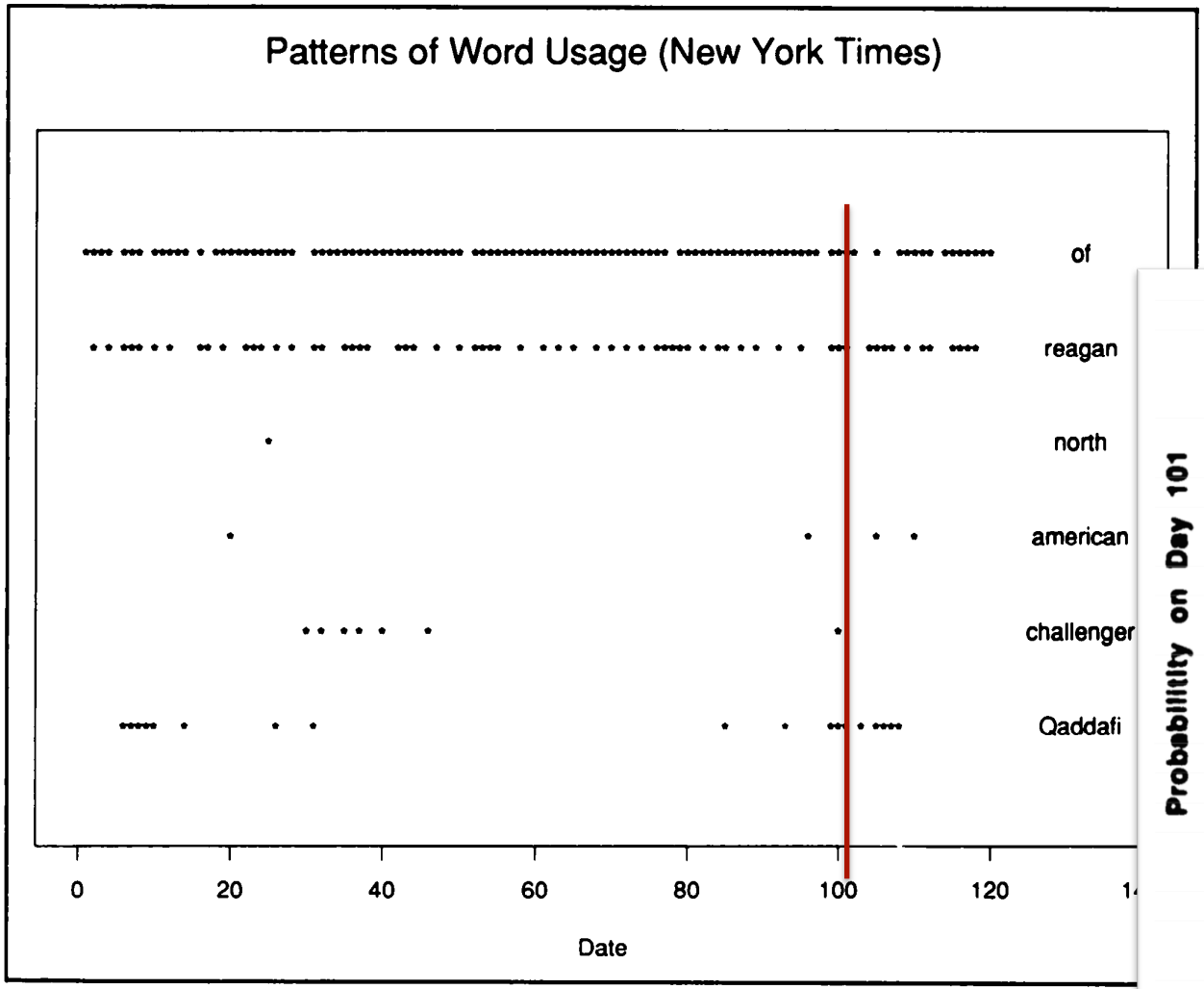
FORM OF THE MEMORY FUNCTIONS

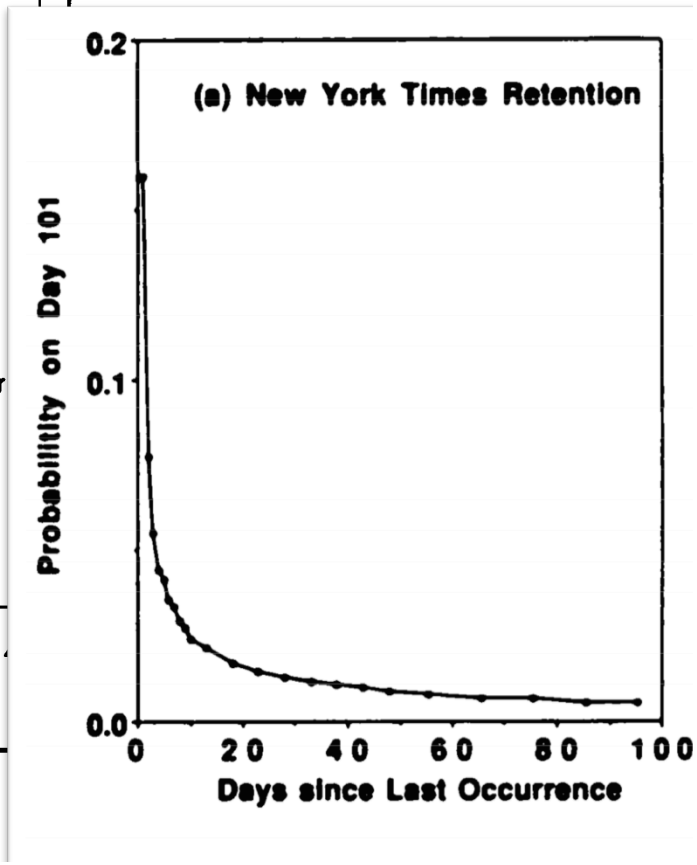
Two of the most basic statistics we might gather about pattern of past use are how often a memory has been practiced and how long it has been since it was last practiced. Learning functions and retention functions to describe these two aspects of human memory have been collected since the original experiments of Ebbinghaus (1885/1964). Figure 1 shows the retention function and practice function obtained by Ebbinghaus.

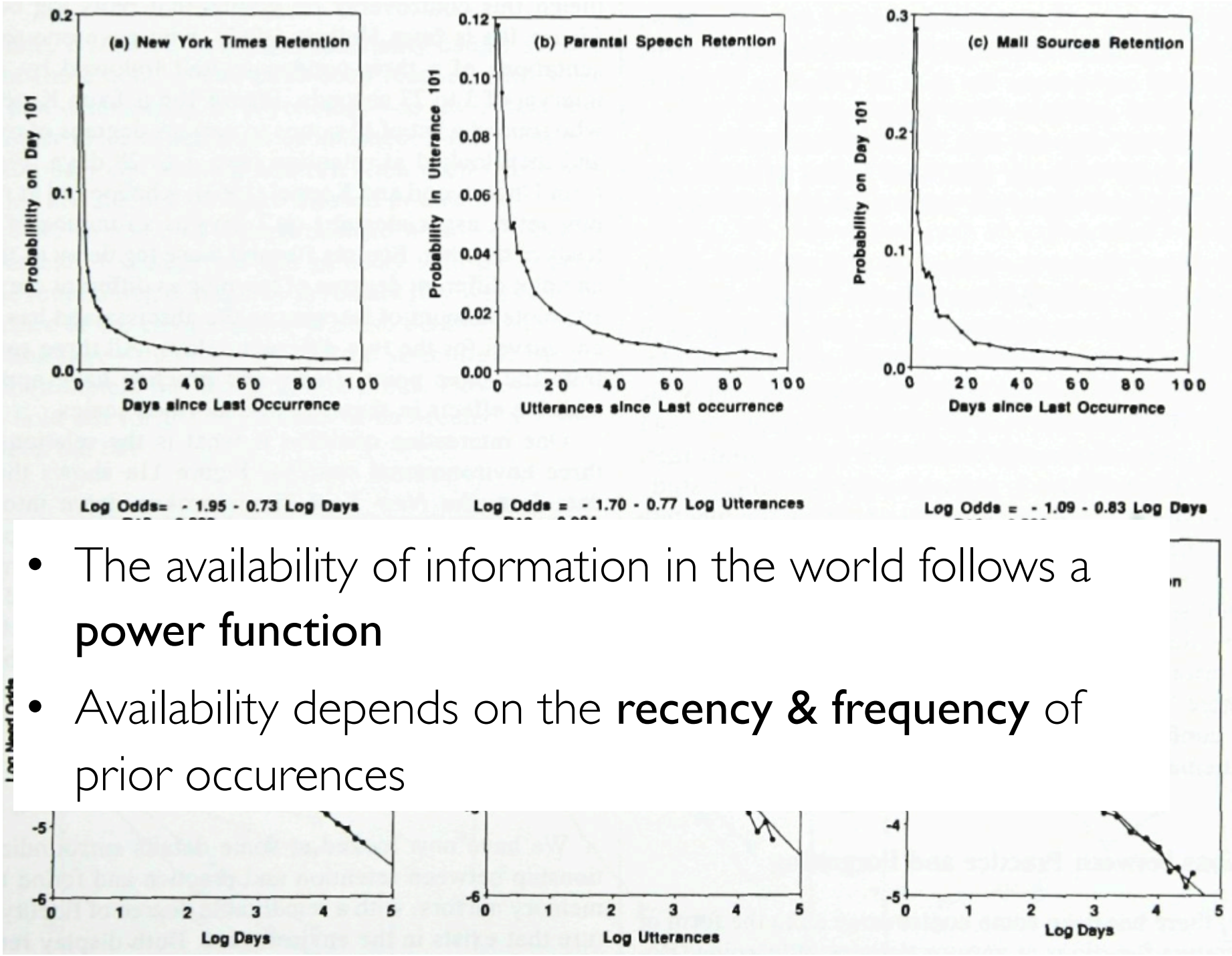
The Retention Function

[illegible] in terms of the percent sav-

How likely is it that a word appears in the headlines of the New York times?



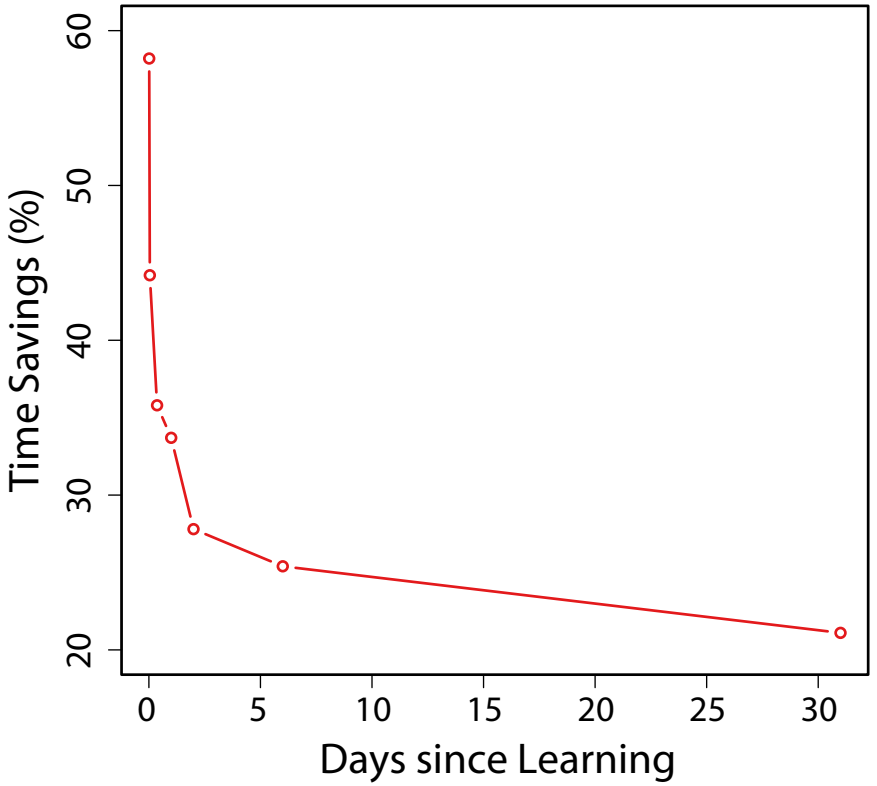


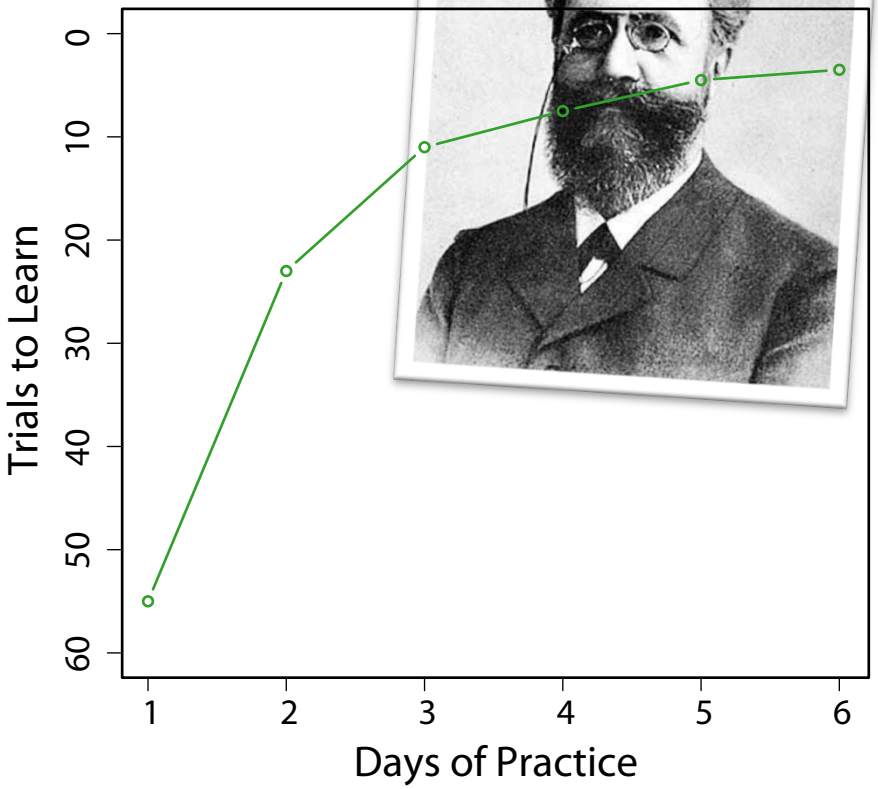


- The availability of information in the world follows a **power function**
- Availability depends on the **recency & frequency** of prior occurences

Availability of information in memory

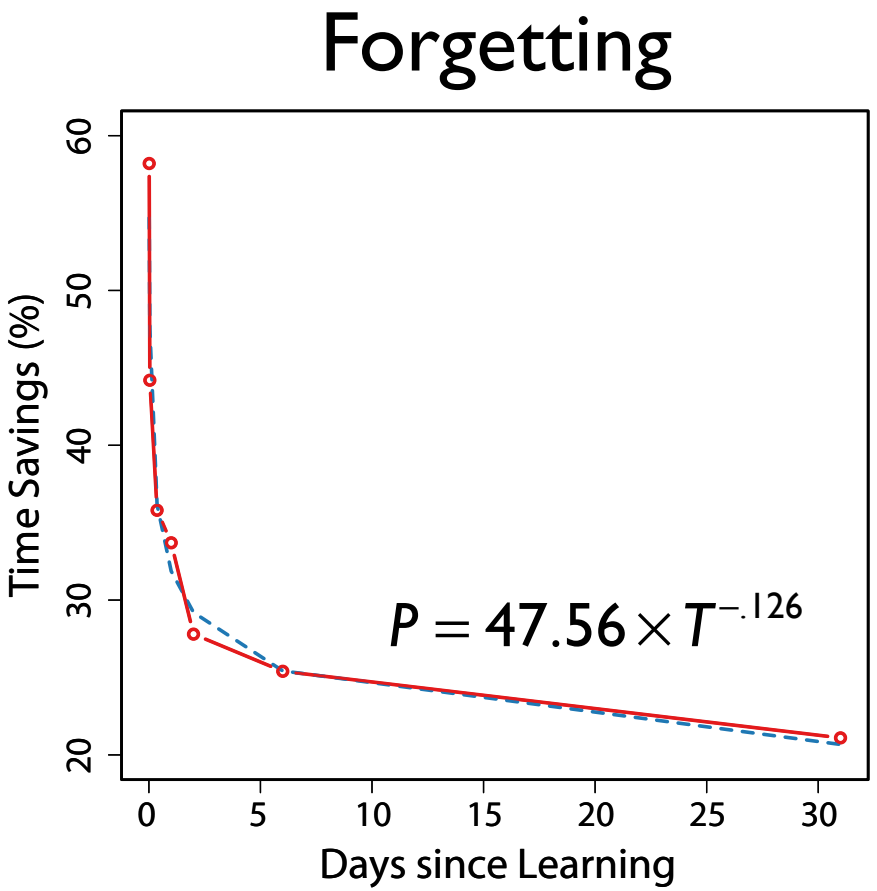
Ebbinghaus (1850-1909)

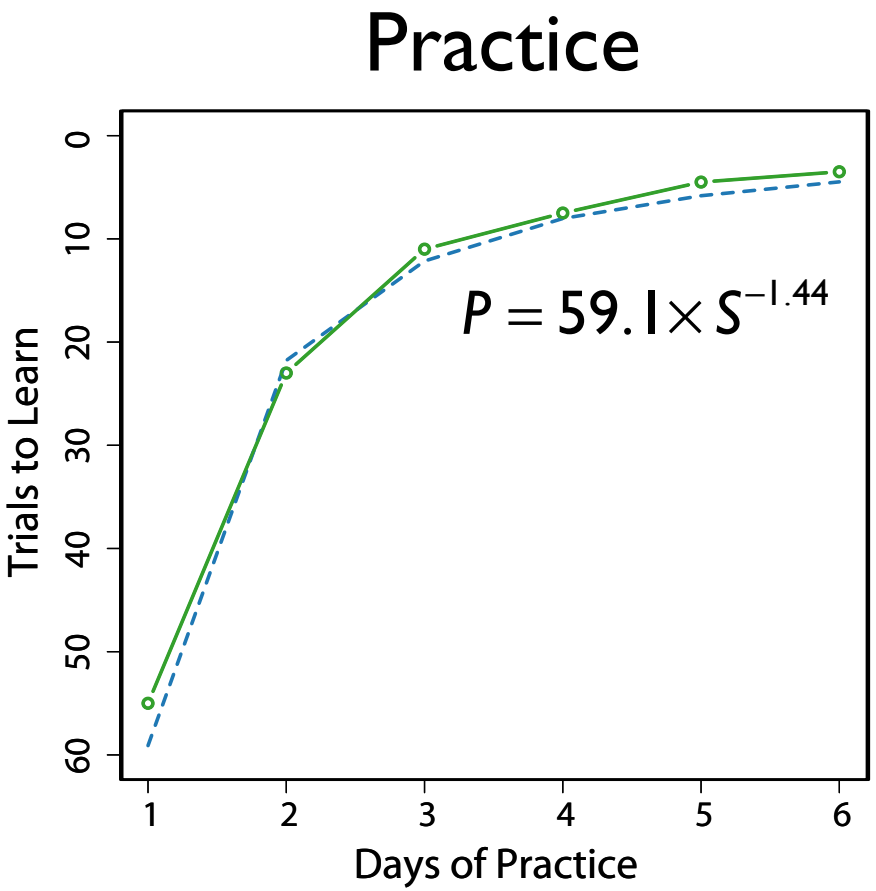




Ebbinghaus, H. (1885/1913)

Availability of information in memory
Ebbinghaus (1850-1909)

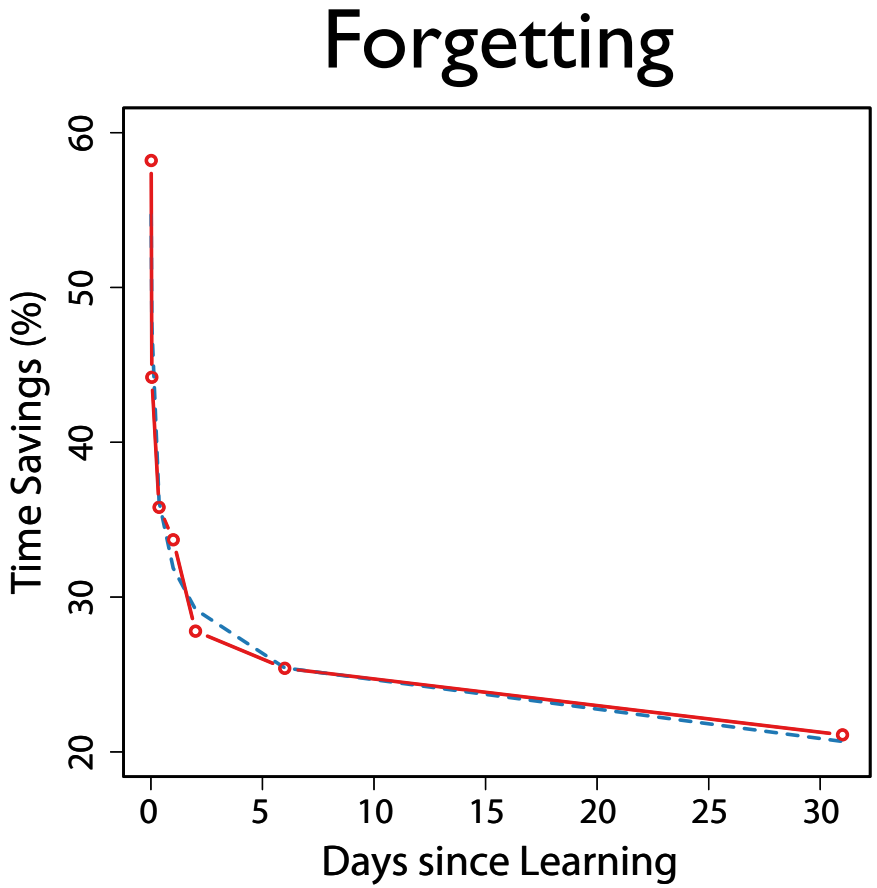




Anderson & Schooler (1991)

Ebbinghaus (1885/1913)

How does forgetting work?



Two theories (mechanisms?) of forgetting

- **Decay**: elements in memory fade away if they are not used
- **Interference**: retrieval of a fact can fail because you cannot find it in between similar facts

Conclusion so far

- The availability of information in the world,
  as well as the availability of information in memory
  follows a power function
- Availability in both cases depends on the recency &
  frequency of prior experiences

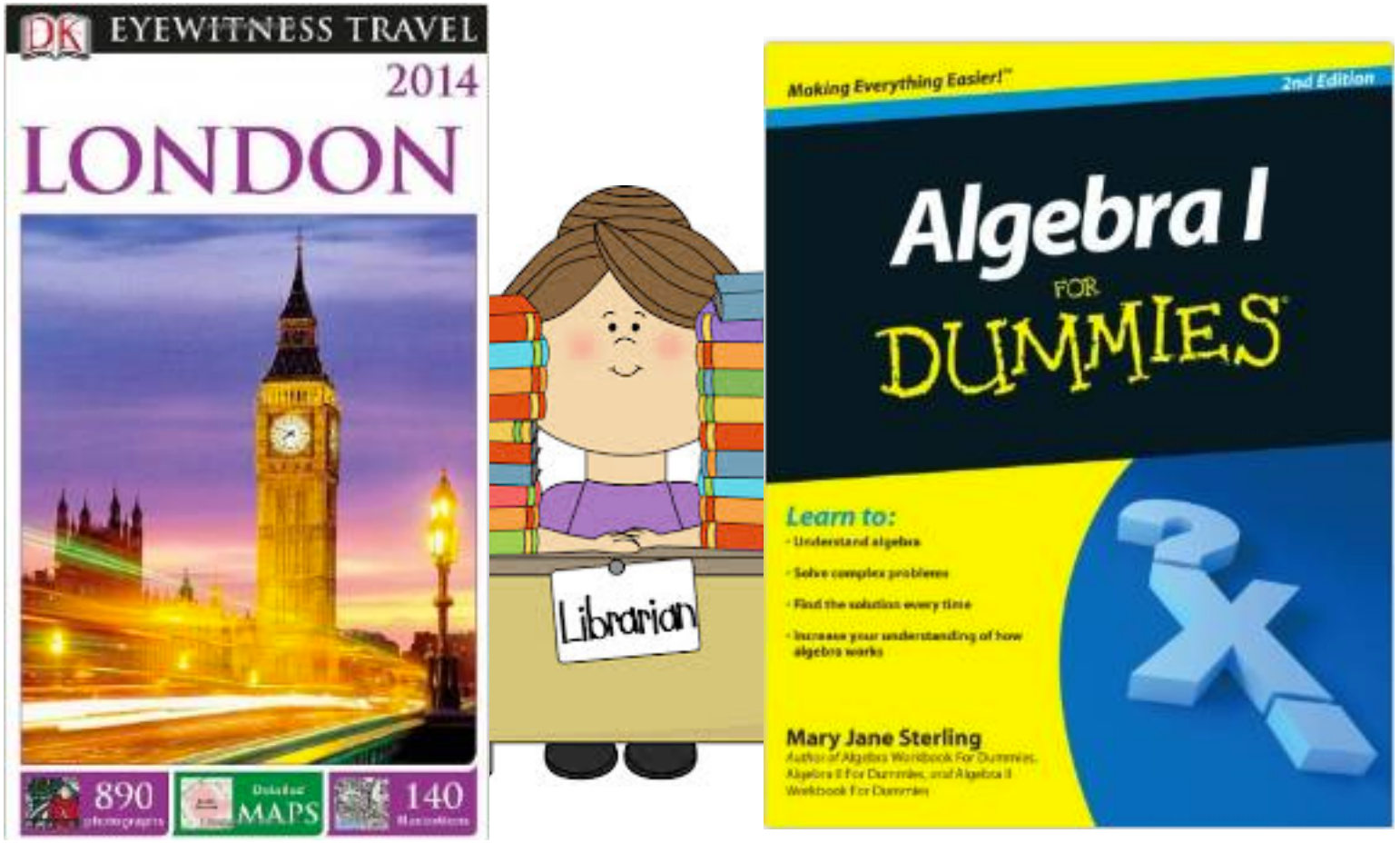
The rational basis for declarative memory – **The importance of current context**

Think of our librarian again.

Assume the new, smaller library specializes on books about traveling.

The rational basis for declarative memory – **The importance of current context**

Which of these books should she keep?



Schooler & Anderson, 1997

COGNITIVE PSYCHOLOGY **32,** 219–250 (1997)
ARTICLE NO. CG970652

The Role of Process in the Rational Analysis of Memory

LAEL J. SCHOOLER

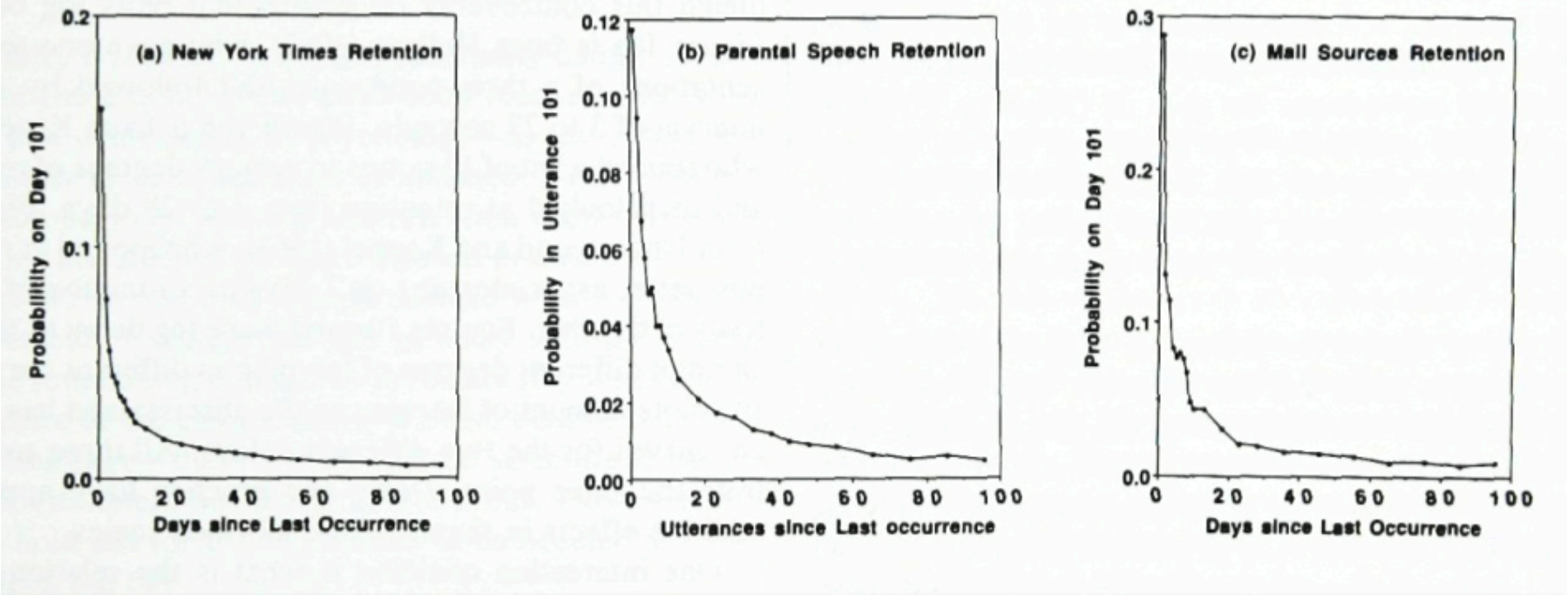
Indiana University

and

JOHN R. ANDERSON

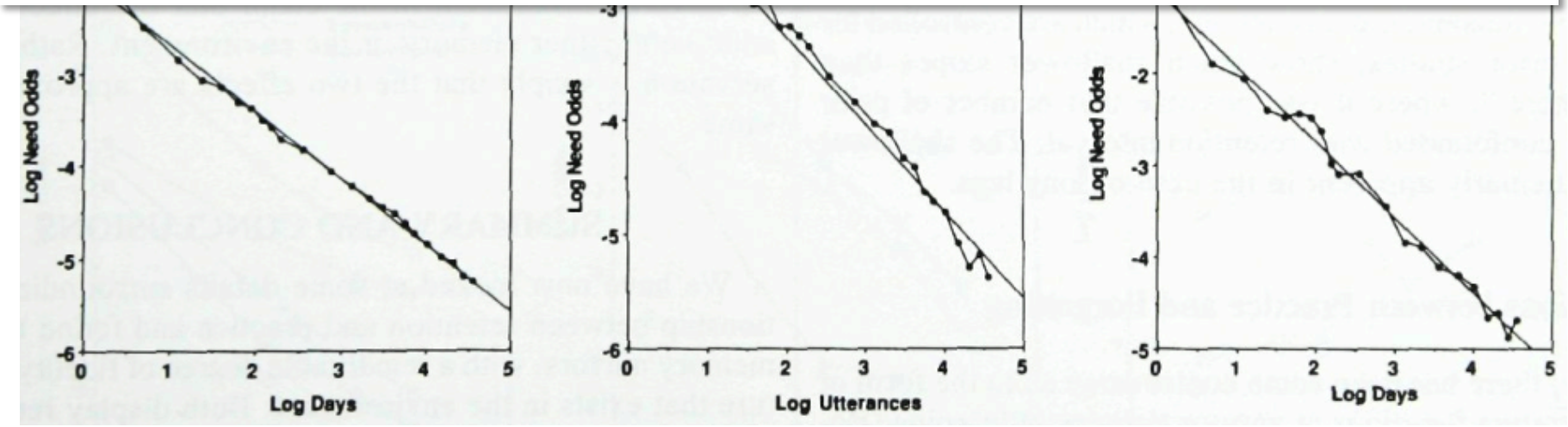
Carnegie Mellon University

The rational analysis of memory (Anderson, 1990) proposes that memory's sensitivity to statistical structure in the environment enables it to optimally estimate the odds that a memory trace will be needed. We have analyzed sources of informational demand in the environment: speech to children and word usage in the front page headlines of the New York Times. In a previous paper (Anderson & Schooler, 1991) [illegible] performance, including recency, also



This data (from the 1991 paper) shows the probability of information being needed as a function of past usage.

Now they ask: How is this probability affected by the current context?



The Effect of Context

1.8% of all New York Times headlines included "AIDS"

p(AIDS) = .018

Associates	p(AIDS|associate)	$\frac{p(AIDS|associate)}{p(AIDS)}$
virus	.75	41.0
spread	.54	29.4
patients	.40	21.8
health	.27	14.6

75% of all headlines that included "virus" also included "AIDS"

"AIDS" is 41 times more likely to occur in a headline that includes "virus" than in one that does not

$$= \frac{.75}{.018}$$

The Effect of Context

<.1% of all parental utterances included "play"

Parental Speech

$p(play) = .0086$

p(play|game)	$\frac{p(play\|game)}{p(play)}$
.41	47.3

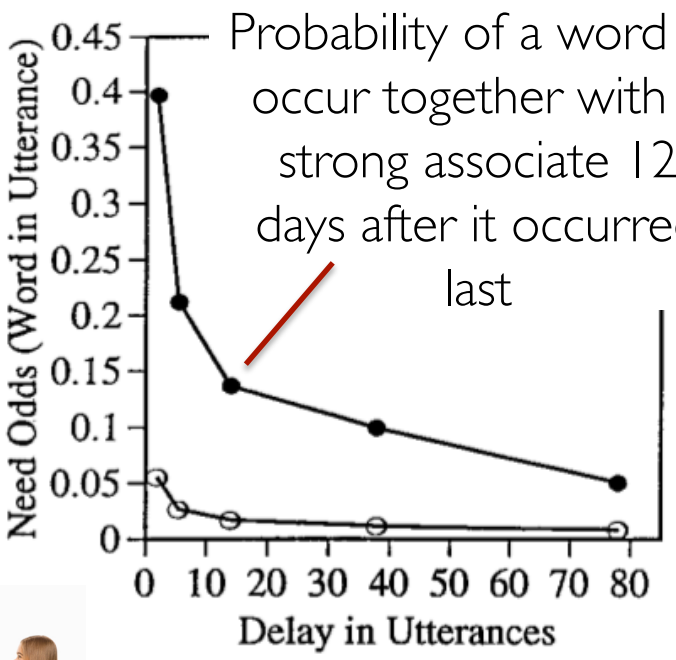
41% of all utt
included "gam
included "play

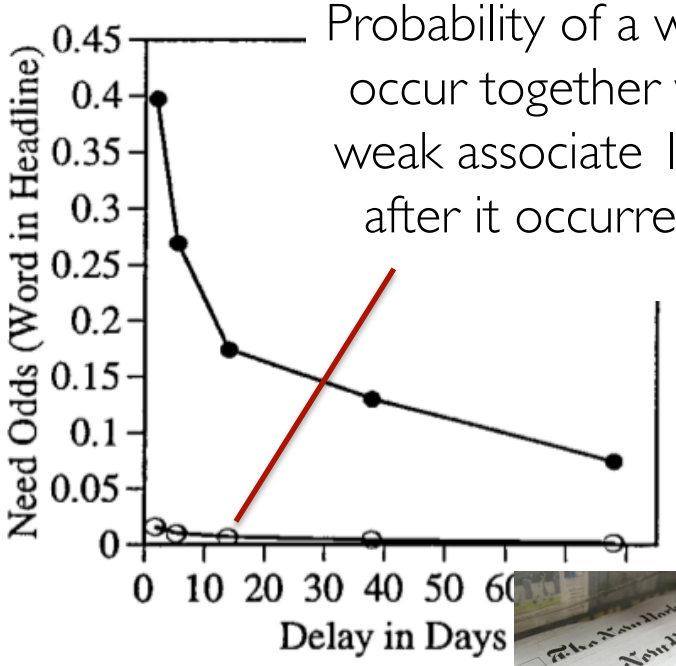
Conclusion:

- The availability of information depends on the context in which it is needed

How do past history and current context work together?

Schooler & Anderson, 1997, estimated recency curves contingent on whether an item co-occurred with a strongly associated item





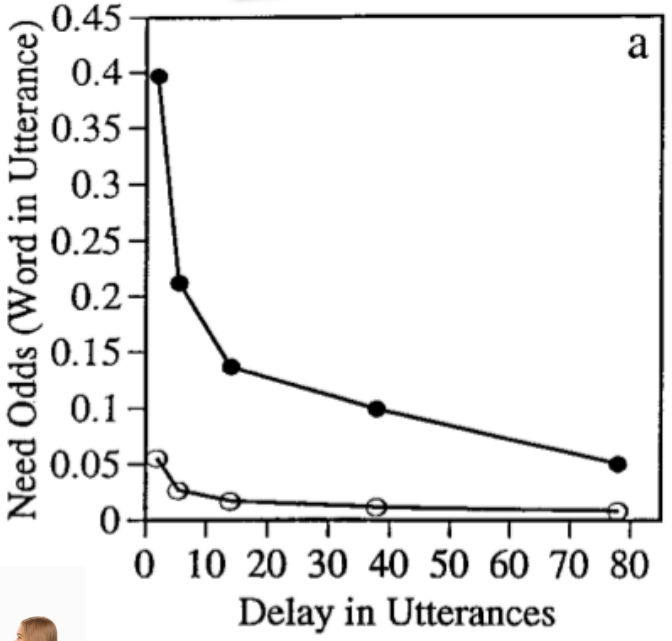
● strong context
○ weak context

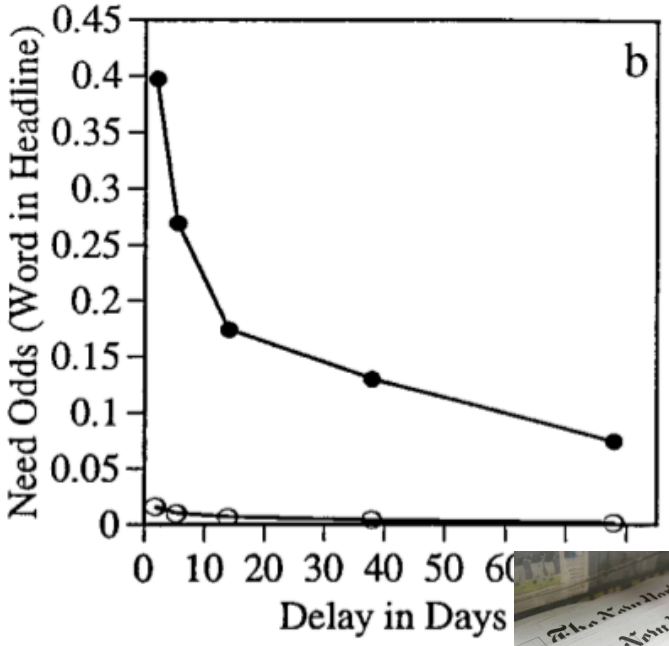
Combining the effects of history and context

Schooler & A… ontingent on whether …

Conclusion:

- The effects of past history and present are additive on a logarithmic scale





● strong context
○ weak context

Conclusion so far

- The probability of a chunk to be needed from memory is based on the combination of its past usage and present context
- If we want to simulate human memory, this ***need probability*** should determine how easy it is to retrieve a chunk from memory

Real-life Analogy...

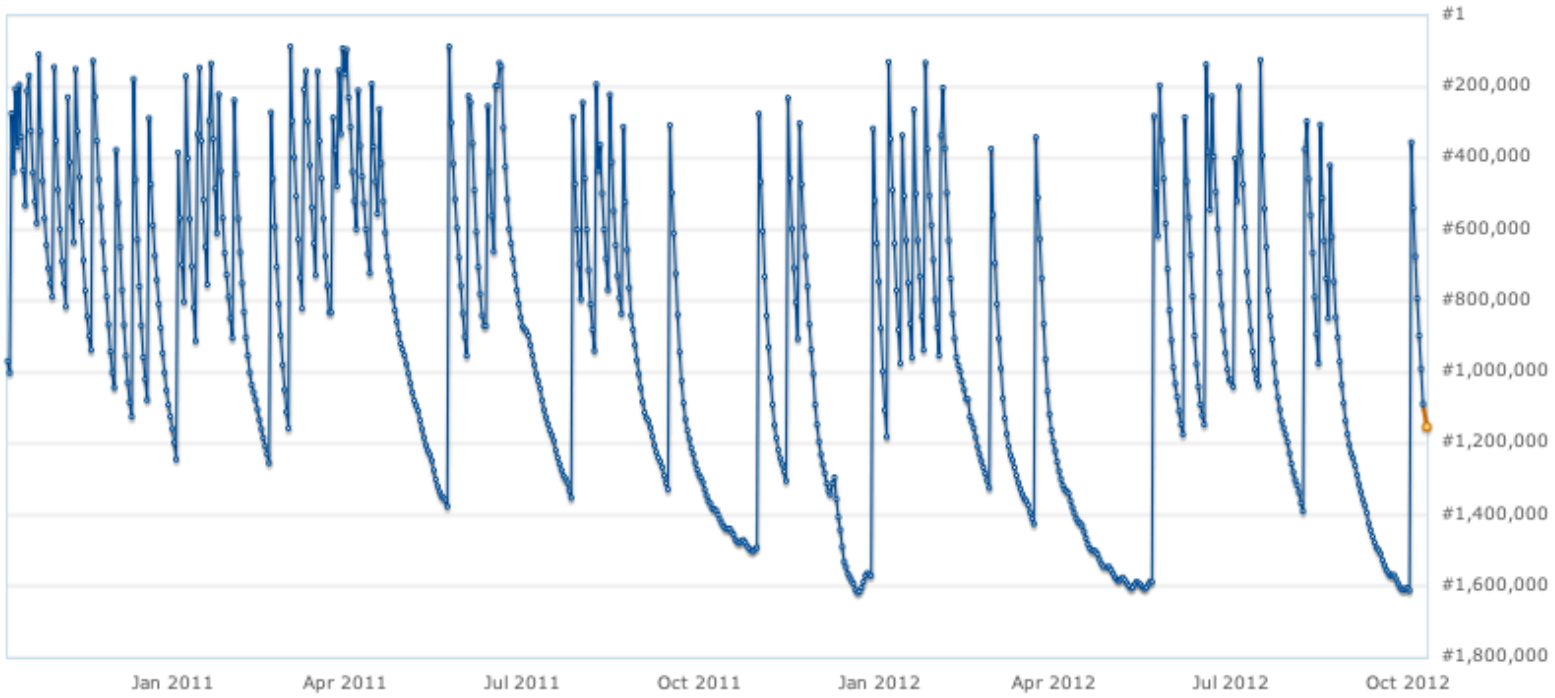
Amazon Bestseller Rank (What's This?)

The Multitasking Mind (Cognitive Models and...

Format: Hardcover, September 2010

As of today: 2 Weeks | 1 Month | 6 Months | 1 Year | 2 Years | **All Available**

This format of "The Multitasking Mind (Cognitive..." is currently ranked **#1,153,039** out of over 8,000,000 books. (updated hourly)



Base-level learning

A single occurrence produces:

$$Odds_j(t) = A(t - t_j)^{-d}$$

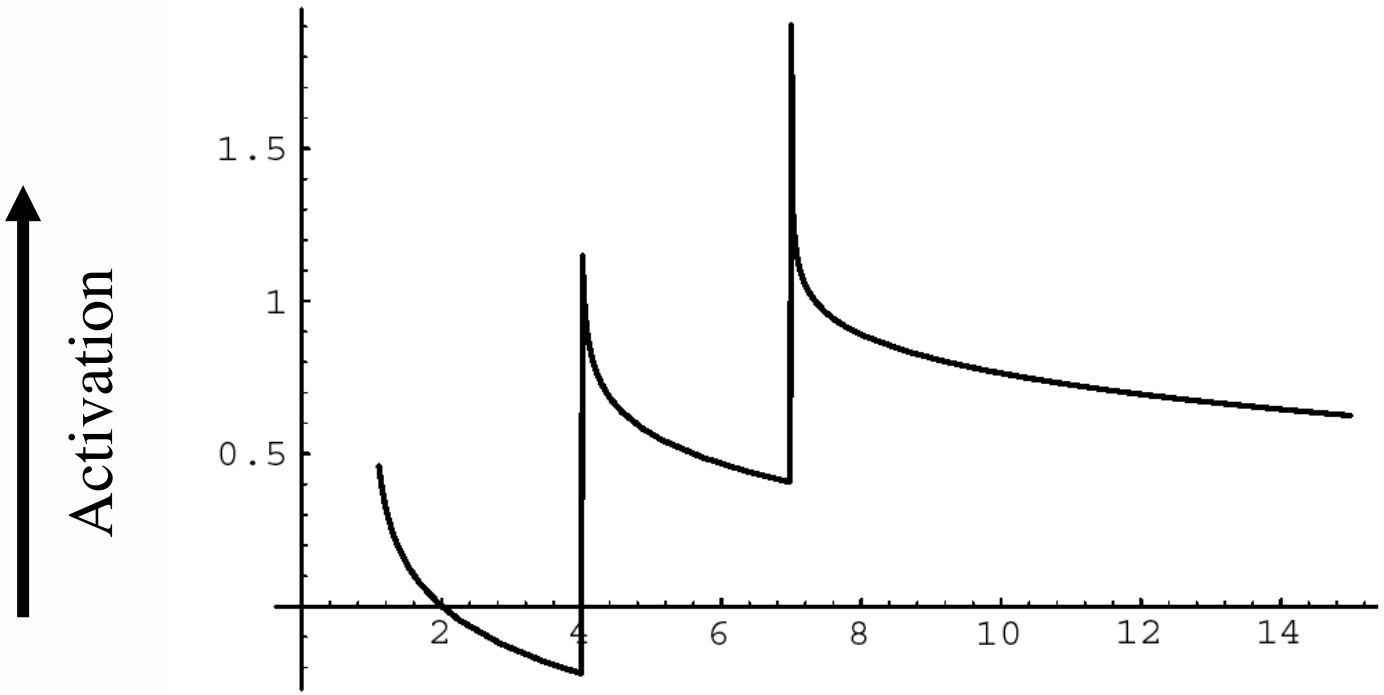
The influence of all occurrences is therefore:

$$Odds(t) = \sum_j A(t - t_j)^{-d}$$

Activation uses log odds, so:

$$B(t) = \log \sum_j (t - t_j)^{-d} + \log A$$

Example of Activation in time



Probability of Recall

Elements in DM can only be retrieved if they are above the retrieval threshold (τ).

On each cycle, logistic noise is added to each element in DM. Because of this, retrieval is probabilistic.

The *s* parameter controls the noise, and is a global ACT-R parameter
The τ parameter is the retrieval threshold, and is also a global parameter

Time to do a retrieval

- Activation determines the time to retrieve a chunk:

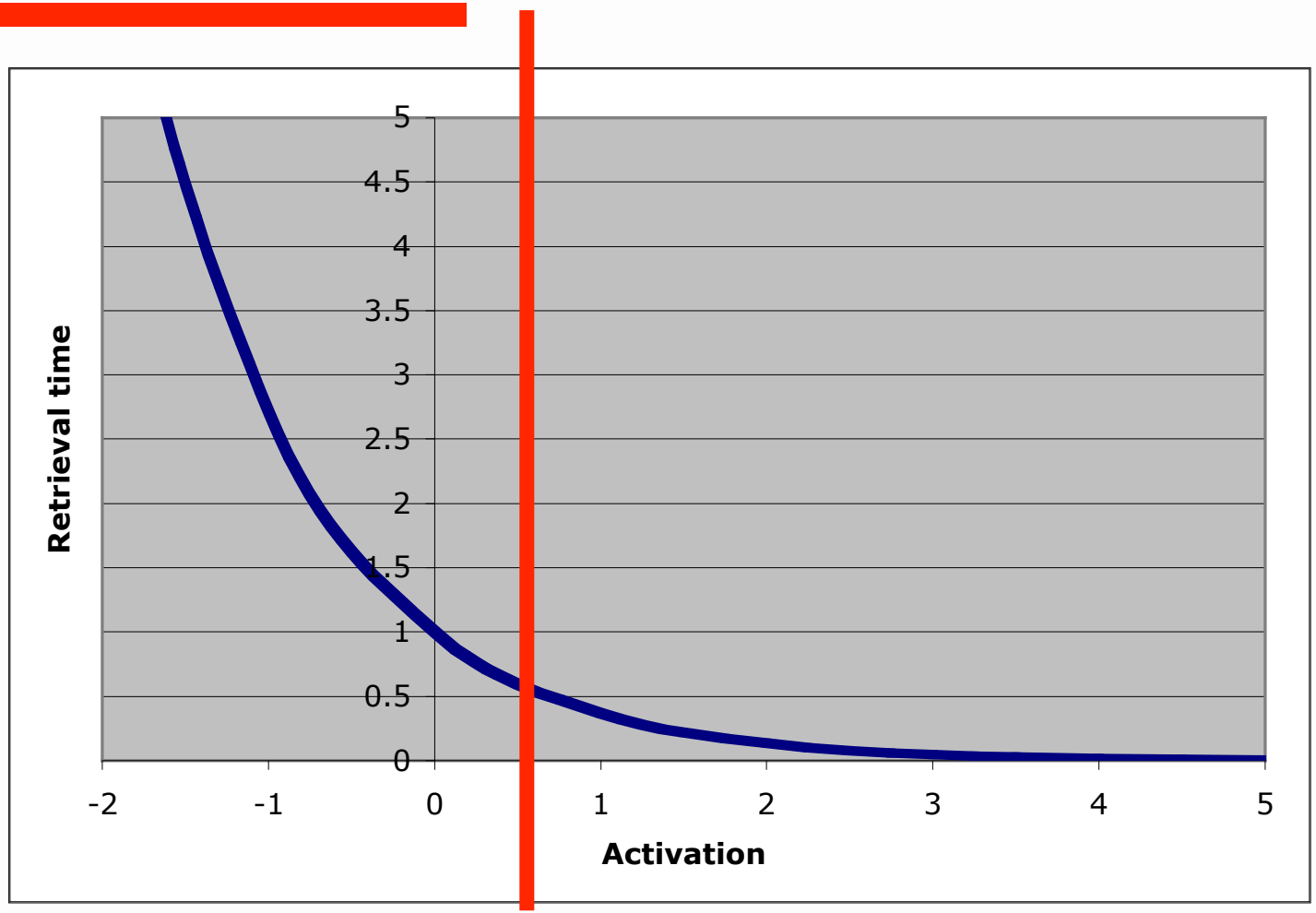
$$time_i = Fe^{-A_i}$$

- If no chunk is retrieved (=retrieval failure), the retrieval threshold determines the time:

$$time = Fe^{-\tau}$$

- F is a global parameter

Graph of retrieval time (F=1)



Instance-based learning

- Simplest example: Alphabet Arithmetic
  - What is J+4?
  - Algorithm/Heuristic is counting
- Learning the English past tense
  - Algorithm/Heuristic is regular rule

Example: Game of Nines

Flavors of IBL

- What is the nature of the instances?
  - State/Actions with evaluations (As in the Gonzalez paper).
  - Only “correct” State/Actions (As in 1-up BlackJack)
- How are instances retrieved?
  - Regular retrieval
  - Partial matching
  - Blending

How to calculate the blend?

- Calculate the activations of all candidate chunks
- Apply mismatch penalties to chunks that do not match completely

- Example: retrieve a blend for long:
  - B(short, 9)=2.0
  - B(short,10)=1.0
  - B(long,16)=2.0
  - B(long,17)=1.5
- Apply penalty (of -2):
  - A(short, 9)=0.0
  - A(short,10)=-1.0
  - A(long,16)=2.0
  - A(long,17)=1.5

How to calculate the blend?

- Now calculate the probability of recall for each of the candidates, using:

$$P_i = \frac{e^{A_i/t}}{\sum_j e^{A_j/t}}$$

- Apply penalty:
  - A(short, 9)=0.0
  - A(short,10)=-1.0
  - A(long,16)=2.0
  - A(long,17)=1.5
- Results in (t=1):
  - p(short, 9)=0.076
  - p(short, 10)=0.0278
  - p(long, 16)=0.558
  - p(long, 17)=0.339

How to calculate the blend?

- Multiply each probability with the slot value, and add it all up to get the blended value

- Probabilities
  - p(short, 9)=0.076
  - p(short, 10)=0.0278
  - p(long, 16)=0.558
  - p(long, 17)=0.339
- 0.076*9 + 0.0278*10 + 0.558* 16 + 0.339 * 17 = **15.65**

Discussion Questions

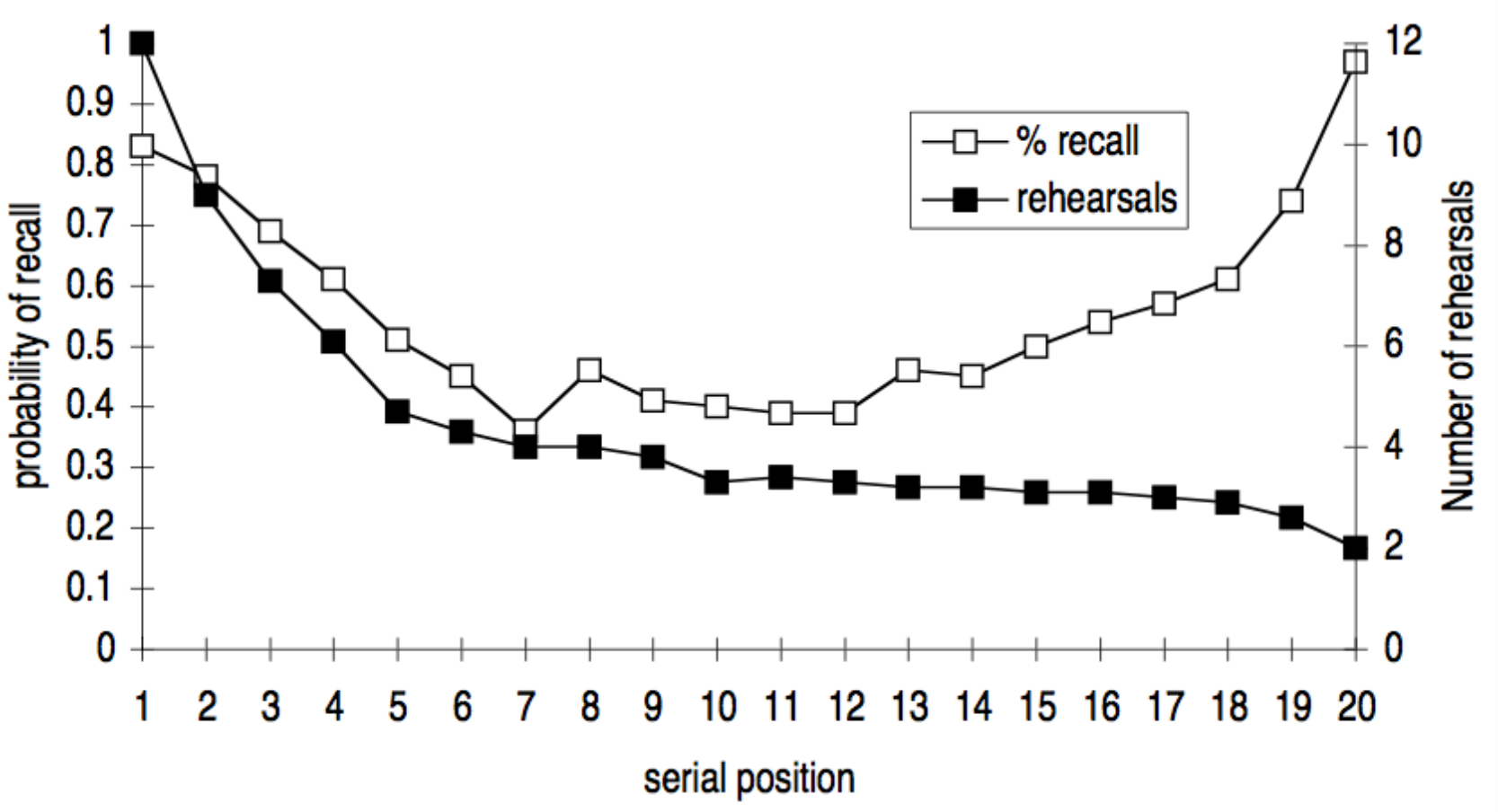
- What happens if the task changes?
- Can the model be improved by modeling time perception explicitly?
- How to deal with delayed feedback?
- Is the dichotomy between instances and heuristics realistic?
- Are people biased towards IBL in favor of “thinking”?

Assignment

- Build a declarative memory in R
- Implement free recall using declarative memory
- Skeleton model is on Nestor
- Next week: model of J&S data

Free recall

- Subjects are given a number of words, more than they can typically remember
- They then have to recall as many of the as possible
- Rundus' experiment: 20 words, 5 seconds between words


probability of recall
Number of rehearsals
% recall
rehearsals
serial position
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
12
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20