**Spacing Effect**

“Spaced study leads to better memory than massed study.”

“Don’t cram before an exam.”

**Study Schedule**
Experimental Paradigm

session 1  session 2  test

time
Experimental Paradigm

session 1  session 2  test

intersession interval (ISI)

retention interval (RI)
Spacing Function

retrieval probability

intersession interval (ISI)
Rich Theoretical Literature Attempts to Explain Spacing Effects
Rich Theoretical Literature Attempts to Explain Spacing Effects

• Encoding variability

• Predictive utility
Rich Theoretical Literature Attempts to Explain Spacing Effects

- Encoding variability
  - Raaijmakkers (2003)

- Predictive utility
  - Staddon, Chelaru, & Higa (2002)
Rich Theoretical Literature Attempts to Explain Spacing Effects

- Encoding variability
- Predictive utility

Encoding Variability
Encoding Variability

A separate trace is laid down for each study episode.

The trace includes not only the item but also a psychological context.

Context wanders gradually over time.
Encoding Variability Explains Forgetting
Encoding Variability Explains Forgetting

Study item at $S$
Encoding Variability Explains Forgetting

Study item at $S$

During retention interval, context wanders
Encoding Variability Explains Forgetting

Study item at $S$

During retention interval, context wanders

Test at $T$

Retrieval success depends on similarity of $c_T$ and $c_S$
Encoding Variability Explains Spacing Effect
Encoding Variability Explains Spacing Effect

Study item at $S_1$
Encoding Variability Explains Spacing Effect

Study item at S1

Study item at S2
Encoding Variability Explains Spacing Effect

Study item at $S_1$

Study item at $S_2$

Test at $T$

Retrieval success at $T$ depends on similarity of $c_T$ to either $c_{S_1}$ or $c_{S_2}$.

Disadvantage for small ISIs: redundancy of $c_{S_1}$ and $c_{S_2}$.
Raaijmakkers (2003)
Raaijmakkers (2003)

Context is represented by pool of binary valued neurons.
Raaijmakers (2003)

Context is represented by pool of binary valued neurons.

Each item to be learned represented by an output neuron.
Raaijmakers (2003)

Context is represented by pool of binary valued neurons. Each item to be learned represented by an output neuron.

Hebbian learning rule
Raaijmakkers (2003)

Context is represented by pool of binary valued neurons.
Each item to be learned represented by an output neuron.

Hebbian learning rule

Output activity at test ~ recall probability

depends on similarity of study and test contexts
Raaijmakkers (2003)

Context is represented by pool of binary valued neurons. Each item to be learned represented by an output neuron.

Hebbian learning rule

Output activity at test $\sim$ recall probability depends on similarity of study and test contexts

Spacing of study $\Rightarrow$ context variability $\Rightarrow$ robust recall
Raaijmakkers (2003): Formal Description

Retrieval at test facilitated when context unit active at both study and test.

Expected output neuron activity ~
$P(\text{retrieval}) \sim P(C_S = 1 \& C_T = 1)$
Raaijmakers (2003): Formal Description

Retrieval at test facilitated when context unit active at both study and test.

Expected output neuron activity ~
\[ P(\text{retrieval}) \sim P(C_S = 1 & C_T = 1) \]

How does context wander over time?

context bits flip from off to on at rate \( \mu_{01} \)
context bits flip from on to off at rate \( \mu_{10} \)

\[ P(C_S = 1 & C_T = 1) = \beta^2 + \beta(1-\beta) \exp(-\alpha \text{ RI}) \]

retention interval

flip rate: \( \mu_{01} + \mu_{10} \)

proportion on: \( \frac{\mu_{01}}{\mu_{01} + \mu_{10}} \)
What It Boils Down To

Forgetting function is exponential
What It Boils Down To

Forgetting function is exponential

Human forgetting functions follow a power law (Wickelgren, 1974; Wixted & Carpenter, 2007):

\[ P(\text{retrieval}) = \lambda (1 + \varphi \text{ RI})^{-\phi} \]
Forgetting function is **exponential**

Human forgetting functions follow a **power law** (Wickelgren, 1974; Wixted & Carpenter, 2007):

\[ P(\text{retrieval}) = \lambda (1 + \varphi \, RI)^{-\phi} \]

**Power law shows scale invariance**

I.e., memory shows same properties at different time scales
Is it a problem that Raaijmakkers’ (2003) model doesn’t show scale invariance?

Yes, spacing effects are scale invariant.

Model has other problems too.

- Many free parameters and ugly hacks
- Doesn’t fit data particularly well

<table>
<thead>
<tr>
<th>Parameters and their meaning</th>
<th>Ranuehart data</th>
<th>Young data</th>
<th>Glenberg data</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$ (Fluctuation parameter)</td>
<td>0.087</td>
<td>0.082</td>
<td>0.013</td>
</tr>
<tr>
<td>$\tau$ (Fluctuation parameter)</td>
<td>0.288</td>
<td>0.150</td>
<td>0.260</td>
</tr>
<tr>
<td>$\sigma$ (Scaling constant for context association)</td>
<td>5.0*</td>
<td>5.0*</td>
<td>5.0*</td>
</tr>
<tr>
<td>$w$ (Probability that a new item enters the STS buffer)</td>
<td>0.766</td>
<td>1.0*</td>
<td>1.0*</td>
</tr>
<tr>
<td>$b$ (Amount of attention information stored on a single study trial)</td>
<td>0.088</td>
<td>0.246</td>
<td>0.732</td>
</tr>
<tr>
<td>$Z$ (Constant representing the interfering effect of other memory traces in sampling)</td>
<td>3.0</td>
<td>2.0</td>
<td>10.0</td>
</tr>
<tr>
<td>$\theta_2$ (Scaling parameter in recovery equation for a test trial)</td>
<td>0.5*</td>
<td>0.300</td>
<td>0.215</td>
</tr>
<tr>
<td>$\lambda$ (Rate of decay from STS)</td>
<td>0.310</td>
<td>0.746</td>
<td>0.800</td>
</tr>
<tr>
<td>$\lambda_{max}$ (Maximum number of retrieval attempts)</td>
<td>3*</td>
<td>3*</td>
<td>3*</td>
</tr>
</tbody>
</table>

*This parameter was not varied but kept fixed in the fitting of the model.
Predictive Utility Theories of Spacing Effects

Suppose that memory
• is limited in capacity, and/or
• is imperfect and allows intrusions.

To achieve optimal performance, memories should be erased if they are not likely to be needed in the future.
Predictive Utility Theories of Spacing Effects

Suppose that memory
- is limited in capacity, and/or
- is imperfect and allows intrusions.

To achieve optimal performance, memories should be erased if they are not likely to be needed in the future.
Rats habituate to a repeated stream of stimuli.

Time for recovery from habituation ~ rate of stimuli

Longer-lasting memory for stimuli delivered at slower rate
Staddon, Chelaru, & Higa (2002)

Each item to be learned represented by memory consisting of **leaky integrators** at multiple time scales.
Staddon, Chelaru, & Higa (2002)

Each item to be learned represented by memory consisting of leaky integrators at multiple time scales.

Memory trace is the sum of the integrator activities.
Staddon, Chelaru, & Higa (2002)

Each item to be learned represented by memory consisting of *leaky integrators* at multiple time scales.

Memory trace is the sum of the integrator activities.

**Memory storage rule**

Integrators with long time constants get activated only when integrators with short time constants have decayed.
Example

10 integrators

Stimulus repeatedly presented at various ISIs

Greater spacing $\Rightarrow$ memory shifts to longer time-scale integrators $\Rightarrow$ more durable memory
Example

10 integrators

Stimulus repeatedly presented at various ISIs

Greater spacing ⇒ memory shifts to longer time-scale integrators ⇒ more durable memory

Model is sensitive to predictive utility

Slower forgetting following longer ISI stimulus sequences.
Limitation of Staddon et al. Model

Model was evaluated only on rat habituation studies, which have many stimulus presentations.

Parameters not sufficiently well specified to model human spacing studies.
Two Models Share Two Key Properties

- Exponential decay of internal representations

![Graph: Retention Interval vs. Retrieval Probability](image1)

- Learning rules for first study are *identical*
Integrating The Two Models

contextual drift (Raaijmakkers)

multiscale representation (Staddon et al.)

multiscale context model (MCM)
In pool $p$, all units flip state at rate $\alpha_p$.

The pools can be different sizes: the relative proportion of units in pool $p$ is $\gamma_p$. 
In pool $p$, all units flip state at rate $\alpha_p$.

The pools can be different sizes: the relative proportion of units in pool $p$ is $\gamma_p$.

Forgetting function is a mixture of exponentials.

$$P(\text{retrieval}) \sim \sum_p \gamma_p \exp(-\alpha_p R_l)$$

Mixture of exponentials are good approximations to human forgetting functions (Wixted).
Forgetting Function

MCM forgetting function is more promising than Raaijmakkers’ forgetting function.

We can constraint MCM parameters, \( \{\alpha_p\} \) and \( \{\gamma_p\} \), to replicate human forgetting functions.
Picking Pool Size ($\gamma$) and Rate ($\alpha$)
Picking Pool Size ($\gamma$) and Rate ($\alpha$)

$$\alpha_p = \mu \nu^p \quad \text{for } p \in [1, N]$$

$$\gamma_p = \omega^p$$

MCM has four free parameters ($\mu$, $\nu$, $\omega$, + one more)
Multiscale Context Model: A Convergence of Theories
# Multiscale Context Model: A Convergence of Theories

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>context drift</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>multiple time-scale</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>representation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>learning rule</td>
<td>X (dependence of</td>
<td>X (cascaded error</td>
</tr>
<tr>
<td></td>
<td>learning on</td>
<td>correction)</td>
</tr>
<tr>
<td></td>
<td>retrieval success)</td>
<td></td>
</tr>
</tbody>
</table>
## Multiscale Context Model: A Convergence of Theories

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>context drift</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>multiple time-scale representation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>learning rule</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>(dependence of learning on retrieval success)</td>
<td></td>
<td>(cascaded error correction)</td>
<td></td>
</tr>
<tr>
<td>variable pool size</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>parameterization of multiscale constants</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>neural characterization</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>constrain parameters via forgetting function</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler (2009), Experiment 1

Forgetting Curve

Percent Recall vs. ISI (days)
Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler (2009), Experiment 1
Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler (2009), Experiment 1
Forgetting Curve

Spacing Curve

RI = 10 days

Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler (2009), Experiment 1
Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler (2009), Experiment 1

Forgetting Curve

Spacing Curve

RI = 10 days

1 day spacing
15 min spacing
MCM Activations Across Time Scales

study sessions spaced 1 day apart

study sessions spaced 15 minutes apart
Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler (2009), Experiment 2

Forgetting Curve

**facts**

**objects**
Forgetting Curve

**facts**

**objects**
Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler (2009), Experiment 2

Forgetting Curve

Spacing Curve

RI = 168 days

Facts

Objects

Percent Recall vs. ISI (days)

Percent Recall vs. ISI (days)
Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler (2009), Experiment 2

**Forgetting Curve**

- **facts**
- **objects**

**Spacing Curve**

- RI = 168 days
Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler (2009), Experiment 2

Forgetting Curve

Spacing Curve

Conditional Spacing Curve

RI = 168 days

facts

objects
Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler (2009), Experiment 2

Forgetting Curve

Spacing Curve

Conditional Spacing Curve

objects

RI = 168 days

facts

RI = 168 days


Forgetting Curve

![Forgetting Curve Graphs](image-url)
Why Are We Proposing Yet Another Model?

Previous models
- have many free parameters, and
- obtain only post hoc fits to data.

Our goal was to develop a truly predictive model.
Predicting Spacing Curve

characterization of student and domain

intersession interval

Multiscale Context Model

predicted recall
Predicting Spacing Curve

Multiscale Context Model

- Characterization of student and domain
- Intersession interval
- Predicted recall
Predicting Spacing Curve

% Recall vs Retention (Days)

- Forgetting after one session
- Intersession interval
- Predicted recall

Multiscale Context Model
Predicting Spacing Curve

Multiscale Context Model

forgetting after one session

intersession interval

predicted recall

Beyond The Spacing Curve

Multiscale study schedule

Context

Model

forgetting after one session

study schedule

predicted recall

lag between session 1 and 2 = 10 days
lag between session 2 and 3 = 20 days
lag between session 3 and test = 50 days

% Recall

Retention (Days)
Beyond The Spacing Curve

predicted recall

Multiscale Context Model

Optimization Routine

forgetting after one session

study schedule

% Recall

Retention (Days)

1 7 14 21 35 70 105

0 20 40 60 80 100
Beyond The Spacing Curve

predicted recall

Multiscale Context Model

forgetting after one session

study schedule

Student

feedback

Online Optimization Routine

optimal study schedule

predicted recall
Simulated student assumptions

- MCM is the correct model of human memory.
- Each item decays at a different rate for each student.
- Scheduler knows only the prior distribution over decay rates.
Scheduler Simulation Details

Comparison of three schedulers

MCM-based

oldest first

worst first
Scheduler Simulation Results

15-25% boost in retention with optimal scheduling
Colorado Optimized Language Tutor (COLT)

Debugged in 2010-2011

Advanced business Spanish course at CU Boulder
Fall 2012

Experiment In Denver area middle school

Collaborator: Jeff Shroyer, member of The Educator Network
Second semester, 8th grade Spanish class, ~ 200 students
Fall 2012

Experiment In Denver area middle school

Collaborator: Jeff Shroyer, member of The Educator Network
Second semester, 8th grade Spanish class, ~ 200 students

Integrating COLT into curriculum

replacing currently used flashcard software
Shroyer restructuring class to do all vocabulary study + testing using COLT
blurs division between practice and quizzes
Fall 2012

Experiment In Denver area middle school

Collaborator: Jeff Shroyer, member of The Educator Network
Second semester, 8th grade Spanish class, ~ 200 students

Integrating COLT into curriculum
replacing currently used flashcard software
Shroyer restructuring class to do all vocabulary study + testing using COLT
blurs division between practice and quizzes

Focus on regular review
current software does not encourage review
Fall 2012

Within-student comparison of three scheduling algorithms

• No review
  (but more time each week for study of new material)

• Generic review for all
  according to teacher’s current educational practice

• MCM selects customized review for each student
Individual Differences

Learning and retention is typically studied using populations of students and items. But individual differences exist and are important.

Distribution of student scores (Japanese-English vocabulary)

Distribution of item scores (Lithuanian-English vocabulary)

Kang, Lindsey, Grimaldi, Pyc, & Rawson (2010)
Individual Differences

Learning and retention is typically studied using populations of students and items. But individual differences exist and are important.

Distribution of student scores (Japanese-English vocabulary)

Distribution of item scores (Lithuanian-English vocabulary)

Challenge: infer a particular student’s state of knowledge for a particular item from very weak feedback.
Item Response Theory

Traditional approach to modeling student and item effects in test taking (e.g., SATs)

\( \delta_i \)  
latent difficulty of item \( i \)

\( \alpha_s \)  
latent ability of student \( s \)

Time invariant theory

\[
\frac{1}{1 + e^{-\delta_i - \alpha_s}}
\]

probability of correct response

student ability – item difficulty
Extending Item-Response Theory To Consider Time

• time to test

• number of study sessions

• spacing of sessions
Extending Item-Response Theory To Consider Time

• time to test
• number of study sessions
• spacing of sessions

incorporate model of memory and forgetting

See Lindsey paper

Considering study schedule yields significant improvement in ability to predict memory state of individual student for specific item.

Bayesian approach leverages data from population of students and items to make predictions about individuals.
Existing Flashcard Software

Many web sites, iPhone apps, etc.

- studyblue.com
- chinglish-online.com
- spaced-ed.com
- smart.fm
- totalrecalllearning.com
- flashcardexchange.com
- supermemo.org
- mnemosyne-proj.org
- anki
Existing Flashcard Software

Many web sites, iPhone apps, etc.

studyblue.com
chinglish-online.com
spaced-ed.com
smart.fm
totalrecalllearning.com

flashcardexchange.com
supermemo.org
mnemosyne-proj.org
anki

All incorporate spacing based on some variant of heuristic system developed by Leitner (1972)

New flashcards start in bin 1
Cards tested correctly promoted to next bin.
Higher bins: longer lag before next review
Cards tested incorrectly demoted.
Existing Flashcard Software

Many web sites, iPhone apps, etc.

- studyblue.com
- chinglish-online.com
- spaced-ed.com
- smart.fm
- totalrecalllearning.com
- flashcardexchange.com
- supermemo.org
- mnemosyne-proj.org
- anki

All incorporate spacing based on some variant of heuristic system developed by Leitner (1972)

- New flashcards start in bin 1
- Cards tested correctly promoted to next bin.
- Higher bins: longer lag before next review
- Cards tested incorrectly demoted.

Goal: study card at the point of *desirable difficulty* (Bjork, 1994), i.e., when the individual is on the verge of forgetting.
Problems With Current Tools

• The point of desirable difficulty depends on specific individual, item, and study history.

  Leitner box is ‘one size fits all’.

  MCM can be tailored to the student, material, and even individual items.

• Optimal spacing depends on window of time over which material needs to be accessible.

  Therefore, can’t prescribe a study schedule without specifying the window.

  MCM can optimize over any window.

Our conjecture

Precise quantitative predictions of forgetting for specific individual and item are needed to obtain the most benefit from a scheduler.
Why Hasn’t Cognitive Science Had A Greater Impact on Education?
Why Hasn’t Cognitive Science Had A Greater Impact on Education?

Most guidance is *qualitative*

“Space your study”

“The harder you work to learn, the better you’ll retain”

“Relate new material to be learned to learner’s existing knowledge”
Why Hasn’t Cognitive Science Had A Greater Impact on Education?

Most guidance is *qualitative*

“Space your study”

“The harder you work to learn, the better you’ll retain”

“Relate new material to be learned to learner’s existing knowledge”

*Quantitative* modeling and prediction can provide specific, customized, detailed guidance.

• particularly useful if there’s significant variability across individuals and materials
To Be Continued...