Forgetting Rate and the Optimization of Retrieval Practice

Matthew J. Hays, Kerry Young, & Robert A. Bjork
University of California, Los Angeles

QUESTION

How should you study a set of to-be-remembered items?

IMAGINE THIS SCENE

You are at a conference.
You've met several new people—most of whom you want to talk to again.
It's awkward to keep looking at the name tags of people you've already met, but you want to retain the pairings of their names and faces.

FRAMEWORK: THE NEW THEORY OF DISUSE

A memory trace (e.g., a name-face pairing) has two indices of strength.

RETRIEVAL STRENGTH (RS)	STORAGE STRENGTH (SS)
An item's accessibility	*An item's overall degree of learning*
Observed as probability of recall	Inferred from resistance to forgetting

EXAMPLES OF ITEMS

ITEM IN MEMORY	Current RS	Current SS
Your hotel room number	High	Low
Easy to recall now, but rapidly forgotten		
Your phone number	High	High
Easy to recall now, and slowly forgotten		
Last year's hotel room number	Low	Low
Difficult to recall now, but rapidly forgotten		
Your high school locker combo	Low	High
Difficult to recall now, but, if I told it to you, it would be slowly forgotten		

INTERACTIONS

High SS slows the loss of RS (as in the case of your locker combination).
High RS reduces the increment in SS yielded by a learning event (e.g., cramming works, but not for long).
Together, these suggest that long-term learning is best achieved by maximizing the interval between learning events (Bjork & Bjork, 1992).
In other words,

"FORGETTING" POTENTIATES SUBSEQUENT LEARNING

WHAT TYPE OF LEARNING EVENT TO USE?

Tests, rather than re-presentations. Why?
1. It's not feasible to ask people to introduce themselves repeatedly.
2. More importantly, successful retrievals (vs. re-presentations) yield superior long-term recall (e.g., Whitten & Bjork, 1977).

This is true *only if retrieval is successful*. When using *re-presentations*, the longer the lag, the better—because maximizing lag maximizes forgetting. However, when using *tests* (without feedback), longer lags are better, but the item must be tested before RS decreases to the point that the item cannot be retrieved; a failed retrieval isn't a learning event at all.

HOW TO SCHEDULE THE TESTS?

```
EXPANDING INTERVAL RETRIEVAL PRACTICE (1-5-9):  P-T-----T---------T
  UNIFORM INTERVAL RETRIEVAL PRACTICE (5-5-5):  P-----T-----T-----T
```

REFERENCES

Berger, S. A., Hall, L. K., & Bahrick, H. P. (1999). Stabilizing access to marginal and submarginal knowledge. *Journal of Experimental Psychology: Applied, 5(4)*, 438-447.
Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35-67). Hillsdale NJ: Erlbaum.
Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London: Academic Press.
Roediger, H. L., III, & Karpicke, J. D. (2005). Expanding retrieval does not improve long-term retention. *Poster presented at the 46th Annual Meeting of the Psychonomic Society*. Toronto, Canada.
Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: The effects of spacing. *Journal of Verbal Learning and Verbal Behavior, 16*, 465-478.

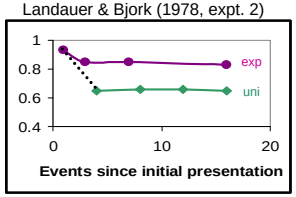
MANY THANKS TO

Nate Kornell (for input and jokes), Dan Fink (for programming and input), Elizabeth Bjork, Matt Makel, Team CogFog, and Team MiniFog

PRIOR STUDIES

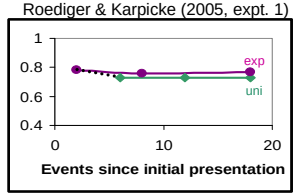
Proportion of items recalled, both during learning and at delayed final tests:

Landauer & Bjork (1978, expt. 2)



FINAL TEST:
30 min: Expanding (.66) > Uniform (.56)

Roediger & Karpicke (2005, expt. 1)

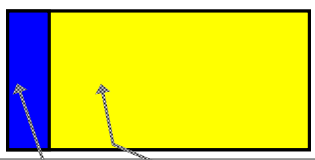


FINAL TESTS:
10 min: Expanding (.71) > Uniform (.62)
48 hours: Expanding (.33) < Uniform (.45)

HYPOTHETICAL SETS OF TO-BE-REMEMBERED ITEMS

DIFFICULT SET
(mostly difficult items)
(e.g., name-face pairings;
e.g., Landauer & Bjork, 1978)

EASY SET
(mostly easy items)
(e.g., GRE words and definitions;
e.g., Roediger & Karpicke, 2005)

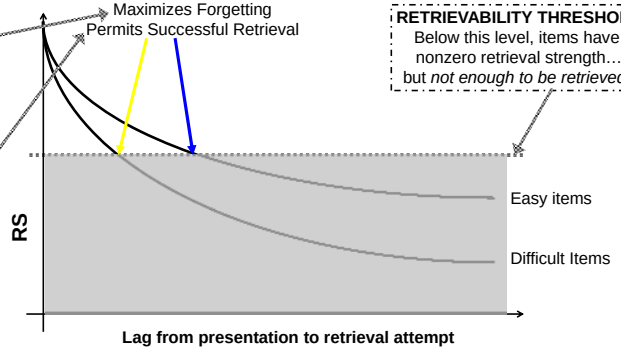


Difficult items (no prior SS): Quickly forgotten
Easy items (some prior SS): Slowly forgotten

RETRIEVAL STRENGTH OF HYPOTHETICAL SETS' ITEMS

Optimal lag to test:
Maximizes Forgetting
Permits Successful Retrieval

RETRIEVABILITY THRESHOLD
Below this level, items have nonzero retrieval strength…
but *not enough to be retrieved*.



Lag from presentation to retrieval attempt

MATERIALS IN THE PRESENT STUDY

Real and fictional trivia questions and answers from Berger, Hall, and Bahrick (1999).

FICTIONAL ITEMS…no prior SS	**REAL ITEMS**…some prior SS
Which comic book character constantly refers to himself as "The Mighty Green One"?	Which *Peanuts* character constantly refers to Peppermint Patty as "Sir"?
Swampman	*Marcie*

(Of course, no participant encountered the real and fictional version of the same question.)

PROCEDURE

After the initial presentation*, an item was tested on one of these schedules:
1. **Repeated** (4) tests at an expanding or uniform interval (same total spacing)

```
EXPANDING(0):  PT-----T--------------T-----------------------------------T
EXPANDING(4):  P----T--------T---------------T-----------------------------T
UNIFORM (15):  P--------------T--------------T--------------T--------------T
```

2. A **single** test with a lag (from presentation* to test) of 0, 4, or 15
3. **Nothing** (presentation* only)

All studied items were then tested at either a 10-minute or 48-hour delay.

*Each item was initially tested and then the correct answer was presented. An item for which a participant could generate the correct answer *before being presented with the answer* was considered "unforgettable" *for that participant*. That is, no matter what the delay (condition), that participant would have been able to correctly answer that question. We omitted unforgettable items from our analyses.

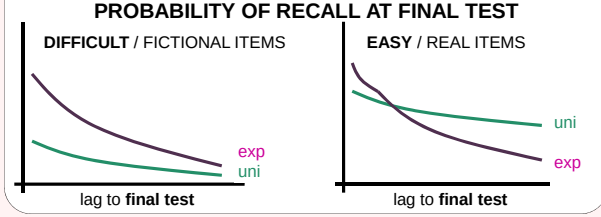
HYPOTHESIS

In many prior experiments in which the learning phase contains multiple tests, performance is flat after the first test; an item that is successfully retrieved from long-term memory will continue to be retrieved throughout training. It may be that little RS is lost after the first test—and little forgetting begets little subsequent learning.

Because of the short initial interval, expanding (vs. uniform) schedules allow retrieval of a greater number of both real and fictional items during learning, yielding superior performance.

Difficult items will be "saved" by the short intervals in the expanding condition (the dotted lines in the PRIOR STUDIES section). However, the items that did not need saving will suffer because they were learned under conditions of inadequate spacing. That is, with little forgetting *after* the first retrieval, forgetting *before* the first retrieval will determine learning.

PROBABILITY OF RECALL AT FINAL TEST



RESULTS AND DISCUSSION

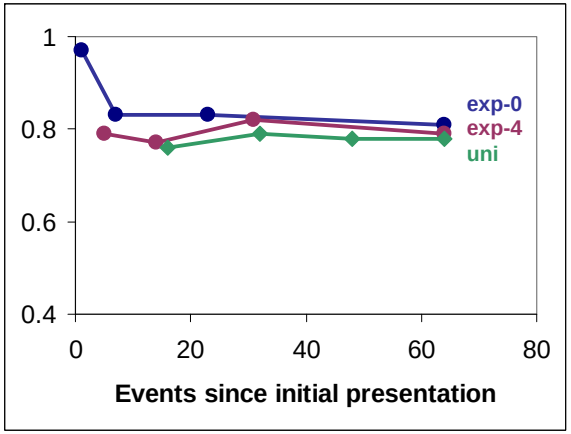
Proportion of items recalled, both during learning and at delayed final tests:

PLEASE SEE BACK OF PAPER

ANSWER

The larger the proportion of items saved by a short initial interval, the more beneficial will be expanding-interval retrieval practice. Optimal retention is achieved by attempting to retrieve *just* before you forget.

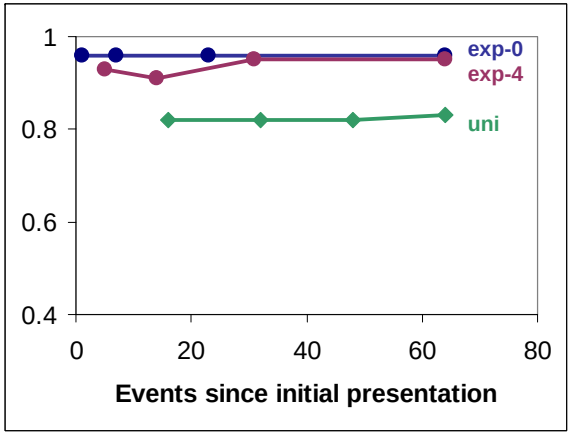
FICTIONAL ITEMS



10min: exp-0 (.78) **=** exp-4 (.75) **>** uni (.59)
48h: exp-0 (.42) **<** exp-4 (.63) **=** uni (.52)

Learning: Similar performance across conditions.
10min: Expanding intervals were superior (*p*=.08).
48h: Expanding intervals were somewhat superior (*p*=.17)...if the 1st retrieval was not from STM.
Additional tests during learning *did improve* recall when the 1st retrieval was not from STM.

REAL ITEMS



10min: exp-0 (.82) **=** exp-4 (.88) **=** uni (.78)
48h: exp-0 (.72) **=** exp-4 (.70) **=** uni (.68)

Learning: Expanding intervals produced significantly higher levels of recall (*p*=.02).
10min: Expanding intervals were superior, but not statistically significantly (*p*=.2).
48h: Similar performance across conditions.
During learning, additional tests *did not improve* recall when the 1st retrieval was not from STM.