

Your Title

Your Name

Your Affiliation

Author Note

Abstract

Keywords:

Your Title

Derivation

We are interested in deriving the simplest model of learning which generates a spacing effect. For the current work, we consider the simple case of two repeated study sessions for the same material, followed by a delayed memory test. Set y as the study separation time between study 1 and study 2 and x as the test delay. The spacing effect can be described simply: the optimal lag between study sessions increases as the delay period of the test increases (Cepeda et al. 2008).

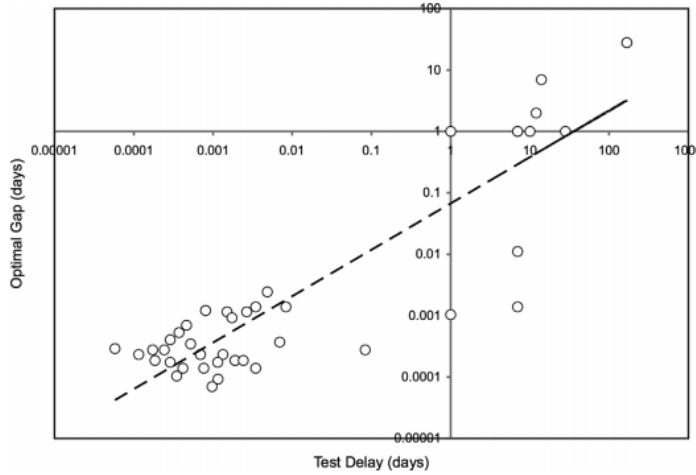


Figure 1. Results from a meta-analysis by Cepeda et al. (2008) demonstrating that the optimal study lag is approximately a linear function of the test delay on a log-log plot

Consider a generic memory model with the following form:

(1) Delta learning rule: When studying an item, a new trace is formed of the memory strength:

$$B_{new} = \delta(B_{max} - B_{current}), \quad (1)$$

where $B_{max} = 1$, δ is the learning rate ($0 < \delta < 1$), and $B_{current}$ is the summed strength of memory traces formed during previous repetitions of the item. Thus, after the very first occurrence of an item, its memory trace has strength $B_1 = \delta$.

(2) Trace-specific forgetting: The strength of each memory trace decreases independently as a monotonic function of time. Two standard such functions are an exponential decay and a power-based decay:

$$\begin{aligned} f_{power}(t) &= (1+t)^{-\gamma} \\ f_{exp}(t) &= e^{-\gamma t}, \end{aligned} \tag{2}$$

where $f(0) = 1$ and $\gamma > 0$ and each memory trace decays independently as a function of time since its formation.

At any point in time, the summed memory strength is the sum of each trace weighted by its decay function:

$$B_{current} = \sum_{k=1}^n B_k f(t_k), \tag{3}$$

with k being the k -th study repetition and t_k is the time elapsed since.

In the special case with two study repetitions separated by an interval y and where memory is tested x units of time after the second session, we have the overall memory strength B at test is a sum of the two memory traces after forgetting:

$$\begin{aligned} B(x, y) &= B_1 f(x+y) + B_2 f(x) \\ &= B_1 f(x+y) + \delta(1 - B_1 f(y))f(x) \\ &= \delta f(x+y) + \delta f(x) - \delta^2 f(x)f(y) \end{aligned} \tag{4}$$

To capture the spacing effect, we are interested in deriving the optimal study separation y that maximizes the memory strength B when the test delay x is fixed. An optimal y that is a monotonically increasing function of x is possible only when the $B(x, y)$ is a convex function of y that has a real solution $y > 0$ to the system of equations involving its first two derivatives:

$$\frac{dB}{dy} = 0, \quad \frac{d^2B}{dy^2} < 0 \tag{5}$$

We get the general form of the derivatives from (4):

$$\begin{aligned} \frac{dB}{dy} &= \delta f'(x+y) - \delta^2 f(x)f'(y) \\ \frac{d^2B}{dy^2} &= \delta f''(x+y) - \delta^2 f(x)f''(y) \end{aligned} \tag{6}$$

Since $0 < \delta < 1$, we can simplify them and have this final system of equations which a model with a spacing effect must satisfy:

$$\begin{aligned} f'(x + y) - \delta f(x)f'(y) &= 0 \\ f''(x + y) - \delta f(x)f''(y) &< 0 \end{aligned} \tag{7}$$

Exponential vs power-based decay

Consider first the exponential-based decay with $f(x) = e^{-\gamma x}$:

$$\begin{aligned} \delta^{-1} \frac{dB}{dy} &= \frac{d}{dy} e^{-\gamma(x+y)} - \delta e^{-\gamma x} \frac{d}{dy} e^{-\gamma y} \\ &= -\gamma e^{-\gamma(x+y)} + \delta \gamma e^{-\gamma(x+y)} \\ &= (\delta - 1) \gamma e^{-\gamma(x+y)} \\ \delta^{-1} \frac{d^2B}{dy^2} &= -(\delta - 1) \gamma^2 e^{-\gamma(x+y)} \end{aligned} \tag{8}$$

Since the learning rate δ is strictly less than 1, then $(\delta - 1)\gamma e^{-\gamma(x+y)} < 0$ and $-(\delta - 1)\gamma^2 e^{-\gamma(x+y)} > 0$. The negative first derivative means that memory strength B will always be highest when y is minimal and there is no spacing between the study items.

Now consider the power-based decay with $f(x) = (1 + x)^{-\gamma}$

$$\delta^{-1} \frac{dB}{dy} = -\gamma(x + y + 1)^{-\gamma-1} + \delta \gamma(x + 1)^{-\gamma}(y + 1)^{-\gamma-1} \tag{9}$$

Setting $\frac{dB}{dy} = 0$, we have:

$$(x + y + 1)^{-\gamma-1} = \delta(x + 1)^{-\gamma}(y + 1)^{-\gamma-1} \tag{10}$$

Both sides are positive, so set $r = -\frac{1}{\gamma+1}$, $-\gamma r = 1 + r$, and raise each side to r degree:

$$\begin{aligned} ((x + y + 1)^{-\gamma-1})^r &= (\delta(x + 1)^{-\gamma}(y + 1)^{-\gamma-1})^r \\ x + (y + 1) &= \delta^r(x + 1)^{1+r}(y + 1) \\ x &= (y + 1)(\delta^r(x + 1)^{1+r} - 1) \\ y + 1 &= \frac{x}{\delta^r(x + 1)^{r+1} - 1} \end{aligned} \tag{11}$$

To simplify the expression, denote $\tilde{y} = y + 1$ and $\tilde{x} = x + 1$, the study and test lag offset by 1. Recognize also that $\delta^r = \delta^{-\frac{1}{\gamma+1}} \geq 1$ is a scaling constant that depends on the learning rate and the decay rate, so just replace it with a combined constant $\alpha > 1$. :

$$\tilde{y} = \frac{\tilde{x} - 1}{\alpha \tilde{x}^{\frac{\gamma}{1+\gamma}} - 1} \quad (12)$$

Since the exponent of \tilde{x} in the denominator is $\gamma/(1 + \gamma) < 1$, as the test lag x increases, so does the optimal study lag y .

Figures below show the optimal offset study gap \tilde{y} in days as a function of offset test delay in days (both measures offset by 1 second as described above). First figure is on the native linear scale, while second figure is on a log10-log10 scale. Both parameters affect the position of the optimal line, but do not change the qualitative results. Increasing the learning during both study sessions makes the optimal study gap longer for the same test delay, although the effect is modest. The forgetting rate has more dramatic effects, but again it just shifts the intercept on the log-log scale. Neither parameter changes the shape of the optimal y as a function of x .

With the exception of $\gamma = 0.01$, all optimal lines are below the diagonal line, meaning that the optimal study gap is less than the test day for most parameter combinations. We included the 0.01 decay rate condition to explore the extremes of the model - that said, it is not a reasonable parameter value and it is far from the range typically used in power-based decay forgetting functions (REFS). All parameter values reproduce the general data pattern from Cepeda (2008). Particular parameter values can be chosen for a perfect fit to the empirical regression line, if desired.

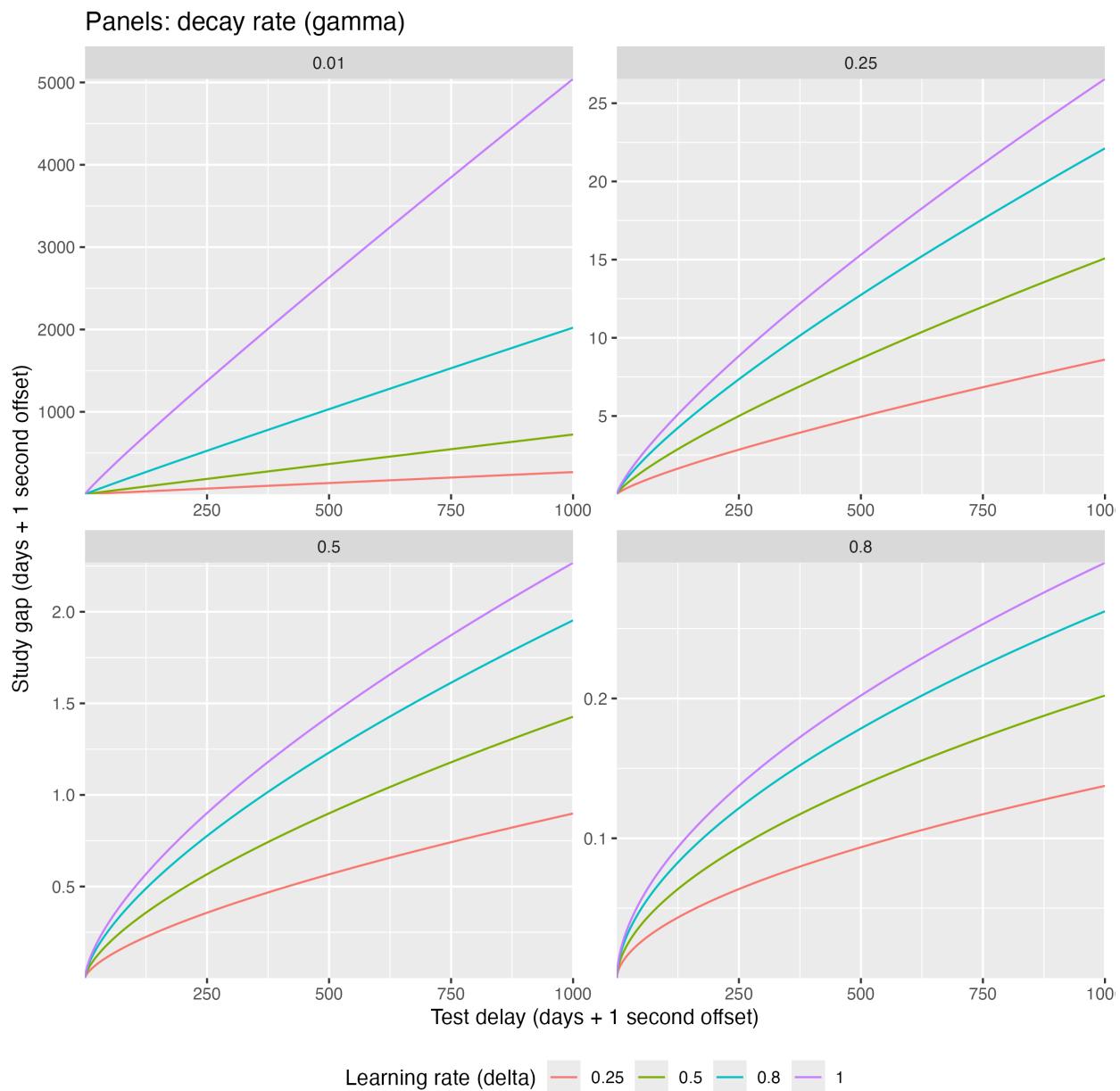


Figure 2. Enter Caption

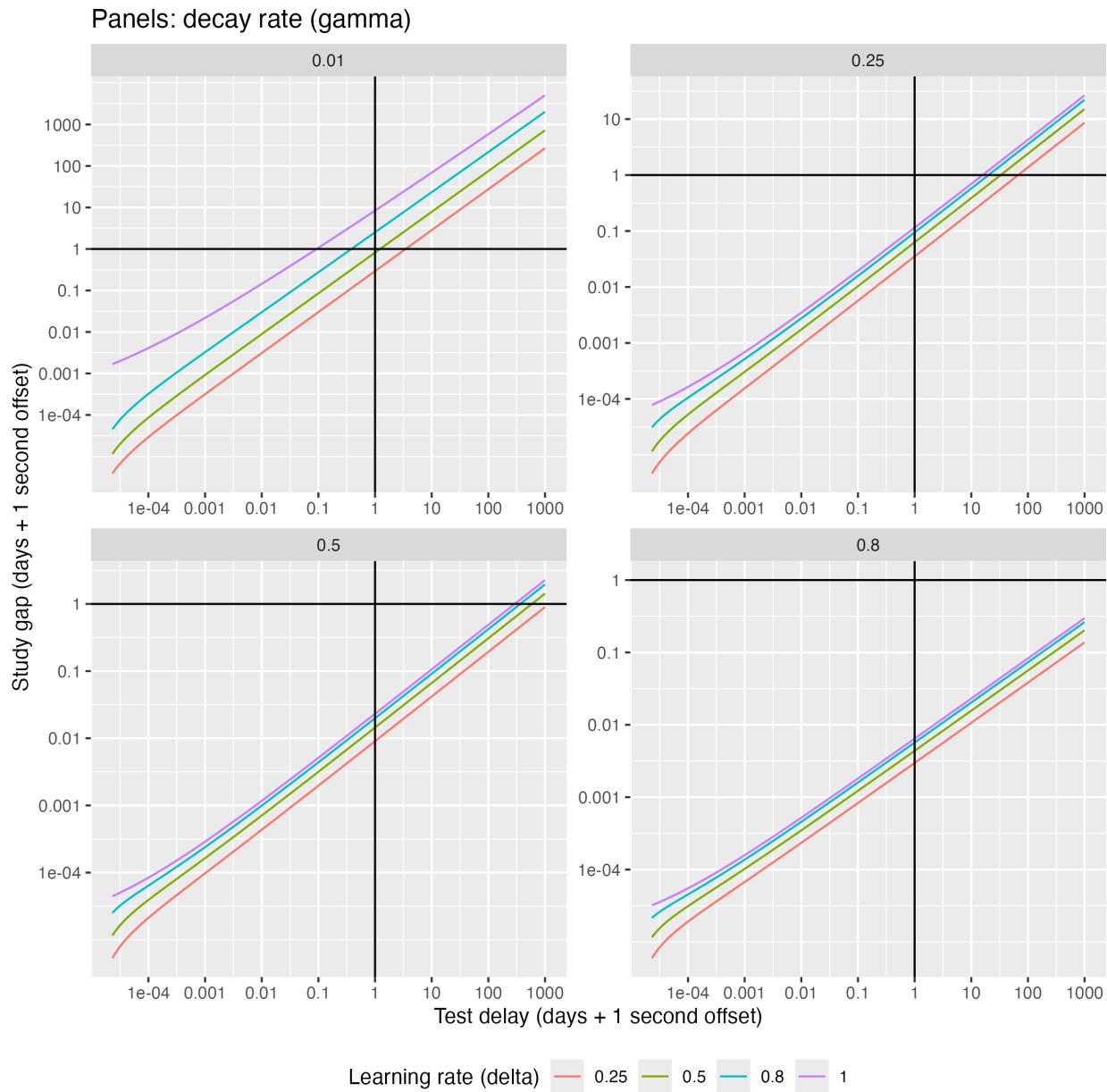


Figure 3. Enter Caption