

Note on the power law of forgetting

Michael J. Kahana
Center for Complex Systems, and
Program in Neuroscience
Brandeis University
E-mail: kahana@brandeis.edu

and

Mark Adler
Department of Mathematics
Brandeis University
E-mail: adler@brandeis.edu

Version: March 12, 2002

Power functions (e.g., $f(t) = at^{-b}$) describe the relationships among many variables observed in nature. One example of this is the power law of forgetting: The decline in memory performance with time or intervening events is well fit by a power function. This simple functional relationship accounts for a great deal of accumulated data. In this note, we consider a simple yet general memory model in which all items decay monotonically in strength, but at different rates. To translate between continuous changes in strength and actual memory for events we assume a simple strength threshold for remembering. We prove a limit theorem for this model: as time grows large and memories decay, the empirical forgetting function approaches a power function under very general conditions. Power forgetting emerges for almost any monotonically decreasing strength function (including linear and exponential cases). We also illustrate by way of simulations that the power function provides an excellent fit to the entire time-course of the forgetting function, not just its limiting behavior.

Key Words: Power law, forgetting, memory models

1. INTRODUCTION

Ever since Ebbinghaus (1885/1913) inaugurated the scientific study of memory, researchers have examined the manner in which memory performance declines with time or intervening events (i.e., the forgetting function). Although it has long been known that forgetting occurs rapidly at first and more slowly as time goes on, it was not until quite recently that the mathematical form of the forgetting function has been precisely established. Wixted and colleagues (Wixted, 1990; Wixted & Ebbesen, 1991, 1997) have demonstrated that the form of forgetting, across various materials and memory tests, is characterized mathematically by a power function. Rubin & Wenzel (1996) compared over a hundred forgetting functions and found that the power function was one of only four that provided a good fit to a wide range of forgetting data. That is, accuracy in a memory task at time t is given by $y = at^{-b}$, where a and b are positive real numbers. This invariance in the form of forgetting suggests a basic law of human memory.

If the nature of the functions that characterize learning and forgetting can teach us about their underlying processes, we ought to take any evidence either for or against a particular function very seriously. For instance, if a model assumes that memory trace strength decreases in an exponential fashion, a seemingly obvious implication is that such a model should be rejected. How could exponentially decreasing trace strength be reconciled with power-law forgetting?

Exponential strength decay is a natural assumption for memory models. If you consider a model like TODAM (Murdock, 1982, 1997), forgetting arises due to a forgetting parameter, α , and also because the variance of the memory increases with list length (but see Murdock & Kahana, 1993,b). In TODAM, the storage equation for item information (Murdock & Lamon, 1988) is given by:

$$\mathbf{m}(t) = \alpha \mathbf{m}(t-1) + p \mathbf{f}(t)$$

Because the memory is premultiplied by α as each new item is learned, recognition performance (as measured by d') should decrease as a function of α to the power of lag (or time).

Exponential decay of trace strength or cue-target strength is common in other models as well. Mensink & Raaijmakers (1988) have a forgetting process based on contextual drift that is *nearly* exponential. The same is true of other models of contextual drift (Murdock, 1997; Howard & Kahana, 2002). Bower's (1967) multicomponent model also assumes exponential strength-decay of the individual components.

Although exponential decay in trace strength is a common assumption of memory models, the data strongly suggest that forgetting obeys a power law. Such power laws are not unique to forgetting. Learning is also well described by a power law. The reduction in reaction times that comes with practice is a power function of the number of training trials (see Anderson, 1995, for a review). Indeed, power laws describe a great many natural phenomena ranging from sensory scaling to the distribution of city sizes. What can we learn from the power-law of forgetting? Does a power-law of forgetting imply that models based on exponential strength decay mechanisms must be rejected?

1.1. Interpretive Problems

Using a computational analysis, Anderson and Tweney 1997 showed that arithmetic averaging of exponential functions can give rise to power functions. They also reported that they were unable to find a general analytic solution to the problem of aggregated exponentials. Indeed, this problem may not have a solution without adding some simplifying assumptions. Anderson & Tweney (1997) concluded that the power law of forgetting may be an artifact of arithmetic averaging (see also Anderson & Tweney, 1998)

Responding to the Anderson & Tweney (1997) critique of power functions, Wixted & Ebbesen (1997) showed that even with geometric averaging of individual subjects' data, a power function still fits the data better than an exponential function. Of course, it is possible that the problem is not at the level of subjects, but rather at the level of items, or even subject-item interactions. Wixted & Ebbesen (1997) acknowledge the possibility that exponential forgetting with variability in forgetting rates across items could give rise to aggregate power functions of forgetting.

Wickens (1998) pursued this point, showing that very different theories of forgetting can give rise to very similar-looking retention functions. In particular, starting with the assumption that individual items are forgotten exponentially, Wickens (1998) showed that models based on (1) heterogeneity of forgetting rates, (2) consolidation of traces over time, and (3) competition among traces for survival, under certain distributional assumptions, can give rise to a Pareto II distribution of item survival. For large t , this description of forgetting simplifies to the power law that has been shown to well describe the empirical data.

Although power-functions provide a good fit to forgetting data for ranges of the forgetting function, there is no sense in which a complete forgetting function, with performance measured from $t = 0$ to $t = \infty$, could be well fit by a power function. This is because a power function implies that the measure of performance tends to infinity as t tends to zero. With percent correct as the dependent measure, performance at $t = 0$ will tend to 1.0, thus “rejecting” a power function.

In the work presented here, we concentrate on the asymptotic properties of forgetting — the behavior of the forgetting function as $t \rightarrow \infty$. In particular we obtain analytic results that show that as $t \rightarrow \infty$, performance is a power function of t for a simple but general class of forgetting models. Simulations support these analytic results, and show that the emergence of power forgetting appears quickly for many different parameter values of the underlying models.

1.2. Power forgetting in a strength-decay to threshold memory model

Memory researchers are interested in knowing the form of changes in memory strength over time — not changes in memory performance over time. But, we can’t measure memory strength directly; rather, we measure performance. In most memory tasks, performance is based on a summary statistic over discrete events (success or failure in recognizing or recalling an item at a given delay between study and test).

Consider the following very simple, yet general, memory model. Items stored in memory have a strength value, S . In the absence of reinforcement but the presence of intervening activity, the strength of each item decays monotonically according to some strength decay function.

Assuming that the strength decays exponentially: $S_i(t) = \alpha S_i(t-1)$ and $0 < \alpha < 1$, therefore,

$$S_i(t) = S_i(t_0)\alpha^t \quad (1)$$

This type of difference equation is characteristic of a number of memory models (e.g., Murdock, 1982, 1997; Mensink & Raaijmakers, 1988). But the form of the forgetting function turns out not to be crucial. The important consideration is that when an item’s strength falls below a fixed threshold, k , the item is forgotten. So long as the strength is greater than k , the item is remembered. We also must assume that strength decays at different rates for different items (cf. Anderson & Tweney, 1997; Wickens, 1998). The basic features of the model are illustrated for an exponential strength decay function in Figure 1.

Here we show that under these conditions, the average forgetting function approaches a power function as t grows large. Before proving a limit theorem under fairly general conditions, we present proofs for two special cases: linear trace decay and exponential trace decay. Simulation results for the case of exponential strength decay show that power forgetting emerges quite rapidly and is not just a property of the tail of the forgetting function (i.e., when accuracy approaches zero at very

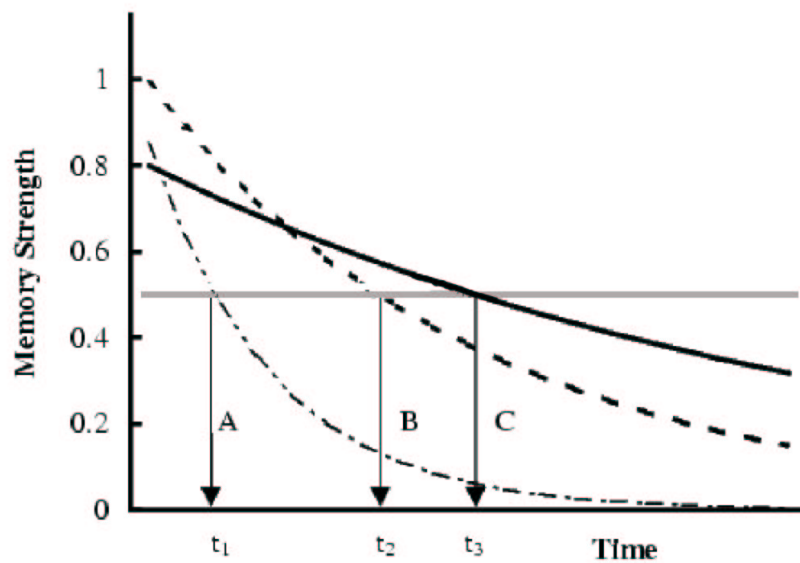


FIG. 1 A simple model of recall. Three curves illustrate the reduction in memory strength for three different items (A, B, and C). Items are assumed to decay at different rates and with different initial strengths. The gray bar indicates the strength threshold for correct recall or recognition. Items are forgotten when their strength falls below the threshold. Three arrows indicate the points in time when items A, B, and C are forgotten. Before t_1 all three items are remembered, two items are remembered until t_2 , one item is remembered until t_3 , and after t_3 none of the three items are remembered.

large t). In the special cases we assume, for convenience, that strength decay rates are normally distributed subject to the constraint that all items do exhibit at least some decay (we use truncated Gaussians). In the general derivation, we show that this assumption is not at all crucial to the proof of the limit theorem.

1.3. Special Case 1: Linear Strength Decay

Consider a linear strength-decay function with independent variable coefficients: $S_i(t) = a_i - b_i t$ with $a_i \sim N(\mu_a, \sigma_a)$, $b_i \sim N(\mu_b, \sigma_b)$, and $\mu_a > 0$, $\mu_b > 0$. Next, define a counter of retrievable memories, $r_i(t)$, such that:

$$r_i(t) = \begin{cases} 1, & S_i(t) > k \\ 0, & S_i(t) \leq k \end{cases}$$

This is a key feature of the model. Items are retrieved if strength is greater than k ; otherwise, they are forgotten.

Observe that: $\frac{\sum_{i=1}^N r_i(t)}{N} \rightarrow E(r_i(t))$, as $N \rightarrow \infty$ by the Law of Large Numbers, because the $r_i(t)$ are identically distributed, independent random variables. The expectation of $r_i(t)$, $E(r_i(t))$ is our forgetting function, averaged across many discrete items. Computing this expectation will give us the form of the forgetting function under the conditions outlined above. To compute $E(r_i(t))$ we proceed as follows:

$$\begin{aligned} E(r_i(t)) &= 1 \cdot P(r_i(t) = 1) + 0 \cdot P(r_i(t) = 0) \\ &= P(r_i(t) = 1) \\ &= P\left(\frac{a_1 - k}{b_1} > t\right) \\ &= \int_{y \geq 0} \int_{x=ty}^{\infty} \frac{\exp\left(\frac{-(x-(\mu_a-k))^2}{2\sigma_a^2}\right)}{c\sigma_a} \cdot \frac{\exp\left(\frac{-(y-\mu_b)^2}{2\sigma_b^2}\right)}{\sigma_b} dx dy, \end{aligned}$$

where we let $x = a_1 - k$ and $y = b_1$, and have constructed Gaussians about these points. Note that we are using a truncated Gaussian about b_i , if $b_i \geq 0$, as seen by the lower limit of the outer integral ($y \geq 0$), and $c = F(\frac{-\mu_b}{\sigma_b})F(\frac{-\mu_a}{\sigma_a})$, with F defined below.

Now we set $\gamma = \frac{x-(\mu_a-k)}{\sigma_a}$ and $\eta = \frac{y-\mu_b}{\sigma_b}$ to simplify our computations. To make a change of variables, however, we must modify the limits of integration. We see that $ty = \gamma = \frac{k-\mu_a}{\sigma_a} + \frac{t\eta}{\sigma_a}$ where $y = \eta\sigma_b + \mu_b$. In which case we can write:

$$\begin{aligned} E(r_1(t)) &= \int_{\frac{-\mu_b}{\sigma_b}}^{\infty} \int_{\gamma=(\eta\frac{\sigma_b}{\sigma_a} + \frac{\mu_b}{\sigma_a})t + \frac{k-\mu_a}{\sigma_a}}^{\infty} \frac{\exp\left(\frac{-(\gamma^2+\eta^2)}{2}\right)}{c} d\gamma d\eta \\ &= \int_{\frac{-\mu_b}{\sigma_b}}^{\infty} \frac{1}{c} F\left(\left(\eta\frac{\sigma_b}{\sigma_a} + \frac{\mu_b}{\sigma_a}\right)t + \frac{k-\mu_a}{\sigma_a}\right) \exp\left(\frac{-\eta^2}{2}\right) d\eta \\ \text{where,} \quad F(z) &\equiv \int_z^{\infty} \exp\left(\frac{-\alpha^2}{2}\right) d\alpha = \sqrt{2\pi}(1 - \text{c.d.f. gaussian}(z)) \end{aligned}$$

Thus, $F(-\infty) = 1$, and $F(z) \leq \exp\left(\frac{-z^2}{2}\right)$ for large $z > 0$, which is easily integrable. Now setting $\eta' = (\eta\frac{\sigma_b}{\sigma_a} + \frac{\mu_b}{\sigma_a})t + \frac{k-\mu_a}{\sigma_a}$ we find that:

$$\begin{aligned}
 E(r_1(t)) &= \frac{1}{t(\frac{\sigma_b}{\sigma_a})} \int_{\frac{k-\mu_a}{\sigma_a}}^{\infty} \frac{1}{c} F(\eta') \exp\left(\frac{-\left(\frac{\eta'}{t\frac{\sigma_b}{\sigma_a}} + \frac{\mu_a-k}{t\sigma_b} - \frac{\mu_b}{\sigma_b}\right)^2}{2}\right) d\eta' \\
 &= \frac{\exp\left(-\frac{\mu_b^2}{2\sigma_b^2}\right)}{t(\frac{\sigma_b}{\sigma_a})} \int_{\frac{k-\mu_a}{\sigma_a}}^{\infty} \frac{1}{c} F(\eta') d\eta' \rightarrow \frac{c'}{t}, \text{ as } t \rightarrow \infty, \\
 \text{where, } c' &= \frac{\exp\left(-\frac{\mu_b^2}{2\sigma_b^2}\right)}{c\frac{\sigma_b}{\sigma_a}} \int_{\frac{k-\mu_a}{\sigma_a}}^{\infty} F(\eta') d\eta'.
 \end{aligned}$$

1.4. Special Case 2: Exponential Trace Decay

Following the same construction as in the linear case, we define a counter of retrievable memories:

$$r_i(t) = \begin{cases} 1, & a_i e^{-b_i t} > k \\ 0, & \text{otherwise} \end{cases}$$

Further, let $x = a_i$ and $y = b_i$. Then the probability of retrieving a memory is given by:

$$E(r_i(t)) = P(r_i(t) = 1) = P(a_i > k e^{b_i t}) = \int_0^{\infty} \int_{x=k e^{y t}}^{\infty} \frac{e^{-\frac{(x-\mu_a)^2}{2\sigma_a^2}} e^{-\frac{(y-\mu_b)^2}{2\sigma_b^2}}}{c \sigma_a \sigma_b} dx dy$$

where, $c = F(-\frac{\mu_a}{\sigma_a})F(-\frac{\mu_b}{\sigma_b})$. Setting $\gamma = \frac{x-\mu_a}{\sigma_a}$ and $\eta = \frac{y-\mu_b}{\sigma_b}$ then gives

$$\int_{-\frac{\mu_b}{\sigma_b}}^{\infty} \int_{\gamma=(c_1 e^{t c_3} + c_2)}^{\infty} \frac{e^{-\frac{(\gamma^2 + \mu^2)}{2}}}{c} d\gamma d\eta$$

where $\gamma = \frac{k e^{y t} - \mu_a}{\sigma_a}$ and $y = \eta \sigma_b + \mu_b$, and where $c_1 = \frac{k}{\sigma_a}$, $c_2 = -\frac{\mu_a}{\sigma_a}$, and $c_3 = \eta \sigma_b + \mu_b$. We then obtain

$$\int_{-\frac{\mu_b}{\sigma_b}}^{\infty} \frac{1}{c} F(c_1 e^{t c_3} + c_2) e^{-\frac{\eta^2}{2}} d\eta,$$

where $F(\cdot)$ is as defined previously. Setting $t c_3 = t(\eta \sigma_b + \mu_b) = \eta'$ gives

$$\frac{1}{t \sigma_b} \int_0^{\infty} \frac{1}{c} F(c_1 e^{\eta'} + c_2) e^{-\frac{(\frac{\eta'}{t \sigma_b} - \frac{\mu_b}{\sigma_b})^2}{2}} d\eta'.$$

Thus, as $t \rightarrow \infty$,

$$E(r_i(t)) = \frac{e^{-\frac{\mu_b^2}{2\sigma_b^2}}}{t \sigma_b} \int_0^{\infty} \frac{1}{c} F(c_1 e^{\eta'} + c_2) d\eta' \rightarrow \frac{c'}{t}$$

$$\text{where, } c' = \frac{e^{-\frac{\mu_b^2}{2\sigma_b^2}}}{c \sigma_b} \int_0^{\infty} F(c_1 e^{\eta'} + c_2) d\eta'.$$

1.5. Discussion of the special cases:

In the two special cases presented above, we saw that as $t \rightarrow \infty$ a threshold model in which item strength decays either linearly or exponentially (with time) yields a hyperbolic relationship between recall probability and time. Although this is a special case of the power law of forgetting, the exponents for the limiting case of $t \rightarrow \infty$ are exactly 1, whereas in actual experiments, the exponent that fits the full range of participants' performance is often significantly less than one.

To see that the model does actually produce exponents that are less than one consider the following analysis. Recall that the forgetting function, given by $E(r_i(t))$, can be rewritten as follows:

$$E(r_i(t)) = \frac{c_1}{t} \int_{c_2}^{\infty} \mathcal{F}(\eta') \exp\left(-c_3 \frac{\eta'^2}{t} + c_4 \frac{\eta'}{t}\right) d\eta'$$

In the equation above, c_1 , c_2 , c_3 and c_4 are all constants. If we use the analytic series, $e^x = \sum_{n=0}^{\infty} x^n/n!$, for $x \in \mathcal{C}$, then we can rewrite the above expression as:

$$\begin{aligned} E(r_i(t)) &= \frac{c_1}{t} \int_{c_2}^{\infty} \mathcal{F}(\eta') \left(1 + \sum_{i=1}^{\infty} \frac{a_i(\eta')}{t^i}\right) d\eta' \\ &= \frac{c_1}{t} \left[\int_{c_2}^{\infty} \mathcal{F}(\eta') d\eta' + \sum_{i=1}^{\infty} \frac{A_i}{t^i} \right] \end{aligned}$$

Next, setting $c' = c_1 A_0$, and $b_i = A_i/A_0$, we have:

$$E(r_i(t)) = \frac{c'}{t} \left(1 + \frac{b_1}{t} + \frac{b_2}{t^2} + \dots\right)$$

If we define

$$\theta(t) = \frac{\ln\left(1 + \frac{b_1}{t} + \frac{b_2}{t^2} + \dots\right)}{\ln(t)}$$

then we have

$$E(r_i(t)) = \frac{c'}{t^{1-\theta(t)}}$$

Because $\theta(t) < 1$ and $\theta(t) \rightarrow 0$ as $t \rightarrow \infty$, we can see how the exponent of the power function varies with t in a systematic manner. For smaller values of t , the exponent will be less than one, but it will approach 1 as t grows large.

The analysis presented above shows that the simple decay-to-threshold model gives rise to power forgetting even when the strength decay functions are linear or exponential. In the next section we consider the centrality of our assumptions in producing the “power-law” behavior at large t . In particular, we generalize our analytic results to a very a broad class of strength decay functions and to non-parametric variability in the decay parameters.

1.6. General Derivation

Consider a strength-decay function of a general form: $S_i(t) = H(a_1^{(i)}, a_2^{(i)}, \dots, a_n^{(i)}, t)$, and assume the following conditions:

1. $H = H(a_1, a_2 e_2(t), a_3 e_3(t), \dots, a_n e_n(t))$ with

$$\mathcal{E}(t) := \prod_{i=2}^n e_i(t) = t^r \quad \text{with} \quad e_i(t) \rightarrow \infty \text{ as } t \rightarrow \infty$$

Remark 1: This form of the strength decay function is very general. Consider the case of the exponential decay common to many memory models, $S_i(t) = a e^{-bt}$. In this case, $a_1 = a$, $a_2 = b$, and $e_2(t) = t$. Similarly, the form of H also accommodates linear strength decay.

2. H is globally monotonic increasing in its first argument $a_1^{(i)}$ for all values of the other arguments, monotonic decreasing in t , and smooth (l continuous derivatives) in all arguments.
3. $a_j^{(i)}$ has distribution function $f_j(a_j^{(i)})$ with f_j independent of i , just depending on j , such that $f_j(x) = 0$ for $x < 0$,
4. all random variables are independent.

Remark 2. Condition 2 implies that $H \geq k \iff a_1^{(i)} \geq g(a_2^{(i)} e_2(t), \dots, a_n^{(i)} e_n(t), k)$, and $H \geq k \iff \phi(a_1^{(i)}, a_2^{(i)}, \dots, a_n^{(i)}, k) \geq t$ for some smooth function ($l-1$ continuous derivatives) in all arguments by the global monotonicity and implicit differentiation.

Set $F_1(x) = \int_x^\infty f_1(y) dy \rightarrow 0$ as $x \rightarrow \infty$. Define a counter of retrievable memories, $r_i(t)$, such that:

$$r_i(t) = \begin{cases} 1, & S_i(t) > k \\ 0, & S_i(t) \leq k \end{cases}$$

Again, this is a key feature of the model. Items are retrieved if strength is greater than k ; otherwise, they are forgotten. As before, observe that $(1/N) \sum_{i=1}^N r_i(t) \rightarrow E(r_i(t))$, as $N \rightarrow \infty$, by the Law of Large Numbers, because the $r_i(t)$ are identically distributed, independent random variables.

The expectation of $r_i(t)$ is our forgetting function, averaged across many discrete items. Computing this expectation will give us the form of the forgetting function under the general conditions outlined above. To compute $E(r_i(t))$ we proceed as follows:

$$\begin{aligned} E(r_i(t)) &= 1 \cdot P(r_i(t) = 1) + 0 \cdot P(r_i(t) = 0) \\ &= P(r_i(t) = 1) \\ &= P(S_i(t) \geq k) \\ &= P(\phi(a_1, a_2, \dots, a_n, k) \geq t) \quad (*) \\ &= 1 - F_\phi(t) \end{aligned}$$

where we have set all $a_l^i = a_l$, and where F_ϕ is a c.d.f. of the random variable $\phi(a_1, a_2, \dots, a_n)$. Note that we have used in the above that

$$S_i(t) \geq k \iff \phi(a_1, a_2, \dots, a_n, k) \geq t$$

for some function ϕ by *Remark 2*.

Analytic solutions for the special cases of linear and exponential forgetting functions (shown earlier) suggested that power forgetting emerges as a general property

of this model. Here we approach the general solution by asking the prospective question: When does

$$F_\phi(t) \simeq 1 - \frac{c'}{t^r}, \quad r > 0,$$

for t large? In other words, what are the conditions that would result in power-law forgetting?

This requires that the density distribution function

$$f_\phi(t) = F'_\phi(t) \sim \frac{c'r}{t^{r+1}}.$$

This is really a condition on the function $\phi(a_1, a_2, \dots, a_n, k)$.

Further analysis proceeds as follows: Using (*) and Remark 2 we compute in general that:

$$\begin{aligned} 1 - F_\phi(t) &= \int_0^\infty \int_0^\infty \dots \left(\int_{z_1 \geq g(z_2 e_2(t), \dots, z_n e_n(t), k)} f_1(z_1) dz_1 \right) \prod_2^n f_i(z_i) dz_i. \\ &= \int_0^\infty \int_0^\infty \dots \int_0^\infty F_1(g(z_2 e_2(t), \dots, z_n e_n(t))) \prod_2^n f_i(z_i) dz_i. \end{aligned}$$

Setting $z_l e_l(t) = z'_l$ gives

$$1 - F_\phi(t) = \frac{1}{\mathcal{E}(t)} \int_0^\infty \int_0^\infty \dots \int_0^\infty F_1(g(z'_2, \dots, z'_l)) \prod_{l=2}^n f_l\left(\frac{z'_l}{e_l(t)}\right) dz'_2 \dots dz'_n := \frac{q(t)}{\mathcal{E}(t)},$$

Using $e_l(t) \rightarrow \infty$ as $t \rightarrow \infty$ (so as $t \rightarrow \infty$, $f_l\left(\frac{z'_l}{e_l(t)}\right) \rightarrow f_l(0^+)$), and using Lebesgue's dominated convergence theorem and Assumption 3 below, we see that as $t \rightarrow \infty$, $1 - F_\phi(t)$ approaches:

$$\left(\frac{\prod_2^n f_i(0^+)}{\mathcal{E}(t)} \right) \int_0^\infty \dots \int_0^\infty F_1(g(z'_2, \dots, z'_l)) dz'_2 dz'_3 \dots dz'_n := \frac{q(\infty)}{\mathcal{E}(t)} := q(\infty) t^{-r}$$

by our hypothesis, with,

$$q(\infty) := \prod_2^n f_i(0^+) \int_0^\infty \dots \int_0^\infty F_1(g(z'_2, \dots, z'_n)) dz'_2 \dots dz'_n.$$

In order to obtain the limit above, we assumed that

$$\int_0^\infty \dots \int_0^\infty F_1(g(z'_2, \dots, z'_n)) dz'_2 \dots dz'_n < \infty \quad (2)$$

and the mild assumption that there exists an M such that $|f_i(x)| \leq M$ for all i and x (so that we can use dominated convergence).

The rate of convergence of $q(t) \rightarrow q(\infty)$ is controlled by the decay rate of $F_1(g(z'))$ as $z \rightarrow \infty$. For very small t , it is likely that the strengths of the individual items are all $> k$. This would mean that nothing has been forgotten yet. As t grows, and strengths begin to drop below the threshold, k , power forgetting may appear quite quickly.

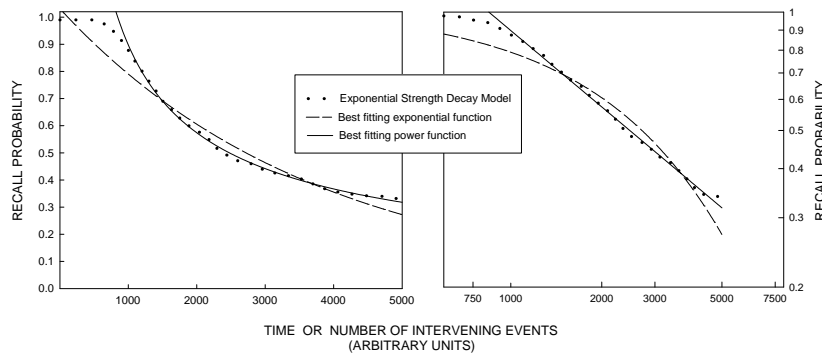


FIG. 2 Simulation example #1: Forgetting function for an exponential strength decay model. Left panel plots data in linear coordinates; Right panel plots data in log-log coordinates. The following parameter values were used: $k = 0.001$, $a \sim N(2.5, 1.0)$, $b \sim N(.003, .004)$. For this simulation, a and b are truncated normals, with 500 items being forgotten at different rates. Exponential and power functions were fit to data in the range of $t = 1000$ to 5000 . The best fitting power function is given by $P(\text{recall}) = 77t^{-.64}$ with $r^2 = 0.999$. The best fitting exponential function is given by $P(\text{recall}) = 1.03e^{-.0003t}$ with $r^2 = 0.95$.

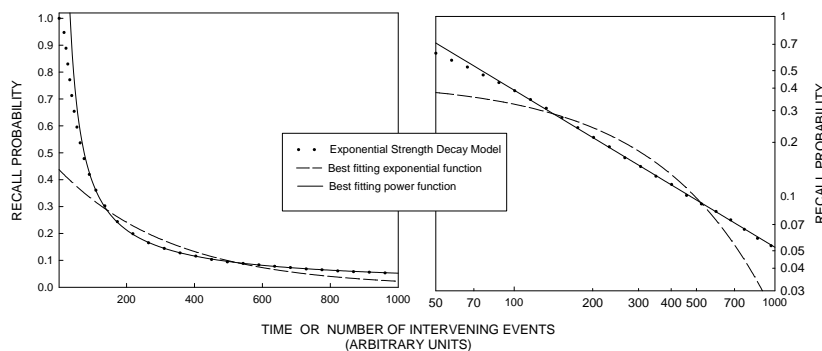


FIG. 3 Simulation example #2: Forgetting function for an exponential strength decay model. Left panel plots data in linear coordinates; Right panel plots data in log-log coordinates. The following parameter values were used: $k = 0.2$, $a \sim N(3.5, 0.5)$, $b \sim f(.06)$ where f denotes an exponential distribution. For this simulation, 12,000 items are forgotten at different rates. Exponential and power functions were fit to data in the range of $t = 80$ to 1000 . The best fitting power function is given by $P(\text{recall}) = 21t^{-.87}$ with $r^2 = 0.999$. The best fitting exponential function is given by $P(\text{recall}) = 0.44e^{-.003t}$ with $r^2 = 0.92$. Identical fits were obtained with the following parameter values: $k = 1$, $b \sim f(.025)$.

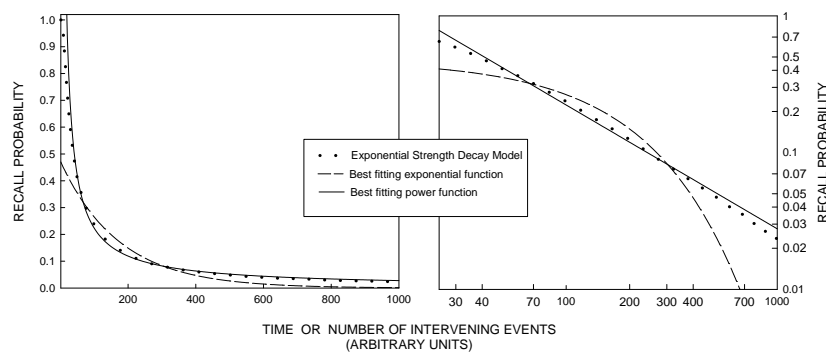


FIG. 4 Simulation example #3: Forgetting function for an exponential strength decay model. Left panel plots data in linear coordinates; Right panel plots data in log-log coordinates. The following parameter values were used: $k = 0.02$, $a \sim N(1.0, 0.2)$, $b \sim f(0.06)$ where f denotes an exponential distribution. For this simulation, 10,000 items are forgotten at different rates. Exponential and power functions were fit to data in the range of $t = 35$ to 1000. The best fitting power function is given by $P(\text{recall}) = 14.4t^{-.91}$ with $r^2 = 0.995$. The best fitting exponential function is given by $P(\text{recall}) = 0.47e^{-.006t}$ with $r^2 = 0.93$. Consistent with the theoretical analysis, the power function that best fit data in the range of $t = 150$ to 1000 was given by $P(\text{recall}) = 29.3t^{-1.03}$ with $r^2 = 1.000$.

1.7. Simulations

To examine how rapidly power forgetting emerges in our model we simulated the special case of exponential trace decay. In particular, $r(t) = E(a_i e^{-b_i t} > k)$, where a and b are random variables. In Simulation 1, a and b were truncated Gaussians ($a_i > 0$ and $b_i > 0$). This follows the proof of the special cases. In Simulations 2 and 3, a and b were drawn from an exponential distribution. This avoids the problem of having to truncate negative values. In addition to the parameters that determine the distributions of a and b , the model has one additional parameter: the forgetting threshold, k .

Figures 2-4 show three simulation runs with different parameter values. For each simulation we have plotted the models' predictions in both linear and log-log coordinates (a power function appears as a straight line in a log-log plot). The parameter values and r^2 s are given in the figure legends.

These simulations show that a simple decay-to-threshold model with exponentially decaying memories can produce aggregate power functions. These results are typical of many simulations run with other parameter values. Although the power function provides an excellent fit once performance drops significantly below ceiling, the power function does not provide an adequate fit to the early portion of our model's forgetting functions¹. Like experimental subjects, our model dictates that performance in the first few moments after learning remains at ceiling. Because $at^{-b} \rightarrow \infty$ as $t \rightarrow 0$ the power function is unable to fit this aspect of either the human data or the model results. It is only once forgetting is well underway that the power form emerges. This is consistent with our analytic results showing that power forgetting emerges as t grows large.

The simulations described above all assumed exponential strength decay of individual memories. Under these assumptions, the exponent of the power function that best fits the data should approach 1 as $t \rightarrow \infty$. Figure 5 replots the simulation results shown in Figure 4 and shows separate power-function fits to different ranges of the forgetting function. Consistent with the analytic results, the best fitting exponents decrease monotonically to -1 as we fit later regions of the forgetting curve.

Although our model assumes a constant threshold, it is mathematically indistinguishable from a model that assumes different thresholds for different items. To eliminate a free parameter, we allowed the variability in the y -intercept of the strength decay function forgetting function to subsume variability in the threshold (across items, but not time).

1.8. Conclusions

We have shown a striking example of how tricky it is to infer psychological process from the form of the retention function. Moreover, our analytic results suggest that there is something special about the power function as a description of the forgetting process. It may be that power functions arise from aggregate forgetting data under fairly general conditions. If this is true, then the challenge for memory theory is to explain violations of the power law of retention. One limitation of our analysis is that it does not readily handle RT data (this is because

¹Consider the simulation shown in Figure 3. For these parameters, performance is at or near ceiling for small t . Although a power function accounts for nearly 100% of the variance when fit to the data from $t = 1000$ to $t = 5000$, the same power function does not provide a good fit when the data from $t = 0$ to $t = 1000$ are included.

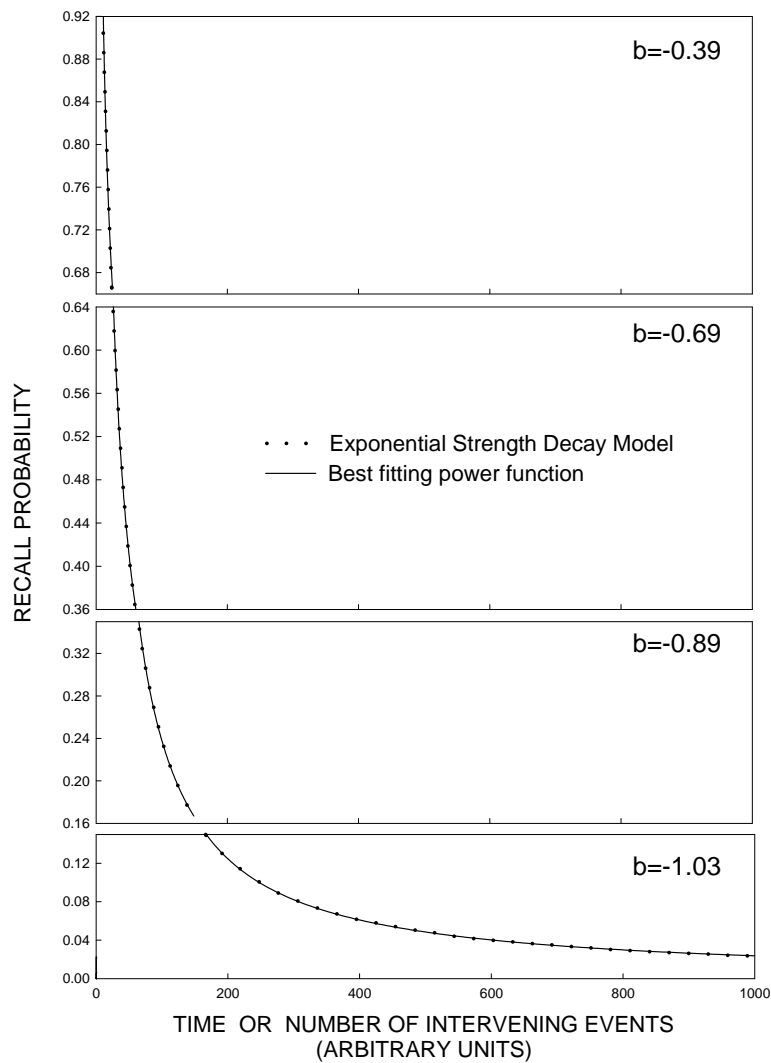


FIG. 5 Replot of simulation example #3: A separate power function was fit to each of four regions of the forgetting function. The best fitting exponent, b is given in each panel.

we assume a threshold that translates memory strength into binary responses). For learning data, the decrease in RT with practice obeys a power law (see Anderson, 1995 for a review). Regarding forgetting data, the increase in correct recognition RT with study-test lag is often linear (see Murdock, 1974, for a review) and the increase in inter-response times with output position is typically exponential (e.g., Murdock & Okada, 1970; Rohrer & Wixted, 1994; Wixted & Ebbesen, 1997). These findings are not addressed by our analysis.

In this note, we have shown how a simple model that assumes monotonic strength decay at the level of individual items coupled with a discrete retrieval threshold predicts aggregate power forgetting functions. This exploration illustrates some of the dangers inherent in inferring mental organization from the mathematical form of the behavioral data. The important question for theories of memory to address is how forgetting is affected by experimental manipulations and not what mathematical form the forgetting process assumes.

ACKNOWLEDGMENTS

The authors acknowledge support from National Institute of Health grant MH55687 to Brandeis University. Correspondence concerning this article should be addressed to Michael Kahana, Volen National Center for Complex Systems, MS 013, Brandeis University, Waltham, MA 02254-9110. E-mail: kahana@brandeis.edu.

REFERENCES

- Anderson, J. R. (1995). *Learning and memory: An integrated approach*. New York: Wiley.
- Anderson, R. B. & Tweney, R. D. (1997). Artifactual power curves in forgetting. *Memory & Cognition*, 25, 724–730.
- Anderson, R. B. & Tweney, R. D. (1998). The power law as an emergent property. *Unpublished Manuscript*.
- Bower, G. (1967). A multicomponent theory of the memory trace. In *The Psychology of Learning and Motivation : Advances in Research and Theory*, K. W. Spence & J. T. Spence, eds., vol. 1, New York: Academic Press, pp. 229–325.
- Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.
- Howard, M. W. & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46.
- Mensink, G.-J. M. & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, 95, 434–455.
- Murdock, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626.
- Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, 104, 839–862.

- Murdock, B. B. & Kahana, M. J. (1993). Analysis of the list strength effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 689–697.
- Murdock, B. B. & Kahana, M. J. (1993b). List-strength and list-length effects: Reply to Shiffrin, Ratcliff, Murnane, and Nobel. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 1450–1453.
- Murdock, B. B. & Lamon, M. (1988). The replacement effect: Repeating some items while replacing others. *Memory & Cognition*, 16, 91–101.
- Murdock, B. B. & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, 86, 263–267.
- Rohrer, D. & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, 22, 511–524.
- Rubin, D. C. & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, 103, 734–760.
- Wickens, T. D. (1998). On the form of the retention function: Comment on Rubin and Wenzel (1996): A quantitative description of retention. *Psychological Review*, 105, 379–386.
- Wixted, J. T. (1990). Analyzing the empirical course of forgetting. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 16, 927–935.
- Wixted, J. T. & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, 2, 409–415.
- Wixted, J. T. & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, 25, 731–739.