Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions

JOHN T. WIXTED and EBBE B. EBBESEN
University of California, San Diego
La Jolla, California

Wixted and Ebbesen (1991) showed that forgetting functions produced by a variety of procedures are often well described by the power function, at^{-b}, where a and b are free parameters. However, all of their analyses were based on data arithmetically averaged over subjects. R. B. Anderson and Tweney (1997) argue that the power law of forgetting may be an artifact of arithmetically averaging individual subject forgetting functions that are truly exponential in form and that geometric averaging would avoid this potential problem. We agree that researchers should always be cognizant of the possibility of averaging artifacts, but we also show that our conclusions about the form of forgetting remain unchanged (and goodness-of-fit statistics are scarcely affected by) whether arithmetic or geometric averaging is used. In addition, an analysis of individual subject forgetting functions shows that they, too, are described much better by a power function than by an exponential.

The idea that the power law of forgetting may be an artifact of averaging data over subjects is one that we have encountered often since reporting our findings on the form of forgetting several years ago (Wixted & Ebbesen, 1991). Although we investigated the possibility of averaging artifacts in great depth at that time, we did not actually report the results of that investigation. R. B. Anderson and Tweney's (1997) recommendation that researchers seriously consider the possibility of averaging artifacts is one that we fully endorse, because, as they show, such artifacts can arise under certain conditions (viz., when high measurement error is combined with floor effects). In hindsight, we also recommend that researchers report the outcomes of their inquiries into the possibility of averaging artifacts lest they arouse the suspicions of a naturally skeptical audience. We failed to do that in our earlier report. The purpose of the present article is to address this issue in some detail, in part by presenting a detailed analysis of individual subject forgetting functions.

Another reaction we have encountered frequently, one not specifically mentioned by R. B. Anderson and Tweney (1997), concerns why anyone should care whether or not the form of forgetting is consistent across different types of memory tests and stimulus materials. The fact that the form of forgetting may be a stable property of memory is

interesting in its own right to some (Rubin & Wenzel, 1996). Nevertheless, to others, an apparent empirical regularity in the absence of detailed theoretical considerations is little more than a mathematical curiosity (e.g., Simon, 1992). In this article, we also consider both the practical utility and the theoretical implications of the fact that forgetting is generally well described by functions that involve time raised to a power.

The Effects of Averaging Over Subjects

Although Ebbinghaus (1885/1913) initiated inquiry into the subject more than a century ago, only a few investigations into the mathematical form of forgetting have ever been performed. In most of the relevant studies, only a single procedure was used to generate a forgetting function. Ebbinghaus, for example, used savings for nonsense syllables and found that a complex logarithmic function described his data. Wickelgren (1970, 1972) generally relied on recognition tests and found that an exponential function described forgetting from short-term memory but that a power function described forgetting from long-term memory. Rubin (1982) used autobiographical memory tests in which subjects were asked to supply dates associated with personal memories. The power function described performance on this task better than the exponential. In the animal memory literature, White (1985) employed the delayed matching-to-sample task and found that the exponential often provided a good fit to these data.

In an effort to test whether the mathematical form of forgetting might be the same across tasks and species, Wixted and Ebbesen (1991) collected forgetting functions for the short-term recall of words, long-term recognition of faces, and delayed-matching-to-sample in pigeons. In addition, like J. R. Anderson and Schooler (1991), we reanalyzed the famous savings function reported long ago by Ebbinghaus. Six different commonly used functions were fit to these data, and a clear pattern emerged. Specifically, the power and logarithmic functions described the data accurately (with the edge going to the former), but the exponential and hyperbola performed poorly.

Wixted and Ebbesen's (1991) analyses were all performed on data arithmetically averaged over subjects (with the exception of Ebbinghaus's savings function). We analyzed averaged data because the resulting forgetting functions were very smooth and were therefore particularly suitable for discriminating between rival mathematical models. However, researchers have long known that averaging over subjects can produce a group function with mathematical properties that are not representative of the individual subject data (e.g., Estes, 1956; Sidman, 1952). R. B. Anderson and Tweney (1997) observed further that arithmetic averaging of exponential functions in particular can yield a group curve that is fit better by the power function than by the exponential if the data are variable and near the floor. At the very least,

R. B. Anderson and Tweney show that arithmetic averaging exaggerates the advantage of the power function over the exponential in terms of the percentage of data variance accounted for. Thus, fitting a curve based on arithmetic averaging may not provide a fair test of the hypothesis that forgetting functions are truly exponential in form.

An exponential function has the form ae^{-bt}, and a power function has the form at^{-b}, where a and b represent constants and t represents time. Geometric averaging of either function preserves its form in a group function (Estes, 1956). Thus, as R. B. Anderson and Tweney (1997) noted, one way to evaluate whether or not averaging artifacts were responsible for the findings that we reported is to reanalyze the data by using geometric averaging. In the following section we do just that. A second and more direct way to address the issue is to fit data separately for individual subjects, and the results of such an analysis are also presented below.

Geometric averaging. One experiment that we reported (Wixted & Ebbesen, 1991) involved long-term recognition of previously studied faces. Subjects studied 40 faces, followed by a retention interval ranging from 1 h to 2 weeks for different groups of subjects, followed by a yes/no recognition test involving the 40 studied faces randomly intermixed with 40 new faces. Figure 1 presents the recognition data from this experiment. We reported these data in 1991, except that we now use d' rather than percent correct as the dependent measure. The upper panel of the figure shows the values obtained by using arithmetic averaging over subjects. The best fitting power function is shown as a solid curve and the best fitting exponential is shown as a dashed curve.

One question that often arises at this point concerns whether or not the dashed curve really represents the best fitting exponential. This question arises because the exponential "curve" looks more like a straight line. In fact, it is the best fitting exponential according to a least squares criterion. That is, the two parameters of the exponential were adjusted to minimize the sum of squared deviations between the predicted and obtained points. Any adjustment of the exponential's free parameters to allow for a more apparent curvilinear shape yields deviations from the data greater than those shown in Figure 1 (i.e., the percentage of data variance accounted for decreases).

The percentage of data variance accounted for by each function is shown in the legend. The power function accounts for almost all of the variance, whereas the exponential accounts for less than 80%. More importantly, the exponential exhibits obvious systematic deviations from the data. The question of interest is whether those deviations were caused by arithmetic averaging over subjects.

The lower panel of Figure 1 shows the same data, but this time geometric averaging over subjects was used. That is, each data point in the figure was obtained by taking the nth root of the product of n d' scores (where n is the number of subjects constituting the mean). This
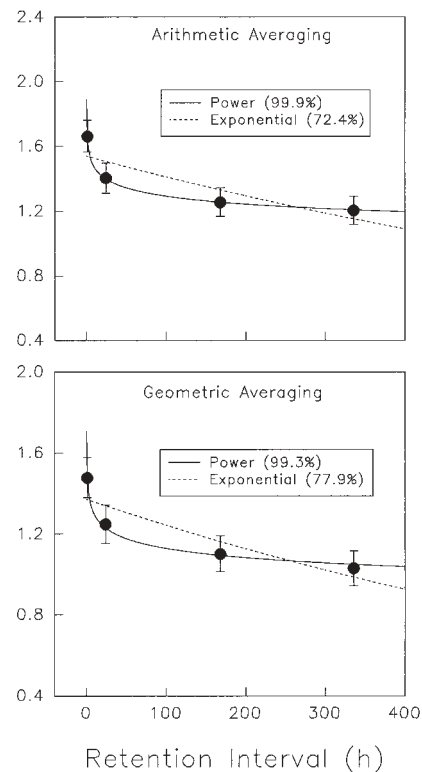

Figure 1. Upper panel: mean d' scores arithmetically averaged over subjects for each retention interval in the face recognition experiment. Lower panel: mean d' scores geometrically averaged over subjects for each retention interval in the face recognition experiment. The solid curve in both graphs represents the best fitting power function, and the dashed curve represents the best fitting exponential.

method of averaging would produce an exponential group function if the component functions were exponential. Once again, the best fitting power function is shown as a solid curve and the best fitting exponential is shown as a dashed curve. As R. B. Anderson and Tweney (1997) have noted, the advantage of the power function over the exponential as measured by the difference in the percentage of data variance accounted for by the two functions is not quite as great when geometric averaging is used. Nevertheless, the advantage of the power function is still large, and the deviations exhibited by the best fitting exponential are still considerable. Thus, for these data at least, the conclusion that the power function provides a much better fit than the exponential does not appear to be the result of an averaging artifact.

Wixted and Ebbesen (1991) also reported the results of an experiment that used free recall instead of recognition. Subjects were exposed to lists of six words, followed by a filled retention interval during which they were required to overtly rehearse distractor words, followed by a free recall period during which they were asked to recall as many of the original words as possible in any order

that they wished. The retention interval ranged from 2.5 to 40 sec across trials, and two learning conditions were used (fast vs. slow presentation rate).

The upper panel of Figure 2 presents the values obtained by using arithmetic averaging over subjects and learning conditions. Once again, these data reveal the obvious advantage of the power function. The lower panel of Figure 2 shows the same data, but this time geometric averaging over subjects and conditions was used. As before, the advantage of the power function declines ever so slightly, but its clear superiority remains plain to see.

Individual subject analyses. Each group of subjects in the face recognition experiment was tested following a single retention interval (i.e., a between-subjects design was used). As a result, it was not possible to analyze individual forgetting functions. Subjects in the free recall experiment, however, were run for six sessions each (three involving a fast presentation rate and three involving a slow presentation rate). Because all subjects were exposed repeatedly to each retention interval, it was possible to analyze the shape of each subject's forgetting function to determine whether or not the mean function was representative of their performance.

Figure 3 shows the data produced by the 8 subjects who participated in the recall experiment. The filled circles

Figure 3. Proportion correct for 8 subjects in the free recall experiment. The solid curves represent the best fitting power functions for each subject in each condition. Filled circles, slow presentation rate; open circles, fast presentation rate.
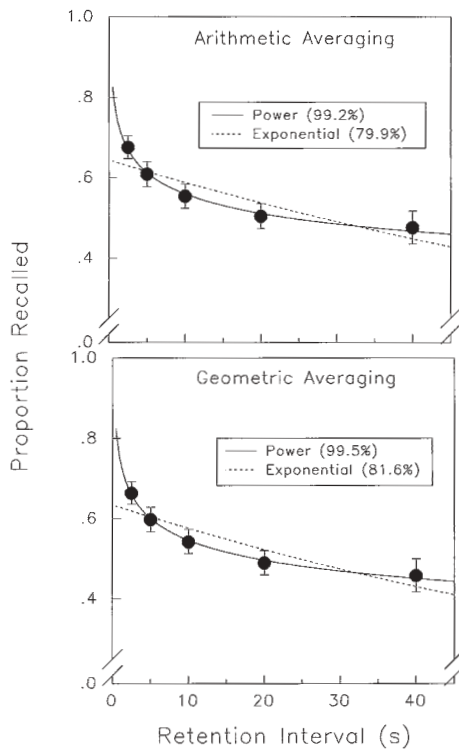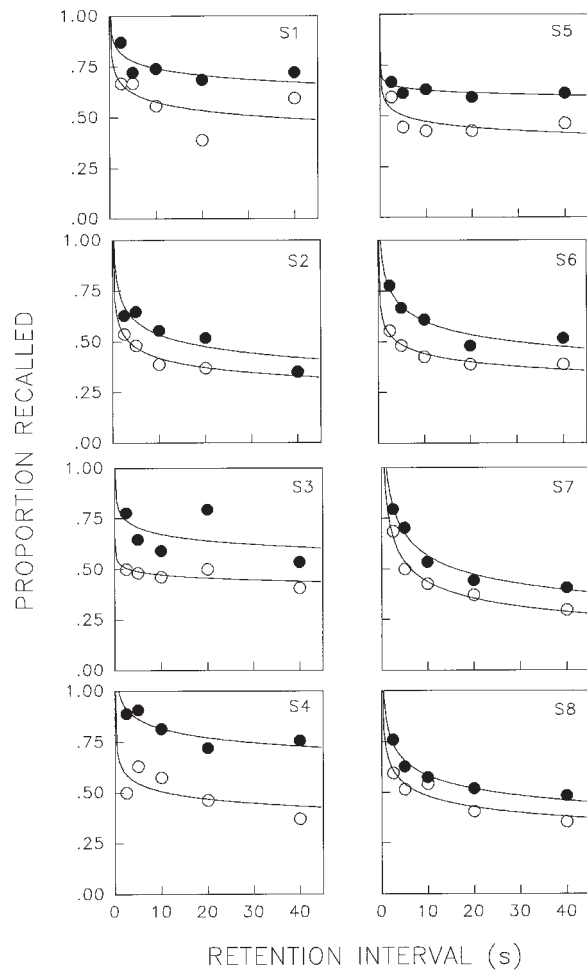

Figure 2. Upper panel: mean proportion correct recall arithmetically averaged over subjects for each retention interval in the free recall experiment. Lower panel: mean proportion correct recall geometrically averaged over subjects for each retention interval in the free recall experiment. The solid curve in both graphs represents the best fitting power function, and the dashed curve represents the best fitting exponential.

represent performance in the three sessions involving the slow presentation rate (high degree of learning), and the open circles represent performance in the three sessions involving a fast presentation rate (low degree of learning). The solid curves represent the best fitting power functions. Although some subjects produced variable data, an examination of this figure suggests that the power function does a reasonably good job of capturing the trends in the data.

Table 1 shows the percentage of data variance accounted for by the power function and by the exponential for the data shown in Figure 3. Of the 16 separate fits (2 for each subject), the power function outperformed the exponential 11 times and the exponential outperformed the power function 5 times (although 2 of those were virtual ties). Of the 5 cases in which the fit of the exponential function exceeded that of the power function, a straight line actually provided the best fit in 4. Thus, a more accurate summary is that the power function fit best in 11 cases, the straight line in 4, and the exponential fit best in 1 case.

Table 1
Percentage of Variance Accounted for
by Exponential and Power Functions for the
Individual Subject Forgetting Functions Shown in Figure 3

Subject	Short		Long	
	Exponential	Power	Exponential	Power
1	14.1	36.9	26.6	56.9
2	69.0	94.1	96.2	79.1
3	62.5	46.2	23.7	21.5
4	68.5	41.0	60.1	77.2
5	12.4	42.8	28.6	54.1
6	63.5	93.1	65.3	90.8
7	77.3	97.3	80.9	97.8
8	88.4	86.2	72.0	96.0
M	57.0	70.9	56.7	71.7

The mean percentage of data variance accounted for was about 71% for the power function and about 57% for the exponential. These data suggest that the form of the mean function is, for the most part, representative of the individual subjects. Thus, while we endorse R. B. Anderson and Tweney's (1997) recommendation that researchers look into the possibility of averaging artifacts (especially when variable data near the floor are averaged together), we conclude that such effects played a minimal role in the analysis offered by Wixted and Ebbesen (1991).

We turn now to a closer look at the power function parameter estimates obtained from fitting the individual subject data. The ability to analyze these estimates is one of the practical advantages of quantitatively analyzing forgetting functions (see White, 1985, for a particularly instructive example of this approach).

Individual Subject Parameter Estimates

For each of the fits shown in Figure 3, the two parameters of the power function (at^{-b}) were adjusted to minimize the sum of squared deviations. Table 2 presents the parameter estimates obtained for each subject. Note that the rate of forgetting (captured by the b parameter) varied considerably across subjects, ranging from a low of .05 (Subject 3, fast condition) to a high of .30 (Subject 7, fast condition). An examination of the forgetting functions produced by these 2 subjects (which are shown in Figure 3, open circles) clearly reveals the differences in their rates of forgetting. The a parameter also varied considerably across subjects, ranging from a low of .53 (again, Subject 3, fast condition) to a high of 1.0 (Subject 7, slow condition). This parameter represents the estimated level of performance after one unit of time (i.e., after a 1-sec retention interval in this case).

These parameters reveal several interesting facts. First, the means of the parameter estimates (shown at the bottom of Table 2) almost exactly match the parameter estimates obtained from fitting the power function to the recall data arithmetically averaged over subjects. Figure 4 shows the mean data and corresponding fits. These are the same data shown in Figure 2, except that now the arithmetically averaged data are shown separately for the two learning conditions. The parameter values obtained from fitting the mean data are shown in the figure. Those values are very similar to the mean values obtained from individual subject fits presented in Table 2. Thus, in this case at least, arithmetic averaging across subjects did not distort either the form or the statistical properties of the individual subject forgetting functions.

A second interesting observation about the parameter estimates shown in Table 2 is that both the degree of learning (captured by the a parameter) and the rate of forgetting (captured by the b parameter) are stable within subjects. The correlation between the values of b obtained in the high and low degree of learning conditions (which were run in separate sessions) was .911. One would expect the rate of forgetting to be a stable property of the subject, and the results shown in Table 2 confirm this expectation. The within-subjects correlation in the a parameter was .679.

A third interesting observation is that the correlation between the parameters a and b across subjects appears to be positive. That is, a high value of a (indicating a high degree of learning) is associated with a high value of b (indicating a faster rate of forgetting). The correlation between a and b is .834 in the low degree of learning condition and .573 in the high degree of learning condition. Note that this is in contrast to the relationship between a and b across conditions. That is, although a increases significantly with study time (from 0.66 to 0.86), the value of b stays essentially constant (0.14 and 0.13, respectively). Across subjects, by contrast, the two parameters appear to be positively correlated. This correlation was unexpected and should probably be replicated before one takes it seriously. Nevertheless, if this is a replicable result, the theoretical implications would be intriguing. It may, for example, represent differences in the strategies used by subjects to memorize list items. Some subjects may concentrate on a few of the words, memorizing them well. For them, overall degree of learning would be low, but so would the rate of forgetting (because the items are well encoded). Other subjects may attempt to memorize all of the items, producing a greater number of more vulnerable traces. For them, degree of learning would be high, but the items would be lost more quickly.

A Note on Odds Versus Percent Correct

In the preceding analysis, the proportion of items recalled (which can range from a high of 1.0 to a low of 0)

Table 2
Power Function Parameter Estimates for the
Individual Subject Data Shown in Figure 3

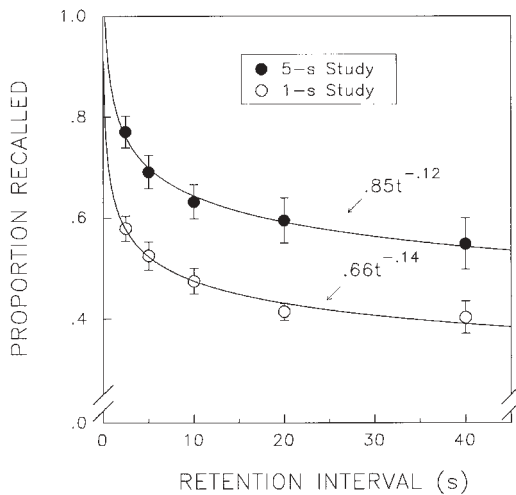| Subject	Short		Long	
	a	*b*	*a*	*b*
1	0.74	0.11	0.87	0.07
2	0.62	0.17	0.79	0.17
3	0.53	0.05	0.79	0.07
4	0.65	0.11	0.98	0.08
5	0.58	0.09	0.67	0.03
6	0.61	0.14	0.89	0.17
7	0.87	0.30	1.00	0.26
8	0.71	0.17	0.86	0.17
M	0.66	0.14	0.86	0.13

Figure 4. Mean proportion correct recall for the fast and slow conditions of the free recall experiment. The solid curves represent the best fitting power function.

was fit by a power function (which has a possible range of infinity to 0). That is, as t approaches 0, at^{-b} approaches infinity. Although a retention interval of 0 may not be physically possible, it might make more sense to use a dependent measure that could, in principle, become infinitely large if t did approach 0. Whereas percent correct (p) cannot, odds can. The odds of being correct (o) is $p/(1-p)$. Thus, instead of using the equation

$$p = at^{-b}, \tag{1}$$

as we have done in the past, one might be better off using

$$o = at^{-b}, \tag{2}$$

as J. R. Anderson and Schooler (1991) have suggested. Since $o = p/(1-p)$, the power function for proportion correct works out to be

$$p = at^{-b}/(1 + at^{-b})$$

or, after some rearrangement,

$$p = a/(a + t^b), \tag{3}$$

where a now represents the *odds* of being correct following one unit of time. This version of the power function is constrained to range between a maximum of 1.0 (as t approaches 0) and a minimum of 0 (as t approaches infinity).

In practice, the performance of the power function is about the same whether Equation 1 is used or Equation 2 (or, equivalently, Equation 3) is used. Thus, for example, when Equation 2 is fit to the forgetting functions shown in Figure 4 (using odds as the dependent measure), it accounts for 97.2% and 97.1% of the variance for the high and low degree of learning conditions, respectively. The reason is that, whether proportion correct or odds is used, the forgetting function exhibits a steep drop initially followed by a long, slow decay that begins well above 0. Under those conditions, just about any function involv-

ing t raised to an exponent will do a good job of fitting the data (the simple power function that we used here is only one of many with that property). Moreover, when the individual subject correlational analyses just described are repeated using odds as a dependent measure, the same conclusions are reached. That is, both the degree of learning (a) and the rate of forgetting (b) are stable across subjects, and the correlation between a and b across subjects (but not across conditions) is positive.

Note that for recognition, the use of d' has the appropriate possible range (infinity to zero). Thus, fitting the simple power function to d' does not raise the same conceptual problem that one encounters with percent correct. The point here is that, for conceptual consistency, the range of the mathematical function used to describe the data should correspond to the possible range of the dependent measure.

Averaging Exponential Recall Latency Functions

Up to this point, we have argued that individual subject forgetting functions (as well as forgetting functions averaged over subjects) are described more accurately by the power function than by the exponential. Might the power function be so inherently flexible that it could outperform (or, at least, rival) the exponential when the data are truly exponential in form? That is the issue we address in this section.

In a separate line of research, we have also analyzed the mathematical form of recall functions, which represent the number of items recalled during each second of the recall period that follows list presentation (Rohrer & Wixted, 1994; Wixted & Rohrer, 1993). Typically, recall begins at a rapid rate and then tapers off to zero according to a curvilinear function of time (in the same way that forgetting functions do).

In one of the experiments reported by Rohrer and Wixted (1994), subjects were exposed to a free recall task much like the one described earlier, but the retention interval was not manipulated. Each list of five words was followed by a short distractor task and then by a 20-sec recall period. The number of items recalled during each second of the recall period was measured using a voice-activated relay. Figure 5 shows the results for 7 subjects, as well as the group average. Note that the complete data sets for these subjects would include one or two low points in the first few seconds of the recall period (i.e., the functions actually rise rapidly and then fall for the remainder of the recall period). The ex-Gaussian distribution describes these full data sets accurately (see, e.g., Rohrer & Wixted, 1994; Wixted & Rohrer, 1993). For the sake of simplicity, however, only the tails of the recall distributions are shown in Figure 5 (i.e., the first one or two low points have been excluded).

The solid curves in Figure 5 represent the best fitting exponential functions of the form ae^{-bt}. The dashed curve for the mean plot represents the best fitting power function. Clearly, these data, at both the individual and group levels, are very well described by the exponential.
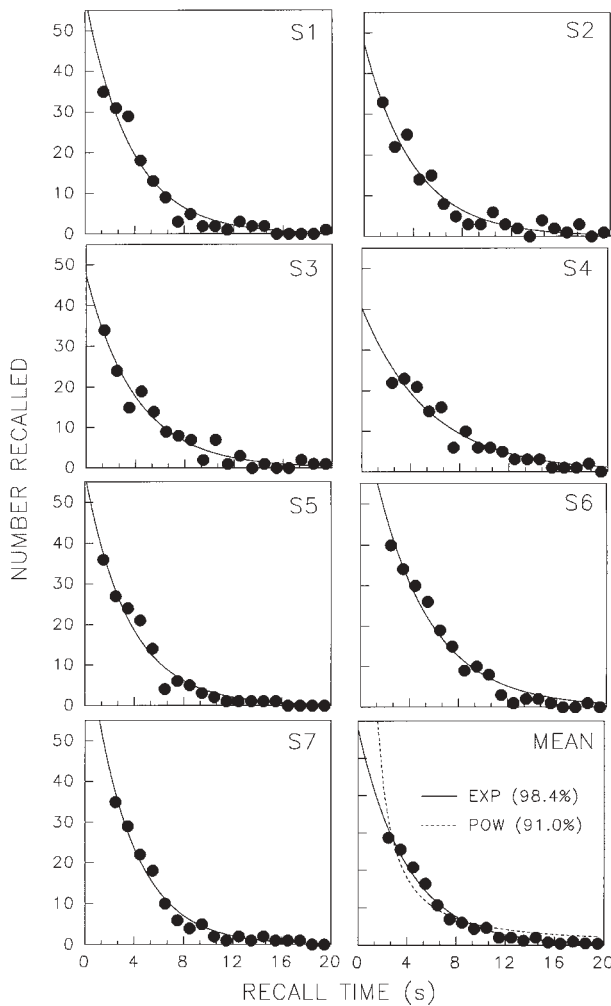
Figure 5. Number of items recalled in each 1-sec bin of a 20-sec recall period for 7 subjects. The solid curves represent the best fitting exponentials. The dashed curve in the mean graph represents the best fitting power function.

The value of *a* ranges from a low of 40.4 to a high of 78.8. The value of *b* ranges from a low of .183 to a high of .301. Note that variability in the *a* parameter does not introduce any shape distortion in the average function, whereas variability in the *b* parameter could. Although the same subjects produced the low and high values for *a* and *b*, overall the correlation between the two parameters is low ($r = .21$). As before, the average of the parameter estimates based on the individual subject fits corresponds closely to the parameter estimates obtained from fitting the data averaged over subjects. The values of *a* and *b* obtained from fitting the exponential to the group function in Figure 5 were 58.0 and 0.25, respectively. As shown in Table 4, the mean values of *a* and *b* obtained from the individual subject fits were 57.8 and 0.25, respectively. Thus, while averaging over subjects certainly can create distortions, under the right conditions it can also generate a fairly representative function.

Averaging Over Items

Averaging over subjects is apparently not responsible for the observation that the mathematical form of forgetting is adequately characterized by a simple power function. What about averaging over items? That is, perhaps the probability of recalling individual items declines according to exponential functions with different rates of decay. If so, the aggregate function may be something other than an exponential.

In general, the issue of averaging over items is much the same as averaging over subjects (i.e., one is still arithmetically averaging exponentials). One relatively minor difference is that, in a list memory experiment, the individual items all have different retention intervals (i.e., lags) associated with them. That is, in the absence of rehearsal, the last item of the list is associated with the shortest retention interval, whereas the first item is associated with the longest. In the fits described thus far, the nominal retention interval has referred to the time since the last item was presented.

To investigate the effects of averaging over items associated with different rates of decay and different retention intervals, we performed simulations similar to those described by R. B. Anderson and Tweney (1997). The simulations were modeled after the free recall experiment described above (i.e., six-item lists with nominal retention intervals ranging from 2.5 to 40 sec). In the simulations, the actual retention intervals were item spe-

Table 3 shows the percentage of data variance accounted for by the exponential and by the power function for each subject. The exponential provides a better fit in every case. Thus, the advantage of the power function over the exponential in describing forgetting functions is not merely the result of the inherent flexibility of the power function. When the data are very nearly exponential in form (as is apparently the case for recall curves), the power function performs rather poorly. These data also underscore a point that will be made more fully in the next section—namely, that in the absence of extreme variability, averaging together exponential functions yields a group function that is true to form.

Although the average recall latency function retained the exponential form of the component functions, the estimates for *a* and *b* from fitting the group data may be very different from the average *a* and *b* parameter estimates obtained from the individual subject fits. Table 4 presents the relevant parameter estimates for each subject.

Table 3
Percentage of Variance Accounted for
by Exponential and Power Functions for the
Individual Subject Recall Functions Shown in Figure 5

Subject	Exponential	Power
1	96.0	83.9
2	94.4	87.7
3	95.4	89.9
4	93.5	82.5
5	96.8	86.6
6	97.5	87.7
7	98.2	92.1

Table 4
Exponential Function Parameter Estimates for the
Individual Subject Data Shown in Figure 5

Subject	a	b
1	58.6	0.28
2	47.4	0.25
3	47.4	0.24
4	40.4	0.18
5	57.0	0.28
6	75.3	0.22
7	78.8	0.30
M	57.8	0.25

cific. Thus, for example, if the nominal retention interval was 2.5 sec, the actual retention interval was 2.5 sec for the last item in the list, 3.5 sec for the second to last item, 4.5 sec for the next to last item, and so on. The probability of recalling an item was assumed to decline according to an exponential function, but the decay constants were variable. Thus, the probability of recalling a particular item after time t was set to $e^{-\lambda t}$, where λ and t were different for every item in the list.

The values of λ were drawn from several kinds of distributions, and the resulting functions were used to compute the average probability of recalling an item after t units of time. Thus for example, six values of λ were first drawn from a parent distribution of some kind. The probability of recall following a particular nominal retention interval was then computed as follows:

$$ p = \frac{1}{6} \sum_{i=1}^{6} e^{-\lambda_i (t + t_i)}, $$

where t represents the nominal retention interval (e.g., 2.5 sec) and t_i represents an additional unit of time determined by the item's serial position (i.e., t_i was 0 for the last item in the list, 1 for the second-to-last item, etc.). This was repeated for nominal values of t equal to 2.5, 5, 10, 20, and 40 sec (which are the actual retention intervals used in the free recall experiment). The resulting simulated forgetting functions were then fit by a power function and by an exponential.

We selected values of λ or $1/\lambda$ (hereafter represented by τ) from various distributions. If the parent distribution for λ was either normal or exponential in form, we were unable to produce an average power function like those shown in Figure 4 (even when extreme variability was introduced), although we could occasionally produce an aggregate curve that was slightly better fit by the power function than by the exponential. Note that the outcome of interest is not merely an aggregate function being better fit by the power function than by the exponential. Instead, we attempted to create an aggregate function that was simultaneously very well fit by the power function and rather poorly fit by the exponential (because that is the pattern observed in the real data).

We also tried drawing values of τ from normal and exponential distributions and still could not produce what appeared to be a power function. Thus, it is not simply the

case that the power law of forgetting results from averaging variable exponentials. A particular kind of variability appears to be required. Indeed, if the values of τ were drawn from a Weibull distribution with extreme variability, it was possible to produce a forgetting function that was very nicely fit by the power function (and poorly fit by the exponential). Figure 6 shows a forgetting function produced in this way. The values for τ were drawn from the following Weibull distribution, which in cumulative form is written:

$$ P(\tau \le x) = 1 - e^{-\frac{(x-\alpha)^\beta}{\gamma}}, $$

with parameters set to $\alpha = 2$, $\beta = 0.2$, and $\gamma = 4$ (see Johnson & Kotz, 1970, for details about this distribution). This distribution yields a mean value of τ equal to 34.13 h (such that the mean half-life of an item is 23.6 h) and a standard deviation of 540.8 h. The mean seems high for a simulated short-term recall task because this is a highly skewed distribution. The median value of τ was only 2.76 min (corresponding to a median half-life of 1.92 min). Thus, half the items on the list have a short half-life (less than 2 min), and the other half have a longer half life. Some have an extremely long half-life, which accounts for the fact that the function appears to be approaching an asymptote. The role of the Weibull distribution in producing a power law in behavioral data is well known (cf. Indow, 1993; Logan, 1995).

The mere fact that values of τ are drawn from a Weibull distribution does not ensure that the aggregate function will be well fit by the power function and poorly fit by the exponential. Many other settings for the Weibull parameter values can be found that produce an aggregate function that is still better fit by the exponential than by the power function. Note that adding Gaussian error (to represent measurement error) did not have much of an effect unless floor effects were at issue. Also, the fact that
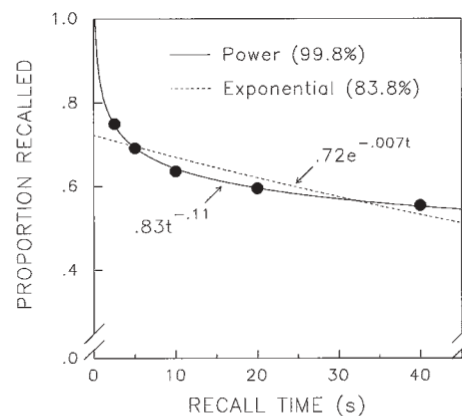

Figure 6. Hypothetical forgetting function produced by averaging exponential decay functions with decay constants drawn from a Weibull distribution. The solid curve represents the best fitting power function, and the dashed curve represents the best fitting exponential.

items were associated with different lags had minimal effect. The most important factor was an extremely variable distribution of τ values.

It may be possible to find other distributions of λ (or τ) that will generate power forgetting functions from individual items that decay according to the exponential. Indeed, a bimodal distribution can probably do the job with less variability. Nevertheless, extreme variability in decay constants is needed. That kind of variability does not appear to exist across subjects, at least not for the subjects studied here. If one is willing to assume extreme variability in the forgetting rates of individual items (perhaps distributed according to the Weibull), however, then exponential forgetting at the level of items cannot be ruled out.

Do individual items have different rates of decay? Almost certainly. However, as indicated above, that fact by itself would not lead one to predict that an aggregate power function will result. To take a concrete example, if values of τ are drawn from an exponential distribution with a mean and standard deviation of 21 min, the resulting aggregate function is still much better fit by the exponential than by the power function. This occurs even though 5% of these simulated items have a half-life of less than 45 sec and another 5% have a half-life of greater than 43 min.

How could one ever test whether or not individual items decay exponentially at extremely different rates? One way might be to compare the form of forgetting following lists of heterogenous items (e.g., common words) with the form of forgetting following lists of more homogeneous items (e.g., nonsense syllables). Presumably, the items in the more homogeneous list would be associated with less variable decay constants. As such, the aggregate function should be correspondingly closer to an exponential. A direct comparison of that kind has not been performed, although Wickelgren (1970) once argued that forgetting proceeds according to an exponential, on the basis of experiments in which the to-be-remembered material consisted of digits (which are presumably more homogeneous than faces and words). Similarly, the well-known forgetting function reported by Peterson and Peterson (1959, Figure 3), which involved recall of consonant syllables over intervals ranging up to 18 sec, is better fit by the exponential than by the power function. On the other hand, six additional forgetting functions were also reported in Table 1 of that article. Of those, five were better fit by the power function (in the form of Equation 3), and the sixth was a virtual tie, slightly favoring the exponential (99.9% vs. 99.4% of the variance accounted for). The mean percentage data variance accounted for by the exponential for the six functions shown in Table 1 of that article was 96.3%. The corresponding value for the power function was 98.9%. Most of those forgetting functions (unlike those that represent the retention of meaningful stimuli) were approaching the floor. Thus, averaging artifacts of the kind reported by R. B. Anderson and Tweney (1997) may account for the apparent slight superiority of the power function in describing Peterson and Peterson's data. Whether the exponential would exhibit a clear advantage

using geometric averaging is unclear. Nevertheless, if this question is pursued, researchers would probably do well to heed the recommendation made by R. B. Anderson and Tweney (i.e., use geometric instead of arithmetic averaging).

Theoretical Considerations

Although not everyone would agree, we take it for granted that, in the absence of averaging artifacts, the mathematical form of behavioral data is theoretically significant. The theoretical significance of the exponential form of recall latency distributions has long been recognized (see, e.g., McGill, 1963). In this section, we consider two general accounts of the theoretical significance of the mathematical form of forgetting.

Environmental adaptation. J. R. Anderson and Schooler (1991) argued that memory processes, including the dynamics of forgetting, reflect an adaptive response to environmental demands. In everyday interaction with the environment, individuals encounter a variety of events that place a demand on memory to access stored information. For example, encountering the word *fire* in a newspaper headline places a demand on the memory system to access that word's meaning and, perhaps, some associated knowledge. Because some words occur frequently in the environment, their corresponding memories are very likely to be needed during the course of a day. By contrast, the memories of words that occur infrequently are unlikely to be needed any time soon. The adaptive memory system, it is argued, calibrates the accessibility of individual memory traces according to how likely they are to be needed at a particular time (a construct referred to as "need odds"). The less needed a memory is, the less accessible its trace will be.

Empirically, the odds that a memory will be needed in the natural environment (e.g., the odds that the word *fire* will appear again in a headline) declines as a power function of time since last use. Thus, the optimal memory system should apply the same rule to memorized list items and make them less accessible as a power function of time as well. Had the power law of forgetting actually been an artifact of averaging over subjects, and had individual subject forgetting functions instead been exponential, this theory would have been disconfirmed. As it stands, the theory remains viable (unless the power law is shown to be an artifact of averaging over items).

Consolidation. As noted by Wickelgren (1972, 1974), reconciling exponential decay with the idea that memories consolidate over time would be difficult indeed. The exponential implies that the proportional rate of decay is independent of how much time has elapsed. Given exponential forgetting, if a trace loses 50% of its strength in the first hour, it will lose 50% of its remaining strength in the next hour. That is to say, its resistance to decay will have remained constant from the first hour to the second hour.

The same idea can be expressed in another way in order to illustrate how exponential forgetting implies the absence of consolidation. Imagine that after a list of items has been

learned, all of the list memory traces have a momentary probability of failure equal to *f*. That is, in any given second of time, the probability that the trace will "fail" (i.e., be forgotten) is *f*. Owing to chance alone, the item may fail in the first second or it may not fail until thousands of seconds have elapsed. If the probability of failure does not change with time (i.e., if the traces do not become more resistant to decay with time), the forgetting function will be exponential in form. If the traces do become more or less resistant to decay over time, the forgetting function will be something other than exponential in form. Indeed, the power function, logarithmic function, and exponential power function, all of which have been shown to describe the course of forgetting better than most other functions (Rubin & Wenzel, 1996), all imply that the probability of failure *decreases* with time, which is to say that they are all consistent with the notion of consolidation.

Although it is possible that individual items decay with constant probability, the extant data are easier to reconcile with the idea that resistance to decay increases with the passage of time, as Wickelgren once argued. Such a conclusion fits with independent evidence suggesting a role for long-term consolidation processes (e.g., Squire, 1987; Squire, Slater, & Chace, 1975). Nevertheless, more work on the form of forgetting is needed before one can confidently rule out the possibility of exponentially decaying items. Also, although the issue has been dealt with in detail elsewhere, it should be acknowledged that the rejection of any mathematical form depends on the assumption of an appropriate measurement scale (e.g., Loftus, 1985; Wixted, 1990). Nevertheless, for the moment, the possibility that the power law of forgetting results from averaging individual subject exponential forgetting functions can be safely ruled out.

REFERENCES

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, **2**, 396-408.

Anderson, R. B., & Tweney, R. D. (1997). Artifactual power curves in forgetting. *Memory & Cognition*, **25**, 724-730.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College Press, Columbia University. (Original work published 1885; reprint of translation published by Dover, New York, 1964)

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, **53**, 134-140.

Indow, T. (1993). *Retention curves of artificial and natural memory: Tight and soft models* (Tech. Rep. No. MBS 93-11). Irvine: University of California, Irvine, Institute for Mathematical Behavioral Sciences.

Johnson, N. L., & Kotz, S. (1970). *Continuous univariate distributions I: Distributions in statistics*. New York: Wiley.

Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 397-406.

Logan, G. D. (1995). The Weibull distribution, the power law, and the instance theory of automaticity. *Psychological Review*, **102**, 751-756.

McGill, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 309-360). New York: Wiley.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, **58**, 193-198.

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, **22**, 511-524.

Rubin, D. C. (1982). On the retention function for autobiographical memory. *Journal of Verbal Learning & Verbal Behavior*, **21**, 21-38.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of forgetting. *Psychological Review*, **103**, 734-760.

Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin*, **49**, 263-269.

Simon, H. A. (1992). What is an "explanation" of behavior? *Psychological Science*, **3**, 150-161.

Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.

Squire, L. R., Slater, P. C., & Chace, P. M. (1975). Retrograde amnesia: Temporal gradient in very long-term memory following electroconvulsive therapy. *Science*, **187**, 77-79.

White, K. G. (1985). Characteristics of forgetting functions in delayed matching-to-sample. *Journal of the Experimental Analysis of Behavior*, **44**, 15-34.

Wickelgren, W. A. (1970). Time, interference, and rate of presentation in short-term recognition memory for items. *Journal of Mathematical Psychology*, **7**, 219-235.

Wickelgren, W. A. (1972). Trace resistance and decay of long-term memory. *Journal of Mathematical Psychology*, **9**, 418-455.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition*, **2**, 775-780.

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., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1024-1039.