Memory & Cognition
1997, 25 (5), 724-730

Notes and Comment

Artifactual power curves in forgetting

RICHARD B. ANDERSON and RYAN D. TWENEY
Bowling Green State University
Bowling Green, Ohio

Recent studies of the mathematical relationship between time and forgetting suggest that it is a power function rather than an exponential function, a finding that has important theoretical consequences. Through computational analysis and reanalyses of published data, we demonstrate that arithmetic averaging of exponential curves can produce an artifactual power curve, particularly when there are large and systematic differences among the slopes of the component curves. A series of simulations showed that the amount of power artifact is small when the slopes of the component curves are normally or rectangularly distributed and when the performance measure is noise free. However, the simulations also showed that the artifact can be quite large, depending on the shape of the noise distribution and restrictions in the performance range. We conclude that claims concerning the form of memory functions should consider whether the data are likely to contain artifact caused by averaging or by the presence of range-restricted noise.

Cognitive and behavioral scientists have often directed their efforts toward describing the mathematical relationships among psychological variables. Such descriptions are integral to cognitive theories in a variety of domains, and include Marr's (1982) account of visual edge detection, exponential descriptions of forgetting (Wickelgren, 1970), power law descriptions of sensation (Stevens, 1971), and power law accounts of skill acquisition (Logan, 1988; Neves & J. R. Anderson, 1981; Newell & Rosenbloom, 1981). For some time, the power function has been accepted as a correct description of skill acquisition (see, e.g., Newell & Rosenbloom, 1981; but see Heathcote & Mewhort, 1995, for a contrasting view). In addition, several recent studies appear to have extended the power law to memory performance, showing that recall and recognition decline as a power function of time (e.g., J. R. Anderson & Schooler, 1991; R. B. Anderson, Tweney, Rivardo, & Duncan, in press; Rubin, 1982; Wixted & Ebbesen, 1991). In the present paper, we revisit an issue, raised by Estes (1956) and others, concerning the possibility that performance curves might posseses artifactual characteristics resulting from particular data-analytic methods. Specifically, we explore the conditions under which a power curve might result from the arithmetic averaging of exponential curves or from the presence of range-restricted noise in the performance measure.

The authors thank Sean Duncan for his assistance in coding the data and running the simulations. We also thank David Rubin, John Wixted, Geoffrey Loftus, and an anonymous reviewer for helpful suggestions on earlier drafts of the manuscript. Correspondence concerning this article should be addressed to R. B. Anderson, Department of Psychology, Bowling Green State University, Bowling Green, OH 43403 (e-mail: randers@trapper.bgsu.edu or tweney@bgnet.bgsu.edu).

EXPONENTIAL AND POWER FUNCTIONS

Functional descriptions of forgetting often concern the relationship between time and a performance measure such as recall or recognition. Figure 1 shows performance as a power function and as an exponential function of time. Equation 1 gives the mathematical form of the exponential function,

$$P = Ae^{-Bt}, \tag{1}$$

where P is the performance measure, t is time, e is the base of natural logs, and A and B are parameters. Like all exponential curves, the one shown in Figure 1 has the following property. Given an arbitrary time interval k, the ratio of performance at time t, (P_t), to performance at a previous time, (P_{t-k}), is equal to the ratio of performance at time $t + k$, (P_{t+k}), to performance at time $t(P_t)$:

$$(P_t)/(P_{t-k}) = (P_{t+k})/(P_t). \tag{2}$$

In the continuous case, the derivative with respect to time is $P' = -ABe^{-Bt} = -BP$, since the derivative of an exponential function is itself an exponential. Thus, according to an exponential description of forgetting, the memory store loses a constant proportion of its contents over each fixed time interval. In Figure 1, there is a 40% loss from Time 1 ($P = 60$) to Time 2 ($P = 36$) and from Time 2 to Time 3 ($P = 21.6$). The exponential forgetting function has natural appeal, because many processes in nature (e.g., radioactive decay) manifest the same exponential property.

Figure 1 also shows performance as a hypothetical power function of time,

$$P = At^{-B}. \tag{3}$$

In contrast to exponential functions, power functions have the following property:

$$(P_t)/(P_{t-k}) < (P_{t+k})/(P_t). \tag{4}$$

Here, the derivative with respect to time is $P' = -ABt^{-B-1} = -APt^{-1}$. That is, the rate of change declines with time, rather than remaining constant: For the power function in Figure 1, there is a 40% loss from Time 1 to Time 2, but only a 26% loss from Time 2 to Time 3.

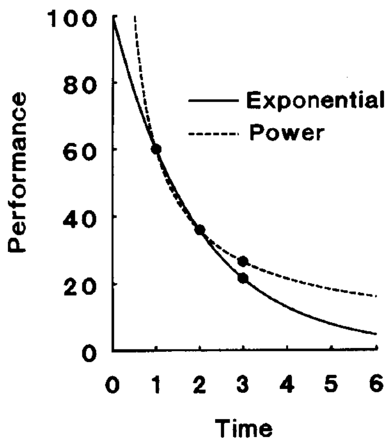

Figure 1. Exponential and power curves showing hypothetical relationships between time and memory performance. Markers show shared and nonshared points for the two curves.

Exponential Forgetting Versus Power-Law Forgetting

Though forgetting appears to follow a power law, there is evidence suggesting that individual memory traces decay exponentially, and that the power law describes forgetting at the aggregate level only. According to Jost's law, the decay rate for a memory trace depends on the age of the trace: Given two memories of equal strength, the younger memory decays more rapidly than does the older one (Jost, 1897, as cited in McGeoch, 1942, p. 140). Simon (1966) noted that Jost's law is incompatible with exponential forgetting but is compatible with a function wherein memories lose a decreasing proportion of their strength over time (as in Equation 4). Simon argued that such nonexponential forgetting could result from the summation of exponential functions. He reasoned that a complex memory trace might have multiple components, each with its own exponential decay function. Consequently, the decay function for the entire memory representation should equal the sum of the component functions, which, according to Simon, must have the property described in Equation 4. As previously noted, power functions in memory are characterized by a loss proportion that decreases over time. Thus, although Simon's account does not specifically predict a power law, it does imply that differential forgetting rates across items could enhance the fit of the aggregate curve to a power function.

Simon's (1966) argument has an important methodological implication for evaluating power law theories. As had been noted by Sidman (1952) and further developed by Bakan (1954) and Estes (1956), because the summation of exponentials can yield a nonexponential function, the arithmetic averaging of data across trials, subjects, or other variables may distort the data and may therefore bias conclusions about the form of forgetting (The arguments of Bakan and Estes are summarized in Appendix A). If the bias favors a power function, then it may cast doubt on theoretical accounts of mental processes and structures implicated in power law learning and forgetting. The following sections extend this point by exploring the conditions under which arithmetic averaging of exponentials can produce a function that not only is nonexponential, but also has an artifactually high goodness-of-fit to a power function.

EXAMPLES OF ARTIFACT DUE TO AVERAGING

Table 1 contains examples of how arithmetic averaging of exponential curves yields an aggregate curve that not only deviates from an exponential function, but bears resemblance to a power function. Because the geometric average of exponentials is itself always an exponential (see Estes, 1956), we used the geometric average as a standard for measuring the distortion caused by arithmetic averaging. Curve fitting for the data in Table 1 was done by log transforming performance and time (for the power fits) or just performance (for the exponential fits), and then fitting the transformed data to a linear function by using the least squares method. In the case of aggregate curves, averaging was done prior to transformation. In Example 1, the individual curves *J* and *K* are perfectly exponential. Hence, for each curve, R^2 is 1.0 for the exponential fit but only .97 for the power fit. Moreover, the R^2 values (equal to 1.0) for the geometric average of curves *J* and *K* are equal to the R^2 values for the individual curves. In contrast, arithmetic averaging produces a different pattern of fits: The exponential fit drops from 1.0 to .84, but R^2_{pow} is .93. Thus, the goodness of the power fit relative

Table 1
Power Fits (R^2_{pow}) and Exponential Fits (R^2_{exp}) for Arbitrary Exponential Curves That Have Been Geometrically and Arithmetically Averaged

Curve	Time 1	Time 2	Time 3	R^2_{pow}	R^2_{exp}	$(R^2_{pow} - R^2_{exp})$
		Example 1				
J	95.00	90.25	85.73	.97	1.00	−.02
K	75.00	3.75	.18	.97	1.00	−.02
Mean						
Geometric	84.41	18.39	4.01	.97	1.00	−.02
Arithmetic	85.00	47.00	42.96	.93	.84	.09
		Example 2				
J	65.00	48.75	36.56	.97	1.00	−.02
K	75.00	3.75	.18	.97	1.00	−.02
Mean						
Geometric	69.82	13.52	2.62	.97	1.00	−.02
Arithmetic	70.00	26.25	18.37	.98	.93	.05
		Example 3				
J	80.00	48.00	28.80	.97	1.00	−.02
K	70.00	52.50	39.37	.97	1.00	−.02
Mean						
Geometric	74.83	50.20	33.67	.97	1.00	−.02
Arithmetic	75.00	50.25	34.08	.98	.99	−.01

Note—In each example, curves *J* and *K* are exponential functions with arbitrary parameters. Both performance and time are measured in arbitrary units. Although the table shows performance values rounded to the nearest two decimal places, the functions were fit to performance values specified to a precision of eight significant digits.

to that of the exponential fit ($R^2_{pow} - R^2_{exp}$) is greater for arithmetically averaged curves (+.09) than for perfect exponentials or for their geometric average (−.02). Examples 2 and 3, which also contain exponential component curves, show results similar to those of Example 1. It is worth noting that arithmetic averaging produces a positive $R^2_{pow} - R^2_{exp}$ only when the component curves have sufficiently different shapes (in an extreme case of two component curves having identical shapes, arithmetic averaging simply reproduces the component curves). Thus, in Example 3, two very similar curves yield an aggregate curve with an $R^2_{pow} - R^2_{exp}$ that is negative, regardless of how the averaging is performed. But even here, R^2_{pow} is greater following arithmetic averaging than following geometric averaging. We were unable to find a simple analytic solution for the exact form of the composite curve (see Appendix B). Nonetheless, the examples show that distortions are possible.

Rubin (1982) argued that forgetting closely approximates a power function, even when the data are not averaged over subjects or over experimental conditions. However, we think that his data still allow the possibility that arithmetic averaging exaggerates $R^2_{pow} - R^2_{exp}$. In one study (Rubin, 1982, Experiment 3), subjects used the words *paper*, *plant*, *wine*, *hospital*, and *fire* as cues for retrieving autobiographical memories. The subjects subsequently estimated a creation date for each memory retrieved, allowing Rubin to compute the number of retrieved memories as a function of autobiographic time. He constructed five separate curves representing responses to each of the five cues. The curves were then fit to power functions, yielding an R^2_{pow} that ranged from .91 to .99. Clearly, the data were well described by power functions. Thus, the goodness of the power fits could not have been an artifact of averaging across experimental materials (in this case, cue words). However, it is not clear that Rubin's power fits were better than exponential fits (none were provided in Experiment 3). In addition, it is possible that across-subject aggregation distorted the curves. In a similar experiment (Rubin, 1982, Experiment 4), data were arithmetically combined across cue words but were kept separate for each of 7 subjects. The values for R^2_{pow} ranged from .55 to .96, with all but one greater than .86. Clearly some subjects in Experiment 4 produced data that closely matched a power function. But again, it is not clear whether those power fits were higher or lower than exponential fits (which were not reported), and here it may still be the case that across-cue aggregation distorted the curves.

Function distortion can result from averaging power curves, as well as from averaging exponential curves. Table 2 shows that the geometric average of two power curves is itself a power curve, and that the resulting power fit is .02 higher than the exponential fit. But the table also shows that arithmetic averaging of power curves with different slopes can yield a power fit that is artifactually high relative to the exponential fit. Thus, arithmetic averaging can exaggerate the relative goodness of the power fit, even when the component curves are themselves power curves.

Table 2
Power Fits (R^2_{pow}) and Exponential Fits (R^2_{exp}) for Arbitrary Power Curves That Have Been Geometrically and Arithmetically Averaged

Curve	Time 1	Time 2	Time 3	Least Squares Fits R^2_{pow}	R^2_{exp}	$(R^2_{pow} - R^2_{exp})$
		Example 1				
J	90.00	81.11	76.32	1.00	.97	.02
K	15.00	8.03	5.58	1.00	.97	.02
Mean						
Geometric	36.74	25.53	20.63	1.00	.97	.02
Arithmetic	52.50	44.57	40.95	.99	.96	.03
		Example 2				
J	90.00	85.14	82.42	1.00	.97	.02
K	25.00	12.94	8.80	1.00	.97	.02
Mean						
Geometric	47.43	33.19	26.93	1.00	.97	.02
Arithmetic	57.50	49.04	45.61	.99	.95	.04

Note—In each example, curves *J* and *K* are power functions with arbitrary parameters. Both performance and time are measured in arbitrary units. Although the table shows performance values rounded to the nearest two decimal places, the functions were fit to performance values specified to a precision of eight significant digits.

It is not always possible to determine whether a published data set has been distorted via arithmetic averaging. If data have been averaged prior to being linearized (i.e., prior to being log transformed), distortion may have occurred, but if they have been linearized prior to being averaged (e.g., J. R. Anderson & Schooler, 1991; Neves & J. R. Anderson, 1981; Newell & Rosenbloom, 1981), no distortion will have occurred (the arithmetic average of two straight lines is a straight line). Even when data have been linearized prior to across-subject averaging, there has usually been some arithmetic aggregation (e.g., averaging across trial position) prior to linearization. The widespread use of arithmetic averaging at various stages in the data aggregation process suggests that power fit exaggeration may be present in many studies.

The practical significance of the power artifact hypothesis depends, of course, on whether the effect can be detected in actual experimental data. We addressed this question directly by reanalyzing published data sets for which the Brown–Peterson (Brown, 1958; Peterson & Peterson, 1959) method had been used to examine the role of proactive (intertrial) interference in short-term forgetting. These data suit our purpose because a separate retention curve was reported for each trial position. Consequently, we were able to use two methods of averaging across trial position, arithmetic and geometric averaging, to determine whether the arithmetic method would exaggerate the relative goodness of the power fit. For example, Keppel and Underwood (1962, Experiment 1) were interested in the buildup of proactive interference over the course of an experiment. Consequently, they aggregated the data across subjects and constructed separate retention curves for Trial 1, Trial 2, and Trial 3. (Note that the assignment of retention interval to trial number was counterbalanced across subjects.) In our reanalysis of the data, we computed both the arithmetic and the geometric averages of Trials 1, 2, and 3, at each retention interval. The quantity $R^2_{\text{pow}} - R^2_{\text{exp}}$ was then computed for both sets of means. Fit exaggeration was assessed with the method illustrated in Table 1. We expected to find only moderate exaggeration of the power fits, because we could only examine a portion of the potential artifact—that portion due to averaging across trial position.

Table 3 shows the effect of geometric and arithmetic averaging on the quantity $R^2_{\text{pow}} - R^2_{\text{exp}}$ for the data sets. As expected, $R^2_{\text{pow}} - R^2_{\text{exp}}$ was generally higher for arithmetically than for geometrically averaged data; only Noyd's data (as cited in Bennett, 1975) produced a contrary pattern of fits. Moreover, in the case of Keppel and Underwood's data (1962, Experiment 1), the choice of averaging method determined which function yielded the better fit: Arithmetic averaging yielded a power fit that was higher than the exponential fit, whereas geometric averaging produced the opposite pattern. The fact that most of the data sets favored a power function, irrespective of the averaging method, does not establish that forgetting follows a power law. As noted earlier, the data sets contained performance values that had already been averaged across subjects and other conditions, and consequently they may have contained artifact that could not be examined in the present study. However, the fact that power fit exaggeration replicates across studies indicates that the effect has measurable consequences for empirical research.

Boundary Conditions for the Averaging Artifact

Under what conditions is it likely that power fit exaggeration has occurred? We noted earlier (and in Appendix A) that no general analytic solution exists. That is, there is no unique set of parameters that will permit a power function to precisely equal the sum of two or more exponentials. Bakan (1954) and Estes (1956) explored the conditions under which an aggregate function would, in general, produce a function of the same form. By using series expansions, Estes showed that a sum of exponentials cannot be equal to an exponential itself (see Appendix B).

Table 3
Power Fits (R^2_{pow}) and Exponential Fits (R^2_{exp}) for Six Data Sets Averaged Geometrically and Arithmetically Across Trials

Data Set	Averaging Method	Least Squares Fits R^2_{pow}	R^2_{exp}	$(R^2_{\text{pow}} - R^2_{\text{exp}})$
Keppel & Underwood (1962)				
Experiment 1	geometric	.97	.99	−.01
	arithmetic	.99	.92	.07
Experiment 2	geometric	.99	.93	.06
	arithmetic	.99	.92	.07
Loess (1964)				
Experiment 1	geometric	.91	.98	−.06
	arithmetic	.94	.99	−.04
Experiment 2	geometric	.99	.96	.02
	arithmetic	.99	.96	.03
Noyd (see Bennett, 1975)	geometric	.95	.84	.11
	arithmetic	.96	.85	.11

Note—Fits are rounded to two decimal places.

However, his solution also does not establish that a power function represents a better fit than the exponential. Newell and Rosenbloom (1981) used Laplace transforms, in an attempt to find the aggregated form of a series of exponentials. They also failed to find a general solution, although they did succeed in establishing that a rectangular distribution of beta weights among a set of summed exponentials would be likely to yield a curve that closely resembled a power curve. The argument is summarized in Appendix C, which also describes our attempts, with the use of Laplace transforms, to find conditions under which aggregated exponentials resemble power curves. We confirmed Newell and Rosenbloom's point concerning rectangular distributions of beta coefficients, and we also found that bimodal distributions had the same effect.

Having confirmed that summed exponentials can approximate a power function, we examined the relationship between particular parameter values of the exponential functions and the magnitude of the power curve artifact. Specifically, we conducted a computational analysis of the effect of averaging two exponential curves on the amount of power artifact. We generated pairs of exponential curves whose parameters differed by varying amounts and calculated $R^2_{\text{pow}} - R^2_{\text{exp}}$ for the arithmetic average of the curves. Figure 2 shows the results. The beta parameter, B, varies from .01 to .99, in increments of .01 for B_1 and in increments of .14 for B_2. The alpha parameters, A_1 and A_2, both equal 1.0. Simulations not reported here indicated that varying A_1 and A_2 had little impact on the results. The lowest point on each curve represents a condition in which B_1 and B_2 are equal, and in which the best fitting exponential function for the averaged curves accounts for 10% more of the variance than does the best fitting power function. For moderate differences between B_1 and B_2, $R^2_{\text{pow}} - R^2_{\text{exp}}$ increases as a function of $B_1 - B_2$. However, the function departs from monotonicity when the beta difference is extreme.

The computational analysis showed that averaging exponential curves sometimes produces power curves. In addition, the reanalysis of published data sets shows that moderate power fit exaggeration can occur in real-world environments characterized by noisy data. However, it is not clear whether highly exaggerated power fits—ones in which exponential functions are artifactually transformed into power functions—can occur only with large systematic differences in B (as is suggested by the computational analysis), nor whether extreme exaggeration may occur through random sampling from a single population of exponentials. In addition, it is not clear whether extreme exaggeration occurs when P (the performance measure) contains noise, as would be the case in an actual experiment. To examine these issues, we conducted a series of simulated experiments.

An exponential forgetting curve was generated for each of 20 simulated subjects. For each curve, the y intercept (A_i) and the slope (B_i) were randomly determined by fixing A_i at 1.0 and B_i at .2, and adding normally distributed noise (with a mean of 0 and a standard deviation of .1) to each of the two parameters. Performance (P_i) was

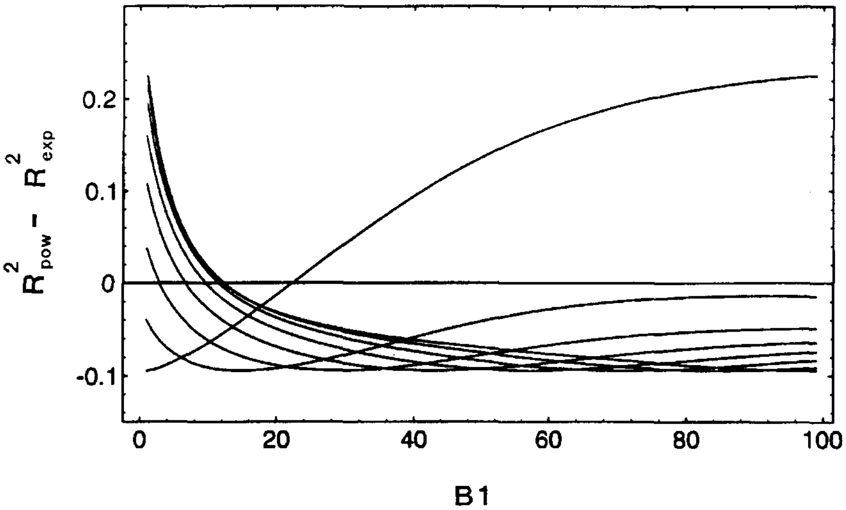

Figure 2. $R^2_{pow} - R^2_{exp}$ as a function of B_1 and B_2 for arithmetically averaged exponential curves. B_1 varies from .01 to .99 by .01. B_2 (not labeled) varies from .01 (bottom left curve) to .99 (top left curve) by .14. $A_1 = A_2 = 1.0$ for all curves.

calculated for each of 20 retention intervals ranging from 1 to 20 and was then arithmetically averaged across all 20 subjects. The entire procedure was repeated 100 times, yielding data from 100 simulated experiments. Each aggregate curve for each experiment was fit to exponential and power functions according to the previously described method (see Table 1). Contrary to our expectations, R^2_{pow} was greater than R^2_{exp} in only 5% of the experiments. Altering the simulations by choosing different retention intervals (e.g., 2, 4, 8, 16, and 32) did not appreciably change this result, nor did varying the amount of added noise or the starting values of A_i and B_i.

Using Laplace transformations, Newell and Rosenbloom (1981) argued that rectangular distributions of exponentials can result in aggregate power curves (see Appendix C). Thus, we conducted another set of 100 simulations using the previously described method, but with rectangularly rather than normally distributed A and B parameters. There was no change in the results; R^2_{pow} exceeded R^2_{exp} in only 5% of the cases. These data suggest that extreme power fit exaggerations do not readily occur when alpha and beta differences arise from random processes, such as the random sampling of subjects. However, we have not tried all possible noise distributions, and therefore we cannot rule out random sampling as a major source of artifactual power functions.

Taken together, the simulations, computational analysis, and reanalysis of published data suggest that averaging artifact is problematic only when the decay parameters of the component curves differ greatly and systematically. In the experiment simulations, data were averaged over decay parameters that were randomly distributed, as might be the case in across-subject averaging. However, the reanalysis of Keppel and Underwood's (1962) Experiment 1 data (Table 3 in the present paper) involved averaging across trial position, a factor known to have large systematic effects on the forgetting rate (e.g., Keppel & Underwood, 1962). Thus, when averaging artifact occurs, it is more likely to involve averaging across systematic factors such as trial position, rather than random factors such as subject.

Range-Restricted Noise: Another Source of Artifact

In an actual experiment, P would itself contain some random error. Thus, we ran four sets of 200 simulations, using the same method described previously but with the following addition: Besides adding normally distributed noise to A_i and B_i (simulating the effect of random sampling of subjects), we added similar noise to each subject's performance (P_i) at each retention interval. In practice, P_i can never be negative, so we truncated P_i from the distribution whenever the noise would have driven it below zero. The result was a set of 20 curves (in each experiment) with noise added to each point and with the amount of truncation increasing as retention time increased. In essence, the noise introduced a floor effect.

For each set of simulated experiments, the noise distribution (before truncation) for P had a mean of 0.0 and a standard deviation of 0.05, 0.10, 0.15, or 0.20. Note that truncation causes the variance of resulting signal+noise distribution to decline as the length of the retention interval increases. The simulation results were striking: R^2_{pow} exceeded R^2_{exp} in up to 97% of the cases, depending on the noise level. Table 4 summarizes the data from all four sets of experiments. The extreme fit exaggeration was due to the floor effect introduced by the noise, rather than to the noise itself: An additional set of simulations incorporated the unrealistic assumption of nontruncated distributions (performance was allowed to fall below zero) and yielded few spurious power functions.

Though the term *floor effect* is appropriate, the empirical forgetting function can be distorted even if it does not closely approach zero. The amount of truncation, and

Table 4
Effect of the Standard Deviation of the Performance Noise Distribution on the Amount of Power Fit Exaggeration

Noise *SD*	Mean R^2_{pow}	Mean R^2_{exp}	Mean $(R^2_{\text{pow}} - R^2_{\text{exp}})$	Frequency of $(R^2_{\text{pow}} > R^2_{\text{exp}})$*
.05	.92	.97	−.05	0
.10	.95	.94	.01	123
.15	.95	.91	.04	185
.20	.94	.87	.06	194

Note—R^2_{pow} is the proportion of variance accounted for by the best fitting power function. R^2_{exp} is the proportion of variance accounted for by the best fitting power function. *Two hundred simulated experiments were conducted at each noise level, with each experiment yielding a value for R^2_{pow} that was higher or lower than R^2_{exp}.

thus the amount of function distortion, depends on the existence of particular episodes of extreme forgetting and does not require extreme forgetting at the aggregate level. For example, in one simulated experiment, the aggregate performance curve matches a power function, even though performance remains above 0.18. Overall, the results show that artifactual power functions can readily occur when the performance measure is not noise free.

SUMMARY AND CONCLUSIONS

We have shown that power functions in memory performance may be an artifact of arithmetic averaging or of noisy performance data. Thus, the exponential function may, in some cases, be a better description of forgetting. Our reanalyses of forgetting curves showed that power-fit exaggeration does occur in published studies. Computational analyses showed that arithmetic averaging of exponential curves can produce spurious power functions, and our computer simulations indicated that although extreme function distortion is unlikely when A_i and B_i are randomly distributed (either normally or rectangularly) and the performance measure is otherwise noise free, such artifact is very likely when the performance measure contains noise. It is therefore possible that disparate findings concerning the form of forgetting—either exponential (e.g., Wickelgren, 1974) or power (see, e.g., Wixted & Ebbesen, 1991)—have resulted, in part, from across-study variation in the amount and kind of noise present in the performance measure. In addition, because the noise distorts the memory function by means of a floor effect, it is likely that the amount of distortion will depend on the degree to which performance approaches zero. It should be noted that the exponential function is not the only one susceptible to distortion via arithmetic averaging (Estes, 1956). Thus, an exaggerated power fit does not, by itself, establish the exponential function as the true form. Nonetheless, the presence of an exaggerated power fit should raise questions concerning the true form of the curve.

Although the power law has been ubiquitous in sensation and in skill acquisition and has been offered as a general description of forgetting (J. R. Anderson & Schooler, 1991; R. B. Anderson et al., in press; Rubin, 1982; Wixted & Ebbesen, 1991), the present findings warrant a reevaluation of the power law's status both as a description of behavior and as a constraint on psychological theory. Power law descriptions should only be accepted when there is sufficient assurance that the fits are not the result of artifact.

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., Rivardo, M., & Duncan, S. (in press). Need probability affects retention: A direct demonstration. *Memory & Cognition*.

Bakan, D. (1954). A generalization of Sidman's results on group and individual functions and a criterion. *Psychological Bulletin*, **51**, 63-64.

Bennett, R. W. (1975). Proactive interference in short-term memory: Fundamental forgetting processes. *Journal of Verbal Learning & Verbal Behavior*, **14**, 123-144.

Boas, M. L. (1983). *Mathematical methods in the physical sciences* (2nd ed.). New York: Wiley.

Brown, J. A. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, **10**, 12-21.

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

Heathcote, A., & Mewhort, D. J. K. (1995, November). *The laws of practice*. Poster presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning & Verbal Behavior*, **1**, 153-161.

Loess, H. (1964). Proactive inhibition in short-term memory. *Journal of Verbal Learning & Verbal Behavior*, **3**, 362-368.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492-527.

Marr, D. (1982). *Vision*. New York: W. H. Freeman.

McGeoch, J. A. (1942). *Psychology of human learning: An introduction*. New York: Corigmans, Green.

Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 57-84). Hillsdale, NJ: Erlbaum.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.

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

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

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

Simon, H. A. (1966). A note on Jost's law and exponential forgetting. *Psychometrika*, **31**, 505-506.

Stevens, S. S. (1971). Neural events and the psychophysical law. *Science*, **170**, 1043-1050.

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

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, **2**, 409-415.

APPENDIX A
Solutions of Aggregated Exponentials

We noted in the text that there is no unique value of A_3 and B_3 such that the following equation is satisfied:

$$A_1 e^{-B_1 t} + A_2 e^{-B_2 t} = A_3 e^{-B_3 t}.$$

However, first letting $A_1 = A_2 = 1.0$, the following expression for B_3 satisfies the equation above:

$$B_3 = (1/t)\,[B_1 t + B_2 t - \ln\,(e^{B_1 t} + e^{B_2 t})].$$

Note, however, that the expression is itself a function of t; there is not one set of parameters that satisfies the equation across different values of t.

APPENDIX B
Series Expansions of Aggregated Exponentials

Bakan (1954) first suggested a criterion for deciding whether or not a given function would result in a function of the same form when averaged. He expanded a quadratic equation using a Maclaurin series around $t = 0$ and showed that the resulting series for an averaged set of quadratics depended only on the averages of the parameters of the original equations. In other words, the coefficients of the averaged equations were simply the averages of the coefficients of the original equations; hence the averaging process did not change the shape of the function. The same was not true for an exponential equation. Bakan found that the coefficients for a set of averaged exponentials included not just the averaged coefficients of the original equations but higher order terms consisting of products of the coefficients as well. Hence the averaging process returned a function whose form differs from that of the original exponentials.

Power functions cannot be analyzed by using Bakan's (1954) approach, since they are not defined at $t = 0$. Thus, Estes (1956) generalized Bakan's result by using Taylor's series expansions, which can be taken around any number, not just 0. He refined Bakan's criterion for "averageability" as follows: Given a function $p = f(t, A, B, C, \ldots)$, where $A, B, C, \ldots$ are parameters, the averaged function $p = f(t, \overline{A}, \overline{B}, \overline{C}, \ldots)$, where $\overline{A}, \overline{B}, \overline{C}, \ldots$ are the average values of the parameters, will have the same shape as the original functions, provided that all of the second and higher order derivatives of the function with respect to the parameters are zero. Thus a set of quadratic equations $p_i = A_i + B_i t^2$ can be averaged, because the second derivatives with respect to A_i and B_i are all zero. An exponential cannot be averaged, however. If $p = A_i e^{-B_i t}$ represents an exponential, then the derivative of p with respect to B_i is given by $p' = -A_i t e^{-B_i t}$, the second derivative is given by $p'' = A_i t^2 e^{-B_i t}$, the third by $p''' = -A_i t^3 e^{-B_i t}$, and so on. Since the derivatives do not vanish, the Taylor expansion includes terms with coefficients other than the averaged coefficients of the original equations, and the resulting curve is therefore not exponential.

APPENDIX C
Laplace Transforms

Newell and Rosenbloom (1981) first noted the applicability of Laplace transforms to the issue of the aggregation of functions. Consider, first, the general logic of transforms, exemplified by the familiar use of Fourier series in the analysis of time-dependent data. A Fourier series in effect provides a series of sine and cosine functions that, when summed, approximate a periodic function. By adding sufficiently many terms to the series, one can approximate to an arbitrary degree of error any periodic function meeting certain criteria of analyticity. In the continuous case, Fourier series can be generalized to Fourier transforms, in which an infinite number of sine and cosine terms are used and the integral from 0 to infinity replaces the sum from 1 to n of the separate terms. Even nonperiodic functions can then be represented.

A Laplace transform is similar, in that it represents an infinite series of exponential terms that are used to represent an arbitrary function. In particular, if $p(t)$ is a function meeting certain analyticity criteria, its Laplace transform, $L(p)$, is a new function in the variable s, given by

$$L(p) = \int_0^\infty p(t) e^{(-st)} dt.$$

For the case of the exponential function,

$$L(p) = L(Ae^{-Bt}) = \int_0^\infty Ae^{-Bt} e^{(-st)} dt = A/(s+B).$$

For power functions, the result is more complex:

$$L(At^{-B}) = \int_0^\infty At^{-B} e^{(-st)} dt = [A\Gamma(1-B)]s^{-(1-B)},$$

where $\Gamma(B)$ is the gamma function of B, a kind of generalized factorial product defined over the continuous number line (not just integers). Note also that a Laplace transform is a linear operator; $L(Cp) = CL(p)$ and $L(p + r) = L(p) + L(r)$ for any constant C and any functions $p(t)$ and $r(t)$.

A Laplace transform of a function is, in effect, a way of representing the function as an infinite sum of exponentials. For the present argument, this means in particular that any power function can be represented as an infinite sum of exponentials (Boas, 1983). Newell and Rosenbloom (1981) noted that a Laplace transform is like a set of weighted exponentials; they singled out the special case in which the exponent, B, of the power function is 1.0 (i.e., where the power function is a hyperbolic function). In this instance, since $1 - B = 0$, the Laplace transform simplifies to the form $A\Gamma(B)$, which in turn is simply equal to the constant A, since $\Gamma(B) = 1! = 1.0$ for $B = 1$. Because Laplace transforms are linear operators, the Laplace of an average is equal to the average of the Laplaces of the individual functions. We explored the conditions under which the Laplace of an average of exponentials would resemble the Laplace of a power function. We were not able to find a general solution, but we did determine that, as Newell and Rosenbloom suggested, a rectangular distribution of weights among an average of exponentials did in fact produce a curve in the transform space that was similar to the curve for the transform of a single power function. Furthermore, we found an even closer match by using a bimodal distribution of exponentials. For example, we created a bimodal distribution of 10 exponentials with A coefficients all equal to 1.0 and exponents as follows: 0.1, 0.1, 0.1, 0.4, 0.5, 0.6, 0.9, 0.9, 0.9. We then took the Laplace transform of the average of the 10 functions, which gives a series of additive terms of the form $1/(B_i + s)$. Plotting this against successive plots of the transform of a power function for different values of the power function exponent allows an informal judgment of the similarity between the two. In general, the curves are very similar throughout most of the range of s values. This result confirms the generality of our two-term computational solution in the text.

(Manuscript received December 13, 1995;
revision accepted for publication May 30, 1996.)