Toward an explanation of the power law artifact: Insights from response surface analysis

IN JAE MYUNG, CHEONGTAG KIM, and MARK A. PITT
Ohio State University, Columbus, Ohio

The power law ($y = ax^{-b}$) has been shown to provide a good description of data collected in a wide range of fields in psychology. R. B. Anderson and Tweney (1997) suggested that the model's data-fitting success may in part be artifactual, caused by a number of factors, one of which is the use of improper data averaging methods. The present paper follows up on their work and explains causes of the power law artifact. A method for studying the geometric relations among responses generated by mathematical models is introduced that shows the artifact is a result of the combined contributions of three factors: arithmetic averaging of data that are generated from a nonlinear model in the presence of individual differences.

A measure of scientific advancement in psychology is the discovery of lawful, predictable behavior. Such discoveries are particularly satisfying when the operation of the underlying mental process that mediates between stimulus and response can be described using mathematical functions. One of the functions that has enjoyed considerable popularity, especially among cognitive psychologists, is the power function ($y = ax^{-b}$). The power function has been shown to provide an excellent description of human behavior in a variety of fields, such as psychophysics (Stevens, 1971), memory (Wixted & Ebbesen, 1991), skill learning (Newell & Rosenbloom, 1981), and judgment and decision making (Stevenson, 1993). For example, research on human memory has explored the rate at which information is forgotten over time (Ebbinghaus, 1964; Wickens, 1998; see Rubin & Wenzel, 1996, for a review). In a typical experimental setup (Peterson & Peterson, 1959), listeners are presented a list of items (e.g., nonsense syllables) and afterward are asked to recall them at varying time delays. Recall is good immediately after learning, but drops precipitously thereafter before leveling off to a roughly constant rate. This forgetting curve is often best approximated by a power function (R. B. Anderson & Tweney, 1997; Wixted & Ebbesen, 1997). Why does the power function provide such a good description of so many seemingly disparate psychological processes? Is its ubiquity evidence of a general property of a subset of cognitive processes (J. R. Anderson &

Schooler, 1991), or could it be due to some artifact (Estes, 1956)? Concerns of the latter type were recently rekindled in a study on the mathematical form of the forgetting function (R. B. Anderson & Tweney, 1997). These authors were concerned about the superior data-fitting ability of the power function relative to the exponential function. Simulations showed that occurrence of the artifact depended on data averaging, whether arithmetic,

$$y_{arith} = \frac{\sum_{i=1}^m y_i}{m},$$

or geometric,

$$y_{geom} = \exp \left(\frac{\sum_{i=1}^m \ln y_i}{m} \right). \quad (1)$$

Specifically, they showed that data generated by either function were fit best by a power model when the data of simulated subjects were arithmetically averaged, a phenomenon we refer to throughout the paper as the *power law artifact*. Geometrically averaging the data sometimes eliminated the artifact, improving the fit of the exponential model relative to the power model.

However, Wixted and Ebbesen (1997) and R. B. Anderson and Tweney (1997) showed that the artifact cannot be due solely to the use of an improper averaging method, because if it were, geometric averaging should always improve the fit of the exponential model. Wixted and Ebbesen (1997) reanalyzed two data sets from a previous study on the form of the forgetting function (Wixted & Ebbesen, 1991). The type of averaging made very little difference. Not only did the power function fit the arithmetically and geometrically averaged data better than the exponential function, but it did so by virtually the same

This paper is based on a presentation at the thirty-first annual meeting of the Society for Mathematical Psychology in August 1998. This research was partly supported by NIMH Grant MH57472 to I.J.M. and M.A.P. We thank Richard Anderson, Greg Ashby, Lester Krueger, David Rubin, and two anonymous reviewers for providing many invaluable suggestions and feedback on earlier versions of this paper. Correspondence should be addressed to I. J. Myung, Department of Psychology, Ohio State University, 1885 Neil Avenue, Columbus, OH 43210-1222 (e-mail: myung.1@osu.edu).

magnitude using both averaging methods. R. B. Anderson and Tweney (1997; Table 3) found a similar outcome when comparing the effects of the two averaging methods using data from five published experiments. Only on occasion did geometric averaging have a nonnegligible effect on fit. Clearly there is more to the artifact than just averaging.

Further simulations by R. B. Anderson and Tweney (1997) identified a few conditions necessary to observe the artifact, but did not pinpoint its cause. For example, when random noise was added to the parameters of exponential forgetting curves of simulated subjects, no power artifact was obtained, with an exponential function providing the best fit to the data 95% of the time. However, adding noise to the data points of the simulated subjects did yield the artifact.

In the present paper, we provide some closure on the preceding work by shedding new light on the power law artifact. A method for studying the geometric relations among mathematical models is introduced that clearly reveals why the artifact occurs. In short, the method, dubbed *response surface analysis* (RSA; Bates & Watts, 1988), shows that the artifact is a result of the combined contributions of three factors: *arithmetic averaging* of data that are generated from a *nonlinear model* in the presence of *individual differences*. We begin by briefly illustrating the power law artifact in the context of a forgetting experiment. RSA is then described and ways of avoiding the artifact are discussed.

THE POWER LAW ARTIFACT

For simplicity of illustration, we describe the artifact under the clearest case, using one-parameter models for two data points. (For a more comprehensive treatment with two parameter models with many data points, see R. B. Anderson and Tweney, 1997.) The power and exponential models are defined, respectively, as

$$y = (t + 1)^{-\theta}$$

and

$$y = \exp(-\theta t), \quad (2)$$

where y is the response probability (e.g., proportion recall), t is the retention interval, and $\theta (\geq 0)$ is the parameter. Note that for $t = 0$, y is equal to 1 (i.e., perfect recall), regardless of the parameter value.

The power law artifact was explored by generating 10 simulated forgetting curves (each with a different parameter value) and then calculating the arithmetic average of these curves. Each "curve" consisted of just two points, corresponding to two different retention intervals ($t = 1$ and 4). To simulate probabilistic responses in a forgetting experiment, a given curve was constructed by using an exponential or a power function to generate a set of 50 binary values (0 or 1) where the probability of a "1" was equal to the value of y given a particular t and θ value. The proportion of 1s constituted the proportion

recalled at the particular t . By this method, 10 "sample" curves were generated, 1 for each of the 10 different parameter values (θ).¹

The 10 sample curves were *arithmetically* averaged, and then both exponential and power models were fitted to the averaged curve. Fit was measured using mean squared error ($MS_e = \sqrt{\text{sum of square error/number of observations}}$). The data-fitting process was repeated 100 times with data generated from the exponential model and 100 times with data generated from the power model. A similar analysis was conducted for data generated by a random model, where the probability, y , was uniformly distributed over the interval $[0,1]$ for all values of t . The top panel of Table 1 (Arithmetic Averaging) shows the mean MS_e and the percentage of time each model provided the best fit to the data. The data-generating models are listed along the rows, with the model being fitted listed in the columns.

The power law artifact is clearly visible in the data in both measures. The power model always fitted data generated by a power function better than the exponential model, with fits being on average four times better (.027 vs. .103). The power model also fitted the exponential data best 90% of the time, with fits on average being more than twice as good (.024 vs. .053). The prevalence of the artifact is further revealed when the models are fitted to the arithmetic mean of individual data sets generated from the random model. One might expect that both models would fit random data equally poorly, with each providing the best fit approximately half of the time. What is found instead is the same superiority of the power model. As with the first two data sets, the power model provided a better fit, although the fits of both models are not good, and it did so 99% of the time. In the next section we explain why the power law artifact occurs and why there is no exponential law artifact.

Table 1
Mean Square Error Fits of the Power and Exponential Models

Data Generated From:	Model Fitted			
	Power		Exponential	
	MS_e	PT	MS_e	PT
Arithmetic Averaging				
Power model	.027	100%	.103	0%
Exponential model	.024	90%	.053	10%
Random model	.140	99%	.214	1%
Geometric Averaging				
Power model	.020	91%	.079	9%
Exponential model	.096	2%	.021	98%
Random model	.152	97%	.217	3%
Individual Analysis				
Power model	.030	72%	.055	28%
Exponential model	.055	22%	.025	78%
Random model	.189	74%	.206	26%
Stochastic Parameter Approach				
Power model	.003	92%	.008	8%
Exponential model	.009	2%	.002	98%
Random model	.022	53%	.021	47%

Note—Each cell contains the mean square error of estimate ($MS_e = \sqrt{\text{sum of square error/number of observations}}$) and the percent of time (PT) the fitted model provided a superior fit.

RESPONSE SURFACE ANALYSIS

RSA is a method for studying geometric relations among responses generated by a mathematical model, often used in nonlinear regression (Bates & Watts, 1988). For a model with p parameters and n observations, the *response surface* is defined as a p -dimensional surface, formed by all possible response vectors that the model can describe. The response surface is embedded in the n -dimensional *data space*, which is the set of all possible response vectors that could be generated independently of any model. The response surface is a hyperplane for a linear model but may be curved when the model is nonlinear. Many statistical concepts, including those of least squares estimation, have informative geometric interpretations in terms of response surfaces. In particular, the effects of averaging on model fit can be seen quite clearly when averaged and individual data are plotted in the space of response surfaces.

Given two retention intervals, t_1 and t_2 , the response surface is a curve or line in a two-dimensional data space composed of (y_{t_1}, y_{t_2}) , the values of which are produced by varying the full range of the parameter θ . In essence, a model is represented graphically as a plot (i.e., y_{t_1} vs. y_{t_2}) in the data space.

To understand how a model is captured in the data space, consider the three forgetting curves of the power model in the left panel of Figure 1, each of which fits 1 subject's data at Times 1 and 4. Note that the model parameter is different in each case. A response curve is cre-

ated by plotting the y values at t_1 against the corresponding y values at t_2 for the full range of the parameter. For example, for Subject 1, the y values at $t_1 = 1$ and $t_2 = 4$ are 0.47 and 0.17, respectively. These two values are then represented as a single point (0.47, 0.17) on the (y_{t_1}, y_{t_2}) plane in the right panel of Figure 1. Data points corresponding to the other 2 subjects are also indicated. The (y_{t_1}, y_{t_2}) points obtained by varying the full range of the parameter form a continuous curve, which is called the response curve of the power model.

The equation that describes this relationship can be derived analytically. To show, first we note

$$y_{t_1} = (t_1 + 1)^{-\theta} \rightarrow \ln y_{t_1} = -\theta \ln(t_1 + 1)$$

$$\rightarrow \theta = -\frac{\ln y_{t_1}}{\ln(t_1 + 1)}$$

$$y_{t_2} = (t_2 + 1)^{-\theta} \rightarrow \ln y_{t_2} = -\theta \ln(t_2 + 1). \quad (3)$$

Then, by replacing θ in the bottom equation with the result in the top equation, we obtain

$$\ln y_{t_2} = \left[\frac{\ln(t_2 + 1)}{\ln(t_1 + 1)} \right] \ln y_{t_1}, \quad (4)$$

which can be rewritten:

$$y_{t_2} = y_{t_1}^{\ln(t_2 + 1)/\ln(t_1 + 1)}. \quad (5)$$

Note that the parameter θ has been removed from this equation. The model is now parameter free, having been redefined as the relationship between two y values in-

Figure 1. The left panel shows power forgetting curves of 3 hypothetical subjects generated from Equation 2 with different values of the parameter θ . The right panel shows the response curve of the power model. The data points corresponding to those of the 3 subjects are indicated.

stead of a parameter and a y value. Each point on the response curve describes the relationship between two y values that are themselves perfectly described by a power function. Thus, each point on the response curve can be thought of as a different power function, in the sense that its parameter value is unique. The response curve itself represents the complete power model; that is, the response curve is a visual description of the model with no error/noise. Any point that falls on the curve can be perfectly fit by a power function. Fit will decrease as a point (i.e., another function) moves further from the response curve.

The response curve for the exponential model is given by

$$y_{t_2} = y_{t_1}^{t_2/t_1} \quad (6)$$

and is plotted in Figure 2 along with the response curve of the power model. Note that for $t_1 = 1$ and $t_2 = 4$, the exponential response curve, $y_{t_2} = y_{t_1}^4$, is more bowed than the power response curve, $y_{t_2} = y_{t_1}^{2.32}$, and positioned in a slightly different part of the data space of possible response

functions. This relationship between the two models is fixed. It does not depend on the particular choices of t_1 and t_2 and holds for any $t_2 > t_1 > 0$. The relationship is the key to explaining the power law artifact.

Recall that the power law artifact refers to the observation that not only does the averaging of power data yield a power fit but the averaging of exponential data also yields a power fit. To show why this artifact occurs, examine Figure 2 once again. The 10 filled circles denote the data points for individual subjects whose data are described perfectly by the power model, each with a different value of the parameter (listed in note 1). Similarly, the 10 open circles are data points generated by the exponential model. The arithmetic means of the two data sets are represented by the triangles (filled for the power model and open for the exponential model).

Note that the arithmetic means are not located on their respective response curves. This outcome is not surprising given the fact that an arithmetic mean represents the center of gravity of data points being aggregated. Only

Figure 2. Response curves of the power and exponential models in Equation 2 for two observations made at $t_1 = 1$ and $t_2 = 4$. The filled circles represent data points of 10 hypothetical subjects generated from the power model. The open circles denote data points from the exponential model. The triangles and rectangles denote the corresponding arithmetic and geometric means, respectively. The dots scattered around a point on the exponential response curve indicate subject variation due to sampling error.

with a linear model (a straight response line) will the arithmetic average also fall on the response curve. Because neither the power model nor the exponential model are linear with respect to the parameter, the arithmetic means fall away from the response curves, to different degrees depending on the model.

The location of the arithmetic means in relation to the two curves clearly shows why arithmetic averaging of exponential data yields a power fit instead of an exponential fit. The arithmetic mean for the exponential model, as well as for the power model, is closer to the power response curve than the exponential response curve, which suggests that both arithmetic means should be fitted best by the power model. This is indeed the case. When the exponential arithmetic mean was fitted to the two models, the MS_{e} s were 0.025 and 0.051 for the power and exponential models, respectively. The fits to the power arithmetic mean were 0.029 and 0.104, respectively.

It should be clear from this example that the power law artifact arises in part from the relative positions of the models' response curves and arithmetic averaging. Because the power and exponential response curves are concave inward in the data space, their arithmetic means will always be located in the area to the left of the curves. Because the power response curve is positioned to the left of the exponential response curve, an area in which the arithmetic mean for the exponential model happens to fall, the exponential arithmetic mean will often be fit better by the power model, thereby producing the power law artifact.

The reason the power curve is often closest to the exponential arithmetic mean is that the extent to which the mean departs from the exponential curve depends on the degree of curvature. The more bowed the response curve, the further the mean will be located from that curve. This relationship is evident in Figure 2. The exponential response curve is more bowed than the power response curve. Consequently, the exponential arithmetic mean is further from the exponential curve than the power arithmetic mean is from the power curve.

In short, the response curve of the power model is situated in data space in such a way that it captures better the arithmetic averaging effects of exponential data than the exponential model itself. Because of its relative location in data space, the arithmetic mean of the power model can never be closer to the exponential curve than to the power curve. Thus their relative positions also explain why an exponential law artifact can never be observed. Furthermore, the positional relationship explains why the power model fitted random data better than the exponential model (Table 1: Arithmetic Averaging). Random data do not occupy any particular region in Figure 2, but instead fill up the entire data space evenly. Every point in the two-dimensional space defined by (y_{t_1}, y_{t_2}) where $0 \leq y_{t_1}, y_{t_2} \leq 1$ has an equal chance of being observed. The arithmetic mean of such data will tend to fall near the center of the graph, $(y_{t_1}, y_{t_2}) = (0.5, 0.5)$, which is closer to the power curve than the exponential curve. This is

why the power model will generally provide the best fit to random data.

These insights into the relationship between the two models and the reasons for the superiority of the power model highlight the usefulness of RSA as a tool for understanding model behavior, not just providing a graphical explanation of the power law artifact.

CONDITIONS FOR OBSERVING THE ARTIFACT

In this section, we briefly discuss the three conditions that are necessary for the artifact to be found: individual differences, arithmetic averaging, and nonlinear models.

The first, which has not been discussed up to this point, is that the artifact depends on there being performance differences among individuals that result in a range of parameters being chosen for the model. If all subject data are fit best using a single parameter so that there is no variability in the data other than sampling error (i.e., one function fits all subject data well), then the arithmetic mean will be located on (or close to) a point on the response curve of the model, with individual subject data varying around this point. The small dots on the exponential curve in Figure 2 depict this situation; they represent subject variation due to sampling error only, which will be canceled out by arithmetic averaging and, therefore, no artifact will be observed. To demonstrate the validity of this claim, we conducted a simple simulation of this situation.² An artifact failed to emerge: The power and exponential models fitted their own data best 95% and 99% of the time, respectively.

The necessity of individual differences for observing the artifact explains why R. B. Anderson and Tweney (1997) failed to find one when random noise was added to the model parameters to generate forgetting curves of simulated subjects, even when the data were averaged arithmetically across subjects. It may also account for why the artifact has not always been found in forgetting experiments. For example, when Wixted and Ebbesen (1997) reanalyzed data from a previous study, the type of averaging (arithmetic or geometric) made little difference, probably because there were few or no individual differences in the data.³

Another necessary condition for the power law artifact, or generally speaking an averaging artifact, is that models under consideration be nonlinear in the sense that the dependent variable (y) is related nonlinearly to the parameter (θ)—not to the independent variable (x). For example, both the power and exponential models are nonlinear, whereas the linear model, $y = 1 - \theta t$, and the logarithmic model, $y = 1 - \theta \log(t + 1)$, are both linear. The reason for this nonlinearity condition is that an artifact due to averaging can occur only with *curved* response surfaces. The response surface of a nonlinear model is always curved.

However, even with nonlinear models and data that contain individual differences, an averaging artifact may

not be found. If the data points on the response curve are relatively linear (i.e., flat), any averaging artifact may be inconsequential. For example, in Figure 2, the arithmetic mean of the four to six data points at the upper rightmost portion of the exponential response curve will fall very close to the exponential curve.

The above explanation of the power law artifact for the one-parameter and two-observation case can be generalized to multiparameter and multiobservation situations. We do not have analytic proofs for this claim; however, it seems intuitively clear that the same logic and explanation will apply to such cases. For instance, Figure 3 shows the response surfaces obtained for the two-parameter power and exponential models with three data points observed at $t = 1, 4, 10$:

$$y = \theta_1(t + 1)^{-\theta_2} \text{ (power model)}$$

and

$$y = \theta_1 \exp(-\theta_2 t) \text{ (exponential model),} \quad (7)$$

where $\theta_1, \theta_2 \geq 0$. Note in the figure that both the power and the exponential response surfaces are curved, and further, that the latter is more curved than the former—the

two conditions required for the occurrence of the power law artifact. It is important to note here that curved response surfaces are obtained because of the nonlinear relationship between θ_2 and y in Equation 7 for both the power and exponential models. On the other hand, θ_1 is linear in y for both models and therefore would play no role in producing the power law artifact. Similarly, adding a third parameter θ_3 to both models as an additive intercept would not change the relationship between models.

To summarize, we have shown that the power law artifact can occur whenever *arithmetic averaging* is employed to aggregate data that are generated from a *nonlinear model* in the presence of *individual differences*. These are three necessary ingredients for the artifact. If the model is linear, arithmetic averaging will not produce the artifact, even in the presence of individual differences. Averaging is problematic only when it is used in conjunction with nonlinear models. In such cases, one can avoid artifactual effects by using nonarithmetic averaging. The model under consideration will influence the particular nonarithmetic averaging technique that is used, which is the final topic of the paper.

Figure 3. Response surfaces of the two-parameter power and exponential models in Equation 7 for three observations made at $(t_1, t_2, t_3) = (1, 4, 10)$. The surfaces shown represent partial response surfaces created by varying values of the two parameters over the ranges of $0.8 \leq \theta_1 \leq 1.2$ and $0 \leq \theta_2 \leq 50$ for both models.

HOW CAN THE ARTIFACT BE AVOIDED?

There are at least three ways to avoid the artifact. Each is discussed briefly.

Nonarithmetic Averaging

If geometric averaging prevents the power law artifact, then the artifact present in the data in the top panel of Table 1 should disappear when the simulations are rerun using geometric averaging. This is what is shown in the second panel from the top in Table 1 (Geometric Averaging). Although the two models did not fit their own data perfectly (due to sampling error from use of a finite sample size—50), the power model no longer fitted the exponential data well. As should be the case, each model fitted the data best when it was generated by the same model. As should also be the case, the power model fitted geometrically averaged random data better than the exponential model. As noted, this is because of the relative locations of the two response curves in data space (Figure 2).

Use of geometric averaging on data generated from power and exponential models has the property of being a model-preserving averaging function. That is, the method of averaging is appropriate for the type of nonlinear models being evaluated and thus will not yield spurious results. The general form of the averaging function for one-parameter models is discussed in the Appendix, along with an example of its derivation for the power model. For each of the six two-parameter models used in Wixted and Ebbesen (1991), which have been put forth as models of forgetting, the corresponding model-preserving averaging function is shown in Table 2.

Individual Analysis

In individual analysis, a model is fitted to each individual's data separately, instead of fitting data arithmetically or geometrically averaged across individuals. The resulting fits may then be analyzed to assess an overall goodness of fit of each model. For instance, individual fits of a model may be aggregated to produce an overall fit measure, which is then compared with that of another competing model. Alternatively, individual fits of two

models may be contrasted to yield the percentage of time in which one model provides a better fit than the other.

In an individual analysis of the simulated data, both power and exponential models were fitted to each of the 10 sample curves separately, and 10 individual MS_e s were obtained for each model. This process was repeated 100 times with data generated from each of the models—power, exponential, and random. Results are shown in Table 1 (Individual Analysis). The outcome is similar to the geometrically averaged data; each model fitted its own data best and the power model provided the better fit to the random data. One obvious difference is that model recovery rates were notably lower—by approximately 20 percentage points.

Stochastic Parameter Approach

This is a statistical method for analyzing data in the presence of individual differences (for technical reviews of the method, see Bryk & Raudenbush, 1987, 1992; Cudeck, 1996; Kim, 1998). The approach assumes that each individual data set is a realization of the common population (i.e., model) but with a different value of the model's parameter, and further, that individual parameter values are random samples of a fixed but unknown probability distribution. In other words, in this approach it is assumed that individual data can be represented as random dots dispersed strictly along the model's response surface. It is in this sense that the approach takes full account of the information provided in the curvature of the response surface. A primary goal of the approach is to estimate the mean and variance structures of the parameter distribution. Fitting stochastic parameter models requires use of statistical software packages such as HLM (Bryk, Raudenbush, & Congdon, 1992) and PROC MIXLIN in the SAS programming language (Vonesh & Chinchilli, 1997).

An important difference between the stochastic parameter approach and individual analysis is that the former is a model-based, confirmatory technique whereas the latter is an exploratory technique. In individual analysis, no assumptions are made about whether individuals are from the same population (i.e., model) or whether individual parameter estimates are generated from a single probability distribution. Rather, the best combination of

Table 2
Model-Preserving Averaging Operation for the
Six Forgetting Models of Wixted and Ebbesen (1991)

Model	Model Equation	Model-Preserving Averaging Operation
Linear	$y_{t,i} = \theta_i - \lambda_i t$	$y_{t,arth} = \theta_{arth} - \lambda_{arth} t$
Logarithmic	$y_{t,i} = \theta_i - \lambda_i \ln(t)$	$y_{t,arth} = \theta_{arth} - \lambda_{arth} \ln(t)$
Power	$y_{t,i} = \theta_i t^{-\lambda_i}$	$y_{t,geom} = \theta_{geom} t^{-\lambda_{geom}}$
Exponential	$y_{t,i} = \theta_i \exp(-\lambda_i t)$	$y_{t,geom} = \theta_{geom} \exp(-\lambda_{geom} t)$
Exponential power	$y_{t,i} = \theta_i \exp(-2\lambda_i \sqrt{t})$	$y_{t,geom} = \theta_{geom} \exp(-2\lambda_{geom} \sqrt{t})$
Hyperbolic	$y_{t,i} = 1/(\theta_i + \lambda_i t)$	$y_{t,harm} = 1/(\theta_{harm} + \lambda_{harm} t)$

Note—The subscripts, *arth*, *geom*, and *harm*, represent the arithmetic, geometric, and harmonic mean of individual values of the corresponding variable or parameter.

model and parameter value that provides the best fit, whatever that may be, is assumed to be the one that generated the sample data.

A stochastic parameter analysis of the simulated data was conducted as follows. The power model was fitted to the 10 sample curves *simultaneously* under the assumption that all belong to the model but with different parameter values, and an overall MS_e of the model as well as individual parameter estimates is obtained. Similarly, the exponential model was fitted to the same data, and its MS_e was obtained. This process was repeated 100 times with data generated from each of the models—power, exponential, and random. Results are shown in Table 1 (Stochastic Parameter Approach). Similar to the geometrically averaged data, the two models, power and exponential, are almost perfectly recovered. Unlike the geometrically averaged and individual analysis data, however, they fit the random data equally well and equally often.

DISCUSSION

The preceding analyses show that the power law artifact arises from a set of co-occurring conditions: the relationship between the models (i.e., their response curves), the type of averaging being used, and the presence of individual differences. If any one of these is not met, the artifact is less likely to be found. Individual differences can be detected by inspecting the individual model fits prior to aggregating the data. Averaging is the only condition that is directly under the researcher's control. To this end, methods for ensuring correct averaging were presented. Of course, their use does not imply that the power model will no longer reign supreme, only that any supremacy will not be due to an artifact of averaging.

The relationship between the models is probably the most difficult condition to evaluate. Response surface analysis provides a method of graphically describing this relationship that clearly reveals why the power model consistently yielded a superior fit. The relative positions of the power model and exponential model in the data space of possible response curves ensures the power model's success. The power model just happens to be positioned in the region of data space where arithmetically averaged exponential data will fall.

An artifact will be seen with any model whose response curve falls to the left of the exponential model. For example, the response curve (actually a straight line) of the logarithmic model falls in between the power and exponential response curves, suggesting that a "logarithmic artifact" will be found. Because of its close proximity to the exponential response curve, the magnitude of the artifact will be less than that found with the power model.

Despite many explicit warnings about its potential misleading effects (see, e.g., Ashby, Maddox, & Lee, 1994; Estes, 1956; Melton, 1936; Siegler, 1987; Singh, 1996), arithmetic averaging across subjects' data is still a common practice among researchers carrying out analyses. Arithmetic averaging can improve statistical inference,

but not always. If there exist individual differences (i.e., individual data are samples from the same population but with different values of the parameter), arithmetic averaging across individual data sets is not justified unless the model is linear. When a nonlinear model is tested, arithmetic averaging can create an artifactual effect that has no relation to the underlying process that is being modeled. The power law artifact, in this sense, is an unfortunate example of the misuse of statistics: Linear techniques are used not only for linear models in data analysis but also for nonlinear models in data fitting.

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.
- ASHBY, F. G., MADDOX, W. T., & LEE, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, *5*, 144-151.
- BATES, D. M., & WATTS, D. G. (1988). *Nonlinear regression analysis and its applications*. New York: Wiley.
- BRYK, A. S., & RAUDENBUSH, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147-158.
- BRYK, A. S., & RAUDENBUSH, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- BRYK, A. S., RAUDENBUSH, S. W., & CONGDON, R. (1992). *Hierarchical linear model and nonlinear modeling with HLM/2L and HLM/3L programs*. Chicago: Scientific Software.
- CUDECK, R. (1996). Mixed effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research*, *31*, 371-403.
- EBBINGHAUS, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans). New York: Dover. (Original work published in 1885)
- ESTES, W. K. (1956). The problem of inference from curves based on group data. *Psychological Review*, *53*, 134-140.
- KIM, C. (1998). *Incorporating individual differences in mathematical psychology*. Unpublished doctoral dissertation, Ohio State University.
- MELTON, A. W. (1936). The end-spurt in memorization curves as an artifact of the averaging of individual curves. *Psychological Monographs*, *47* (2, Whole No. 202).
- NEWELL, A., & ROSENBLUM, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognition 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., & WENZEL, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*, 734-760.
- SIEGLER, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250-264.
- SINGH, R. (1996). Subtractive versus ratio model of "fair" allocation: Can the group level analyses be misleading? *Organizational Behavior & Human Decision Processes*, *68*, 123-144.
- STEVENS, S. S. (1971). Neural events and the psychophysical law. *Science*, *170*, 1043-1050.
- STEVENSON, M. K. (1993). Decision making with long-term consequences: Temporal discounting for single and multiple outcomes in the future. *Journal of Experimental Psychology: General*, *122*, 3-22.
- VONESH, E. F., & CHINCHILLI, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. New York: Marcel Dekker.
- WICKENS, T. D. (1998). On the form of the retention function: Comment on Rubin and Wenzel (1996): A qualitative description of retention. *Psychological Review*, *105*, 379-386.

- 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.

NOTES

1. The 10 parameter values used to generate the simulated data were as follows: For the power model, $\theta = 0.07, 0.15, 0.24, 0.36, 0.49, 0.64, 0.85, 1.1, 1.5, 2.1$, and for the exponential model, $\theta = 0.03, 0.065, 0.11, 0.165, 0.23, 0.32, 0.46, 0.63, 0.9, 1.3$.
2. The simulation paralleled that used to generate the data in Table 1, except that for each model the 10 forgetting curves were generated from a single value of the parameter ($\theta = 0.24$ for the power model and $\theta = 0.11$ for the exponential model) rather than 10 parameters.
3. Individual differences are not the only factor that can yield an averaging artifact. Any factor that creates performance variation from different values of the parameter of a model can potentially produce an averaging artifact. For example, R. B. Anderson and Tweney (1997) pointed out and demonstrated (see their Table 3) that averaging across trial position could also produce an artifact. This is also true for variation across items.

APPENDIX

This appendix describes the general form of the one-parameter model-preserving averaging function and its derivation for the power model.

The general form of a one-parameter model for individual i ($i = 1, \dots, m$) may be written as

$$y_{t,i} = f(t, \theta_i), (i = 1, \dots, m),$$

where $y_{t,i}$ is the response at time t for individual i and θ_i is the parameter value that generates $y_{t,i}$ for that individual. Suppose that there exists a function h such that $h^{-1}(y_{t,i}) = \theta_i$ for all t and i . Then it can be shown that the model equation is preserved under the following averaging operation:

$$y_{t,av} = h \left(\frac{\sum_{i=1}^m h^{-1}(y_{t,i})}{m} \right).$$

in the sense that

$$y_{t,av} = f(t, \theta_{arth}),$$

where θ_{arth} is the arithmetic mean of individual θ_i s, that is, $\theta_{arth} = \sum_i \theta_i / m$. For the linear model, the above equation rep-

resents arithmetic averaging (i.e., $h(x) = h^{-1}(x) = x$), whereas for the power model or the exponential model, it represents geometric averaging (i.e., $h(x) = \exp(x)$ or $h^{-1}(x) = \ln(x)$). In Figure 2, the filled square represents the geometric mean of the power data and the open square denotes the geometric mean of the exponential data. Each square falls on the response curve of the corresponding model, indicating that the geometric mean of individual data points generated from one of the models can be described best by that same model.

It is important to note that there does not exist a single averaging function that preserves all forms of model equations. Instead, the model-preserving averaging function will generally be different for different models.

According to the power model, a forgetting curve at time t for individual i follows the equation

$$y_{t,i} = \theta_i(t+1)^{-\lambda_i}, (i = 1, \dots, m),$$

where θ_i and λ_i are two positive parameters to be estimated from the data. By applying the logarithmic function to both sides of the above equation, we get

$$\ln y_{t,i} = \ln \theta_i - \lambda_i \ln(t+1), (i = 1, \dots, m)$$

Summing across individuals and then dividing by m yields

$$\frac{\sum_{i=1}^m \ln y_{t,i}}{m} = \frac{\sum_{i=1}^m \ln \theta_i}{m} - \frac{\sum_{i=1}^m \lambda_i}{m} \ln(t+1).$$

Applying the exponential function to both sides of the equation produces

$$\exp \left(\frac{\sum_{i=1}^m \ln y_{t,i}}{m} \right) = \exp \left(\frac{\sum_{i=1}^m \ln \theta_i}{m} \right) \exp \left(- \frac{\sum_{i=1}^m \lambda_i}{m} \ln(t+1) \right).$$

This equation can be rewritten in terms of geometric and arithmetic means as

$$y_{t,geom} = \theta_{geom} \exp[-\lambda_{arth} \ln(t+1)] = \theta_{geom}(t+1)^{-\lambda_{arth}},$$

yielding the desired result.

(Manuscript received December 28, 1998;
revision accepted for publication August 12, 1999.)