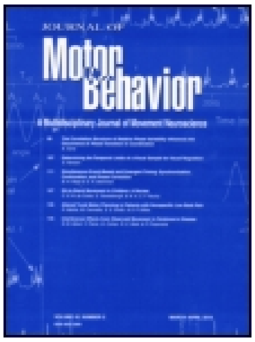


Journal of Motor Behavior

ISSN: 0022-2895 (Print) 1940-1027 (Online) Journal homepage: https://www.tandfonline.com/loi/vjmb20

Beyond Curve Fitting? Comment on Liu, Mayer-Kress, and Newell (2003)

Andrew Heathcote & Scott Brown

To cite this article: Andrew Heathcote & Scott Brown (2004) Beyond Curve Fitting? Comment on Liu, Mayer-Kress, and Newell (2003), Journal of Motor Behavior, 36:2, 225-232, DOI: 10.3200/JMBR.36.2.225-232

To link to this article: https://doi.org/10.3200/JMBR.36.2.225-232

Published online: 07 Aug 2010.

Journal of Motor Behavior, 2004, Vol. 36, No. 2, 225–232

Beyond Curve Fitting? Comment on Liu, Mayer-Kress, and Newell (2003)

Andrew Heathcote
School of Behavioural Sciences
University of Newcastle, Australia

Scott Brown
Department of Cognitive Science
University of California, Irvine

ABSTRACT. Y.-T. Liu, G. Mayer-Kress, and K. M. Newell (2003) fit learning curves to movement time data and suggested 2 new methods for analyzing learning. They claimed that the methods go "beyond curve fitting." However, in neither their curve fitting nor their new methods is measurement noise accounted for, and therefore they produce inefficient and biased results. Using the data of Liu et al., in which variance caused by learning is small relative to the level of noise for most participants, the present authors demonstrate those problems and provide better alternatives that are more noise tolerant, more powerful, and go beyond curve fitting without displaying the extreme bias produced by the methods of Liu et al.

Key words: averaging, learning curve, model testing, motor learning

Liu, Mayer-Kress, and Newell (2003) examined learning in a discrete movement task. Participants were required to produce a 5° elbow flexion in 125 ms. The data consisted of the movement times (MT) for 8 participants over 200 trials. Changes in MT caused by learning were small relative to the level of noise (see Figure 1): For Participants C, D, G, and H, only the first-trial MT was greater than were later MTs; for Participants E and F, the MTs were greater on only the first two trials; and MT for Participant A was less for the first trial than on later trials. Participant B showed greater learning, but even that was not substantial; MTs were greater only on the first six trials than were MTs from later trials. Evidently, any analysis of learning for those data must tolerate substantial noise.

Noisy learning curves such as those of Liu et al. (2003) are usually modeled by an equation of the following form: $\text{MT} = f(t) + \varepsilon(\boldsymbol{\theta})$, where $f(t)$ is a deterministic function of practice trials (t), and $\varepsilon(\boldsymbol{\theta})$ is a random variable with zero mean, that is, $E(\varepsilon) = 0$, and with parameter vector $\boldsymbol{\theta}$. That form is implicitly assumed by Liu et al.'s use of least squares regression, which is optimally efficient if ε is normal and is asymptotically efficient even when ε is nonnormal, under mild regularity conditions (see Jennrich, 1969). However, Liu et al. did not fit MT, but instead fit absolute error, $AE = |\text{MT} - 125|$.

If the aim of fitting is to estimate $f(t)$, then least squares regression on AE is clearly inappropriate. When performance is variable, the mean value of AE must be greater than zero, even when mean MT is exactly 125 ms. As can be seen in Figure 1, performance was clearly variable. The bias induced by using AE is a function of both the level of variability and the distance between $f(t)$ and its asymptote. Larger variability leads to greater average AE values, and, as $f(t)$ approaches asymptote, the bias increases, systematically distorting the shape of the estimated function. In short, the function estimated by least squares regression on AE is not $f(t)$ but $f(t) + B(t,\boldsymbol{\theta})$, where $B(t,\boldsymbol{\theta})$ is the (trial and noise-dependent) bias.

We strongly recommend that researchers avoid the use of AE when their data contain noise, which is always the case for response time measures such as MT (cf. Luce, 1986). In the following sections, we report least squares regression results for the data of Liu et al. (2003) and analyze on the basis of MT rather than AE the two new methods that they proposed. Our analyses of their two new methods for going beyond curve fitting reveal further problems caused by the failure of Liu et al. to account for the effects of noise. First, however, we examine the statements of Liu et al. concerning the most powerful method in a researcher's armory for minimizing the effects of noise: averaging.

Correspondence address: Andrew Heathcote, School of Behavioural Sciences, Aviation Building, University of Newcastle, NSW, 2308, Australia. E-mail address: andrew.heathcote@newcastle.edu.au

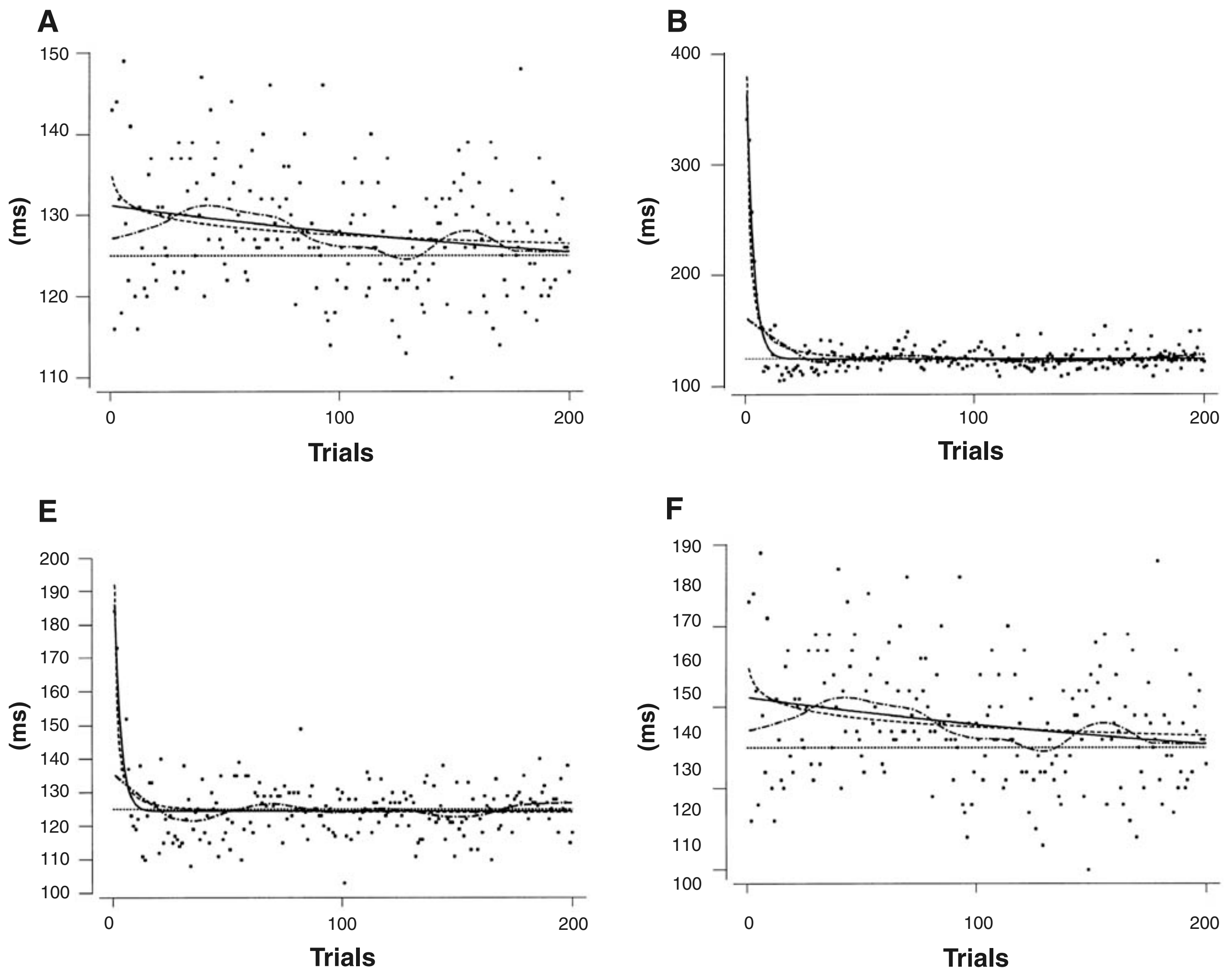
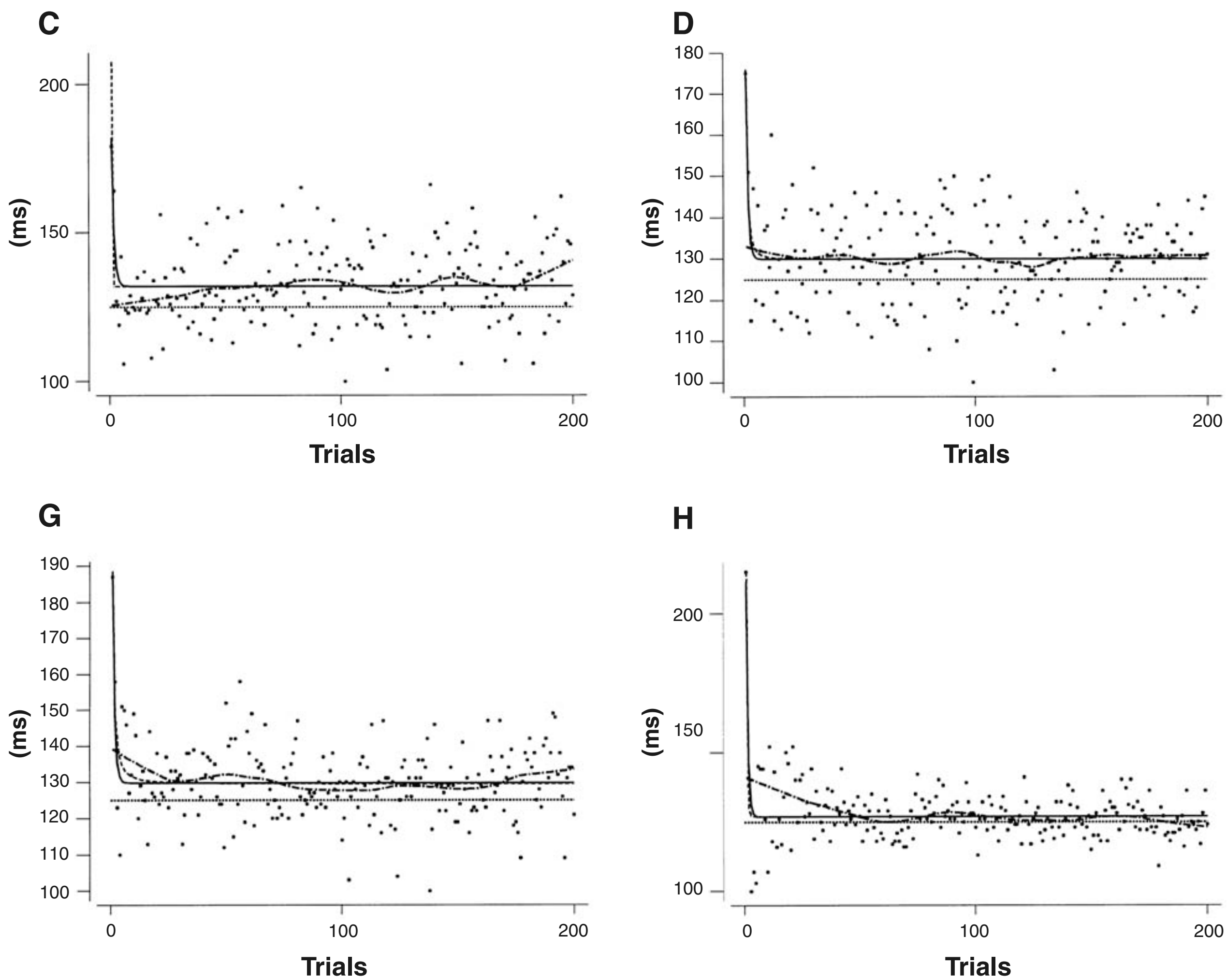


FIGURE 1. MT (movement time) data (dots) for each participant (A . . . H; see ordinate labels) of Liu, Mayer-Kress, and Newell (2003), locally weighted scatterplot smoother (LOWESS) smooths (dash-dot lines), and best-fitting exponential (solid lines) and power (dashed lines) functions. The target MT was 125 ms (horizontal dotted lines).

Averaged Learning Curves

We agree with the caution of Liu et al. (2003) concerning the averaging of learning curves across participants. Brown and Heathcote (2003a) provided a simple proof that an arithmetic average of curves has the same form as the component curves if and only if the component curves are linear in all parameters that vary across components. The core of that proof has been known since at least 1821, when Cauchy published it; the interested reader is referred to Aczel (1966) for extensions (e.g., averages other than arithmetic). Heathcote, Brown, and Mewhort (2003) provided simple graphical and inferential methods of checking whether the linearity condition holds in noisy data. Their analysis of data from Heathcote, Brown, and Mewhort's (2000) survey of learning studies suggested that that condition rarely holds in practice. Brown and Heathcote showed that averaging across exponential functions that differ in their rate parameters by as little as a single order of magnitude can produce an average curve better fit by a power function. Hence, curves produced by averaging across participants will usually not be characteristic of participant curves, and in any attempt to adjudicate between nonlinear models of participant learning one should not rely on data averaged over participants.

However, we disagree with Liu et al. (2003) with respect to averaging over trials. Brown and Heathcote (2003a) proved that averaging across trials has no effect on the form of exponential functions, producing a change only in the (linear) scale parameter. Because only a linear parameter is changed, the trial average function remains exactly exponential; therefore, trial averaging will not cause a bias against the exponential form. Brown and Heathcote also showed that the bias induced by trial averaging of power



functions is usually negligible, except in a region of extreme curvature, and even then only when the number of trials averaged is large relative to the extent of that region. Brown and Heathcote's analysis of trial averages of data from the survey of Heathcote et al. (2000) showed that there was little effect on the relative fit of power and exponential functions. Trial averages are very useful because of their ability to reduce noise. For example, Brown and Heathcote's trial averages produced R^2 values comparable with averages across participants with only a minimal risk of the distortion caused by participant averages. Noise causes discrimination between exponential and power functions to fail, approaching no discrimination with extreme noise (Brown & Heathcote, 2003b). Hence, trial averaging can improve discrimination between curve forms.

Trial averages, and generalizations of trial averages usually referred to as *smooths* (e.g., Bowman & Azzalini, 1997), are particularly useful for exploratory data analysis and graphical display of trends in noisy data—hence, their popularity among researchers. The thick wavy lines in Figure 1 show broad trial averages (encompassing 50 trials) obtained using the widely available locally weighted scatterplot smoother (LOWESS; Cleveland, 1979). The smooth for Participant A contradicts the claim of Liu et al. (2003) that Participant A showed no learning. Although weak relative to the extent of the noise, there is a clear downward trend in the data, which we confirmed by performing the statistical test reported in the next section. Participant C, in contrast, showed a clear upward trend, which might arouse suspicion that fatigue effects were present. It is only with the aid of the trial averages that such trends become evident.

The smooths in Figure 1 also illustrate a weakness of trial averages; they are unable to follow the fast changes evident in early trials. Clearly, one must choose the widths of trial averages to suit the rate of change in the data, and trial averages may not be suitable for all data. The data of Liu et al. (2003) provide a very difficult case for the analysis of learning both because noise was high and because, for all but Participant A, learning was largely confined to the first few trials. However, trial averages remain useful even in cases such as those because they enable one to check for slow variations in the right tail of the learning function that might confound fits of parametric curves such as the power and exponential, which make a strong assumption of constancy in the right tail.

Curve Fitting

In Table 1, we provide the proportion of variance accounted for by least squares regressions on the MT data for the exponential ($A_E + B_E e^{-rt}$) and power ($A_P + B_P t^{-c}$) functions. Those equations have two linear parameters that quantify the asymptote (A) and the scale of change (B) of the learning function, and one nonlinear parameter, the exponential rate (r) and the power exponent (c). Unlike the *AE* regressions of Liu et al. (2003), the exponential is favored in the majority of participants. Because of the possible fatigue effect for Participant C revealed by the smooths, we used the entire data set (200 trials) rather than the method of Liu et al. (they used the last 100 trials to estimate A). When only the first 50 and first 10 trials were fit for Participant C, the fit of the exponential function remained superior. For all fits, the proportion of variance accounted for by both curves was significant, with $p < .001$ for all but Participant C (ps = .001 and .047 for the exponential and power functions, respectively) and Participant A (ps = .016 and .04 for the exponential and power functions, respectively). The latter result contradicts the statement of Liu et al. that Participant A showed no learning and confirms the gradual change made evident by the LOWESS plot in Figure 1. The thin, smooth curves in Figure 1 plot the best-fitting exponential (solid line) and power (dotted line) functions.

Liu et al. (2003) claimed that there was “no significant difference in the percentage of variance accounted for” (p. 197) by the power and exponential functions. However, they did not report any inferential tests of that difference and so had no basis for their statement. One can perform inferential tests by using the nested model technique of Heathcote et al. (2000). In that test, one must fit a four-parameter function that has both power and exponential components: the asymptote power exponential (APEX) function, that is, $A + Be^{-\alpha t}t^{-\beta}$. The APEX function *nests* (i.e., has as a special case) both the power function (when $\alpha = 0$) and an exponential function (when $\beta = 0$). Nesting allows one to test the significance of the power and exponential components by using an F test:

$$F_{(f-r,N-f-1)} = \frac{(N-f-1)\left(R_F^2 - R_R^2\right)}{(f-r)\left(1-R_F^2\right)}.$$

The subscripts F and R refer to the full (i.e., APEX) and reduced (i.e., power or exponential) models, with $f = 4$ and $r = 3$ *df*, respectively, corresponding to the number of parameters estimated for each model. N is the number of data points. The results of those tests are presented in Table 1; $R^2_{\text{APEX}} - R^2_{\text{E}}$ tests the power component (i.e., the loss of fit when the power component is omitted), and $R^2_{\text{APEX}} - R^2_{\text{P}}$ tests the exponential component (i.e., the loss of fit when the exponential component is omitted). Highly significant evidence for a purely exponential component was obtained for 4 participants, whereas the power component never approached significance, except marginally for Participant H. Because of its faster approach to asymptote, Figure 1 reveals that among the participants with highly significant differences, Participants B, E, and F favored the exponential function. In addition, because the power function overestimated the first point, Figure 1 shows that C favored the exponential function.

TABLE 1. Fits of the Exponential (Subscript E), Power (Subscript P), and APEX (Subscript APEX) Functions to Movement Time Data From All 200 Trials for Each Participant

Participant	R^2_{E}	R^2_{P}	R^2_{APEX}	$R^2_{\text{APEX}} - R^2_{\text{E}}$	F	p	$R^2_{\text{APEX}} - R^2_{\text{P}}$	F	p
A	.051	.041	.051	.000	0.00	.997	.010	2.01	.157
B	.843	.733	.843	.000	0.00	.996	.110	136.05	.000
C	.082	.040	.082	.000	0.00	1.000	.042	9.00	.003
D	.098	.097	.098	.000	0.00	1.000	.001	0.23	.633
E	.398	.356	.398	.000	0.00	1.000	.042	13.61	.000
F	.357	.309	.357	.000	0.00	1.000	.048	14.49	.000
G	.167	.170	.170	.003	0.63	.428	.000	0.00	1.000
H	.346	.359	.358	.011	3.46	.064	.000	0.00	1.000

Note. F ratios, $df = (1, 195)$, tested the R^2 difference in the preceding column. APEX = asymptote power exponential function.

Beyond Curve Fitting

We agree with Liu et al. (2003) that it is desirable to have tests that go beyond curve fitting in the sense of looking at more than relative goodness-of-fit of parametric models. However, we conceptualize that approach somewhat differently from Liu et al. as being about nonparametric analyses, that is, analyses that are not prefaced on a particular parametric model or finite set of parametric models. The "fat-tails indicator" (FTI) of Liu et al., in contrast, is based on parametric (i.e., power and exponential) model estimates. Liu et al. proposed a measure called absolute FTI (λ_{abs}) for discriminating between power and exponential learning curves and then extended it to a relative FTI (λ_{rel}) test, but they "do not claim . . . that this . . . is optimal in any sense" (p. 203).

To test the bias and efficiency of λ_{abs}, we performed a simulation study using 10,000 noisy exponential curves and 10,000 noisy power curves, created by adding normally distributed deviates with mean zero and standard deviation 10 to the fits reported by Liu et al. (2003) for their Participant B (see Figure 2 in Liu et al.).[1] We compared the model discrimination performance of λ_{abs} against simply picking the model with the best fit (i.e., highest R^2 value). According to the suggestions of Liu et al., we counted $\lambda_{abs} \leq 0$ as exponential and $\lambda_{abs} > 0$ as power.

The best-fit test correctly identified power curves in 96% of simulated decisions, whereas λ_{abs} identified the power curves at close to chance levels (48%). For the exponential curves, λ_{abs} was 100% correct and the best-fit test was 98% correct. Those results indicate a strong exponential bias in λ_{abs}. The relative FTI (λ_{rel}) of Liu et al. (2003) was even worse: It was provably biased against power functions. Given a noiseless power function, the power function estimated from the data by any competent method is exactly the same as the data. That equality makes the numerator in Equation 6 of Liu et al. equal to zero, and therefore λ_{rel} always incorrectly chooses an exponential model for pure power curves! We confirmed, using the same simulation method applied to λ_{abs}, that the extreme bias also applies to noisy curves; λ_{rel} always classified the power data as exponential. Clearly, the FTI methods of Liu et al. should be avoided. In contrast, a simple comparison of goodness-of-fit performs well, and we recommend its use.

Liu et al. (2003) suggested a second method of going beyond learning curves that is truly nonparametric—their discrete proportional error change measure (R_n). R_n is very similar to the relative learning rate (RLR) measure that Heathcote et al. (2000) used to characterize differences and similarities among continuous parametric learning curves with apparently unrelated forms. Heathcote and his colleagues' measure produced simpler results for continuous curves, and the same conclusions apply to R_n; therefore, we use RLR here: $\text{RLR}(t) = f'(t)/[f(t) - f(\infty)]$. The prime indicates differentiation with respect to t, and we assume that f must be once differentiable and strictly monotonic. Constant RLR is a defining feature of the exponential function, $\text{RLR} = r$; therefore, average RLR directly estimates the exponential rate parameter. For a power function, RLR decreases to zero hyperbolically—$\text{RLR}(t) = c/t$. Therefore, detecting a significant decrease in RLR with trials favors a power function, although not uniquely.

Two problems arise when estimating RLR from noisy data. One problem is paramount for asymptotic samples, in which the true RLR denominator approaches zero. As the true denominator approaches zero, even very small perturbations caused by noise can produce very large fluctuations in RLR estimates. Any attempt to summarize the behavior of RLR can become dominated by those effects, obscuring useful information about changes in MT with practice in early trials. That behavior is evident in Figures 3 and 4 of Liu et al. (2003), in which they used a linear regression to summarize the behavior of the RLR estimates. They reported that slopes were not significantly different from zero and concluded that learning was exponential because RLR did not change with practice. However, it is evident from their Figure 4 that because their estimates were swamped by asymptotic noise, their estimates did not differ significantly from zero overall.[2]

The second problem with the direct calculation of R_n for each pair of trials (and the corresponding calculation of RLR) in Liu et al. (2003) is that it relies on an accurate estimate of $f(\infty)$. Even a small error in that estimate (as might have happened because of fatigue for Participant C) can produce extremely large distortions in RLR estimates. That approach also relies on having a long and stationary set of measurements of asymptotic performance, which is not always available or even practicable. Unfortunately, the two problems tend to compound each other; measurement of asymptotic performance improves the asymptote estimate but also increases asymptotic noise.

Estimating the RLR function by using regression rather than by calculating directly can solve both problems. For the exponential function, $-f'(t) = r[f(t) - f(\infty)]$, and thus a linear regression with slope r and intercept $-rf(\infty)$ is predicted in a plot of $-d\text{MT}(t + 1/2)$ against $\overline{\text{MT}(t+(1/2))}$, where $d\text{MT}(t + 1/2) = \text{MT}(t + 1) - \text{MT}(t)$ estimates f', and $\overline{\text{MT}(t+(1/2))}$ (the average value of MT over trials t and $t + 1$) estimates f. The plot for Participant B is shown in Figure 2. A quadratic regression[3] (solid line in Figure 2) reveals a significant linear, $F(1, 196) = 28.6$, $p < .001$, but not a quadratic, $F < 1$, component, with $R^2 = .129$. It is evident from Figure 2 that only the first five points provided much information about the RLR function. The first point is clearly highly influential; when it is removed, the linear regression indicated by the dotted line in Figure 2 is obtained, with $R^2 = .156$. Solving for the parameters, we obtained an asymptote estimate of 124.24 and a rate estimate of 0.3345, both of which are in good agreement with the least squares exponential fit (124.75 and 0.350, respectively).

For the power function, $-f'(t) = c[f(t) - f(\infty)]^{t-1}$. Hence, a linear plot of $d\text{MT}(t + 1/2) \times (t + 1/2)$ against $\overline{\text{MT}(t+(1/2))}$ is predicted, with slope c and intercept $-cf(\infty)$. In Figure 2,

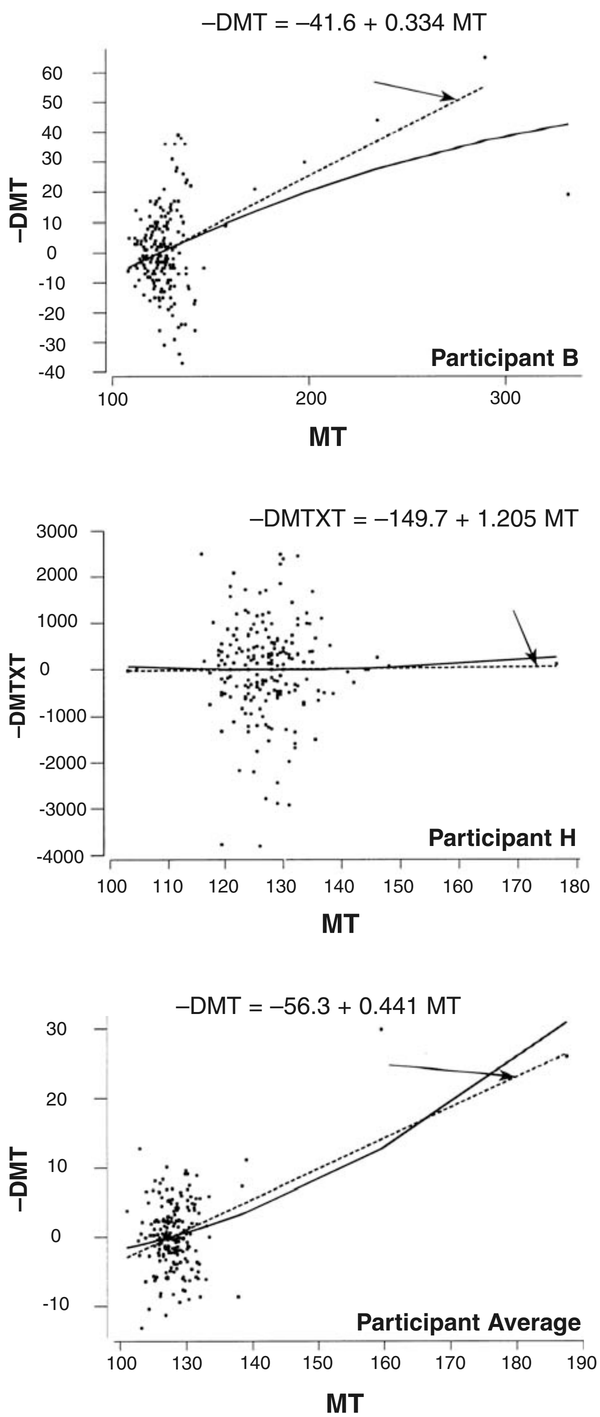


FIGURE 2. Exponential (Participant B, participant average) and power (Participant H) RLR function estimation plots. Dashed lines indicate a best-fitting linear regression and solid lines indicate a best-fitting quadratic regression.

we show the plot for Participant H because that participant had the strongest parametric evidence for a power function. Neither linear nor quadratic components were significant (Fs < 1). The linear slope grossly underestimated the power exponent relative to the least squares power fit (4.824) but only slightly underestimated the asymptote (124.2 vs. 127.0, respectively). Those results do not support the power function as an adequate model of learning, and they indicate that one slow point (calculated from Trials 1 and 2) and two fast points (calculated from Trials 3 and 4 and Trials 4 and 5) were highly influential.

Finally, if one's aim is simply to investigate possible deviations from a reference curve form (e.g., exponential), averaging across participant RLR function estimation plots, for example, over $d\mathrm{MT}(t + 1/2)$ and $\overline{\mathrm{MT}(t + (1/2))}$ for the exponential, is both convenient and mathematically appropriate, because of their linearity. Averaging will distort the exact form of the nonlinearity in the plot for functions of different forms, but deviation from linearity will still be evident, and the reference curve can be rejected without confounding. Note, however, that in that approach it is assumed that each participant has the same curve form, differing only in parameters.

Given that earlier tests favored the exponential for most subjects, we calculated an average RLR function estimation plot for the exponential function (Figure 2). A quadratic regression gave $R^2 = .206$, with a significant linear, $F(1, 196) = 47.8$, $p < .001$, but not quadratic, $F(1, 196) = 2.57$, $p = .11$, components. The best-fitting linear regression line ($R^2 = .195$) had a slope that only slightly overestimated the geometric mean of the rate parameters from the least squares exponential fits (0.412) and the arithmetic mean of their intercepts (127.7 and 126.8, respectively). Given the weakness of learning and high noise levels for most participants, the agreement was surprisingly good. However, it must be acknowledged that nonparametric approaches pay a cost in power relative to parametric approaches, at least when the assumptions of the latter are true. In the data of Liu et al. (2003), which had both little information about the systematic structure of the learning curves and high noise levels, only Participant B provided clear results with RLR plots, and even in that case an influential point had to be censored.

Brown and Heathcote (2002) suggested an alternative nonparametric approach[4] that is more powerful because it makes a slightly more restrictive, but still plausible, assumption about the form of the learning curve: that it is smooth. In that method, one uses trial-averaging techniques (smooths) and bootstrap estimation (e.g., Davison & Hinkley, 1997) to attach probability values to hypotheses of the following form: The regression curve is not significantly different from that of the data-generating model. The method extends Azzalini, Bowman, and Hardle's (1989) approach to model selection by compensating for smoothing bias. By using a Wild bootstrap (Wu, 1986), one can obtain confidence intervals in a completely nonparametric manner (i.e., without assuming any parametric form for the noise; cf. Hardle & Marron, 1991).

We applied Brown and Heathcote's (2002) method to the data of Liu et al. (2003). The analyses returned the inferential probability values shown in Table 2. Higher probability

TABLE 2. Inferential Probability Values From the Method of Brown and Heathcote (2002)

Method	A	B	C	D	E	F	G	H
Exponential	.504	.044	.623	.677	.594	.073	.242	.038
Power	.452	.003	.622	.692	.288	.070	.275	.034

Note. A, . . . , H = Participants A through H.

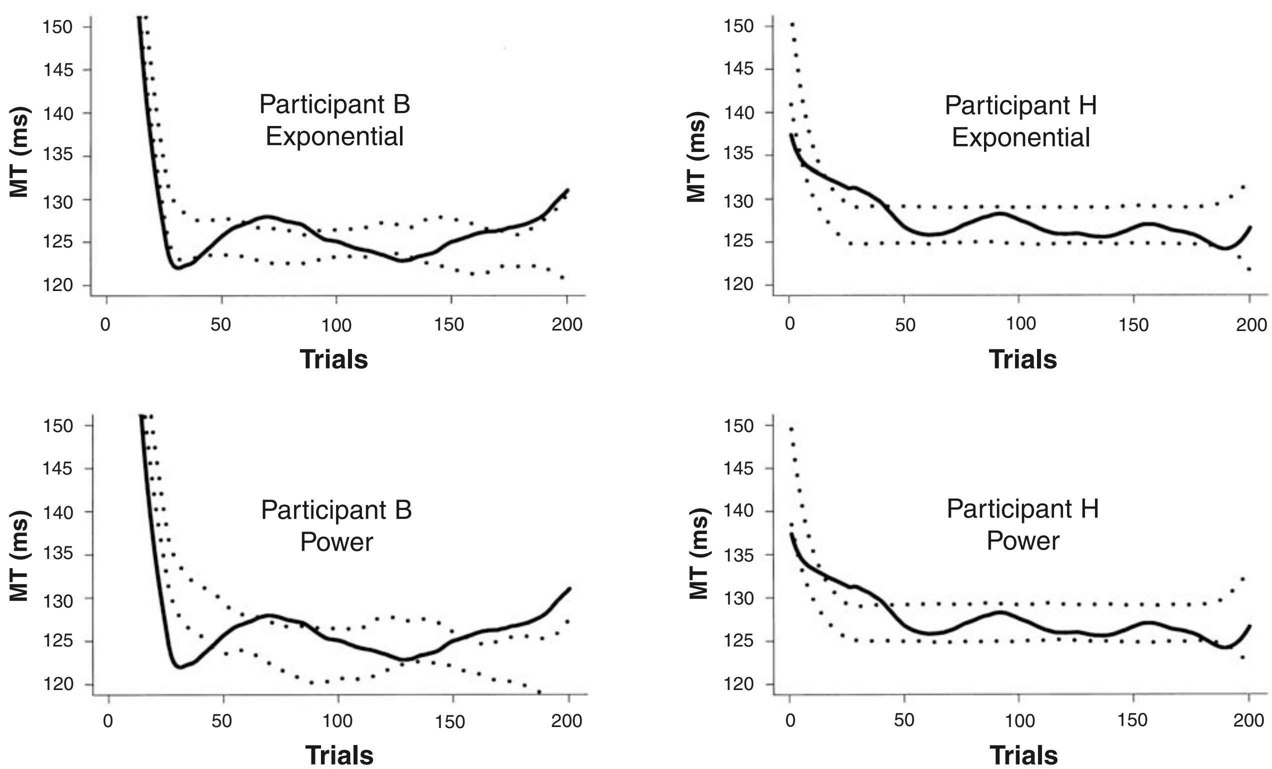


FIGURE 3. Nonparametric regression estimates for Participants B and H (solid lines), along with pointwise 95% confidence intervals (dotted lines) on the location of those estimates under the assumption of either power or exponential models. We modeled error structure by using 5,000 bootstrap samples from the residuals; we calculated smooths by using local linear regression with an Epanechnikov kernel of half-width $h = 2$ (Wand & Jones, 1995).

values indicate greater evidence in favor of the corresponding model. Comparison of power and exponential model probabilities are generally consistent with goodness of fit, although that is not necessarily so (see Brown & Heathcote for an example in which they clearly disagree). Figure 3 shows the smooths and corresponding confidence intervals for Participants B and H. As can be seen in Table 2, the data of Liu et al. (2003) did not have sufficient power to enable them to reject either model, except for those 2 participants. Both models performed fairly poorly for Participant H, and we concluded that neither is entirely adequate in that case. For Participant B, the data in Figure 3 show that the power function badly underestimated early performance and overestimated later performance. The exponential model showed some overestimation around Trial 75 but was generally much better. Figure 3 illustrates the value of confidence intervals generated from Brown and Heathcote's technique—the actual type of misfit can be identified, even in noisy data.

Discussion

Liu et al. (2003) suggested that traditional methods of discriminating between learning curve models are inadequate, and they proposed two new methods for going beyond curve fitting. However, by using absolute error rather than the raw movement time data in their implementation of curve fitting, they produced biased results. When we corrected that error, we found that, despite the high noise levels and rapid learning in the data of Liu et al., properly constructed and efficient statistical tests do clearly adjudicate in favor of an exponential function for half of Liu et al.'s participants. For the remaining participants, the tests also conveyed important information—that the data were too noisy to provide clear evidence either way.

The mistake of Liu et al. (2003) in assuming that Participant A showed no significant learning illustrates the importance of both formal statistical testing and trial averages in detecting slow trends obscured by high levels of noise. We also reviewed evidence that appropriate trial averages cause little or no distortion in curve form for the sorts of smooth models considered by Liu et al. Hence, although we agree that participant averages should be avoided, we differ from Liu et al. in recommending trial averaging as a useful method for investigating learning curves.

Neither of the methods of Liu et al. (2003) for going beyond curve fitting are useful. The FTI is extremely biased in favor of an exponential model. Therefore, it is no surprise that that method provided purportedly clearer evidence in favor of the exponential model than curve fitting did. Their R_n measure did not suffer from bias but was extremely intolerant of noise and had little power to discriminate between curve forms. Liu et al. claimed that failure to find significant decreases in R_n with practice favors an exponential function. That conclusion is invalid because it affirms the null hypothesis with a test that demonstrably has no power.

Our analyses have brought to light serious flaws in the methods of Liu et al. (2003). However, better methods are readily available. In the domain of curve fitting, the standard approach of choosing the model with the best least squares fit vastly outperformed the FTI method, and the nested-model testing method of Heathcote et al. (2000) was shown to provide a powerful inferential test of curve form. The method of Brown and Heathcote (2002), and the linear RLR plot method suggested here, fulfill the desire of Liu et al. to go beyond curve fitting in that they provide information about distinctive properties of, and systematic deviations from, learning curve models in a truly nonparametric manner. Despite the fact that our conclusions about their movement time curves do not differ from those of Liu et al. (we support an exponential model for most participants), we strongly recommend that researchers do not adopt their methods.

NOTES

1. That participant was chosen because that individual displayed the clearest evidence of learning in Figure 1. Apart from an overall decrease in discrimination, we obtained a similar pattern of results by using a wide range of learning function parameters and levels of noise (*SD* = 5, 10, 15, 50), and similarly for estimated parameters from Participants A and D in Liu et al. (2003).

2. By definition, any decreasing positive function must have RLR > 0, although that might be difficult to detect in a power function as RLR approaches zero with practice. Hence, failure to find RLR estimates significantly greater than zero could be interpreted as evidence against an exponential learning curve because its RLR remains greater than zero even with extended practice. We prefer an interpretation in terms of lack of power because of the very large fluctuations evident in RLR estimates.

3. One can also use higher order polynomials, or alternative sets of basis functions. Strictly, univariate regression is inappropriate, because $\overline{\mathrm{MT}(t+(1/2))}$ is measured with error. However, $\overline{\mathrm{MT}(t+(1/2))}$ necessarily has less error than $d\mathrm{MT}(t+1/2)$. Hence, the univariate regression is approximately correct, and the approximation is convenient because univariate regression software is widely available.

4. MatLab code for performing that analysis is available from http://www.newcastle.edu.au/school/behav-sci/ncl/software_repos.html.

REFERENCES

Aczel, J. (1966). *Lectures on functional equations and their applications*. London: Academic Press.

Azzalini, A., Bowman, A. W., & Hardle, W. (1989). On the use of non-parametric regression for model checking. *Biometrika, 76,* 1–11.

Bowman, A. W., & Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations*. Oxford: Clarendon Press.

Brown, S., & Heathcote, A. (2002). On the use of nonparametric regression in assessing parametric regression model. *Journal of Mathematical Psychology, 46,* 716–730.

Brown, S., & Heathcote, A. (2003a). Averaging learning curves across and within participants. *Behaviour Research Methods, Instruments & Computers, 35,* 11–21.

Brown, S., & Heathcote, A. (2003b). Bias in exponential and power function fits due to noise: Comment on Myung, Kim and Pitt. *Memory and Cognition, 31,* 656–661.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74,* 829–836.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.

Hardle, W., & Marron, J. S. (1991). Bootstrap simultaneous error bars for non-parametric regression. *Annals of Statistics, 19,* 778–796.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). Repealing the power law: The case for an exponential law of practice. *Psychonomic Bulletin and Review, 7,* 185–207.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2003). *Average curves and repeated-measures analysis of variance*. Manuscript submitted for publication.

Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics, 40,* 633–643.

Liu, Y.-T., Mayer-Kress, G., & Newell, K. M. (2003). Beyond curve fitting: A dynamical systems account of exponential learning in a discrete timing task. *Journal of Motor Behavior, 35,* 197–207.

Wu, C. F. J. (1986). Jackknife, bootstrap, and other resampling methods. *Annals of Statistics, 14,* 126–1343.

Luce, R. D. (1986). *Response times*. New York: Oxford University Press.

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. London: Chapman & Hall.

Submitted July 25, 2003
Revised October 20, 2003