The Law of Practice and Localist Neural Network Models

Andrew Heathcote and Scott Brown

Department of Psychology, The University of Newcastle, Australia

University Avenue, Callaghan, 2308, NSW, Australia

heathcote@psychology.newcastle.edu.au and sbrown@psychology.newcastle.edu.au

http://psychology.newcastle.edu.au/

Abstract

An extensive survey by Heathcote, Brown, and Mewhort (in press) found that the Law of Practice is closer to an exponential than a power form. We show that this result is hard to obtain for models using leaky competitive units when practice affects only the input, but that it can be accommodated when practice affects shunting self-excitation.

In a recent survey, Heathcote, Mewhort and Brown (in press) analyzed the form of the Law of Practice in 7910 practice series from 475 subjects in 24 experiments using a broad range of skill acquisition paradigms. When the practice series were not averaged over subjects or conditions, an exponential function (mean response time, $RT = A + Be^{-\alpha N}$, where A is asymptotic RT, B it the amount that learning decreases RT, and N is practice trials) provided a better fit than a power function ($RT = A + BN^{-\beta}$) for the majority of cases in every paradigm. The defining property of an exponential function is that its relative learning rate, $RLR = -{}^{dRT}\!/_{dN}/(RT - A)$ equals a constant (α). In contrast, the power function's RLR decreases hyperbolically to zero, $RLR = \beta/N$. Previous findings in favor of a power function (e.g. Newell & Rosenbloom, 1981) used practice series averaged over subjects and/or conditions. When exponential practice series with different rates (α) are averaged, the RLR of the average decreases, because fast learners (with large α) control the rate of change early in practice, while slow learners (with small α) dominate later in practice (see Brown & Heathcote, in preparation, for detailed analyses of averaging effects). As theories of skill acquisition model the behavior of individuals, not averages, Heathcote, et al. concluded that the

"Law of Practice" is better characterized by an exponential than a power function. Hence, the power function prediction made by Page's model does not accord with recent empirical results.

We believe that an exponential law of practice is extremely difficult to obtain using Page's approach to practice effects in competitive leaky integration networks (Equation 5). To see why, consider the time (t) it takes the activation ($x(t)$) of a leaky integrator ($dx/dt = I - kx$, where I is input and k is leakage rate and $x(0) = 0$) to reach a criterion χ.

$$t = \frac{1}{k}\ln\left(\frac{I}{I - k\chi}\right) \qquad (1)$$

The *RLR* of (1) with respect to I decreases to zero. If we assume, as Page does, that practice decreases t by increasing I, the *RLR* of (1) with respect to N will decrease to zero unless $I(N) \geq O(N^2)$ for large N. Such a faster than linear increase in input is difficult to justify. The increase of I with N is slower than linear for Page's "noisy-pick-the-biggest" model. Even if all instances, rather than just the maximally activated instance, were to contribute to I, the increase would be only linear. Page's simulation results (Figure 6) indicate that the same power like effects of increasing I apply to the time it takes competing leaky integrators to pass an activation criterion.

However, competitive leaky integrators can account for Heathcote et al's (in press) findings if practice alters shunting terms, such as the weights of self-excitatory connections[1]. Consider a two-unit system of the type discussed by Usher and McClelland (1995), with normalized inputs I and $(1 - I)$ and linear threshold transfer functions:

---

[1] We also obtained an exact exponential result for inputs that 1) increase with practice according to a learning rule like Page's Equation 2 ($I = M(1-e^{-\lambda N})$), 2) are non-stationary (decreasing with presentation time, t, as $I = 1/(t + (b - cI))$, $b/c > M$), and 3) have a shunting effect on a single unit's activation ($dx/dt = (U - x)I$). We will not pursue this model here, as it is very different from Page's approach (but see Heath, 1992, and Smith, 1995, for more on non-stationary inputs, and Heathcote, 1998 for more on shunting inputs).

$$^{dx_1}\!/_{dt} = I - (k - \varepsilon)x_1 - \delta x_2 \qquad (2)$$

$$^{dx_2}\!/_{dt} = 1 - I - (k - \varepsilon)x_2 - \delta x_1 \qquad (3)$$

A response is made when the activation of one unit exceeds a criterion, χ. Assume that as practice proceeds, the self-excitatory weight, ε, approaches the leakage rate, k, using a weight-learning rule like Page's Equation 2:

$$^{d\varepsilon}\!/_{dN} = \lambda(k - \varepsilon) \qquad (4)$$

In simulations with gaussian noise added to (2) and (3) at each step of the integration (Page's N_1 term in his Equation 5) and larger values of *I* so errors did not occur, learning series were consistently better fit by an exponential than a power function. Insight into this result can be gained from the analytic result for the one unit case (i.e. Equation 2 with competitive weight, $\delta = 0$, which was also better fit by the exponential in simulations):

$$t = \frac{1}{k} e^{\lambda N} \ln\left(\frac{I}{I - k\chi e^{-\lambda N}} \right) \qquad (5)$$

For a linear Taylor approximation to (5), *RLR* decreases marginally with *N*, but asymptotically approaches λ rather than zero. Heathcote, et al. (in press) found that an APEX function ($RT = A + Be^{-\alpha N}N^{-\beta}$), which has a *RLR* that decreases to an asymptote greater than zero, consistently fit slightly better than an exponential function. We found the same pattern of fit to our simulation results for both the one and two-unit models. The parameter estimates for these fits also concurred with the survey results. Estimates of the power function *A* parameter were implausibly small (as *N* increases *t* approaches χ/I for the linear Taylor approximation to (5), whereas most power function *A* estimates were zero). Fits of a power function with an extra parameter (*E*) to account for prior practice ($RT = A + B(N+E)^{-\beta}$) produced implausibly large *B* estimates, mirroring Heathcote et al's findings with the survey data.

Given limited space it is not possible to quantitatively examine this type of model further (see Heathcote, 1998, for related findings and Heathcote & Brown, in preparation, for a detailed analysis). However, the findings presented are sufficient to demonstrate that Heathcote et al's (in press) results are not incompatible with the overall localist neural network approach. Indeed, learning in shunting connections, both self-excitatory and competitive, provides an adaptive mechanism for consolidating and differentiating local response representations (cf. Usher and McClelland, 1995, who note that the “units” in such models may correspond to collections of neurons bound together by mutually excitatory connections). Reduced leakage with practice can also explain Jamieson and Petrusik’s (1977) finding (cited in Usher & McClelland, 1995) that the difference between error and correct RTs decreased with practice. As leakage approaches zero, a leaky integrator approximates a classical diffusion process, for which error and correct RTs are equivalent.

References

Brown, S., & Heathcote, A. (in preparation). The effects of averaging on the determination of curve form.

Heath, R. A. (1992). A general nonstationary diffusion model for two-choice decision-making. Mathematical Social Sciences, 23, 283-309.

Heathcote, A. (1998). Neuromorphic models of response time, Australian Journal of Psychology, 50, 157-164.

Heathcote, A., & Brown, S. (in preparation). Neuromorphic models of the Law of Practice.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (in press). The Power Law repealed: The case for an Exponential Law of Practice. Psychonomic Bulletin & Review.

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.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. Psychological Review, 102, 567-593.