Human learning:
Power laws or multiple characteristic time scales?

Karl M. Newell,   Gottfried Mayer-Kress   Yeou-Teh Liu
The Pennsylvania State University   *Taiwan Normal University*

The central proposal of A. Newell and Rosenbloom (1981) was that the power law is the ubiquitous law of learning. This proposition is discussed in the context of the key factors that led to the acceptance of the power law as *the* function of learning. We then outline the principles of an epigenetic landscape framework for considering the role of the characteristic time scales of learning and an approach to system identification of the processes of performance dynamics. In this view, the change of performance over time is the product of a superposition of characteristic exponential time scales that reflect the influence of different processes. This theoretical approach can reproduce the traditional power law of practice – within the experimental resolution of performance data sets - but we hypothesize that this function may prove to be a special and perhaps idealized case of learning.

The paper of A. Newell and Rosenbloom (1981) was a landmark publication on human learning. The central and fundamental proposal of this paper was that the power law is the ubiquitous law of learning. This idea was inspired by universal properties of phase transitions in physical systems and was the basis for the development of a chunking theory of information processing in human learning.

It is only recently, however, that direct theoretical and empirical challenges and developments have been given to the idea of the power law as a universal law of learning (Gallistel, Fairhurst, & Balsam, 2004; Heathcote, Brown, & Mewhort, 2000; K. Newell, Liu, & Mayer-Kress, 2001). There was a follow-up paper on the chunking theory of learning by Rosenbloom and Newell (1987) but it is as if the broad and thorough analysis by synthesis of A. Newell and Rosenbloom (1981) of extant papers evaluating the functions of learning left little room for doubt about the veridicality of the proposition of the power law as the law of learning. Many authors subsequently endorsed the centrality of the

power law in learning (e.g., Anderson, Fincham, & Douglass, 1999; Ivry, 1996; Logan, 1988; Salmoni, 1989), but scientific activity on this problem was very limited until some 20 years later and the turn into the 21st century. On reflection, it appears, as in other scholarly domains (Krieger, 2006), that the proposal of a universal power law has actually stifled investigation, rather than being a goal and guide to vigorous theoretical and experimental activity.

Here we briefly outline the legacy of the A. Newell and Rosenbloom (1981) paper in regard to the function(s) of learning. Some important challenges to the empirical basis of the power law are then subsequently developed. These themes provide the background to the proposal, based on a landscape model of the dynamics of learning (K. Newell et al., 2001; K. Newell, Liu, & Mayer-Kress, 2003, 2005), of multiple characteristic time scales reflecting the deterministic change of performance dynamics. In this view, the power law is an idealized law or limiting case of learning that in practice is rarely if ever achieved.

Legacy of the A. Newell and Rosenbloom (1981) Paper

The A. Newell and Rosenbloom (1981) paper is a quantitative account of practice and learning effects that provided the basis for the development of their information

processing based chunking theory of learning. At that time, the power law was primarily discussed in regards to learning for the change of performance in perceptual-motor skills but A. Newell and Rosenbloom wanted to generalize this relation to all contexts of human learning – hence, the notion of a ubiquitous law of learning. In the 20th century, learning theory had considered several functions of change including the exponential, hyperbolic and logistic (Lane, 1987; Mazur & Hastie, 1978), but had tended to de-emphasize or even fail to recognize the potential of the power law.

Snoddy (1926) is generally regarded as the first to propose a power law for human learning based upon his non computerized log-log plots of mirror tracing performance over practice trials and days. This data set, together with the now famous log-log plot of factory worker cigar roller performance over 10 million trials and 7 years of practice (Crossman, 1959), are usually two of the key examples promoted by advocates of the power law of learning. Indeed, A. Newell and Rosenbloom gave these two studies a prominent place in their evaluation of experimental data sets that support a power law even though both of these data sets reveal significant limitations for such an inference to be made (K. Newell et al., 2001). Furthermore, it is noteworthy that subsequently Snoddy (1935) had already begun to question the generalizability of the power law as the function of learning in deference to the relevance of the single log plot – the exponential.

A small set of power laws that vary in the number of parameters fitted to the data have been used in the assessment of human learning (Lane, 1987; A. Newell & Rosenbloom, 1981). A. Newell and Rosenbloom (1981) compared fits of a 4 parameter power law and 3 parameter exponential and hyperbolic equations to some 18 different learning data sets. The power law was:

$$T = A + B (N + E)^{-\alpha} \qquad (1)$$

where T is task or performance time, A is the asymptote of learning, B is the performance on the first trial, N is the number of trials, performed after a reference number E reflecting the experience from E previous trials, and α is the exponent of the power law describing the performance. An analysis of the parameters for the best fits as well as the estimate of the fit (r^2) was provided. The primary outcome of the analysis was evidence in terms of percent of variance estimates for the claim of the ubiquity of the power law of learning in all types of human behavior.

The theoretical significance of the determination of a power law of learning is that by definition it leads to the position that in learning the rate of change in performance does not remain constant (that is proportional as in the exponential) over trials. In other words, there is not a constant time scale of change in a power law model of learning. Indeed, by definition a power law has an infinite number of time scales over the range of the phenomena studied (Schroeder, 1991). The A. Newell and Rosenbloom (1981) chunking theory of learning was built in part from Miller's (1956) original notion of a chunk - a structured stored expression of knowledge. This construct was combined with the assumption of an exhaustion model of learning in which there is a diminution of aspects of the learning process so that, in other words, learning slows down from proportional change in the later stages of learning. Thus, the building of the chunks into larger and larger segments of patterned information takes increasingly more time in a relative sense.

In our view, the legacy of the A. Newell and Rosenbloom (1981) paper 25 years later is to be found primarily in the description of the learning process as a power law – in short, the idea that the power law is the universal law of learning. In contrast, the chunking theory of learning that was created around the power law description is much less prominent and influential today in theorizing about learning. However, as hinted earlier, it is only in the recent past that the claim of the power law as the universal law of learning has been subjected to a theoretical and experimental challenge that begins to match the enormity of the original claim of A. Newell and Rosenbloom (1981).

The Generality of the A. Newell and Rosenbloom (1981)

Power Law of Learning

A. Newell and Rosenbloom (1981) reported the modeling of exponential, hyperbolic, and power law functions to a large number of learning data sets. Their analysis by synthesis concluded for the ubiquity of the power law in learning though it should be noted that only a few years earlier Mazur and Hastie (1978) conducted a similar analysis and concluded for the exponential. This contrast in viewpoint points up that there are several factors that mediate the interpretation of existing data sets and importantly, that act as a goal and a guide for the creation of a general theory of learning. In addition, it should be noted that the difference from a percent of variance standpoint of the fits of the candidate functions of learning is often very small. And, even when it is not, it is typically less than the often held working norm in studies of fractals and power laws (Bak, 1996; Bassingthwaighte, Liebovitch, & West, 1994) of observing change over at least three orders of magnitude, due to the limited amount of practice in learning experiments that is often conducted and measured. Here we briefly mention three general factors that have influenced and perhaps led to the broad-based acceptance of the generality of the power law for learning.

Averaged data: A major factor in mediating the interpretation of the laws of learning is that of averaged data (Heathcote et al., 2000; K. Newell et al., 2001). Averaging data over subjects and trials tends to lead to a different function for the change in performance over time than is present in a single individual. In our view, this is because the averaging of data, whether over participants or trials (and most studies have done both) changes the time scale of performance over time or trials. In K. Newell et al. (2001) it was shown that averaging exponentials that have different exponents leads to a function that approximates a power law because it brings in the many different time scales of change (exponents) to the averaged function. It should be noted that most of the data sets analyzed in A. Newell and Rosenbloom (1981) were of group averaged data and they gave no recognition to the problem of averaging data mediating interpretations of the laws of learning.

Model fits: There has been considerable progress since the publication of the A. Newell and Rosenbloom (1981) paper in regard to mediating issues in model testing (Burnham & Anderson, 2002; Pitt & Myung, 2002; Pitt, Myung, & Zhang, 2002; Roberts & Pashler, 2000). The value of a goodness of fit criterion has been challenged as insufficient in decision making regarding the usefulness of a model and this holds particular relevance when the difference in the percent of variance accounted for in exponential, hyperbolic and power law fits is often very small (less than 1%). This leads to a parallel emphasis on the qualitative properties of the fit (Liu, Mayer-Kress, & Newell, 2003) and other criteria that reflect the number of parameters, complexity of the model (Pitt et al., 2002), and what states a model cannot predict. In this regard, it is relevant that A. Newell and Rosenbloom (1981) used a 4 parameter power law but only 3 parameter exponential equation, a factor that probably biased their examination of the best fitting function toward a power law. It is also the case that in this percent of variance fit approach the values of parameters were not used by A. Newell and Rosenbloom (1981) for characterization and they sometimes changed in unrealistic ways, such as in the sign switching from positive to negative or vice-versa, which implies mutually exclusive theoretical explanations.

Theory driven curve fitting: Thurstone (1919) emphasized the importance of rational versus empirical curve fitting in the study of learning curves. This viewpoint provides a most general statement of the value of theory as a goal and guide in the testing of model equations. A. Newell and Rosenbloom (1981) in their analysis of learning data sets were certainly driven a priori by the hypothesis of the power law. It is difficult to determine if the chunking theory of learning was sitting behind so to speak their expectation of the priority of the power law or whether the establishment of the power law preceded the development of the learning theory. Irrespective of their position on this important issue of rational curve fitting, it is safe to say that the general implementation of Thurstone's call for theory as a goal and a guide to the analysis of learning curves has played a secondary role to the percent of variance curve fitting criterion.

In summary, the above represent some of the major factors that mediate the determination of the laws of learning, although there are other contributing issues (Lane, 1987). The net consequence in our view is that there are very few strong examples (if any) of the power law capturing the qualitative and quantitative properties of performance dynamics in learning. Virtually all the studies reporting data consistent with a power law are on averaged data (see A. Newell & Rosenbloom, 1981). Indeed, we are not aware of any published studies showing a power law of motor learning on individual data (see our observation later regarding the Snoddy, 1926 data).

The highlighting of the impact of these issues in assessing the function of learning has led in recent years to challenges to the position of A. Newell and Rosenbloom (1981) that the power law is the ubiquitous law of learning (Gallistel et al., 2004; Heathcote et al., 2000; K. Newell et al., 2001). It is ironic that the challenges to the position of the power law as the ubiquitous law of human learning come at a time when there is increasing evidence for the presence of multiple time scales in both nonliving systems (Bak, 1996; Mandelbrot, 1977) and biological and cognitive processes (Bassingthwaighte et al., 1994; Gilden, Thornton, & Mallon, 1995; Van Orden, Holden & Turvey, 2003; Ward, 2002). It is our position that an understanding of the dynamical basis of a power law and the role of multiple time scales to system output can reveal both the theoretical potential and experimental reality of power-law like human behavior, including that of learning.

Power Laws, Time Scales and System Identification

There has been an increased understanding of the importance of characteristic time scales in the study of human behavior over the last 25 years. The construct of multiple characteristic time scales has become central in dynamical systems and connectionist approaches to behavior and physiology, and more generally neuroscience approaches to human behavior. Indeed, the infinitely many time scales of events and processes are present by definition in 1/f fractal-like processes of living and nonliving systems (Bak, 1996; Bassingthwaite, Libovitch, & West, 1994; Mandelbrot, 1977, Schroeder, 1991).

Time Scales. The phrase or term characteristic time scales

is, however, rarely defined in most instances of its use in the study of human behavior. This oversight is unfortunate because it contributes to the use of different meanings to the phrase characteristic time scales and allows its embellishment to the notion of multiple time scales also to be invoked in less than rigorous ways. The notion of characteristic time scales has an obvious link to the concept of time and for many interpretations of time scale it is simply that – the time or duration of an event, process to unfold or an action.

It is relevant to the focus of this paper that the notion of time scales has rarely been mentioned never mind discussed in the use of the power law to analyze or reflect human learning. Neither have the exponents, that are critical in the context of phase transitions, been used to characterize classes of systems or conditions for learning. Thus, the use of the power law in human learning has been largely that of a curve fitting percent of variance accounted for procedure rather than a theoretically motivated exercise. In the sense then of Thurstone (1919), the curve fitting of the power law since Snoddy (1926) to learning data has been an empirical implementation uninfluenced by the theoretical construct of characteristic time scales and their dynamical, epigenetic, or physiological origin. The theoretical implication of the power law reflecting a scale invariance in the performance dynamics over time has not been approached directly, though A. Newell and Rosenbloom (1981) came as close to these considerations as anyone in the last 75 years of power law analysis of learning through their postulation of the changing chunks of information processed.

In our view, the phrase characteristic time scales holds clues to its use here in a dynamical systems framework to human performance and learning. It directly implies the notion of the different scales of time where the phrase different scales refers to different units that are most appropriate to characterize the passage of time, e.g. micro seconds of chemical processes to the scale of years in the process of development or expert skill learning. Indeed, the concept of time has been refined throughout history in part due to the analysis of new time scales.

The time scale of a day is the most fundamental element of all calendars and is based on the dynamical process of the circadian rhythm of earth rotation. However, not all frameworks of time are based upon the rotation of the earth. The forms of time known as dynamical, atomic and coordinate time are not based on the standard idea of the day (Haliday, Resnick, & Walker, 2005), but they do relate to properties of dynamical systems.

In dynamical systems there are two types of idealized motions that naturally introduce the fundamental concept of time scales. These motions have either periodic oscillations

(Haken, 1983) or growth/decay at a constant rate (Kaplan & Glass, 1995). In the case of oscillatory systems, the intrinsic time scale of the system is the period (inverse of frequency). In growth/decay systems the intrinsic time scale is the inverse of the growth/decay rate that can also be expressed as "doubling" or "half life" time by including a factor of $\log_2$.

These two classes of behavior and combinations thereof are in linear dynamical systems the only forms of movement observed apart from the trivial dynamics of a non-moving fixed point. They are the basis of almost all known clocks: earth rotation, a swinging pendulum, the vibrations of a quartz crystal, and cesium atoms represent clocks based on periodic movements. Clocks based on radioactive decay, such as carbon dating or the shortest possible measurable time scales, the lifetimes of elementary particle, are examples of methods using constant rates of change and exponential functions. Other methods of time keeping, such as water clocks, hourglasses, burning incense and so forth are described not by linear but by differential equations that change with a constant speed, not a constant rate, i.e. the speed of change does not depend on the size of the variable. Clocks at cosmic scales use the redshift of distant galaxies and are also not described by simple exponentials.

For both the oscillatory and growth/decay types of exponential trajectory the performance $x(t)$ at time t is expressed as the real part of a complex exponential function:

$$x(t) = A \, Re(\, e^{(\gamma + i\omega)t}) = A \, e^{\gamma t} \, \text{Cos}(\omega t) \qquad (2)$$

where A is the observed performance at time $t = 0$, the real part γ of the exponent is the relaxation (decay or approach) rate and the imaginary part ω is the frequency of the oscillation. The reciprocal of the growth or decay rate is the intrinsic or characteristic time scale. Thus, a time scale is not simply the duration or time of an event as it often used in a colloquial fashion but significantly for our purposes here the characteristic time duration of an event that arises from a periodic or growth/decay dynamical process.

Equation 2 describes two types of motions in nonlinear dynamical systems (Kaplan & Glass, 1995; Strogatz, 1994). These oscillatory and growth/decay processes are fundamental to describing behavior close to a fixed point. A fixed point is the mathematical concept that is associated with the equilibrium regions of the dynamics. Fixed points correspond to the absence of motion as in a pendulum at rest. Furthermore, and importantly for the implementation of this approach to the change associated with learning and development, the motion close to a fixed point can be approximated to be described by linear dynamical systems. This means that the motion of the trajectory can be characterized by the exponential function of Equation 2.

Expressed another way, the motion close to a fixed point can be completely characterized by the exponents γ and ω. The time scale within a growth or decay process to a fixed point is characterized by the time for the dynamics to double/half the distance to the fixed point:

Here we are especially interested in systems that approach a fixed point according to a distance function given by an exponential:

$$x(t) = A\, e^{\gamma\, t} \qquad (2a)$$

where we assume that the approach rate is positive, *i.e.* $y > 0$. Now we can ask how long it will take until our distance to the goal has decreased by on half. Let us call this time $T_{1/2}$ the half-distance time associated with this characteristic time scale y. That means we have:

$$x(t+ T_{1/2}) = A\, e^{\gamma\, (t+T_{1/2})} = \tfrac{1}{2}\, A\, e^{\gamma\, t} \qquad (3)$$

Solving this equation for $T_{1/2}$ yields:

$$T_{1/2} = \log 2 \,/\, \gamma \qquad (4)$$

This time $T_{1/2}$ is given in standard time units such as "trial number" or "hour". Thus, it provides us with a convenient estimate after how many trials of practice we can expect to be 50% closer to our learning goals or after how many hours of rest we can expect that our performance has dropped by 50% due to warm-up decrement.

These basic assumptions about attractor dynamics and fixed points provide the theoretical basis for the assessment of the characteristic time scales of change in motor learning and development (K. Newell et al., 2001, 2003, 2005). In our approach, the multiple time scales of change in task outcome over time are interpreted to originate from the system's trajectory on an evolving attractor landscape. Different bifurcations between attractor organizations and transient phenomena can lead to exponential, power law or S-shaped learning curves, among other pathways of change. Thus, in our theoretical framework to learning there is not a single law (function) of learning as has been sought in the exponential, logistic, hyperbolic and power law but rather a coherent single set of dynamical principles that can lead to different functions of change in task outcome and limb trajectories. K. Newell et al. (2001, 2003, 2005) outlined more complicated cases of fixed point dynamics, landscape evolution and their relation to the dynamics of motor learning.

System Identification. The notion of characteristic time scales arising from the dynamics provides a principled basis to consider the change in performance over time and the constructs of learning, development and adaptation. A central consequence of this approach is that the characteristic time scales of oscillations and growth/decay processes of a dynamical system allow the system identification of component processes. This engineering based strategy has been used in psychology and biology for the analysis of oscillatory processes at different levels of analysis of the system as a reflection of a range of constructs, but the dynamics of growth/decay processes have been invoked less in behavioral studies. Kaplan and Glass (1995) have proposed that the equation for exponential growth and decay is probably the most important dynamical model in biology.

In the study of brain processes through electroencephalography (EEG) it has long been recognized that the frequency bandwidths (alpha, beta 1, beta 2, gamma, theta, delta) of the EEG signal(s) are indicative of different processes (Neidermeyer & da Silva, 1999; Nunez, 2000). This assessment is based on the higher probability of these frequencies occurring in the composite output of the EEG signal. Thus, frequency decomposition through spectral analysis procedures has become an important strategy for understanding the processes of the brain activity. Similarly, at the behavioral level the analysis of tremor is largely based on the frequency decomposition of the tremor signal (Elble & Koller, 1990). In finger postural tremor, for example, there are two frequency ranges that dominate the signal, the 8-12 Hz bandwidth of physiological tremor and the 20-25 Hz bandwidth of mechanical resonance of the finger system.

The frequency bandwidths of both EEG and EMG signals tend to respectively, speed up and slow down a little in development through the years of childhood and later in the period of old age. This change reflects the imposition of another (slower) time scale of development on the dynamics of moment to moment control processes of brain and behavior. These approaches in the analysis of EEG, EMG and limb kinematics are examples of the characteristic time scales (period) of the signal oscillation being used to identify processes of system regulation.

As we remarked earlier, the time scale that arises from the growth away from or decay rate to a fixed point (where the rate of change is zero) can also be used to identify the dynamic processes of change in learning and development. This alternative strategy of decomposing the processes of performance change uses the real parts of the characteristic exponents provided in Equation 2. This approach to the study of change over time has been used extensively in different branches of biology with population dynamics being an example. Perhaps the best known model of growth dynamics is that of Verhulst (1938) who argued that growth cannot continue indefinitely and that the basic exponential equation required an adjustment. Verhulst proposed the logistic population growth function that has two fixed points but importantly from the theoretical perspective proposed here is still driven by the growth/decay properties
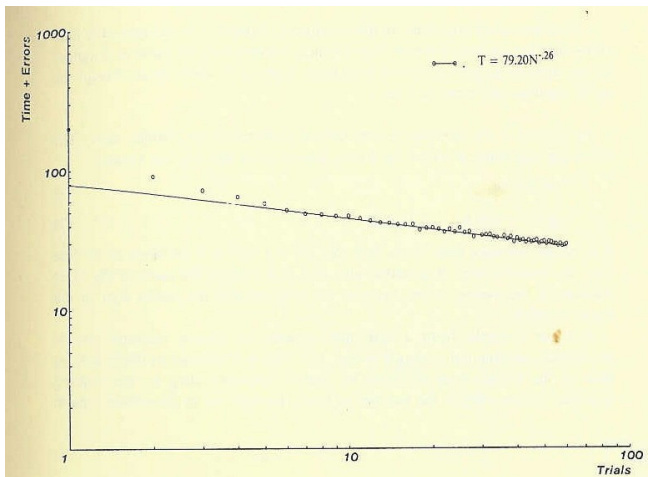
$$T = 79.20N^{-.26}$$

Figure 1. The power law fit of A. Newell and Rosenbloom (1981) to the Snoddy (1926) mirror tracing learning data (reproduced with permission from A. Newell and Rosenbloom, 1981, Figure 1.1). The performance score (p) here is Time + Error whereas Snoddy (1926) and Figures 2-4 here use 1000/T+E which is why the direction of change in performance score as a function of practice is opposite here in Figure 1 to that of Figures 3-4.

of exponentials.

In summary, we are suggesting that the exponential dynamics of growth/decay to a fixed point can be used as the theoretical basis for considering change in human learning. It is not that all functions of performance change over time will be an exponential but rather that exponential change and the associated influences of bifurcations and transient dynamics can be constructed from this principled basis (K. Newell et al., 2001). Thus, in our approach the analysis of the learning curve is driven by the principles of dynamics and not the percent of variance criterion, although both qualitative and quantitative aspects of the model fit to experimental data are important.

Landscape Model of Characteristic Time Scales

The traditional definitions of learning have included a number of constructs and properties but a general and central notion has always been that learning is about the relatively permanent change in behavior (e.g., Hilgard & Bower, 1975). This definition separates the relatively permanent from the relatively transient changes in performance, a statement that is, by definition, a relative one. Furthermore, learning is an inference drawn from the change in performance over time and, therefore, is viewed as an unobservable or latent construct. Thus, curve fitting the performance changes over time is implicitly or explicitly an attempt to capture the relatively permanent change in performance.

The transient changes in performance have often been interpreted in terms of noise-like processes and it is this

assumption that sits behind the experimental analysis strategy of averaging data. That is, averaging data over blocks of trials reduces the trial to trial random-like fluctuations that may mask the revelation of the deterministic processes of learning. Similarly, averaging data over participants reduces the between participant variance and affords a less contaminated assessment of the relatively permanent change over time. Of course, in most learning studies, the data are averaged over both trials and participants thus, as noted above, this leads to the masking or even the mediating of the persistent change.

The general problem is that the performance outcome over time reflects the contribution of many processes to the behavior that themselves may change as a function of learning. The idea that the performance change observed reflects a single deterministic function superimposed by random like processes has been a useful guide to learning theory. But, this is, in effect, a signal plus noise model that appears, as in many other domains too simplistic to capture the potential multiple time scales of processes that influence learning. Indeed, there are several potential factors that can change the relatively permanent and transitory changes of performance dynamics and that can mediate inferences about learning.

The transitory changes could come from a variety of sources such as: short-term shifts in attention or motivation, noise–like properties of the numerous levels of the biological system, warm-up properties at the beginning of a practice session, and fatigue-like properties that produce decrements in performance toward the end of a practice session. These sources of relatively transient change are all mediators of performance and potential maskers of the function of relatively permanent change of learning. Nevertheless, there has been little attempt to distinguish these processes analytically from the relatively permanent change in performance dynamics, the result being that at best these factors are all viewed as contributors to the error term in percent of variance fits of mathematical functions to performance data over time. Another way of saying this is that constructs such as motivation, attention and fatigue are, in effect, seen experimentally as neutral influences in the change in performance over time even though there is agreement that these processes are influencing performance dynamics and its change over time.

It is noteworthy too in the context of our framework that these mediating processes of learning are all seen as not only relatively short term but also reversible. In other words, these effects, except perhaps for the idea of noise like processes, can be reversed usually by the influence of sufficient rest between trials or practice sessions. These changes then are seen as distinct from the processes of

learning but mediators of performance dynamics. Processes that influence performance and not learning have also been called processes of adaptation, in the sense that the relatively permanent changes of learning are generally assumed to engender a structural change in the representational system (Hallett & Grafman, 1997; Shadmehr & Wise, 2005).

Learning, retention and transfer are related constructs that are different aspects of the same principles. Indeed, these constructs have all been related to the core issue of the representation for the action. A central notion here in the context of this paper is the idea of the strength of the memory trace expressed as it is in different theories of learning and memory (Adams, 1971; Anderson, Fincham, & Douglass, 1999; Kandel, 2006; Tulving, 2000).

The candidate short term transient influences on performance dynamics may not necessarily be related to representation and memory processes. For example, the short term shifts in attention and motivation through practice trials are not typically interpreted as due to memory loss or loss of memory trace strength. Similarly, the negative effects of fatigue on performance are not generally understood as a reflection of losses of memory strength. Furthermore, the influences of warm-up decrement at the beginning of a practice period following rest interval also may not be related to memory losses (Irion, 1949; Nacson & Schmidt, 1971), although this performance decrement has also been interpreted on the basis of changes in memory trace strength (Anderson et al., 1999).

It is our position that the influence of these transient adaptive processes to the performance dynamics can be isolated from the relatively permanent changes of learning through the principles of system identification that were outlined earlier. This is based on the theoretical assumption that these processes will each have their own characteristic time scales. We now provide an example of this theoretical approach through a consideration of the characteristic time scales of adaptation and learning associated with the short fast time scale of warm-up decrement and the slow time scale of the relatively permanent change over time.

Characteristic Time Scale Modeling of Learning

It is generally recognized that the original demonstration of power law learning was that of Snoddy (1926). He showed through visual inspection that a straight line on log-log paper could well approximate the change in the group performance data from the trials and days of practice of a mirror tracing task. This data set was one of the primary examples in the A. Newell and Rosenbloom (1981) group of studies used to examine the relative fits of exponential, hyperbolic and power law functions. Figure 1 is a

reproduction from A. Newell and A. Rosenbloom (1981, Fig 1.1) of their power law fit to the Snoddy (1926) data. In their quantitative assessment A. Newell and Rosenbloom (1981) showed that a 4 parameter power law was the best fitting function to the data accounting for 98% of the variance. The formula used was the 4 parameter power law of Equation (1). The function fitting analysis was done on group averaged data[1].

Perusal of Figure 1 shows that the biggest departure of the Snoddy (1926) data from the power law function is on the first few trials of day 1. This significant departure of the fit on the early trials is not uncommon in assessments of the power law function of learning. A. Newell and Rosenbloom (1981) indicated that they omitted some data early and late in analysis of some data sets but we do not know if this was done in their analysis of Snoddy (1926).

The power law function can cover up important details in the performance dynamics that are lost in the log scales of power laws. Figure 2, reproduced from Mayer-Kress, Liu, and Newell (1998) shows another more typical distribution of practice data set analyzed from Snoddy (1926) where 100 subjects practiced 20 trials a day with a 1 min rest between trials for 4 days. The data were fit with a 3 parameter exponential over all trials and by day (top figure 2a) and a 3 parameter power law over all trials and by day (bottom figure 2b). The figures clearly show that there is more detail in the Snoddy data set than is being revealed by a power law. In particular, there are strong trends of warm-up decrement (Adams, 1961) at the beginning of each day that is passed through by the straight line of the power law.

Our theoretical perspective is to model the learning data with a view to testing the contributions of characteristic time scales to the performance dynamics. The main ingredient of a landscape based theory of learning is the identification of behavioral patterns as locations in a landscape and the performance score values as elevation levels within this landscape (Newell et al., 2001, 2003). Following a tradition in the physical sciences we assign a goal state a lowest elevation in the landscape. Figure 3 shows a landscape with two distinct time scales reconstructed from the learning data

---

[1] It should be noted that A. Newell and Rosenbloom (1981) stated that the data in this figure were from a single subject but they did not state explicitly which data set of Snoddy (1926) they were presenting. There are no single subject learning data presented in Snoddy (1926) and our interpretation is that they were presenting a power law function for the data from Table 2 of Snoddy (1926) with 80 subjects having the unusual practice schedule of only 1 trial per day of practice for each of 60 days.
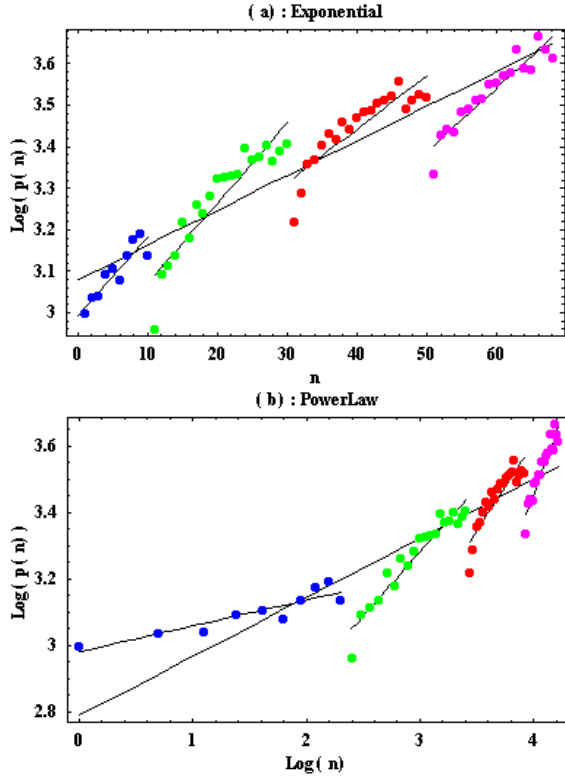
Figure 2. The Snoddy (1926, Figure 4 adapted) data set plotted with exponentials (over sessions and all days) and power laws (over sessions and all days) (both reproduced with permission from Mayer-Kress et al., 1998).
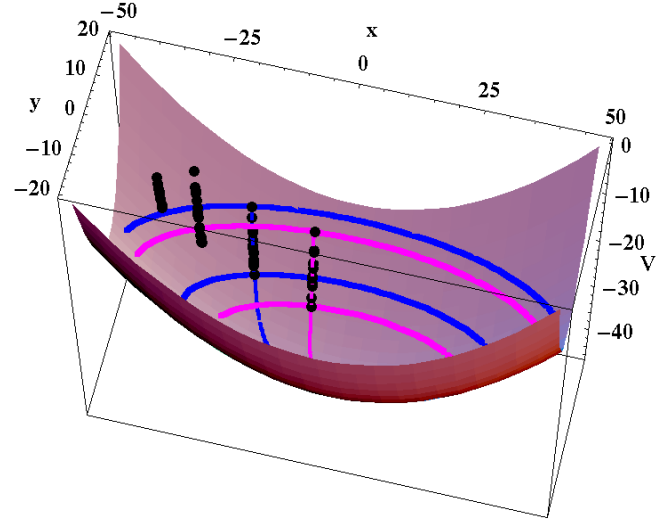

Figure 3. Landscape associated with Snoddy's (1926, Figure 4) score data (black dots) as elevation levels. The four clusters correspond to the four training sessions. The x-behavioral variable corresponds to the slow time-scale (shallow dimension) whereas the y-variable corresponds to the fast time-scale (steep dimension). For sessions 3 and 4 we plot the contour lines of the first and most successful trials. Note, that each contour line illustrates the degree of behavioral degeneracy (redundancy) for the given performance score (reproduced with permission from Newell, Liu & Mayer-Kress, 2005).

of Snoddy (1926). The relatively fast time scale captures the performance dynamics at the beginning of each trial while the relatively slow time scale captures the persistent change in the performance dynamics over the days of practice. Figure 4 shows the same data plotted in a more typical learning curve frame of reference with the power law superimposed for comparison. Clearly, the two time scale model fits the Snoddy (1926) data better than the power law.

We have conducted more formal investigations of the two time scale model in the Snoddy (1926) data set and in other motor learning data sets (Liu, Mayer-Kress, Hong & Newell, 2006). Equations 5 and 6 capture the performance dynamics for the two time scale model. The envelope function in the model is represented as an exponential decay that occurs across all trials, while the fast time-scale changes as the function of the trials within each practice session, j. The two time scale model was compared to other standard models of learning data though Akaike information criterion evaluations (Burnham & Anderson, 2002) and a rigorous theoretical assessment of the exponents and asymptotic values of the equations. The two time scale model was shown to consistently fit the performance dynamics better in both individual and group averaged learning data sets.

Slow time scale:

$$V_{en}(n) = V_{inf} + a_{en}\, e^{-\gamma_{en}\, n} \qquad (5)$$

Fast time scale:

$$V_j(n) = V_{inf} + a_{en}\, e^{-\gamma_{en}\, n} + a_j\, e^{-\gamma_j\,(n - n_j)} \qquad (6)$$

The model describes how a performance related variable $y(n)$ converges to an asymptotic target value V_{inf} as time n – measured in units of trial numbers – increases. a_{en} is the initial distance to the performance goal V_{inf} and $-\gamma_{en}$ the exponent for the envelope slow time scale function whereas a_j determines the amount of warm-up decrement at the beginning of session j, starting with trial number n_j. The exponent for the respective practice session $j = 1, …, 4$ is given by $-\gamma^j$. The model holds that the exponents for the fast time scale are the same on each day, an assumption of invariance that reduces the number of parameters in the model.

The epigenetic landscape framework supplemented with the system identification approach to decomposing the performance dynamics clearly shows in the group averaged data of Snoddy (1926) that the two time scale model fits the data qualitatively and quantitatively better than the power law and other models of learning. As we have stated previously, however, this does not mean that all performance dynamics will be fit by this two characteristic time scale model. It is our postulate that this model will be most relevant when there are no bifurcations in the

performance dynamics.

A question to be pursued in the analysis of learning curves is what further processes could be decomposed from the performance dynamics beyond the characteristic time scale of the warm-up decrement adaptation phase and that of the relatively permanent change over all trials and days. We have mentioned the short term fast time scale and reversible influences of motivation, attention and fatigue. These processes will be difficult to tease out of the performance dynamics and this modeling effort may need to use other ways to analyze the performance data. For example, a consideration of the best score achieved in a learning session to define the dynamics may prove useful, but this just reflects a broader need to go beyond the standard analysis strategy of single function curve fitting to all the performance data.

Finally, we would propose that this approach to the system identification of performance dynamics holds promise for a new approach to the analysis of the effects of practice distributions on learning. This is a topic that has lost impetus since Hull's (1943, 1951) theoretical formations regarding reactive inhibition and learning. The distinguishing of the adaptive and learning effects on practice schedules should provide new ways to consider such old problems as oblivescence and reminiscence (Ballard, 1915), and massed/distributed practice schedules (Schmidt & Lee, 2005), and practice condition effects more generally.

Closing Comments

In our view the determination by A. Newell and Rosenbloom (1981) of the power law as the ubiquitous law of learning is an entry to the problem not the solution. However, it is only in recent years that theoretical and empirical challenges to the power law of learning are being constructed. In our theoretical framework, there is not a single function of performance change over time as has generally been sought in fits of exponential, hyperbolic, logistic, and power law to learning data sets but rather that a coherent set of dynamical principles can lead to different functions of change in task outcome and limb trajectories. Indeed, basic assumptions about attractor dynamics and fixed points provide the theoretical basis for the assessment of the multiple characteristic time scales of change and their supporting processes in motor learning, development and adaptation. Our approach can in principle produce a power law of learning but we hypothesize that this function may prove to be a special and perhaps idealized case of learning, particularly when considered on individual learning data. The condition for power laws to occur would imply the presence of infinitely many time-scales whose distribution is
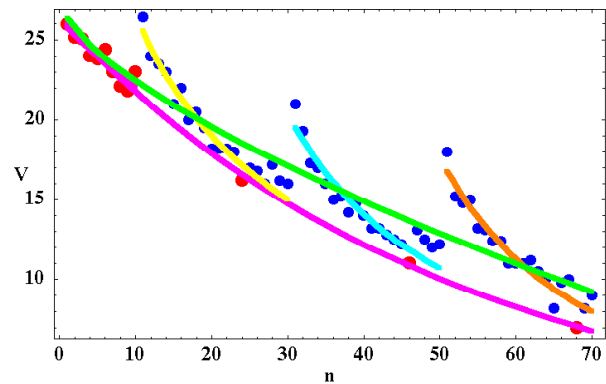

Figure 4. Performance data V_n vs trial number n from Figure 4 of Snoddy (1926). Main figure shows distance to asymptotic score $V_a = 46$ in linear scale. Magenta curve: exponential regression through first ten scores and best score of each session. Yellow, Cyan, Orange: Exponential fit to each of the practice sessions. Green: Power law fit through all data points. Note that the ten first points are not included in the main regression whereas in our model all data points can be accounted for by the model (reproduced with permission from Newell, Liu & Mayer-Kress, 2005).

described by a single parameter ("critical exponent"). This condition is related to phase transitions or bifurcations that occur at critical parameter values and, therefore, does not represent generic behavior of the system.

References

Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behavior*, *3*, 111-150.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1120-1136.

Bak, P. (1996) . How nature works: *The science of self-organized criticality*. New York: Springer-Verlag.

Ballard, P. B. (1915). Oblivescence and reminiscence. *British Journal of Psychological Monographs, #2*.

Bassingthwaighte, J. B., Liebovitch, L. S., & West, B. J. (1994). *Fractal physiology*. New York: Oxford University Press.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodal inference: A practical information-theoretic approach*. New York: Springer.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics, 2*, 153-166.

Elble, R. J., & Koller, W. C. (1990). *Tremor*. Baltimore: Johns Hopkins University Press.

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: implications of a quantitative analysis. *Proceedings of the National Academy of Science USA, 101*, 13124-13131.

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/f noise

in human cognition. *Science, 267*, 1837-1839.

Haken, H. (1983). *Synergetics: An introduction (3ʳᵈ Ed.).* Berlin: Springer-Verlag.

Haliday, D., Resnick, R., & Walker, J. (2005). *Fundamentals of physics (6ᵗʰ Ed.).* New York: Wiley.

Hallett, M., & Grafman, J. (1997). Executive function and motor skill learning. *International Review of Neurobiology, 41*, 297-323.

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

Hilgard, E. R., & Bower, G. H. (1975). *Theories of learning.* Englewood Cliffs, N.J.: Prentice-Hall.

Hull, C. L. (1943). *Principles of behaviour: An introduction to behavior theory.* New York: Appleton-Century-Crofts.

Hull, C. L. (1951). *Essentials of behavior.* New Haven: Yale University Press.

Irion, A. L. (1948). The relation of "set: to attention. *Psychological Review, 55*, 336-341.

Ivry, R. (1996). Representational issues in motor learning: Phenomena and theory. In H. Heuer & S. W. Keele (Eds.). *Handbook of perception and action Volume 2: Motor skills* (pp. 263-330). London: Academic Press.

Kandel, E. R. (2006). *In search of memory: The emergence of a new science of the mind.* New York: Norton.

Kaplan, D., & Glass, L. (1995). *Understanding nonlinear dynamics.* New York: Springer-Verlag.

Krieger, K. (2006). The net reloaded. *New Scientist*, July

Lane, N. E. (1987). *Skill acquisition rates and patterns.: Issues and training implications.* London: Springer-Verlag.

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

Liu, Y-T., Mayer-Kress, G., Hong, S. L., & Newell, K. M. (2006). Adaptation and learning: Characteristic time scales of performance dynamics. *Manuscript under review*.

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

Mandelbrot, B.B. (1977). *The fractal geometry of nature.* New York: Freeman and Co.

Mayer-Kress, G. J., Newell, K. M., & Liu, Y-T. (1998). What can we learn from learning curves? *InterJournal of Complexity, 246*.

Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A re-examination of the learning curve. *Psychological Bulletin, 85*, 1256-1274.

Miller, G. A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Nacson, J., & Schmidt, R. A. (1971). The activity-set hypothesis for warm-up decrement. *Journal of Motor Behavior, 3*, 1-15.

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.

Newell, K. M., Liu, Y-T., & Mayer-Kress, G. (2001). Time scales in motor learning and development. *Psychological Review, 108*, 57-82.

Newell, K. M., Liu, Y-T., & Mayer-Kress, G. (2003). A dynamical systems interpretation of epigenetic landscapes for infant motor development. *Infant Development and Behavior*, 26, 449-472.

Newell, K. M., Liu, Y-T., & Mayer-Kress, G. (2005). Learning in the brain-computer interface: Insights about degrees of freedom and degeneracy in a landscape model of motor learning. *Cognitive Processing, 6*, 37-47.

Neidermeyer, E., da Silva, L. (1999). *Electroencepahalography: Basic principles, clinical applications and related field (4ᵗʰ Ed).* Baltimore: Williams & Wilkins.

Nunez, P. (2000). Toward a quantitative description of large scale neocortical dynamic function and EEG. *Behavioral and Brain Research, 23*, 371-437.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6*, 421-425.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a methods of selecting among computational models of cognition. *Psychological Review, 109*, 472-491.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*, 359-367.

Rosenbloom, P. S., & Newell, A. (1987). An integrated computational model of stimulus-response compatibility and practice. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in theory and research* (Vol 21). (pp. 3-49). London: Academic Press.

Salmoni, A. W. (1989). Motor skill learning. In D. H. Holding (Ed.), *Human skills* (pp. 197-227). New York: Wiley.

Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis (4ᵗʰ Ed.).* Champaign, Ill: Human Kinetics.

Schroeder, M. (1991). *Fractals, chaos, power laws.* New York: W. H. Freeman.

Shadmehr, R., & Wise, A. P. (2005). *The computational neurobiology of reaching and pointing: A foundation for motor learning.* Cambridge, MA: MIT Press.

Snoddy, G. S. (1926). Learning and stability. *Journal of Applied Psychology, 10*, 1-36.

Snoddy, G. S. (1935). *Evidence for two opposed processes in mental growth.* Lancaster, PA: Science Press.

Strogatz, S. H. (1994). *Nonlinear dynamics and chaos.* New

York: Addison-Wesley.

Thurstone, L. L. (1919). The learning curve equation. *Psychological Monographs*, XXVI, Whole No. 114. 1-51.

Tulving, E. (Ed.). (2000). *Memory, consciousness, and the brain*. London: Psychology Press.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132*, 331-350.

Verhulst, P. F. (1838). Notice sur la loi que la population suit dans on accroissement. *Correspondance Mathématique et Physique, 10*, 113-121.

Ward, L. M. (2002). *Dynamical cognitive science*. Cambridge, MA: MIT Press.