Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR

NLR-TP-2000-308

The Power Law of Practice in adaptive training applications

J.J.M. Roessingh and B.G. Hilburn

Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR

NLR-TP-2000-308

The Power Law of Practice in adaptive training applications

J.J.M. Roessingh and B.G. Hilburn

This report is based on a presentation held on the Annual Meeting of the Human Factors Society - Europe Chapter, Martlesham Heath, UK, November 11-12, 1999.

Division: Flight
Issued: June 2000
Classification of title: Unclassified

Abstract

This research paper was inspired by the difference of findings in empirical learning-curve-data collected by ourselves and those of Newell and Rosenbloom (1981), who analysed empirical data-sets from other investigators to prove the ubiquity of the Power Law of Practice. The aim of this paper is to show some fundamental statistical and numerical flaws associated with Power Law analysis in this research literature. Additionally, we present the results from analysis of empirical data that reveals the low predictive value of the Power Law of Practice, using least-squares non-linear regression technique to find the best fitting estimate for the free parameters of the Power Law. These exercises lead us to the recommendation to take account of these flaws, for example, when considering application of the Power Law of Practice in adaptive training or intelligent tutoring.

Contents

1.	**Introduction**	**4**
2.	**Measuring on learning processes**	**5**
3.	**Power Law of Practice**	**6**
4.	**Log-log transformation and goodness-of-fit**	**8**
5.	**The Power Law and the speed of learning**	**12**
6.	**Another empirical test for the Power Law**	**15**
7.	**Conclusion**	**17**
8.	**References**	**18**

(19 pages in total)

1. Introduction

We have been searching for some time for a psychological model that successfully predicts the learning curve for students acquiring complex skills in the aviation domain. An example of this is the prediction of the learning curve of military pilots who learn to fly a specific mission. The model we were looking for would be able to predict the course of trainees' future task performance on the basis of past performance, that is, on the basis of a series of previous trials performed on the task. With future task performance, we mean the performance level that is reached at the end of the training, i.e. after a specified number of training trials or after a specified amount of training time.

When used in an adaptive training device (e.g. a suitably equipped flight simulator), it would be possible, with the use of such model, to optimally adapt the learning environment. We could, for example, adapt the complexity of the training task or we could adapt the feedback to the trainee. An accurate prediction of future task performance could lead to the optimal adaptation. In fact, a predictive model would actually be indispensable in an adaptive training device (or intelligent tutoring device) for training complex skills. After all, from a didactical point of view, the ultimate aim of a training system is to influence future task performance-- that is, to deliver a skilled operator after training. In this respect, a training provider could either attempt to attain the highest skill-level possible for its trainees, given the training time available, or alternatively, attempt to reach a specified skill-level in the shortest training time possible. In either case, the output variable that must be controlled by the adaptive logic is future task performance, rather than momentary task performance. For example, special instructional feedback that is automatically generated on the basis of momentary task performance may temporarily improve this performance, but transfer-of-training to future task performance is by no means guaranteed, other than by trial-and-error, as long as it is unknown how longer-term learning behaviour is affected. When we encountered the well-known "Power Law of Practice" we initially had the impression that we had found an almost perfect model for implementation in an adaptive training device and to predict the future course of performance in skill learning. The Power Law of Practice is a quantitative formulation of the learning curve that predicts the time (speed) to perform a practice trial from the number of trials previously performed. The Power Law in its most simple form is based on two free parameters, that may be interpreted as *learning speed* and *initial trial time*, and can be adjusted for the particular data under examination. Given a history of trial times, those two parameters could be estimated through curve fitting (using regression analysis for finding the best fit) and with the parameter estimations, future trial times could be predicted. The literature (see Newell and Rosenbloom, 1981) reported extremely good fits (most often 90% to 100% of the variability accounted for) of the Power Law model to a wide range of experimental data-series. In fact, unusually good fits for a psychological model, especially for the fuzzy, relatively ill-understood, realm of learning behaviour. If this was true, one could make almost perfect predictions on the basis of the Power Law.

Roessingh (1993) implemented the Power Law model in a training device used for training complex perceptual-motor skills and conducted a practical training experiment with it. Analysis of the resulting data-set revealed the weak predictive power of the Power Law, despite the great deal of empirical evidence (Newell and Rosenbloom, 1981, Logan, 1988, 1992) supporting its use.

2. Measuring on learning processes

The learning process during repetitive practice training is reflected in the performance measured over a limited number of (most often) equal time periods or time intervals. The graphical depiction hereof is generally called the learning curve. Since Bryan and Harter (1899) investigated the learning processes of telegraph operators (the auditory receiving of coded messages and the simultaneous typing thereof), there has been a search for characteristic features of learning curves for complex skills (see e.g. Hilgard, 1962, Pew, 1966, Newell and Rosenbloom, 1981, Spears, 1985, Damos, 1991). Some of those features are:

- slowing down of improvement after a first few trials
- improvement over extensive periods of time
- unstable, variable performance during the early stage of practice
- false asymptotes (plateaux)

Performance data is most often based on a quantitative or quantitised variable, such as, in the case of the telegraph operators, the ‘count’ of correct received Morse code signals per unit time or time period.
With performance measurements on the basis of such counts, the time period over which the count is derived, will also determine the ‘shape’ of the learning process. For example, if we would choose, for the case of a telegraph operator, a time period of two hours over which correct received Morse signals would be counted, then after ten hours of training we would have collected five counts and maybe conclude that the learning curve of a particular operator is steadily rising. While, if we would base our counts on five-minute-periods for the same telegraph operator, the 120 collected data-points might reveal several dips, peaks, plateaux and a bearing towards an asymptote of the learning curve.
From this trivial example, it is obvious that the question “learning or not learning?” from practice can not unambiguously be answered by just analysing the learning curve, since its appearance will be determined by the time intervals or periods over which performance is counted, integrated or averaged.

Curve-fitting techniques have been used in research to discover functional relationships between (change in) performance and practice trials undertaken. For example, Spears (1985) used logistic and exponential curves. These functions seemed to describe the learning curve fairly well. These functions can only be considered as compact descriptions of data; Spears did not assign a psychological meaning to the parameters in the functions and thus did not provide a clarification for the learning process.
The time required to complete a self-paced task is generally seen to decrease with practice. In this case, clarifications *have* been provided for the widely observed regularity in learning processes that time-periods to complete practice trials on a certain task decrease with a *power function* with the number of trials practised. The “power speed-up” in practice-trials has been observed with the learning of various perceptual motor skills, for example in Crossmans’s (1959) data on the manufacture of cigars by female operators using a cigar-making machine. Power speed-up has been observed also in more cognitive oriented skills such as elementary decision making, i.c. the time needed to react to specific events, given a number of alternatives (Seibel, 1963), problem solving (the time needed to finish a solitaire game), recognition (the time needed to recognise word combinations in changing contexts) and many more skills. An inherent restriction of the Power Law is its applicability to *self-paced* tasks only. If the pace with which the fixed task must be performed is dependent on external factors or determined by an external agent, than the Power Law will not apply. For example, an air pilot that flies a standard pattern around an airfield must follow Air Traffic Control directives in terms of speed and position. Trial-times on this task (flying a standard pattern) will be fairly constant and not or only marginally depend on the skill level of the pilot. The Power Law will therefore not apply. Newell and Rosenbloom (1981) provide an overview of clarifications based on cognitive learning theory. Logan (1992) even remarks that every serious learning theory should account for “power speed-up”.

3. Power Law of Practice

The power speed-up has been formulated as the Power Law of Practice. The general form of this model is:

$$T_N = T_A + b\,(N + N_0)^{-a}$$

In which:

- the *independent, discrete variable N*, is the number of trials that has been performed to (successfully) complete specific task (for example, to manufacture a cigar, as in the Crossman-example). Thus N is the count or rank-number of the trial ($N \geq 1$);

- the *dependent variable* T_N [seconds] is the time to perform the N^{th} trial, e.g. the time to manufacture a cigar in the N^{th} trial. It must be noted that the Power Law only considers the times to (successfully) achieve trials on a fixed task. Thus, the Power Law takes into account speed-based performance only, and therefore, successful achievement of a fixed task or trial should be defined such that accuracy is fixed (not variable) from trial to trial. For example, in the case of cigar-manufacturing, it must be specified what accuracy is required in manufacturing a cigar.
- *Parameter* T_A [seconds] is the asymptotic lower bound for the time to perform a trial. That is, after infinity trials it would take a particular operator T_A seconds to manufacture a cigar.
- *Parameter* N_0 is a measure for previous experience, expressed in number of trials. If a learner has already practised with a similar skill this would be taken into account by parameter N_0.
- *Parameter b* [seconds] is that part of the time to perform a trial that can be eliminated through practice. After infinity trials, the trial time would diminish with *b* seconds, when compared with the first trial. If we suppose that N_0=T_A=0, then *b* can be considered as the trial time for the first trial, that is, a measure for initial performance.
- *Parameter a* is a measure for the rate with which trial times diminish (the learning rate parameter). On the basis of the analysis of numerous experimental data sets Newell and Rosenbloom estimated *a*-values to be in the range from zero to 1 (depending on the characteristics of the task and the operator).

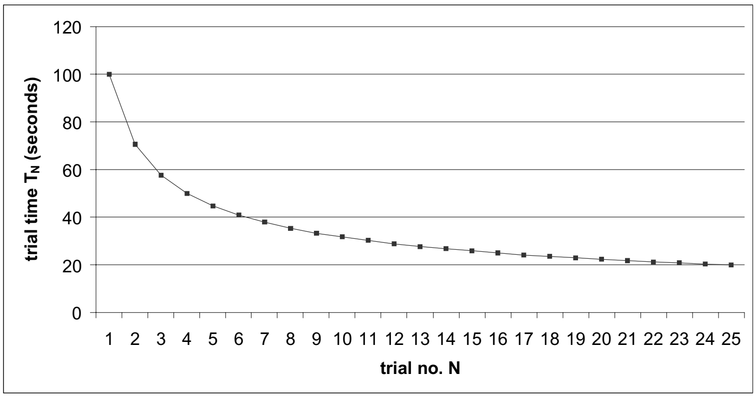


Fig. 1 Power Law: $T_N = b * N^{-a}$ (a=0.5; b=100 s)

To keep the discussion sufficiently basic, we will use the simple (and original) form of the Power Law (see figure 1) throughout this section. That is, without the parameters T_A and N_0. The current discussion is relevant for either form of the Power Law, but since the two extra parameters have additional associated complexity and difficulties, these parameters will not be discussed here.

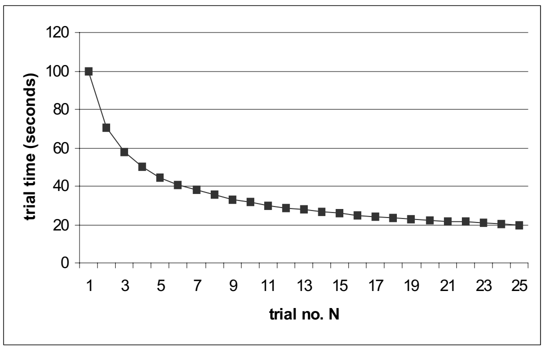
Now that we have introduced the Power Law to the reader, we will specifically address two flaws encountered in the literature. In our view, these flaws have led to overly optimistic conclusions about its ubiquity in skill learning and hence its applicability and importance for cognitive psychology.

4. Log-log transformation and goodness-of-fit

We checked the models' goodness of fit using our own data sets, to help assess the usefulness of the model. The goodness-of-fit of the Power Law, and thereby its validity as a description of learning behaviour, has in the empirical literature been measured by the value of r^2 (that is the squared 'coefficient of correlation', or, when multiplied by 100%, the 'percentage of the variance accounted for' or 'coefficient of determination'). Newell and Rosenbloom but also Anderson (1983) and others have found values of r^2 that are close to 1 (100%) for Power Law curve fitting to experimental learning curve data. Thus, these researchers found an almost perfect fit of the Power Law model to the data for a large number of different skills.
The first question that arises is: what method is used in establishing these high values of r^2 ? The answer is as straightforward as complicating. The Power Law is transformed to a linear function using log-log transformation as follows:

- $\mathbf{T_N = \mathit{b} * N^{-\mathit{a}}}$ the power function can be rewritten as:
- $\mathbf{\log(T_N) = \log(\mathit{b} * N^{-\mathit{a}})}$ or:
- $\mathbf{\log(T_N) = \log(\mathit{b}) - \mathit{a} * \log(N)}$ or:
- $\mathbf{\log(T_N) = \mathit{b}' - \mathit{a} * \log(N)}$

Thus, following this procedure the power function becomes a linear function when plotted in logarithmic co-ordinates (see figure 2), that is a straight line with intercept $b' = \log(b)$ and slope a.



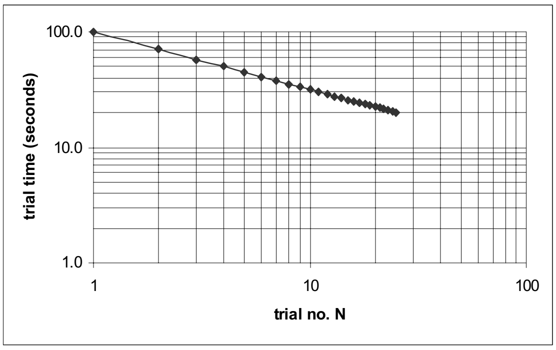


Fig. 2 Power Law (a= 0.5, b=100) plotted in normal (left side) and logarithmic (right side) co-ordinates

Having achieved this, the parameters *a* and log(*b*) can be estimated using the method of least squares, using the logarithmic transformed raw data and assuming a simple linear regression model.
A second question then immediately arises; why should one use log-log transformation? The most probable answer is that to estimate free parameters *a* and *b* in the Power Law: $T_N = b * N^{-a}$ is rather complicated, requiring a non-linear regression technique, where estimation of *a* and *b*' = log(*b*) in the "log-log linear" model is fairly easy when least squares linear regression is used. Thus, the analysis of log-log linear model in combination with least squares linear regression has led to good results in the literature: a high percentage of data variance accounted for.
An additional reason for the log-log linear model could be the easy graphic representation and visual inspection of the deviation of raw data points with respect to the model (see figure 3, bottom).
Newell and Rosenbloom refer to the log-log linear law as if it were identical to the Power Law

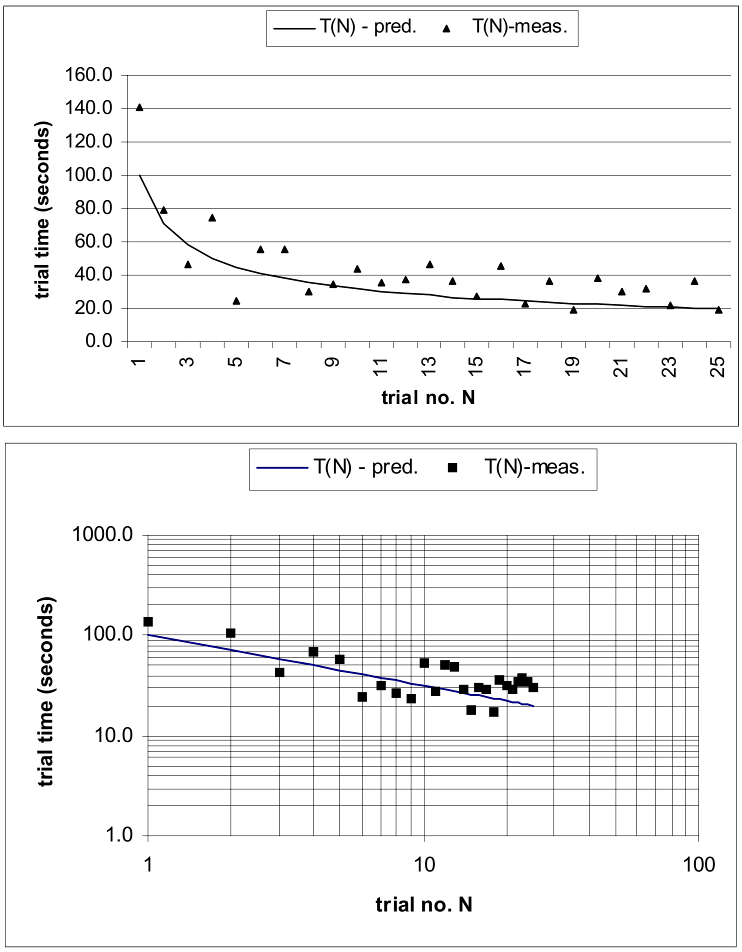


Fig. 3 Observed trial times plotted in normal and logarithmic co-ordinates

('we shall refer to this law variously as the log-log linear learning law or the power law of practice', Newell and Rosenbloom, 1981, page 3). We comment on this practice by considering the consequences of log-log transformation. Our argument is that logarithmic transformation of the observations (trial times) smoothes errors in the observations in a non-linear way (see figure 3). It is easy to contemplate that observations with a positive deviation (or residual) with respect to the best fitting power law curve will shrink more than observations with an equally negative deviation. For example, if the predicted time to perform a trial is 50 [seconds] and we would observe a value of 75 [s], then the relative deviation would be 50%. And if we would observe a value of 25 [s], then the relative deviation would also be 50%. Though, in the log-log world, the predicted value would be $\log(50)\approx1.70$. The observed values would be $\log(75)\approx1.88$, resulting in a relative deviation from the predicted value of 11%, and $\log(25)\approx1.40$, resulting in a relative deviation of 18%.

Thus, through log-log transformation, observations with positive deviations with respect to a predicted mean are more attenuated than observations with negative deviations. It is clear that the distribution of the data set will change through log-log transformation and will in general closer resemble the normal distribution, since trial times (and reaction times and the like) tend to have a skew distribution: it is more likely to observe higher-than-average trial times than lower-than-average trial times.

As a result, if we estimate the best fitting parameters *a* and *b* for the log-log linear model using least square linear regression, it is well possible that these estimates are different from estimates that would have been made for the Power Law model using non-linear least squares regression or another estimation technique. In summary: if the log-log linear model is used in combination with least squares linear regression to estimate *a* and *b,* then estimates for *a* and *b* on the basis of the log-log linear model $\mathbf{\log(T_N)=\log(\mathit{b})-\mathit{a}*\log(N)}$ will not necessarily fit the Power Law $\mathbf{T_N=\mathit{b}*N^{-\mathit{a}}}$, depending on the nature of error variability.

What does the log-log linear law mean? It predicts the logarithm of a trial time T_N, given the logarithm of a trial number N. Since this entails a non-linear deformation of time, an entity that is essentially known or dealt with as being linear in normal life, it is on itself meaningless[1].

We could do an inverse transformation on the predicted value $\log(T_N)$ to make it meaningful. However, as argued, we may have not estimated our parameters *a* and *b* right in the first place. The adverse consequence could be that small errors in the predicted value of $\log(T_N)$ will become large errors in the predicted value of T_N, and in a non-linear way, such that small negative deviations become large negative deviations and small positive deviations become huge positive deviations. Thus, predictions for T_N will not be very reliable, despite the high

[1] Unless one would like to take specific advantage of expressing a functional relationship in its logarithmic form. Multiplicative relationships become additive such that doubling an initial value or cutting it into half amounts to adding or subtracting a fixed number to its logarithm, like, for example, is done with sound intensities.

percentage of variance accounted for by log-log linear model. The question may arise whether Newell and Rosenbloom have estimated parameters *a* and *b* on the basis of non-linear regression, using the Power Law model. It is unlikely that non-linear regression was used, since the co-efficient of determination r^2 is in this case not a reliable measure for goodness-of-fit. Firstly, the model is non-linear in its parameters. Secondly, with this type of data, data-points (trial times) measured at the end of the learning process will be arranged around the flat tail of the Power Law curve. These data-points will therefore not add to the variability accounted for by the model (r^2), since the predicted trial times at the flat tail of the curve are principally invariable, while the measurements will have some variability due to additive error sources. This error will thus predominantly lower r^2. Thus, the longer the learning process, the more data-points will be arranged at the tail of the curve and the lower will be the value for r^2.
In general, the complementary case will also be true. Trial times measured at the start of the learning process, that is, at the steep part of the curve, where trial times have a relatively long duration and will decrease quickly from trial to trial, will yield larger values of variance accounted for (r^2).
In summary: the log-log linear law may well be different from the Power Law in real applications. We conclude that Newell and Rosenbloom probably applied the calculations of r^2 to the log-log linear model and the logarithmic transformed data $\log(T_N)$ and reported high values of r^2 for the log-log linear law, but did not establish goodness-of-fit for the actual Power Law.

5. The Power Law and the speed of learning

In this section we focus on the shape of the learning curve and how this shape is affected by the value of the "learning rate" parameter *a* (see figure 4).

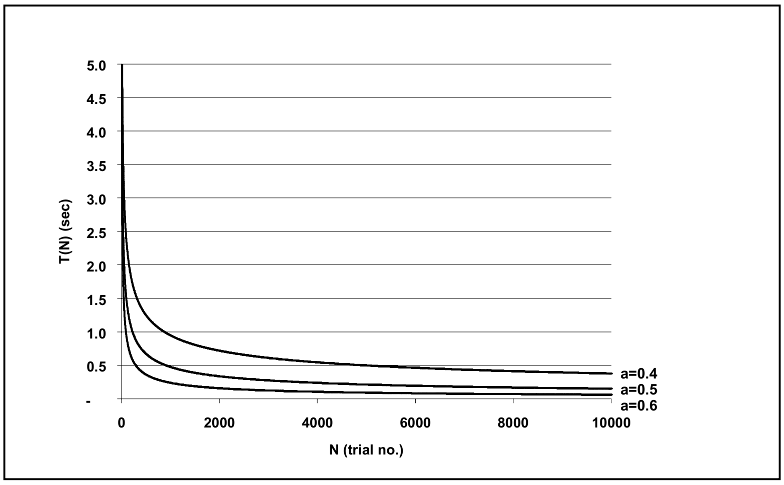


Fig. 4 The Power Law of Practice for three values of the learning rate parameter a (0.4, 0.5, 0.6), parameter b (trial time for the first trial) is 15 sec..

According to Newell and Rosenbloom, the parameter *a* can take any value on the interval [0,1], predominantly depending on task characteristics. A value of zero denotes zero learning from trial to trial and a value of 1 denotes "hyperbolic" learning ($T_N = b/N$).
Estimating the parameters of the Power Law in its simplest and most often used form, that is, without the parameters for asymptote T_A and previous experience N_0, Newell and Rosenbloom found indeed values for *a* on the interval [0,1]. For example, parameter *a* has been estimated as 0.81 in the data from Neisser et al (1963). The percentage of variance accounted for by the simple Power Law was 97.3%.
We are interested in the implications of different values of parameter *a* in the time domain, that is, when the Power Law is plotted as a function of time, rather than as a function of practice trials. Therefore, we will analytically transform the simple Power Law (trial-time as a function of practice trials) into a learning curve in which performance (expressed as completed trials per time interval) is plotted as a function of practice time (time-on-task).
Practice will be considered as a continuous repetition of a fixed task (e.g. the manufacturing of cigars). Thus, total time-on-task is constructed of a series of trial-times T_1, T_2, .., T_{N-1}, T_N that seamlessly link up together. For the total time-on-task T_t this yields:

$$T_t = T_1 + T_2 + \ldots + T_{N-1} + T_N = b\sum_{i=1}^{N} i^{-a} \qquad \textbf{(1)}$$

In which:
i = the rank-number of the trial,
N = the total number of trials that fits in the total time-on-task T_t.

For ease of exposition we will consider i as a continuous variable and replace Eq. (1) by an integral (cf. Newell and Rosenbloom, 1981, page 14), such that:

$$T_t \approx b\int_1^N i^{-a}\,di = \frac{b}{1-a}N^{1-a} - \frac{b}{1-a} \qquad \textbf{(2)}$$

We put that A=(1-a)/b. Eq. (2) will then convert into:

$$T_t \approx \frac{1}{A}(N^{1-a} - 1) \qquad \textbf{(3)}$$

From Eq. (3), we can solve N, the total number of trials that fits into the total time-on-task Tt:

$$N \approx (AT_t + 1)^{\frac{1}{1-a}} \qquad \textbf{(4)}$$

We define $T_t = s \cdot T_s$, that is, we subdivide total time-on-task T_t in s even intervals with duration T_s. Since we are interested in the count or performance f(s) of the trainee during each subsequent time interval, we have to determine the number of trials f(s) that fit on each of the intervals T_s, which can be derived by solving f(s) from Eq. (5), which is a variation of Eq. (1):

$$T_s = b\sum_{i=N-f(s)}^{N} i^{-a} \qquad \textbf{(5)}$$

Analogous to Eq. (4), Equation (5) will then yield the total number of trials N(s) that will fit into s subsequent intervals:

$$N(s) \approx (sAT_s + 1)^{\frac{1}{1-a}} \qquad \textbf{(6)}$$

We can now determine the number of trials that fits on interval T_s, that is, solving f(s) from Eq. (5), by determining N(s) - N(s-1):

$$f(s) \approx [sAT_s + 1]^{\frac{1}{1-a}} - [(s-1)AT_s + 1]^{\frac{1}{1-a}} \qquad \textbf{(7)}$$

The graph of the obtained performance function f(s) (Eq. 7) is a learning curve, i.e. performance as a function of time (-interval), on the basis of the Power Law. Hence, the course of learning over time can be predicted, provided that parameters a and b are known (or estimated). In figure 5 we have sketched three Power Law learning curves for different values of the learning rate parameter a, that is, for a smaller than 0.5, equal to 0.5 and greater than 0.5.

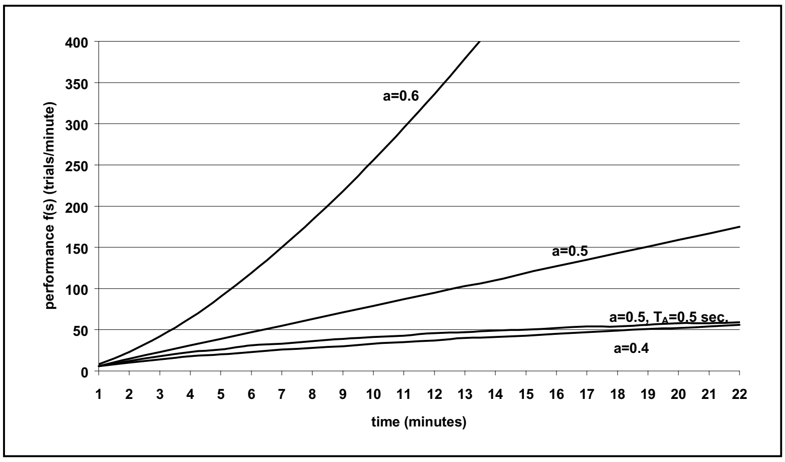


Fig. 5 The Power Law of Practice plotted as a function of time, for three values of the learning rate parameter a (0.4, 0.5, 0.6) and two values for T_A (0.0 sec., 0.5 sec.), b is 15 sec. for all curves. Note the improbable shape of the curve for a > 0.5.

Eq.7 shows that when the learning rate parameter a is made equal to 0.5, that higher order terms of s will neutralise each other and performance f(s) is linear with time (see the curve in figure with a=0.5 and T_A=0.0 in figure 5). When a is made smaller than 0.5, equation 7 shows that performance f(s) decelerates, that is, it takes longer to conduct one more trial on each subsequent time interval T_s. When a is greater than 0.5, the increase of performance per time interval will accelerate. This means that it takes shorter to conduct one more trial on each subsequent time interval T_s.

Thus, the Power Law plotted in time rather than trials predicts global “superlinear” progress for values of parameter a higher than 0.5. Such a progress in learning is opposite to the basic assumptions of learning theories, although it may be observed over short periods of time (as a local characteristic of the curve).

From the analysis it becomes apparent that it is unlikely for the parameter a of the Power Law to have a higher value than 0.5. Nevertheless, Newell and Rosenbloom report estimates for the parameter a larger than 0.5 for some empirical learning data sets.

This analysis provides additional reason to treat analysis (parameter estimation) on the basis of logarithmic transformation of the Power Law with caution. The improbable learning rates

(a > 0.5) that are assumed in literature (for example, Logan, 1992, page 883) cast some doubt upon the overall validity of the Power Law.

Note: when the asymptote parameter T_A is introduced into the model, a clear deceleration of performance in the time domain is observed (see figure 4, a=0.5, T_A=0.5). However, the introduction of this additional parameter precludes linearisation of the Power law through the log-log approach. Using other estimation methods it was found that the cross-correlation of the estimates of each parameter within the extended Power Law is high, resulting in multiple solution sets for given input data. Further analysis with parameter T_A in the time domain is outside the scope of this article.

6. Another empirical test for the Power Law

An experiment, suitable for testing the Power Law model, was conducted by Roessingh (1993). The results presented in this section are based on the data collected in this experiment. The task that was used was the Space Fortress game (Donchin, 1989). This is a laboratory PC-game, specifically defined and designed for training research and well described and examined in literature (Mané and Donchin, 1989, Foss et al., 1989, Frederiksen and White, 1989, Gopher et al., 1989, Gopher et al., 1992). We select this task for Power Law research since the top-level strategy of this task essentially requires speed-based performance ('destroy the fortress as quickly as possible'). Moreover, like the Power Law, this game is generally associated with skill, in particular with perceptual-motor skills.
One trial in this game is defined as one destruction of the space fortress using a manual controlled space ship. With this task and this definition, time-on-task corresponds exactly to the sum of trial times T_N. Thirty-six subjects (male university undergraduates) participated in the experiment. Time-on-task varied between subjects from 2.5 hours to 16 hours, with 20 minute breaks between every 40 minutes time-on-task and maximum 80 minutes per day.
Subjects with fewer than 16 hours time-on-task received similar training on simpler versions (part tasks) of the Space Fortress game, such that all trainees received 16 hours of practice training over 12 days. Subjects were paid Dfl. 540 (245 Euro) plus Dfl. 150 (68 Euro) bonus upon completion of the experiment. Laboratory-equipment consisted of the soft- and hardware needed to host the game, i.e. DOS-PC's of type 80386SX with a joystick to control the space ship and a mouse to conduct additional tasks and additional software to record time-on-task T_t and trial- times T_N.
With respect to the data-analysis we restrict ourselves here with the determination of the accuracy with which the Power Law can be used to predict the number of trials that will fit into

a fixed amount of time-on-task T_t. In this procedure we use non-linear regression to estimate the four parameters *a*, *b*, T_A and N_0 for each individual subject on the basis of all observed T_N-values for this subject. For non-linear least squares regression we use the Marquardt-Levenberg process[2] (Press et al., 1989), an iterative process that is advised (Mooiweer, 1991) when model parameters have a strong cross-correlation.
After convergence of the iterative process we use the estimates of the four parameters, *a*, *b*, T_A and N_O, to numerically solve the variable N (i.e. the total number of trials fitting in the time-on-task Tt) for each subject, using the equation:

$$T_t = T_1 + T_2 + \dots + T_{N-1} + T_N = \sum_{i=1}^{N} (T_A + b(i + N_o)^{-a})$$

Subsequently, we compare this predicted N_{pred} with the actual observed total number of trials N_{obs} that were performed in time-on-task T_t through calculation of the percentual (relative) error in N_{pred} with:

$$Error\ \% = \frac{N_{pred} - N_{obs}}{N_{obs}} x100\,\%$$

The results of this exercise are depicted in the bar-graph of figure 6. Each bar represents the relative error in N_{pred} for one subject. The subjects have been grouped along the horizontal axis according to their part-task experience. For example 10/50 is an indication that 10 ‘units’ previous experience was gained on part-task 1 and that 50 ‘units’ previous experience was gained on part-task 2. Thus, in the 0/0 condition (control condition – 6 subjects), subjects spent 16 hours time-on-task, while in the other conditions subjects spent less time on task (but at least 2.5 hours). Average time-on-task was 10 hours. Average error percentage in N_{pred} is 17%. The error percentage seems to be larger for subjects with a longer time-on-task (left side of the bar-graph), although the estimation/calculation of N_{pred} is based on more data for these subjects. All errors are positive ($N_{pred} > N_{obs}$) which points at a systematic error, either in our estimation or calculation procedure or in the Power Law model.

[2] We found this non-linear least-squares regression technique lacking robustness for this application. Moreover, convergence of the iterative process was problematic, making the process rather unsuitable for automatic adaptive training. An improved method is under study.

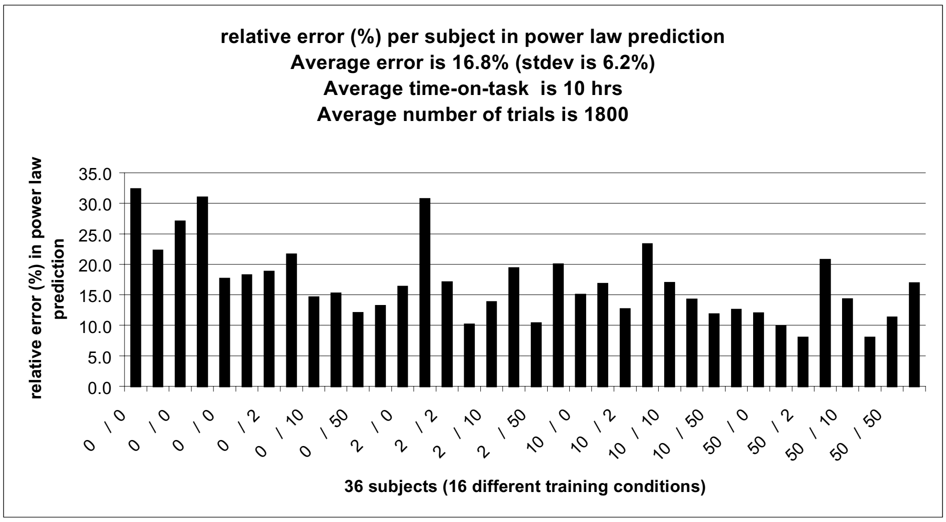


Fig. 6 Errors in power law predictions for the number of trials that can be performed in time-on-task of known duration. Errors are relative to actual observed numbers for 36 subjects

It is important to realise that this analysis concerns individual data (no group- or averaged data) and that through this procedure of calculating accumulated trials N over time-on-task T_t, we eliminated the effects of trial-to-trial variability. It is clear that these results, with considerable errors between predicted and observed values, do not justify the claim of ubiquity. This analysis of empirical data has shown that the Power Law is actually a fairly poor predictor in practical applications with off-the-shelf statistical techniques.

7. Conclusion

For our purposes (adaptive training) we investigated the suitability of the Power Law as a learning curve model. We checked the Power Law on three aspects: (1) goodness-of-fit, (2) learning rate parameter, and (3) predictive value.
Firstly, we challenge its supposed goodness-of-fit as reported in the literature (primarily Newell and Rosenbloom, 1981). We provide a clarification for the unusually high percentages of variance accounted for by the log-log-linear model by considering the characteristics of log-log transformation. It is argued that the log-log-linear model is different from the Power Law model and that the predictions of trial-times using the log-log-linear model is erroneous.
Secondly, we transformed the Power Law into a time curve (performance as a function of time) and conclude that this curve is super-linear (accelerated performance increments, which is contrary to the notion of learning) for values of the learning rate parameter *a* greater than 0.5. Surprisingly, Newell and Rosenbloom established, on the basis of empirical data, that parameter

a could take any value between zero and 1. For example parameter *a* has been estimated as 0.81 in the data from Neisser et al. (1963). The percentage of variance accounted for by the Power Law (or more likely, the log-log linear law) was here 97.3%. This is a second line of evidence that the curve-fitting technique used by Newell an Rosenbloom may not be appropriate (finding improbable estimates for parameter *a*) and that the Power Law does not fit the data very well. Moreover, this result questions the validity of the Power Law, since both theorists (Logan, 1988, 1992, 1995) and empiricists that assumed ‘power speed-up’ have found values for the learning rate parameter greater than 0.5.
Thirdly, with respect to the predictive value of the Power Law, we have found large systematic deviations in its predictions on the basis of empirical data, where we used a non-linear least-squares regression technique, rather than a linear one.
On the basis of our checks on those three aspects the Power Law, we comment on Logan’s (1992) statement that ‘every serious learning theory should predict power speed-up’ by stating that this power speed-up is not necessarily implying the Power Law of Practice in its current form. We were not able to apply the Power Law as could be expected from literature, which restricts its application in adaptive training or intelligent tutoring, where a reliable student learning model is indispensable.

8. References

Anderson, J.R. (1983). *The architecture of cognition,* Harvard University Press. Cambridge, Mass., US.

Bryan, W.L., Harter, N. (1899). Studies on the telegraphic language. The acquisition of a hierarchy of habits. *The Psychological Review, Vol. VI, No. 4.*

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

Damos, D.L. (1991). Examining transfer of training using curve fitting: a second look. *The International Journal of Aviation Psychology, I(1),* (pp. 73-85).

Donchin, E. (1989). The Learning Strategies Project. *Acta Psychologica, Vol. 71,* (pp. 1-15).

Foss, M.A., Fabiani, A., Mané, A.M., Donchin, E. (1989) Unsupervised practice; the performance of the control group. *Acta Psychologica, Vol. 71* (pp. 23-51*).*

Frederiksen, J.R., White, B.Y. (1989). An approach to training based upon principled task decomposition. *Acta Psychologica, Vol. 71* (pp. 89-146).

Gopher, D., Weil, M., Bareket, T. (1992). The transfer of skill from a computer game trainer to actual flight. *Proc. of the Human Factors Society, 36th annual meeting.*

Gopher, D., Weil, M., Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica, Vol. 71* (pp. 147-177).

Hilgard, E.R. (1962). *Introduction to psychology*, New York: Harcourt, Brace and World, Inc. pp. 267-274.

Logan, G.D. (1988). Towards an instance theory of automatization. *Psychological review*, Vol. 95, No. 4 (pp. 492-527)

Logan, G.D. (1992). Shapes of reaction-time distributions and shapes of learning curves: a test of the instance theory of automaticity. *J. of Experimental Psychology; Learning, Memory and Cognition, Vol. 18, No.5,* (pp. 883-914).

Logan, G.D. (1995). The weibull distribution, the Power Law, and the Instance Theory of Automaticity, *Psychological review*, Vol. 102, No. 4 (pp. 751-756)

Mané, A., Donchin, E. (1989). The Space Fortress Game. *Acta Psychologica, Vol. 71* (pp. 17-22).

Mané, A.M., Adams, J.A., Donchin, E. (1989) Adaptive and part-whole training in the acquisition of a complex perceptual-motor skill. *Acta Psychologica, Vol. 71* (pp. 179-196).

Mooiweer, G.D. (1991). *Parameter estimation through curve fitting in process technological equations using the SAS package* (report in Dutch). Reportno. 6750 of the Faculty Chemical Technology, Technical University Eindhoven.

Newell, A., Roosenbloom P. (1981). Mechanisms of skill acquisition and the Law of Practice. In J. R. Anderson (Ed), *Cognitive Skills and their acquisition* (pp. 1-55) Hillsdale N.J: Erlbaum Associates.

Neisser, U., Novick, R., Lazar, R. (1963). Searching for ten targets simultaneously. Perceptual and motor skills, Vol. 17 (pp. 955-961).

Pew, R.W. (1966) Acquisition of hierarchical control over the temporal organization of a skill. *J. of experimental psychology, Vol. 71, No. 5* (pp. 764-771).

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.(1989). *Numerical recipes in Pascal, the art of scientific computing*. Cambridge University Press.

Roessingh, J.J.M. (1993). *Adaptive part-task training for complex perceptual-motor skills*. MSc thesis (in Dutch), Technical University of Eindhoven, Faculty of Technology Management, Eindhoven, The Netherlands.

Spears, W.D. (1985). Measurement of learning and transfer through curve fitting.*Human Factors, Vol. 27(3)*, (pp. 251-266).

Acknowledgement

The authors would like to thank P.J. Hoogeboom (NLR-Flight division) and L.J. Thijs (Technical University Eindhoven) for providing useful comments on the draft of this publication.