

PRACTICE AND FORGETTING CURVES DEDUCED FROM SCALE INVARIANCE

Martin Riopel¹

¹Université du Québec à Montréal (CANADA)

Abstract

Practice plays an important role in e-learning, yet it is still difficult to find a consensual and usable mathematical model to describe the shapes of the learning and forgetting curves. As pointed out by [1], “a fundamental understanding of the distributed practice effect is lacking; many qualitative theories have been proposed, but no consensus has emerged” (p. 236). This paper explores some of the reasoning behind deducing the mathematical properties of the observed practice and forgetting curves. The model it proposes is based on three scale-invariant hypotheses that were shown to be empirically valid in another paper [2], where subjects learned to identify images of constellations using a PHP/HTML web application.

As pointed out by [3], the shape of the forgetting curve is a problem dating back to Ebbinghaus (1885) that remains without a consensual solution. It is proposed that the shape of this curve may be deduced by using scale invariance hypotheses. The logarithm of correctness would thus follow an inverse Weibull cumulative function that, for large t , behaves as a three-parameter power function. This new three-parameter forgetting law is consistent with the results of [3], with the added advantage of being scale-invariant. When fitted to [3]’s data by minimizing the chi-squared statistic, the two curves are almost equivalent, with the new function demonstrating a slightly better fit.

The power law of practice applies to the relationship between response time and number of practice trials. According to [4], however, “the evidence for a power law is flawed, because it is based on averaged data.” Their analysis demonstrates that the exponential function is likely a better choice for individual data. This paper proposes that ordinal invariance may be used with the consequences of composition invariance to deduce another shape for this curve: a four-parameter stretched exponential function. This is compatible with [4]’s analysis, which concludes that the three-parameter exponential function fits practice data almost as well as the more flexible four-parameter power-exponential function. Stretching the exponential by adding a fourth parameter would only improve the fit, potentially rendering the resulting function, deduced from invariance hypotheses, the best choice in all cases. However, empirical verification is still required.

It is interesting to note that scale invariance can be used to deduce some of the properties of two strong and highly dissimilar learning-related effects. The laws proposed in this paper are somewhat different from other laws but still compatible with the most precise existing empirical data. This important achievement in the field of learning sciences could ultimately lead to a complete and coherent proposal for a uniquely defined scale-invariant theory of learning.

Keywords: Astronomy, Practice, Forgetting, Learning Theory.

1 INTRODUCTION

When teachers prepare for their lessons, they have to figure out how long, how many times and exactly when students should be exposed to a particular content. Of course, the precise answers are difficult to find because they depend on a lot of variables but also because research on distributed practice has yet to reach definitive conclusions. More precisely, even if practice plays an important role in learning, yet it is still difficult to find a consensual and usable mathematical model to describe the shapes of the learning, forgetting, and spacing curves related to distributed practice. As pointed out by [1], “a fundamental understanding of the distributed practice effect is lacking; many qualitative theories have been proposed, but no consensus has emerged” (p. 236). Furthermore, “although distributed practice has long been seen as a promising avenue to improve educational effectiveness, research in this area has had little effect on educational practice” (p. 236).

Empirical validation of three scale-invariant hypotheses was presented in another paper [2], where subjects were asked to identify images of constellations. Since these general hypotheses will be used

to deduce the mathematical properties of observed empirical curves, it seems important to present them again in this context.

Let's consider, for example, a student asked many times to learn to identify equivalent constellations. In this particular situation, the first hypothesis (composition invariance) states that, for that student, the probability of success for answers about one constellation should follow the same mathematical curve, apart from a single scaling parameter, than the combined success (product of probabilities) for two or more constellations. The second hypothesis (temporal invariance) states that slowing a sequence of questions cannot change the mathematical curve for correctness, apart from a single scaling parameter. Similarly, the third hypothesis (ordinal invariance) states that, when counting successive occurrences of questions about a given constellation, one is free to choose freely what constitutes a unit (a single occurrence or a group of occurrences) without changing the mathematical curve for correctness, apart from a single scaling parameter. It is important to note that the hypotheses described here for correctness could also be applied to response time.

This paper concentrates on using these three hypotheses to deduce the mathematical properties for the practice effect, the forgetting effect and their combined effects. It also presents some empirical support for the derived laws.

2 METHODOLOGY

The empirical data presented in this paper come from four independent published sources. The first, a study described in [2], involves an online application where subjects were asked to identify a series of constellations and given time to review the correct answers. Intended to validate scale-invariance hypotheses, the experiment used a within-session block design with different scales. The second source of empirical data, described in [3] and related to the forgetting curve, is an experiment designed to measure cued recall and stem completion from one minute to 28 days after study, with more observations per interval per participant than in previous studies. The third source, described in [4] and related to the law of practice, is a rigorous survey that assessed the form of the practice function for individual learners based on 40 sets of data from 13 published and three unpublished sources. The fourth and final source, described in [8] and related to regular spacing, is an experiment performed to investigate how practice and spacing affected the retention of Japanese–English vocabulary paired associates and to validate a general model for both effects. These four data sources will be used to present specific empirical evidence for the theoretically derived laws in each section.

3 RESULTS

3.1 Forgetting curve

As [3] notes, the shape of the forgetting curve is a problem dating back to [5] for which there is still no consensual solution. Here we propose that scale invariance hypotheses can be used to deduce the shape of this curve. As described in [2], composition invariance leads to the following equation for the logarithm of correctness:

$$\log P = \log P_0 \cdot \exp(-\Delta\theta),$$

where P_0 is the reference correctness and θ is the ability of the subject. We can define t as the elapsed time before retrieval and use it to explain the change in ability, such that $\Delta\theta = F(t)$ and F is a monotonically decreasing function of t (because more time leads to less ability). Temporal invariance in this context states that slowing the sequence can change the scale of F , but not its shape. This can be expressed mathematically as

$$F(\alpha t) = \beta F(t),$$

where α is a time scale parameter and β as an ability scale parameter. This equation defines a positively homogeneous function. A general solution in this context is a negative power function for which $\beta = \alpha^{-b}$:

$$F(t) = a t^{-b}.$$

The logarithm of correctness for the forgetting curve is therefore

$$\log P = \log P_0 \cdot \exp(- a t^{-b}),$$

where a , b , and P_0 are positive parameters. The logarithm of correctness follows an inverse Weibull cumulative function [6]. For large t , Taylor's expansion shows that correctness should behave as a three-parameter power function:

$$P \approx P_0 [1 - \log P_0 \cdot a t^{-b}] = P_0 + a' t^{-b}.$$

This new three-parameter forgetting law is compatible with the results of [3] and has the added advantage of being scale-invariant. Fig. 1 compares the new function with the power function when fitted to [3]'s data by minimizing the chi-squared statistic. The two curves are almost equivalent, with the new function, $\chi^2(9) = 0.71$, $p > .999$, demonstrating a better fit than the power function, $\chi^2(9) = 0.84$, $p > .999$. The fit for the exponential function (not shown) is not as strong as that for $\chi^2(9) = 3.03$, $p = .960$.

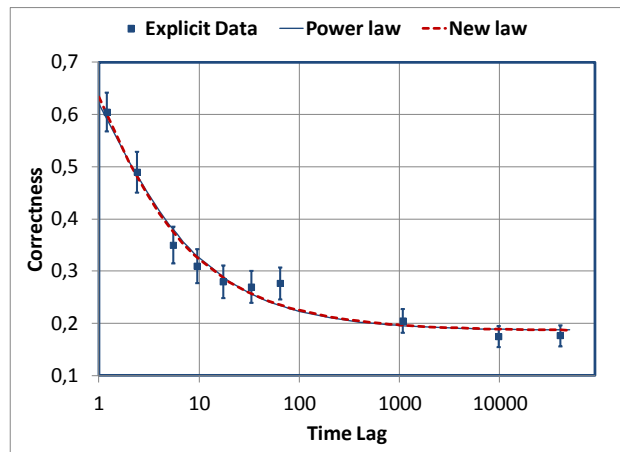


Figure 1. The new law is almost equivalent to the power law for explicit data from [3]. The error bars represent the 95% credible intervals for the population mean retention probability estimates.

It is important to note that [3]'s conclusion concerning the power law was obtained empirically by comparing the three most plausible candidates from previous studies (exponential, Pareto, and power). By contrast, the new function was obtained using invariance hypotheses and a series of logical steps. Within this context, we believe the new function is theoretically stronger and should be preferred.

3.2 Law of practice

The power law of practice was first proposed by [7] and applies to the relationship between response time and number of practice trials. However, according to [4], "the evidence for a power law is flawed, because it is based on averaged data." Their analysis shows that the exponential function (or a combination of exponential and power functions) is likely a better fit for individual data. In this paper, we suggest that ordinal invariance can be used with the consequences of composition invariance to deduce another shape for the practice curve.

Defining n as the number of practice trials, we describe change in ability as $\Delta\theta = G(n)$, where G is a monotonically increasing function of n (because more practice leads to more ability). Ordinal invariance in this context states that counting by groups can change the scale of G , but not its shape:

$$G(\alpha n) = \beta G(n),$$

where α is a practice trial scale parameter and β is an ability scale parameter. A general solution in this context is a positive power function for which $\beta = \alpha^c$:

$$G(n) = a n^c.$$

The logarithm of correctness for practice trials is therefore

$$\log P = \log P_0 \cdot \exp(-a n^c),$$

where a , c , and P_0 are positive parameters. At this point, we can use the second consequence of composition invariance, presented in [2], to link response time to correctness with the equation

$$T = T_{min} - S \cdot \log P = T_{min} + d \cdot \exp(- a n^c),$$

where a , c , d , and T_{min} are positive parameters. This equation shows that the law of practice should behave as a four-parameter stretched exponential function—a result consistent with [4]’s data analysis of 7,910 learning series in 24 experiments from 16 independent sources. They conclude that the three-parameter exponential function is the parsimonious choice for a law of practice that fits almost as well as the more flexible four-parameter power-exponential function. Stretching the exponential by adding a fourth parameter would only improve the fit, potentially rendering this new function, deduced from invariance hypotheses, the best option in all cases. This has yet to be empirically verified with, for example, the 16 sources of [4].

3.3 Regular spacing

Regular spacing refers to practice spread out over regular intervals, such that ability is influenced by both the number of practice trials and the length of the intervals. A coherent formulation for the regular spacing effect should reduce to simpler forms given the proper conditions. Two examples of simpler forms are the law of practice and the law of forgetting, introduced in sections 3.1 and 3.2, which can be expressed for correctness as

$$\log P = \log P_0 \cdot \exp(-\Delta\theta),$$

where $\Delta\theta = a t^{-b}$ for forgetting, $\Delta\theta = a n^c$ for practice, and a , b , c , and P_0 are positive parameters. Since a more general expression must preserve the positive homogeneity of each law, the only possible expression for change in ability is

$$\Delta\theta = a t^{-b} n^c.$$

This equation reduces to the law of forgetting when there is only one practice trial before retrieval ($n = 1$) and to the law of practice when forgetting can be ignored ($b = 0$). When $n = 1$, the regular time interval, t , reduces to retrieval time; when $n > 1$, t acts as a constant parameter for the law of regular spacing. The combined expression for the law of practice and the law of forgetting applies when comparing regular practice sequences with different time intervals. This can be seen in Fig. 2, where data from [2] and [8] have resulted in three regular sequences that show good agreement with the proposed function.

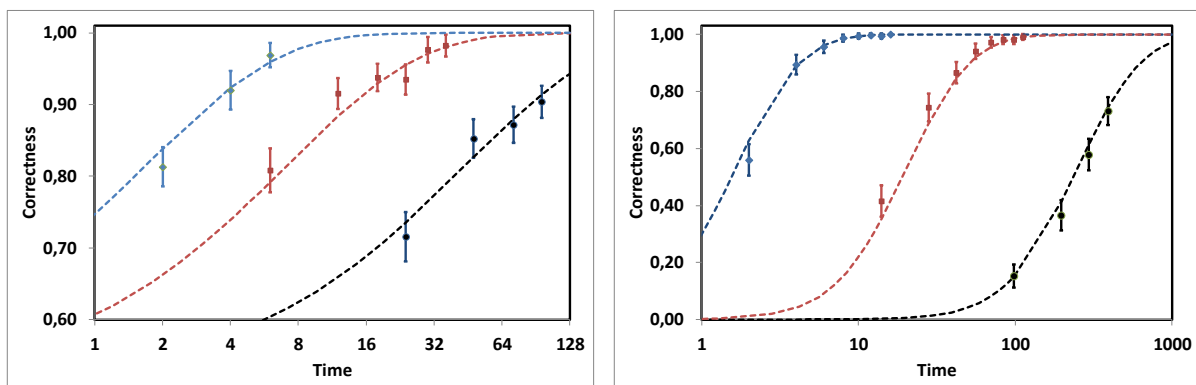


Figure 2. Empirical support for the regular spacing law: the data on the left is taken from [2] for regular intervals of 2, 6, and 24; the data on the right is taken from [8] for regular intervals of 2, 14, and 98; and the dotted lines correspond to the model minimizing the chi-squared statistic. The 2σ error bars represent the 95% confidence interval.

It is important to note that the dotted curves in Fig. 2 were produced with one regular spacing function combining both practice and forgetting. Although the new function was obtained from invariance hypotheses using a series of logical steps, it remains empirically valid.

4 CONCLUSIONS

The results presented in this paper provide empirical support for the shapes of the forgetting curve, the law of practice, and their combined effects, which were deduced from the three scale-invariance hypotheses. It is interesting to note that these hypotheses were successfully used to determine

properties for strong and highly dissimilar learning-related effects. The laws that have been proposed differ slightly from previous laws, while still being compatible with the most precise empirical data available. We believe this is a significant achievement in the field of learning sciences that could lead to a complete and coherent proposal for a uniquely defined scale-invariant theory of learning.

ACKNOWLEDGEMENTS

The author wishes to thank Patrice Potvin, Julien Mercier, Salim Sader, Gabrielle Fortin-Clément, and Reza Shams Latifi for discussions about the ideas presented in this paper.

REFERENCES

- [1] N. J. Cepeda, N. Coburn, D. Rohrer, J.T. Wixted, M. C. Mozer and H. Pashler, "Optimizing distributed practice: Theoretical analysis and practical implications". *Experimental Psychology*, vol. 56, pp. 236-246, 2009.
- [2] M. Riopel, P. Chastenay, G. Fortin-Clément, P. Potvin, S. Masson and P. Charland, "Using invariance to model practice, forgetting and spacing effects", *EDULEARN17 conference (3rd-5th July, Barcelona, Spain)*, 2017.
- [3] L. Averell and A. Heathcote, "The form of forgetting curve and the fate of memories", *Journal of Mathematical Psychology*, vol. 55, no. 1, pp. 25–35, 2011.
- [4] A. Heathcote, S. Brown, S and D. J. K. Mewhort, "The power law repealed: The case for an exponential law of practice", *Psychonomic Bulletin & Review*, vol. 7, no. 2, pp. 185-207, 2000.
- [5] H. Ebbinghaus, *Memory: A contribution to experimental psychology*, Oxford, UK: Dover, 1885/1964.
- [6] M. S. Khan, G. R.. Pasha and A. H. Pasha, "Theoretical Analysis of Inverse Weibull Distribution", *WSEAS transactions on mathematics*, vol. 7, no. 2, pp. 30-38, 2008.
- [7] A. Newell and P. S. Rosenbloom, "Mechanisms of skill acquisition and the law of practice", in *Cognitive skills and their acquisition* (J. R. Anderson Ed.), pp. 1-55, Hillsdale, NJ: Erlbaum, 1981.
- [8] P. I. Pavlik Jr. and J. R. Anderson, "Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect", *Cognitive Science*, vol. 29, pp. 559–586, 2005.