The learning curve

Draft version of:

Ritter, F. E., & Schooler, L. J. (2001). The learning curve. In *International encyclopedia of the social and behavioral sciences.* 8602-8605. Amsterdam: Pergamon. http://www.iesbs.com/

Frank E. Ritter
School of Information Sciences and Technology
Penn State University
504 Rider
120 S. Burrowes St.
University Park, PA 16801
814 865-4453
814 865-5604 (fax)

Lael J. Schooler
Department of Psychology
Penn State University
University Park, PA 16801
814 863-5813

Word count in body of text: 1508 + 2 figures

Abstract (98 words)

Most tasks get faster with practice. This holds across task size and task type. Learning curves plot time to complete a task as a function of practice. Such curves generally follow what is called a power law, thus, they are often said to conform to "the power law of practice". Cognitive psychology has shown that the power law of practice is ubiquitous, and cognitive modeling has explained both the general speedup and variability in performance, which previously was taken to be noise. Research is ongoing to find out why it is ubiquitous and where it does not apply.

The learning curve

Most tasks get faster with practice. This is not surprising because we have all seen this and perhaps know it in some intuitive sense. What is surprising is that the rate and shape of improvement is fairly common across tasks. Figure 1 shows this for a simple task plotted both on linear and log-log coordinates. The pattern is a rapid improvement followed by ever lesser improvements with further practice. Such negatively accelerated learning curves are typically described well by power functions, thus, learning is often said to follow the "power law of practice". Not shown on the graph, but occurring concurrently, is a decrease in variance in performance as the behavior reaches an apparent plateau on a linear plot. This plateau masks continuous small improvements with extensive practice that may only be visible on a log-log plot where months or years of practice can be seen. The longest measurements suggests that for some tasks improvement continues for over 100,000 trials.

There are some related regularities. There is evidence to suggest that standard deviation and skew in performance time also decrease according to a power law, but with worse correlations. Indeed, in some cases the decrease in standard deviations appears to cause the improvement, because the minimum time to perform a task does not change (Rabbit & Banerji, 1989).

The power law of practice is ubiquitous. From short perceptual tasks to team-based longer term tasks of building ships, the breadth and length of human behavior, the rate that people improve with practice appears to follow a similar pattern. It has been seen in pressing buttons, reading inverted text, rolling cigars, generating geometry proofs and manufacturing machine tools (cited in Newell and Rosenbloom, 1981), performing mental arithmetic on both large and small tasks (Delaney, Reder, Staszewski, & Ritter, 1998), performing a scheduling task (Nerb, Ritter, & Krems, 1999), and writing books (Ohlsson, 1992). Further examples are noted in reviews (e.g., Heathcote, Brown, & Mewhort, in press). In manufacturing this curve is called a progress function. You can see it for yourself by taking a task, nearly any task, and timing how long it takes to complete over 10 trials, or better over a hundred trials. For example, try reading this article upside down. The time per paragraph will generally decrease, but with some differences caused by the different words and paragraph lengths.

In general, the more averaging, the smoother the curve. The learning curve appears smoother when the data is averaged across subjects, across tasks, or both. When the tasks are known to vary in difficulty, such as different complex mental arithmetic problems (e.g., 27x5 and 23x28), the learning curve only appears when averaging is performed because the different problems naturally take different times. Even when problems are of comparable difficulty, subjects may use different strategies. For example, arithmetic problems can be solved by two strategies, retrieval and calculation. The power law applies across strategies, but the fit is better to each strategy (Delaney et al., 1998), or even an individual's strategies (Heathcote et al, in press).

Averaging can mask important aspects of learning. If the tasks vary in difficulty, the resulting line will not appear as a smooth curve, but bounce around. Careful analysis can show that different amounts of transfer and learning are occurring on each task. For example, solving the problem 22x43 will be helped more by previously solving 22x44 than by solving 17x38 because there are more multiplications shared between them. Where sub-tasks are related but different, such as sending and receiving Morse code, the curves can be related but visibly different (Bryan & Harter, 1897).

The learning curve has implications for learning in education and everyday life. It suggests that practice always helps improve performance, but that the most dramatic improvements happen first. Another implication is that with sufficient practice people can achieve comparable levels of performance. For example, extensive practice on mental arithmetic (Staszewski reported in Delaney et al., 1998) and on digit memorization have turned average individuals into world class performers.

To sum up, the learning curve is visible with enough aggregation of dissimilar tasks or across similar tasks down to the level of individual subject's strategies.

Mathematical definitions

The shape of the curve is negatively accelerated -- further practice improves performance, but with diminishing returns. Power laws and exponentials are both functions that provide this shape. Mathematical definitions are given in Table 1. The exact quality of the fit depends on innumerable details of averaging, the precise function used, and the scale. For example, the full power law formula is the most precise, but it has additional terms that are difficult to compute; and the asymptote is usually only visible when there are over 1,000 practice trials (Newell & Rosenbloom, 1981). The typical power law formula is simpler, but leaves out previous practice. When using this formula, the constants for a set of data can be easily computed by taking the log of the trial number and log of task time and computing a linear regression. That is, fitting a regression in log-log space.

In general, the power function fit appears to be robust, regardless of the methods used (Newell & Rosenbloom, 1981). However, recent work (Heathcote et al, 2000) suggests that the power law might be an artifact arising from averaging (Anderson & Tweney, 1997), and that the exponential function may be the best fit when individual subjects employing a single strategy are considered. Distinguishing between the power and exponential functions is not just an esoteric exercise in equation fitting. If learning follows an exponential, then learning is based on a fixed percentage of what remains to be learnt. If learning follows a power law, then learning slows down.

Regardless of the functional form of the practice curve, there remain some systematic deviations that cause problems, at the beginning and end of long series. The beginning deviations may represent an encoding process. For example, it may be necessary to transform a declarative description of a task into procedures before actual practice at the task can begin (Anderson & Lebiere, 1998); the residuals at the end may represent approaching the minimum time for a given task as defined by an external apparatus. These effects appear in Figure 1 as well.

Table 1. Functions that fit practice data.

[format according to conventions adopted for encyclopedia]

Time = MinTime +const. (TrialNumber + PrevPractice) ^ -(constant for task)
[Full Power law formula]

Time = constant * (TrialNumber) ^ -(constant for task)
[Typical power law formula]

Time = constant * e ^ -(TrialNumber +PrevPractice)
[Simple exponential formula]

Process based explanations of the learning curve

The power law of learning is an important regularity of human behavior that all theories of learning must address. For example, Logan (1988) suggests that the retrieval of memory traces guides responding in speeded tasks. Response times represent the fastest retrieval time among the memory traces racing to guide the response. The power law falls out of this analysis, because with practice the number of redundant memory traces increases, which, in turn, increases the chances of observing fast retrieval times. Olhsson's (1996) theory notes how the learning curve could arise out of error correction. Different assumptions (e.g., how fast errors are caught) give rise to different curves including the exponential and power law. The learning curve has also been demonstrated for connectionist models as well.

ACT-R (J. R. Anderson & Lebiere, 1998) and Soar (Newell, 1990), two cognitive architectures, generally predict a power law speedup, but for different reasons. ACT-R does this because rules and memory traces are strengthened according to a power law based on the assumption that the cognitive system is adapted to the statistical structure of the environment (J. R. Anderson & Schooler, 1991). Several models in Soar (see related entry) have been created that model the power law (e.g., Nerb et al., 1999; Newell, 1990). These models explain the power law as arising out of hierarchical learning (i.e., learning parts of the environment or internal goal structure) that initially learns low level actions that are very common and thus useful, and with further practice more useful larger patterns are learned but that occur infrequently. The Soar models typically vary from human learning in that they learn faster than humans, and they do not learn for as long a period. Soar and ACT-R also predict variance in the improvement on all tasks due to different amounts of transfer across problems and learning episodes. Figure 2 shows how such a model can predict differential transfer as well as continuous learning as it appears in human data.

Summary

The learning curve is a success story for cognitive psychology, which has shown that learning is ubiquitous and has provided mathematical accounts of the rate. The learning curve is also a success story for cognitive modeling, which has explained the curve and the noise inherent in it partly as differences in transfer of knowledge between tasks and how the curve arises out of mechanisms necessary for processing. The multiple explanations also suggest that there may be multiple ways that the curve can arise.

Related Topics

Soar, ACT-R, learning theories.

References (up to 15 have 14)

1. Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.
2. Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396-408.
3. Anderson, R. B., & Tweney, R. D. (1997). Artifactual power curves in forgetting. *Memory & Cognition, 25*, 724-730.
4. Bryan, W. L., & Harter, N. (1887). Studies in the physiology and psychology of the telegraphic language. *Psychological Review, 4*, 27-53.
5. Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy specific nature of improvement: The power law applies by strategy within task. Psychological Science, 9(1), 1-8.
6. Heathcote, A., Brown, S., & Mewhort, D. J. K. (in press). The Power Law repealed: The case for an Exponential Law of Practice. *Psychonomic Bulletin & Review*.
7. Jones, G., Ritter, F. E., & Wood, D. J. (2000). Using a cognitive architecture to examine what develops. *Psychological Science, 11*(2),1-8.
8. Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.
9. Nerb, J., Ritter, F. E., & Krems, J. (1999). Knowledge level learning and the power law: A Soar model of skill acquisition in scheduling. *Kognitionswissenschaft [Journal of the German Cognitive Science Society] Special issue on cognitive modeling and cognitive architectures, D. Wallach & H. A. Simon (eds.)*,20-29.
10. Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
11. 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.* 1-51. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

12. Ohlsson, S. (1992). The learning curve for writing books: Evidence from Professor Asimov. *Psychological Science, 3*(6), 380-382.

13. Ohlsson, S. (1996). Learning from performance errors. *Psychological Review, 103*(2), 241-262.

14. Rabbit, P., & Banerji, N. (1989). How does very prolonged practice improve decision speed? *Journal of Experimental Psychology: General, 118*,338-345.

15. Ritter, F. E., & Bibby, P. A. (1997). Modelling learning as it happens in a diagramatic reasoning task (Tech. Report No. 45). ESRC CREDIT, Dept. of Psychology, U. of Nottingham.

16. Rosenbloom, P. S., & Newell, A. (1987). Learning by chunking, a production system model of practice. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development.* 221-286. Cambridge, MA: MIT Press.

17. Seibel, R. (1963). Discrimination reaction time for a 1,023-alternative task. *Journal of Experimental Psychology, 66*(3), 215-226.

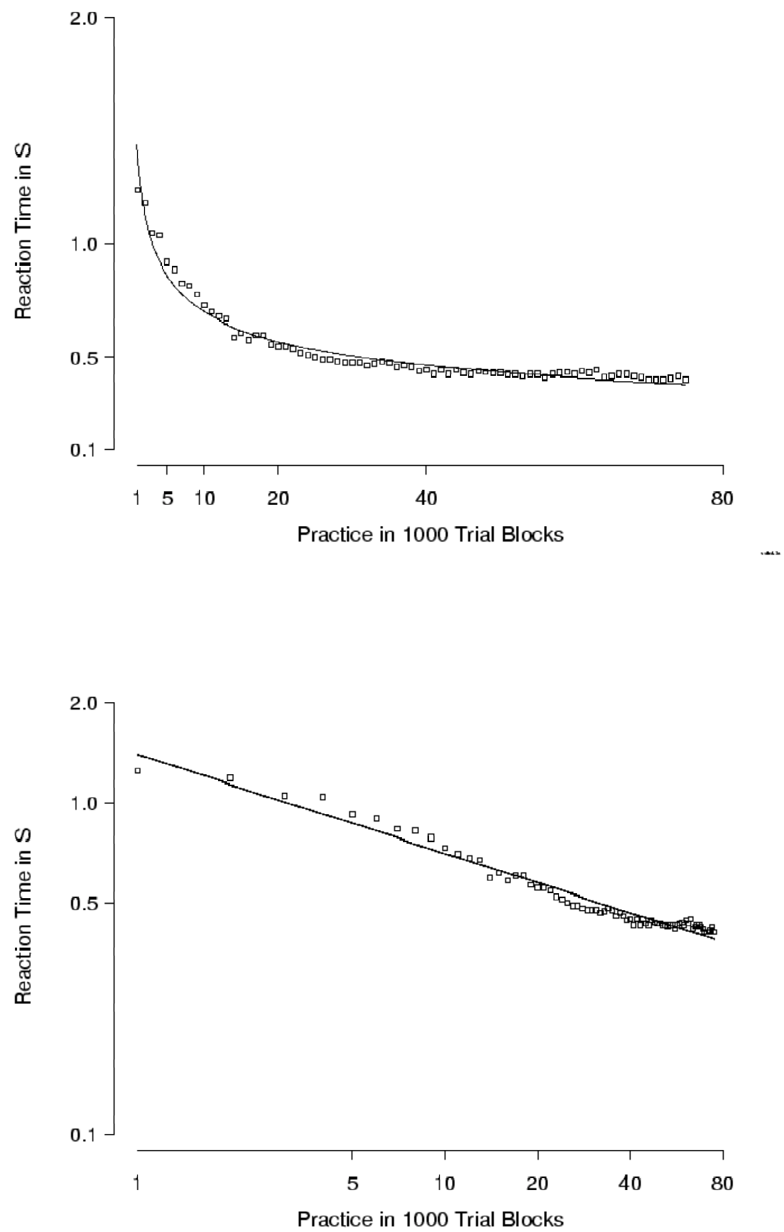

Figure 1 a & b. Figure 1. Time to perform a simple task on a linear and log-log plot as well as a power law fit to the data (adapted from Seibel, 1963). [figures available as postscript files]

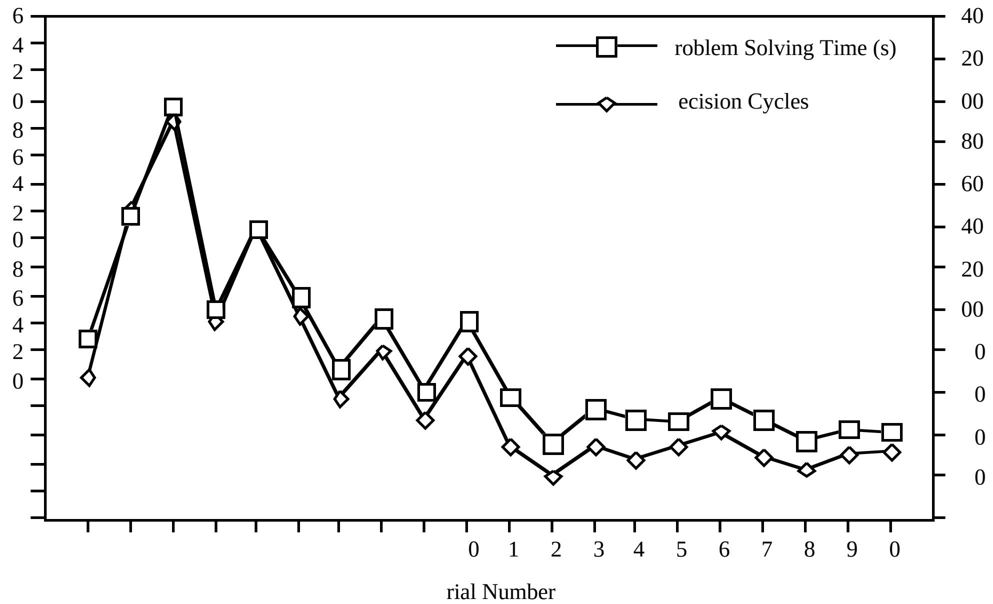

Figure 2. The time to perform a trouble shooting task by both a cognitive model and a subject (Ritter & Bibby, 1997).

Learning curves â€“ the basic idea. Letâ€™s say we have some data and split it into a training set and validation set. We take one single instance (thatâ€™s right, one!) from the training set and use it to estimate a model.Â The error scores will vary more or less as we change the training set. We thus have two error scores to monitor: one for the validation set, and one for the training sets. If we plot the evolution of the two error scores as training sets change, we end up with two curves. These are called learning curves. Learning curve definition: A learning curve is a process where people develop a skill by learning from their | Meaning, pronunciation, translations and examples.Â There is a learning curve in the process of seeking employment. COBUILD Advanced English Dictionary. Copyright Â© HarperCollins Publishers. A learning curve is a graphical representation of how an increase in learning (measured on the vertical axis) comes from greater experience (the horizontal axis); or how the more someone (or thing) does something, the better they get at it. [1]. Fig 1: Learning curve for a single subject, showing how learning improves with experience. Fig 2: A learning curve averaged over many trials is smooth, and can be expressed as a mathematical function.