Cognitive Skills and Their Acquisition

Edited by
John R. Anderson

ROUTLEDGE

Carnegie Mellon Symposia on Cognition Series

COGNITIVE SKILLS AND THEIR ACQUISITION

COGNITIVE SKILLS AND THEIR ACQUISITION

Edited by
JOHN R. ANDERSON
CARNEGIE-MELLON UNIVERSITY

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First Published by

Lawrence Erlbaum Associates, Inc., Publishers
365 Broadway
Hillsdale, New Jersey 07642

This edition published 2011 by Routledge
711 Third Avenue, New York, NY 10017
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Library of Congress Cataloging in Publication Data
Symposium on Cognition, 16th, Carnegie-Mellon University, 1980.
Cognitive skills and their acquisitions.
Bibliography: p.
Includes index.
1. Cognition—Congresses. 2. Learning, Psychology of—Congresses. I. Anderson, John Robert, 1947–
II. Title. [DNLM: 1. Cognition—Congresses. 2. Learning —Congresses. W3 C126B 16th 1980c / BF 311 C6785 1980]
BF311.S83 1980 153.4 80-28702
ISBN 0-89859-093-0

Publisher's Note

The publisher has gone to great lengths to ensure the quality of this reprint but points out that some imperfections in the original may be apparent.

TO THE MEMORY OF OUR COLLEAGUE

LEE W. GREGG

WHO, AS MEMBER AND HEAD OF THE PSYCHOLOGY DEPARTMENT
AT CARNEGIE-MELLON UNIVERSITY, FOR A QUARTER CENTURY
GAVE HIS ENERGIES AND TALENTS UNSELFISHLY TO OUR
EXPLORATION OF THE NEW DOMAINS IN COGNITIVE PSYCHOLOGY.

Contents

Preface

This book is a collection of the papers presented at the Sixteenth Annual Carnegie Symposium on Cognition, held in May 1980. A couple of years earlier Pat Langley and Bob Neches persuaded the psychology department about the need to have a symposium on the topic of learning, and I was recruited to organize the Symposium. I too was persuaded that the time was ripe for a new attack on issues of learning. In separating itself from behaviorism, cognitive psychology had abandoned interest in learning and for 20 years had focused on understanding the performance of established cognitive skills. As a consequence we have acquired a fair knowledge of cognitive mechanisms, but this knowledge of performance is both disjointed and hard to use in educational applications. Learning theory offers both the potential of achieving a new degree of generality in our understanding of human cognition and the possibility of opening the way for applications. Given our much more sophisticated understanding of cognition, we can hope that learning mechanisms can be developed that avoid the weaknesses of past learning theories. The need for a new learning theory and potential of such a learning theory is discussed at length in the Langley and Simon chapter.

In the past few years there has developed a number of exciting new endeavors at understanding learning. I invited what I thought were among the best practitioners of this new approach to learning, and they all accepted. The symposium promised to be exciting and it was. It turned out that our ideas about learning were more similar than I had expected, given what a broad domain of phenomena might fit under the topic. In particular, we were interested in cognitive skills and how they develop in complex environments over long periods of practice. Moreover, we found we were using similar manipulations, choosing similar types of skills for study, finding similar results, and proposing similar theoretical mechanisms.

I have simply ordered the papers into chapters as they were presented in the conference. It seems unwise to try to organize these chapters into subgroups concerned with various topics. The problem is that there are so many shared themes running through these chapters that there are multiple possible classification schemes. I leave it to the reader to decide what the most salient dimensions are. Perhaps, presented in this order the reader can get a sense for some of the excitement we felt at the Symposium presentations as we waited to hear what each presenter had to say.

Newell and Rosenbloom look at the ubiquitous power law that relates time to perform a task to amount of practice. They show how difficult it is to derive the power law but are able to show how it might be derived from a set of assumptions about chunking and variability in the environment. Neves and Anderson look at what we call knowledge compilation processes that are responsible for the specialization of procedures to particular tasks and for the creation of larger operators in a production system framework. This is related to the power law and to various behaviors associated with automaticity. Clayton Lewis presents an analysis of the differences (or lack of differences) between the algebraic skills of expert mathematicians and students of varying strengths. He also shows how various speedups in performance of algebraic operations cannot be explained by composition learning mechanisms assumed both by Newell and Rosenbloom and by Neves and Anderson.

Shiffrin and Dumais provide a review of their extensive research concerned with the development of automaticity and a theoretical discussion of the distinction between controlled and automatic processing. They also describe some experimental investigations of why consistency is critical to the development of automaticity. Chase and Ericsson provide a description of their marvelous subject who has trained his memory span to over 80 digits. They also provide a careful theoretical analysis of the memory skills underlying this feat, challenging some of the established notions about human memory.

Anderson, Greeno, Kline, and Neves provide an analysis of the development of the problem-solving skills that underlie proof generation in geometry. We try to identify the learning processes responsible for the initial understanding and encoding of the skill, its speedup, and the tuning of the proof search with practice. Hayes-Roth, Klahr, and Mostow analyze the kind of learning that is involved in taking advice. Typically, a considerable amount of analysis is required before advice can be converted into an executable procedure. Jeffries, Turner, Polson, and Atwood consider the differences between experts and novices at software design. They argue that experts have acquired a design schema that enables them successfully to decompose a design problem into subproblems.

De Kleer and Brown provide an analysis of the kinds of representations and processes needed to reason about mechanisms. This analysis is critical to understanding how people can learn simply by running mental simulations of processes and trying to troubleshoot problems that come up in the mental simulations. Larkin provides an analysis of the mechanisms underlying solving physics problems and how one becomes more expert in this domain. Critical to her learning analysis is the idea that working-forward productions are created that embellish a problem representation with information that previously had to be obtained by working backward. Rumelhart and Norman are concerned with how new schemata can be created out of analogy to existing schemata. They provide both a discussion of the mechanisms underlying the analogical process and an application of this process to explaining confusions students have in learning to use a text editor.

Finally, we have the chapter by Simon and Langley. Theirs is not a symposium commentary, although the chapter does make comments on the Symposium papers. In their chapter they provide an analysis of why there is this renewed interest in learning, why it is a central problem for cognitive psychology, and what the issues are that need to be addressed in this new work on learning.

These chapters represent some of the best work in the area and contain ideas that should prove to have growing importance in the 1980s. The chapters are well-written, even though I put much pressure on the authors to get them out quickly. So, read and enjoy.

ACKNOWLEDGMENTS

Lee Gregg was very helpful during the early stages of planning this Symposium, and much of the credit for its success goes to him. His loss is profoundly felt in the Department. We are grateful to the Alfred P. Sloan Foundation for a grant that provided a substantial part of the support of this year's Symposium. Vickie Silvis Wille has had a major hand in organizing the Symposium and preparing material for the book. The success of the Symposium and the processing of the book has depended critically on her efforts, and we all owe her our deepest gratitude. She is supported by my NSF Grant BNS78-17463 and by my ONR Contract N00014-77-C-0242.

John R. Anderson

1 Mechanisms of Skill Acquisition and the Law of Practice

Allen Newell and Paul S. Rosenbloom
Department of Computer Science
Carnegie-Mellon University

INTRODUCTION[1]

Practice makes perfect. Correcting the overstatement of a maxim: Almost always, practice brings improvement, and more practice brings more improvement. We all expect improvement with practice to be ubiquitous, though obviously limits exist both in scope and extent. Take only the experimental laboratory: We do not expect people to perform an experimental task correctly without at least some practice; and we design all our psychology experiments with one eye to the confounding influence of practice effects.

Practice used to be a basic topic. For instance, the first edition of Woodworth (1938) has a chapter entitled "Practice and Skill." But, as Woodworth [p. 156] says, "There is no essential difference between practice and learning except that the practice experiment takes longer." Thus, practice has not remained a topic by itself but has become simply a variant term for talking about learning skills through the repetition of their performance.

With the ascendence of verbal learning as the paradigm case of learning, and its transformation into the acquisition of knowledge in long-term memory, the study of skills took up a less central position in the basic study of human behavior. It did not remain entirely absent, of course. A good exemplar of its

[1]This chapter relies on the data of many other investigators. We are deeply grateful to those who made available original data: John Anderson, Stu Card, Paul Kolers, Tom Moran, David Neves, Patrick Rabbitt, and Robert Seibel. We are also grateful to John Anderson, Stu Card, Clayton Lewis, and Tom Moran for discussions on the fundamental issues and, especially, to Clayton Lewis for letting us read his paper, which helped to energize us to this effort.

continued presence can be seen in the work of Neisser, taking first the results in the mid-sixties on detecting the presence of ten targets as quickly as one in a visual display (Neisser, Novick, & Lazar, 1963), which requires extensive practice to occur; and then the recent work (Spelke, Hirst, & Neisser, 1976) showing that reading aloud and shadowing prose could be accomplished simultaneously, again after much practice. In these studies, practice plays an essential but supporting role; center stage is held by issues of preattentive processes, in the earlier work, and the possibility of doing multiple complex tasks simultaneously, in the latter.

Recently, especially with the articles by Shiffrin & Schneider (1977; Schneider & Shiffrin, 1977), but starting earlier (LaBerge, 1974; Posner & Snyder, 1975), emphasis on *automatic* processing has grown substantially from its level in the sixties. It now promises to take a prominent place in cognitive psychology. The development of automatic processing seems always to be tied to extended practice and so the notions of skill and practice are again becoming central.

There exists a ubiquitous quantitative law of practice: It appears to follow a power law; that is, plotting the logarithm of the time to perform a task against the logarithm of the trial number always yields a straight line, more or less. We shall refer to this law variously as the *log-log linear learning law* or the *power law of practice*.

This empirical law has been known for a long time; it apparently showed up first in Snoddy's (1926) study of mirror-tracing of visual mazes (see also Fitts, 1964), though it has been rediscovered independently on occasion (DeJong, 1957). Its ubiquity is widely recognized; for instance, it occupies a major position in books on human performance (Fitts & Posner, 1967; Welford, 1968). Despite this, it has captured little attention, especially theoretical attention, in basic cognitive or experimental psychology, though it is sometimes used as the form for displaying data (Kolers, 1975; Reisberg, Baron, & Kemler, 1980). Only a single model, that of Crossman (1959), appears to have been put forward to explain it.[2] It is hardly mentioned as an interesting or important regularity in any of the modern cognitive psychology texts (Calfee, 1975; Crowder, 1976; Kintsch, 1977; Lindsay & Norman, 1977). Likewise, it is not a part of the long history of work on the *learning curve* (Guilliksen, 1934; Restle & Greeno, 1970; Thurstone, 1919), which considers only exponential, hyperbolic, and logistic functions. Indeed, a recent extensive paper on the learning curve (Mazur & Hastie, 1978) simply dismisses the log-log form as unworthy of consideration and clearly dominated by the other forms.

---

[2]But see Suppes, Fletcher, and Zanotti (1976), who do develop a model yielding a power law for instructional learning, though their effort appears independent of a concern with the general regularity. Unfortunately, their description is too fragmentary and faulty to permit it to be considered further.

The aim of this chapter is to investigate this law. How widespread is its occurrence? What could it signify? What theories might explain it? Our motivation for this investigation is threefold. First, an interest in applying modern cognitive psychology to user-computer interaction (Card, Moran, & Newell, 1980a; Robertson, McCracken, & Newell, in press) led us to the literature on human performance, where this law was prominently displayed. Its general quantitative form marked it as interesting, an interest only heightened by the apparent general neglect of the law in modern cognitive psychology. Second, a theoretical interest in the nature of the architecture for human cognition (Newell, 1980) has led us to search for experimental facts that might yield some useful constraints. A general regularity such as the log-log law might say something interesting about the basic mechanisms of turning knowledge into action. Third, an incomplete manuscript by Clayton Lewis (no date) took up this same problem; this served to convince us that an attack on the problem would be useful. Thus, we welcomed the excuse of this conference to take a deeper look at this law and what might lay behind it.

In the next section we provide many examples of the log-log law and characterize its universality. In the following section we perform some basic finger exercises about the nature of power laws. Then we investigate questions of curve fitting. In the next section we address the possible types of explanations for the law; and we develop one approach, which we call the *chunking theory of learning*. In the final section, we sum up our results.

THE UBIQUITOUS LAW OF PRACTICE

We have two objectives for this section. First, we simply wish to show enough examples of the regularity to lend conviction of its empirical reality. Second, the law is generally viewed as associated with *skill,* in particular, with perceptual-motor skills. We wish to replace this with a view that the law holds for practice learning of all kinds. In this section we present data. We leave to the next section issues about alternative ways to describe the regularity and to yet subsequent sections ways to explain the regularity.

We organize the presentation of the data by the subsystem that seems to be engaged in the task. In Table 1.1 we tabulate several parameters of each of the curves. Their definitions are given at the points in the chapter where the parameters are first used.

Perceptual-Motor Skills

Let us start with the historical case of Snoddy (1926). As remarked earlier, the task was mirror-tracing, a skill that involves intimate and continuous coordination of the motor and perceptual systems. Figure 1.1 plots the log of performance on the vertical axis against the log of the trial number for a single subject.

TABLE 1.1
Power Law Parameters for the (Log-Log) Linear Data Segments

	Power Law $T = BN^{-\alpha}$		
Data Set	*B*	α	r^2
Snoddy (1926)	79.20	.26	.981
Crossman (1959)	170.1	.21	.979
Kolers (1975) - Subject HA	14.85	.44	.931
Neisser et al. (1963)			
Ten targets	1.61	.81	.973
One target	.68	.51	.944
Card, English & Burr (1978)			
Stepping keys - Subj. 14	4.95	.08	.335
Mouse - Subj. 14	3.02	.13	.398
Seibel (1963) - Subject JK	12.33	.32	.991
Anderson (1980) - Fan 1	2.358	.19	.927
Moran (1980)			
Total time	30.27	.08	.839
Method time	19.59	.06	.882
Neves & Anderson (in press)			
Total time - Subject D	991.2	.51	.780
The Game of Stair			
Won games	1763	.21	.849
Lost games	980	.18	.842
Hirsch (1952)	10.01	.32	.932

The first important point is:

- The law holds for performance measured as the *time* to achieve a fixed task.

Analyses of learning and practice are free a priori to use any index of performance (e.g., errors or performance time, which decrease with practice; or amount or quality attained, which increase with practice). However, we focus exclusively on measures of performance time, with quality measures (errors, amount, judged quality) taken to be essentially constant. Given that humans can often engage in tradeoffs between speed and accuracy, speed curves are not definable without a specification of accuracy, implicit or otherwise.[3] As we illustrate later, the log-log law also appears to hold for learning curves defined

[3]Snoddy used an indicator, 1/(time+errors), and we have replotted the figure using time+errors. This strikes the modern eye as incongruous, adding together apples and oranges. In fact, the measure is almost purely performance time. Snoddy was endeavoring to cope with the speed/accuracy tradeoff. He fixed the error rate to be equal to the performance time (in seconds) and had the subject work faster or slower in order to hold the error rate at that level. Thus the error rate bore a fixed average relationship to time; and adding the actual value of the errors to the performance time was a way of compensating for momentary shifts in the speed/accuracy tradeoff.

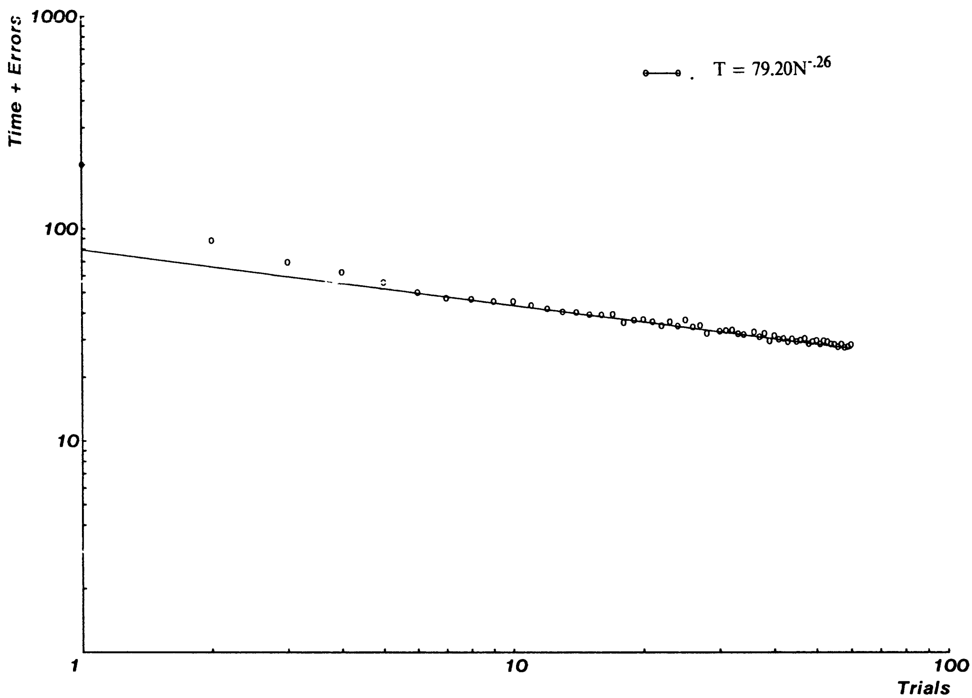

FIG. 1.1. Learning in a mirror tracing task (log–log coordinates). Replotted from Snoddy (1926).

on other performance criteria. Though significant for understanding the cause of the power law, we only note the existence of these other curves.

Several other things can be noted in Fig. 1.1, which show up generally in the other curves.

- The points are sparse at the left and become denser to the right. This arises from taking the log of the trial number. Even when trials are aggregated into blocks, this is usually done uniformly in linear space. Thus, this is just an artifact of the display.
- There is systematic deviation at one end. Here it is the beginning. Snoddy made a lot of this initial deviation, though we need not follow him in this. As we shall see, systematic deviation can occur at either end.
- There is little doubt that the bulk of the curve lies along a line in log–log space. This arises in part because of the relatively large number of points available.[4] The curves are for an individual, not for grouped data. This is not a condition of the law, but it shows that the law holds for individual data.

---

[4]Obvious deviations at the ends of the empirical curves were eliminated before the fits in Table 1.1 were computed. The equations therefore primarily represent this linear portion of the curve. The solid line in Fig. 1.1 (and in the following figures) reflects this fit.

• Data are rarely presented on many subjects, though in some cases such data exists and (apparently) is robust. For instance, Snoddy took his curve as diagnostic and appears to have gathered it on large numbers of individuals, though he never reported any mass of data.

In Table 1.1 we tabulate several critical features of the Snoddy data. The following equations describe the power law in linear and log–log spaces:

$$T = BN^{-\alpha} \quad (1)$$

$$\log(T) = \log(B) - \alpha \log(N) \quad (2)$$

B is the performance time on the first trial ($N = 1$) and α is the slope of the line (i.e., the learning rate). A positive value of α (e.g., .26 for the curve of Fig. 1.1) indicates a decreasing curve, because we have located the minus sign in the equation itself.

Another example from a task that appears to involve intimate motor-perceptual coordination is shown in Fig. 1.2. This is Crossman's (1959) famous data on the manufacture of cigars by female operators using a cigar-making machine. Noteworthy is the number of trials, namely, up to 20 million cigars.

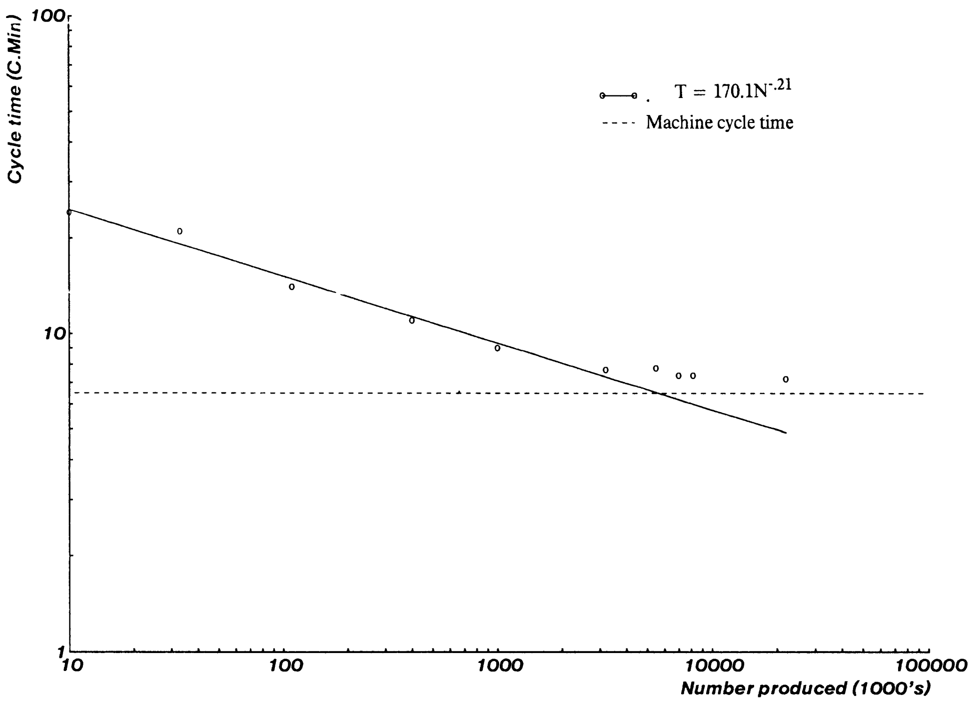

FIG. 1.2. Cross-sectional study of learning in cigar manufacturing (log–log coordinates). Replotted from Crossman (1959).

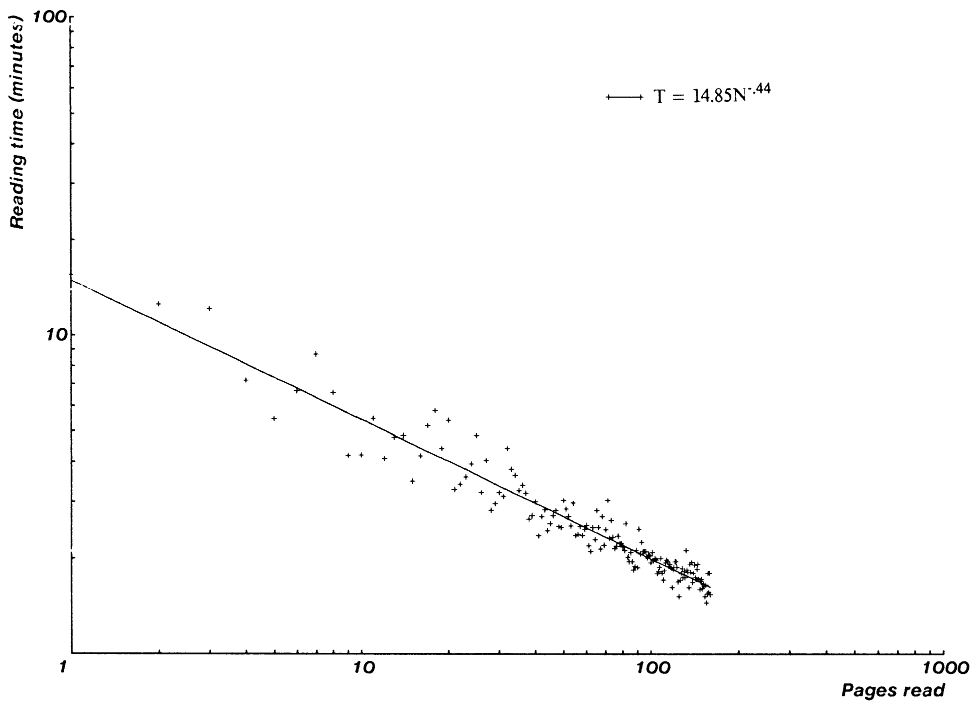

FIG. 1.3. Learning to read inverted text (log-log coordinates). Plotted from the original data for Subject HA (Kolers, 1975).

Also, there is a known lower bound for the performance time, namely, the cycle time of the machine. The curve eventually deviates from the log-log line, flattening out in submission to physical necessity. Still, practice followed the law for almost 3 million trials (and 2 years). Furthermore, additional small improvements continued; and it would be foolish indeed to predict that no further improvements would occur. Crossman's data differs from all other data in being cross-sectional (i.e., different individuals make up each point).

Perception

Figure 1.3 shows the data from one subject (of eight) in Kolers's well known studies on reading graphically transformed text (Kolers, 1975). Here, the transformation is inversion of each line around its horizontal axis. The task of the subject is to read many pages of such text for comprehension. Reading in general is a complex task, but the difficulties here are clearly strongly perceptual, being caused primarily by the perceptual transformation. Without inversion, reading is much faster and improves hardly at all (though we don't show Kolers's control data on this). In any event, as the figure shows, learning is log-log linear.

Figure 1.4 shows some data replotted from a paper by Neisser, Novick, & Lazar (1963). The task consisted of finding any of multiple targets in pages of

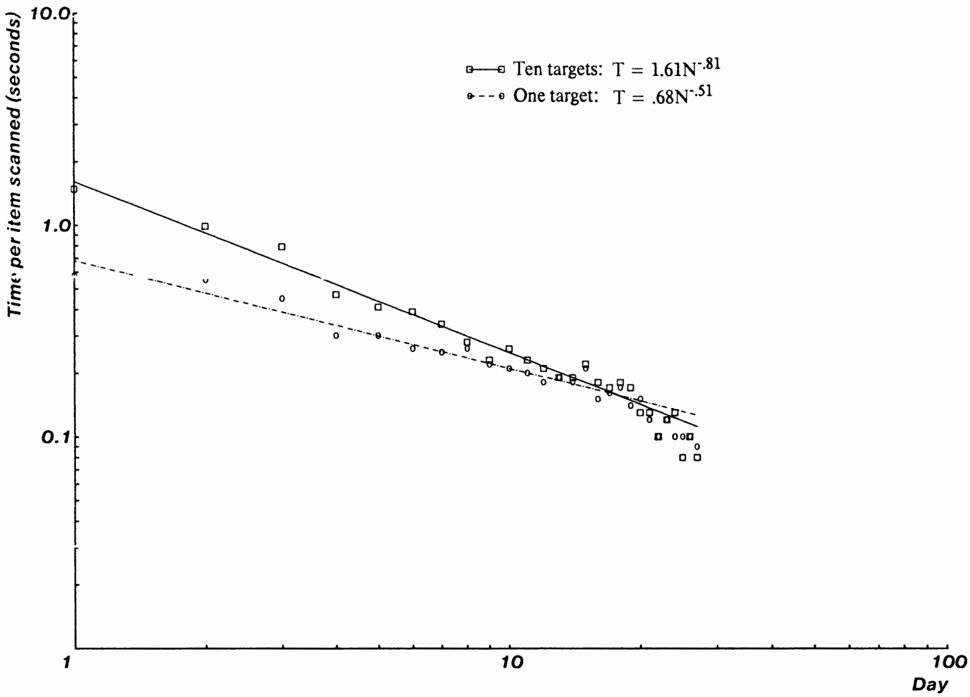

FIG. 1.4. Learning to scan for visual targets (log-log coordinates). Replotted from Neisser, Novick, & Lazar (1963).

letters. The result was that, with practice, identification time becomes essentially independent of the size of the target set. As Fig. 1.4 shows, this data also follows the log-log law, though there seems to be a slight drop at the end. These two curves (scanning for one target and for ten targets) represent the two bounding conditions of the five used in the experiment. Each curve is the average of six subjects. One of the reasons for exhibiting these particular curves is to point out that much learning data in the literature fits the log-log law, even though it has not been plotted that way.

Motor Behavior

Figure 1.5 is from a task where a subject sees a target mark appear on a video terminal and has to position the cursor at that mark (Card, English, & Burr, 1978). Four different pointing devices were used: a mouse, which permits a smooth pointing motion isomorphic to the motion of the cursor; a joystick; a set of stepping keys; and a set of text keys, which allow movement by paragraph, word, etc. Some of these devices are well described by Fitts's law (Fitts, 1954); some have a different structure. The two curves in Fig. 1.5 show the mouse and stepping key data for one subject, averaged over blocks of 20 trials (excluding errors). For all of the devices, the total performance time follows the law, though the degree of variability increases as one moves from the Fitts's law devices (the mouse) toward the other ones.

Elementary Decisions

Figure 1.6 is from a task designed by Seibel (1963) to probe the dependence of reaction time on the number of alternatives. It followed in the wake of the work by Hick (1952), Hyman (1953), and others showing that choice RT was linear in the information (bits) required to select the response, at least for small ensembles (up to 3 or 4 bits). The subject's 10 fingers rested on 10 response keys (shaped to fit the natural position of the resting hand) and looked at 10 stimulus lights that were configured isomorphically to the keys. A subset of the lights would turn on, and the subject was to strike the corresponding keys. There are 1023 ($2^{10} - 1$) different subsets of the lights; hence, the arrangement achieves a choice RT task of 10 bits. For our purposes what is interesting is that the learning over a large number of trials (40,000) was log-log linear, though at the end the curve flattens out. This is data for a single subject, averaged over blocks of 1023 trials; approximately the same behavior was shown by each of three subjects.

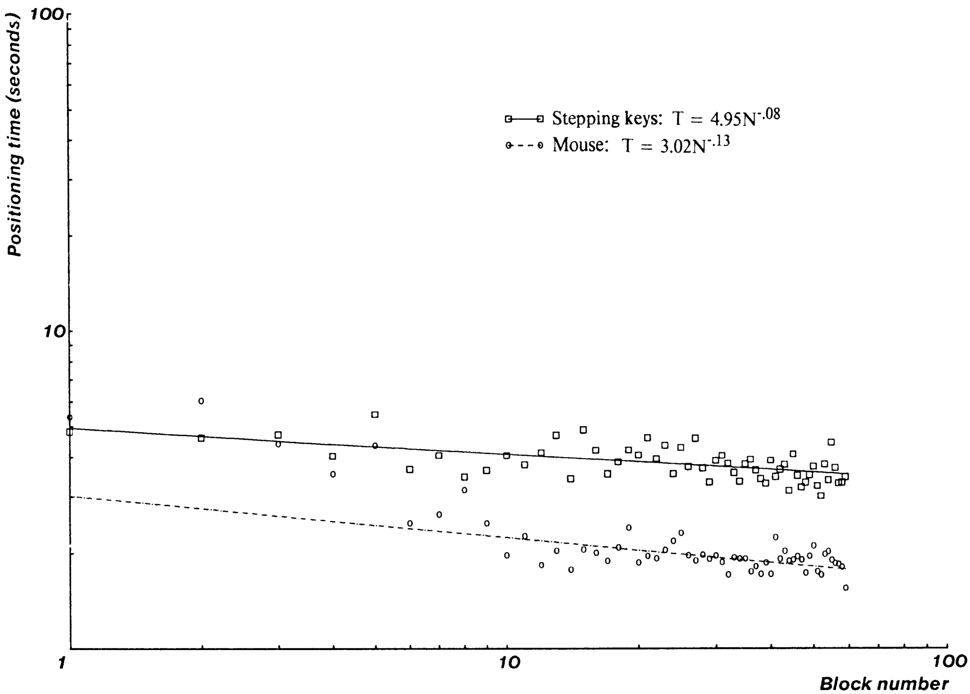

FIG. 1.5. Learning to use cursor positioning devices (log-log coordinates). Plotted from the original data for Subject 14 (Card, English, & Burr, 1978).

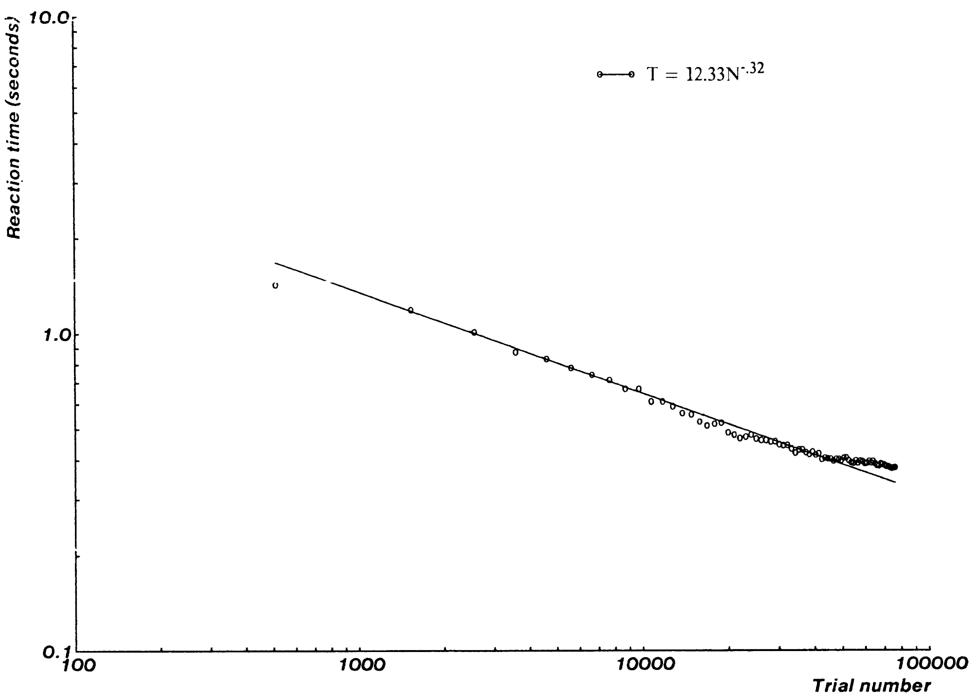

FIG. 1.6. Learning in a ten finger, 1023 choice task (log-log coordinates). Plotted from the original data for Subject JK (Seibel, 1963).

Memory

Figure 1.7 is from some unpublished work of John Anderson (1980). It shows learning performance in a task that would appear to stress mostly memory, though of course it has both a perceptual and a motor response aspect. The task is an old-new judgment on a set of simple sentences, such as "The doctor talked to the lady." There is a fixed population of grammatical subjects, objects, and verbs; a subset of these are seen initially, and then sets of the originals plus distractors (made from the same populations) are shown repeatedly. After awhile of course a subject has seen both the targets and the distractors several times. The figure shows that the reaction time to make the memory judgment follows the log-log linear law.

Complex Routines

Figure 1.8 is from some work done in connection with a general attack on understanding user-computer interaction (Moran, 1980). A specific, complex on-line editing task of completely rearranging a given sentence of three clauses is being performed repeatedly. The task is absolutely identical each time (i.e., the same sentence). Thus we are seeing a subject simply follow an internally familiar, complex plan. The top curve is the total time to perform the task. The lower

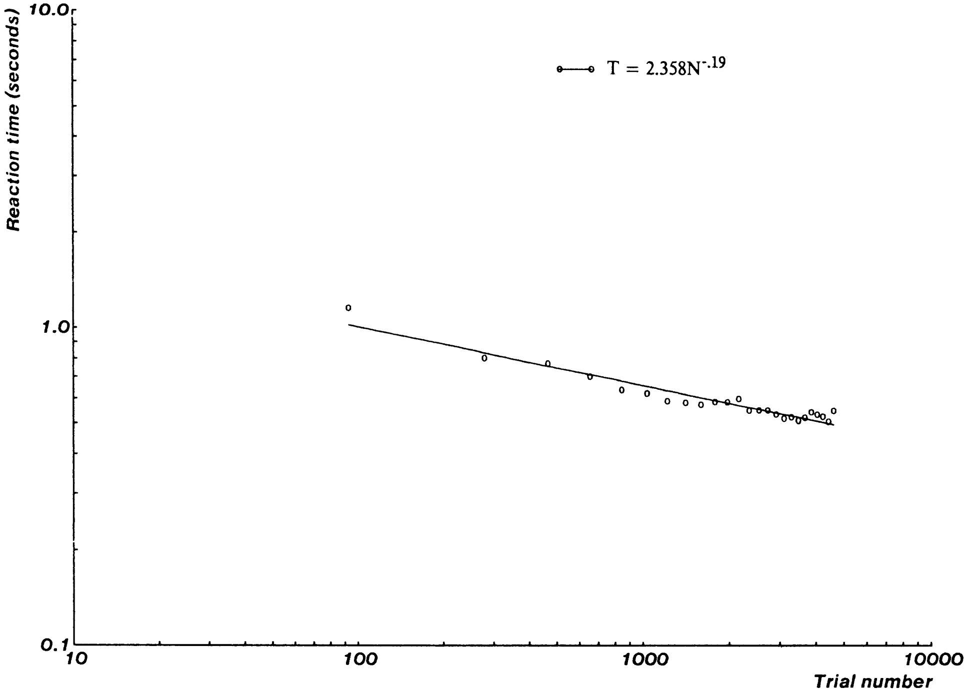

FIG. 1.7. Learning in a sentence recognition task (log–log coordinates). Plotted from the fan 1 data of Anderson (1980).

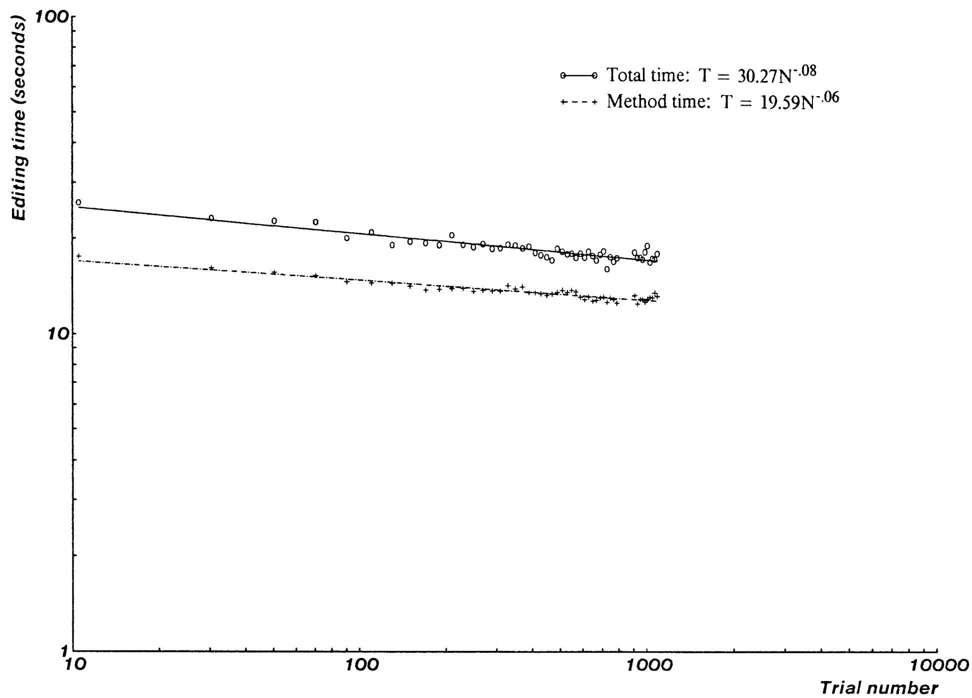

FIG. 1.8. Learning of a complex on-line editing routine (log–log coordinates). Plotted from the original data of Moran (1980).

curve shows the execution time attributable to the specific method being used, computed according to a model based on the keystroke sequence (Card, Moran, & Newell, 1980b). It decreases only if the subject makes some improvement that changes the number of keystrokes rather than decreasing think time. Both curves show log-log linear practice effects.

Figure 1.9 shows a more complex cognitive task (Neves & Anderson, in press), but one that still can be considered as evolving toward a complex routine. The task is to find the rule justifying each step in a proof in a simple formal proof system, taken to mirror the typical proof system of synthetic geometry. The subject faces a display that shows (on request) the lines of the proof, the axioms, or the theorems that are applicable to derive new steps in the proof. He must assign to each step whether it is an axiom or which rule is used in its derivation. As the figure shows, the time to perform this task follows the log-log linear law.

Problem Solving

Figure 1.10 shows our own small addition to the population of tasks known to follow the log-log linear law. As the ubiquity of the law became clear, it seemed that it was miscast as something applying only to perceptual and motor skills, but

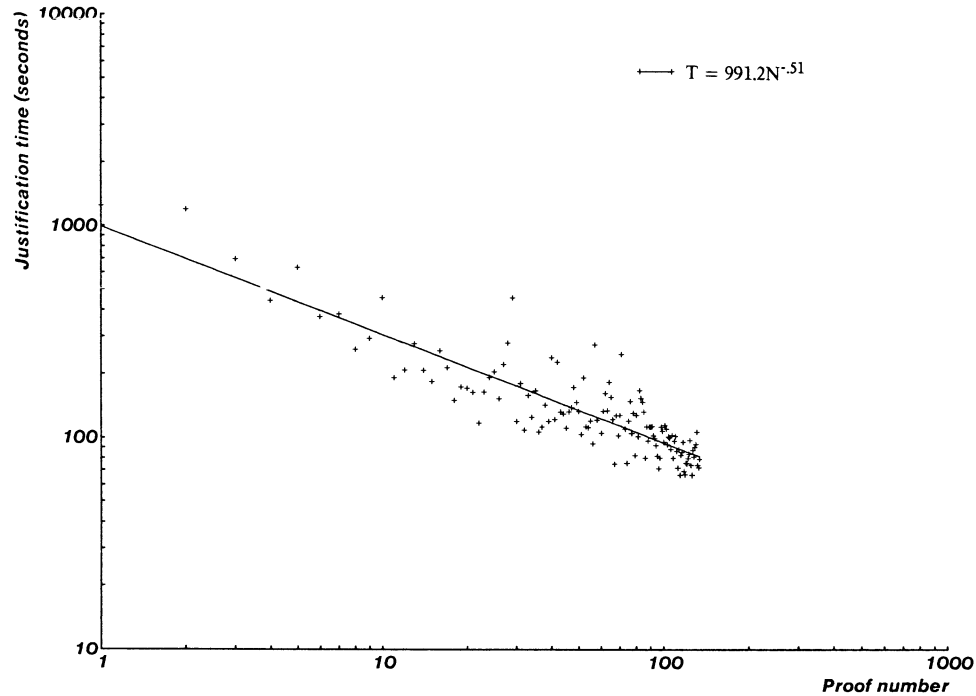

FIG. 1.9. Learning in a geometry proof justification task (log-log coordinates). Plotted from the original data (Neves & Anderson, in press).

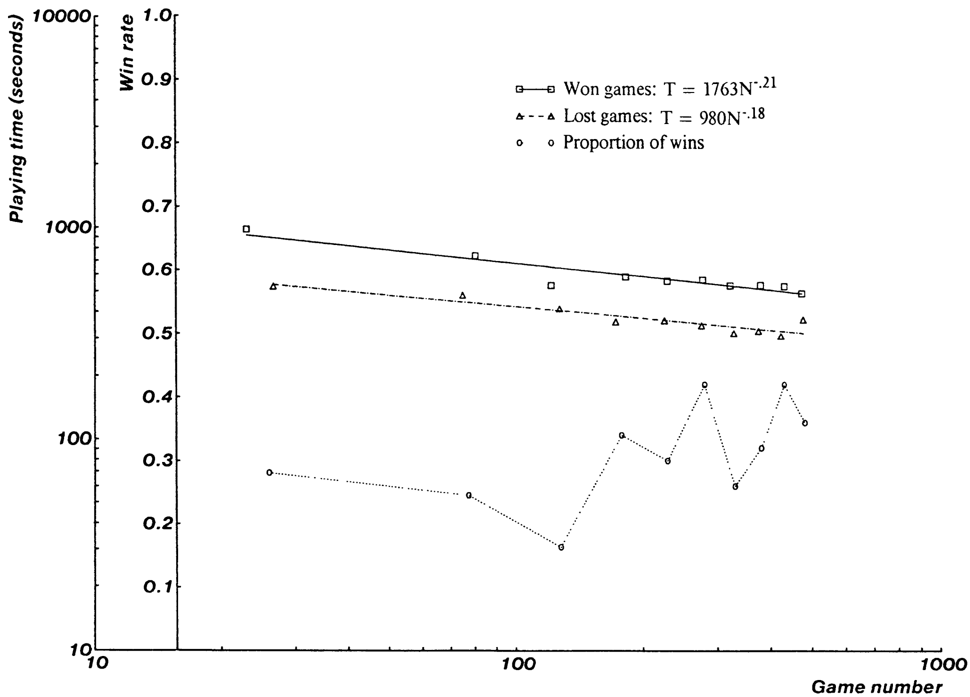

FIG. 1.10. Learning in the card game Stair (log-log coordinates).

rather it applied to all forms of mental behavior. To test whether the law applied to problem solving tasks, we had a single subject play 500 hands of a game of solitaire called *Stair*.

Stair involves laying out all 52 cards face up from a shuffled deck, in 8 *columns* (four with 7 rows, four with 6 rows). There are also four *spots* (initially empty), each of which can hold only a single card. The aim is to build four *stacks*, ace to king, one for each suit, by moving cards around under typical solitaire constraints. A card in a spot or at the bottom of a column may be moved: (1) to a spot, if it is empty; (2) to a stack, if the card is the next in order building up; or (3) to the bottom of another column, if the card is the next lower in the same suit (e.g., the six of spades appended to the seven of spades).

The game can be seen to be one of perfect information—all cards are faceup. The shuffled deck simply picks out one of the possible initial conditions at random. From that point no further chance element enters. Whether the game can be won or not, or how many cards can be moved to the stacks, is derivable from the initial configuration. The subject, whose ability to calculate ahead is of course limited, may create a partial plan and then proceed to execute it; in doing so, he may make irrevocable moves that lose him the possibility of winning. But such failures all arise, as in chess or checkers, because of his limited problem-solving ability. Although this task certainly has a strong perceptual component (and a weak motor component), it is to be classed as fundamentally an intellectual task, in the same way as games such as chess and checkers or problems such as the *traveling salesman problem*.

Turning to the figure, the top curve shows the time for games that the subject won; the lower curve shows the time for games that the subject lost; at the bottom the proportion of games won is shown. The points are averaged over 50 games. There is of course only one series of trials, since all games, won or lost, contribute to practice. Each group of 50 games is therefore split between the two curves before being averaged. Both curves essentially follow the log-log linear law. In general it takes longer to win than to lose, since losing involves becoming stuck after a relatively small number of cards has been played to the stack, whereas winning always involves working through all 52 cards (though the tail end goes rapidly).

The issue of the speed-accuracy tradeoff reveals itself in this data. Clearly, the subject is applying various criteria of certainty to his play. He could conceivably, as a strategy choice, study each initial layout for 5 hours before making his first move or play impulsively with no contemplation at all. In fact, the subject felt he had little genuine control of the speed-accuracy tradeoff, partly because the complexity of the initial position made it unclear whether an apparently lost game was just a bad layout or was due to a failure to spend enough time analyzing. Note that the most deviant point from the log-log line (at 150–200 trials) corresponds to the lowest win frequency.

Other Tasks and Measures

The story does not quite end at this point. Learning in other tasks and measured on other criteria seems to follow the log-log law. We give here a couple of examples.

Figure 1.11 is reproduced from Stevens and Savin (1962). It plots eight tasks with various response measures in log-log space. The criteria are all oriented to increase with practice. The plot is actually of the *cumulated* responses (i.e., the integral of the usual curve). This is just the same as the usual power law, because the integral of a power law is a power law (though integration tends to smooth the curve, helping to account for the lovely appearance of the curves, in addition to the relatively large numbers of subjects).

$$\int_1^N Bx^{-\alpha}\,dx = B(1-\alpha)^{-1}(N^{1-\alpha}-1) \tag{3}$$

Some of these curves are time curves (actually, amount accomplished per unit time, to make them positive curves); but several are not (e.g., 1 is the number of correct anticipations in learning nonsense syllables, 2 is the time on target in a pursuit tracking task; 3 is the number of balls thrown into a target area; 4 is the num-

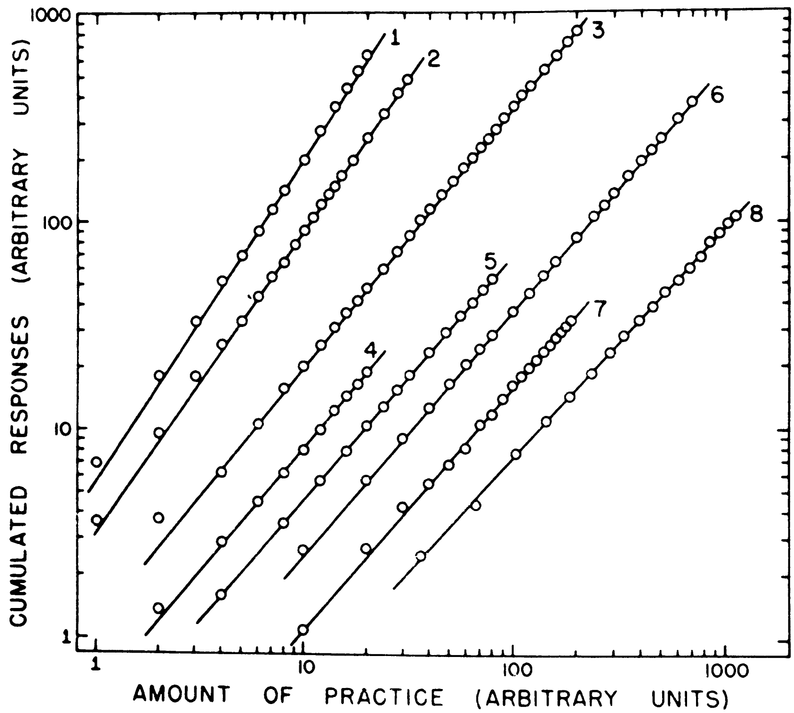

FIG. 1.11. Eight cumulated response practice curves (log-log coordinates). Figure from Stevens & Savin (1962). Copyright 1962 by the Society for the Experimental Analysis of Behavior, Inc.

ber of correct responses in an animal experiment in learning a maze, and so on.)

As a second type of example, it has long been known in industrial engineering that the so-called learning curve for production of manufactured projects was log-log linear. In part this comes of various simple rules of thumb (e.g., "... each time the quantity of [air]planes is doubled, the cumulative average man-hours per plane will be [reduced by] 80%" [Rigon, 1944]). However, Fig. 1.12 shows an empirical curve from machine tool manufacture (Hirsch, 1952). Notice that the index of performance is not time but cost.

Summary

We have shown some 12 diverse examples of the log-log linear law of practice for trials versus time. From Table 1.1 we can make one more particular point:

- The learning rates, α, are all less than 1.

Our main point is that the law is ubiquitous when one measures the log of performance time against the log of trial number. Where the general impression

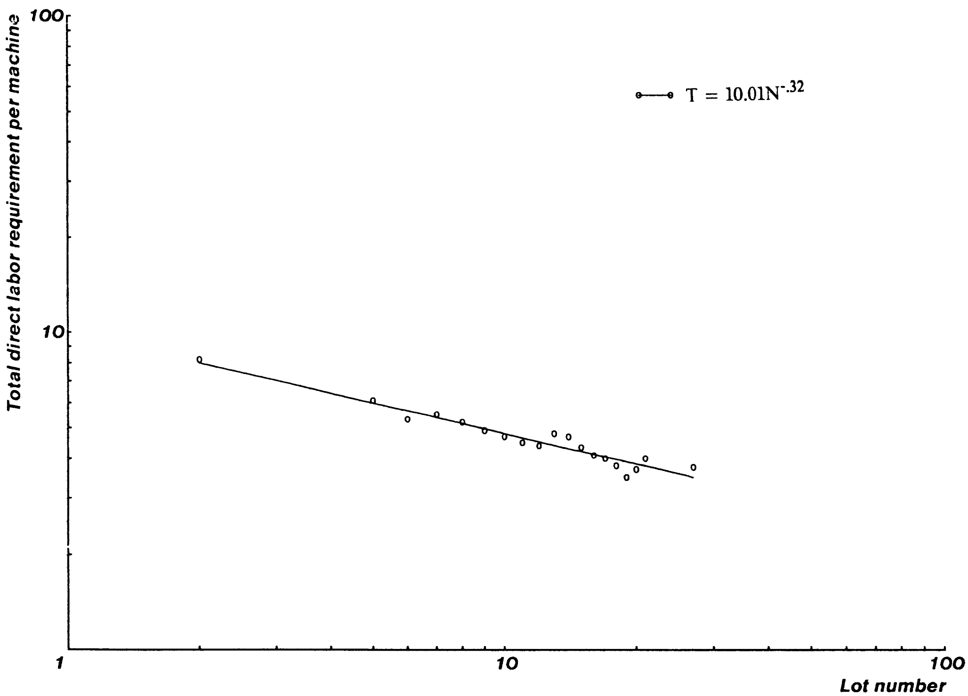

FIG. 1.12. The effect of practice on direct labor requirement in machine production (log-log coordinates). Replotted from Hirsch (1952).

seems to have been that the law showed up in perceptual-motor behavior, we think it is clear that it shows up everywhere in psychological behavior—at least it cannot easily be restricted to some part of the human operation.

Our proposition on ubiquity is extended, perhaps beyond our druthers, to learning curves involving other measures of performance and even to tasks possibly (but not certainly) beyond the pale of individual human behavior. We do not however claim that all learning is log-log linear. Nor do we claim that practice always leads to learning.

We do not wish to assert that such an effect stems from a single cause or mechanism. Indeed, its ubiquity might seem to indicate multiple explanations. We do wish to make one general comment about the regularity and what might be expected from understanding it. Its widespread occurrence implies that it depends on quite general features of the learning situation or of the system that learns. If we develop a theory that depends on detailed perceptual or motor mechanisms, we shall just create trouble for the more cognitive instances or vice versa.

One is immediately reminded of other examples of ubiquitous regularities and their explanation. The *normal distribution,* which arises out of the independent additive combination of many small increments, is the most well known.

Another, usually known as *Zipf's law,* gives the distribution for items according to their rank order, which is common to word frequencies, city sizes, incomes, and many other ordered phenomena (Simon, 1955). Consistently, highly general stochastic models underly these various phenomena. They explain the regularity but leave open the detailed mechanisms that produce the stochastic processes.

Thus, in searching for an explanation for this regularity, we should expect at best to find some such general considerations. Though it will not tell us in detail about the learning mechanism, it may still tell us something worth having.

BASICS ABOUT POWER LAWS

In this section we present some general perspectives on power laws and what they mean.

Differential Forms and Rates of Change

We start with the power law and its equivalent log–log form:

$$T = BN^{-\alpha} \tag{4}$$

$$\log(T) = \log(B) - \alpha\log(N) \tag{5}$$

It is instructive to see this in terms of the local rate of learning, dT/dN.[5]

$$\frac{dT}{dN} = -\alpha BN^{-\alpha-1} \tag{6}$$

$$= -\frac{\alpha T}{N} = -\left(\frac{\alpha}{N}\right)T \tag{7}$$

$$= -\alpha B^{-1/\alpha}T^{1+1/\alpha} \tag{8}$$

Now, one baseline form for learning is exponential. It can arise, for instance, from any mechanism that is completely local. If there is something that learns on each local part of a performance, independent of any other part, then the change in T (the sum of the changes to each part of T) is proportional to T:

$$\frac{dT}{dN} = -\alpha T \tag{9}$$

$$T = Be^{-\alpha N} \tag{10}$$

Comparing this differential form to that of the power law, shows that power-law learning is like exponential learning in which the instantaneous rate α' decreases with N, that is,

$$dT/dN = -\alpha' T \tag{11}$$

where $\alpha' = \alpha/N$

[5]For ease of exposition we treat the trial number N as a continuous variable. In fact, nothing material depends on it; we could work with finite differences throughout, at the cost of added complexity.

Both the exponential and the power function are monotonically decreasing functions that asymptote at 0. The decreasing rate of learning in the power function leads to its approaching asymptote much more slowly. Figure 1.13 shows these two curves in linear coordinates, with identical initial values (B = 1). This corresponds to $N = 0$ for the exponential, and $N = 1$ for the power. Thus, one way to think of power law learning is that it is a learning process in which some mechanism is slowing down the rate of learning.

Not every scheme of slowed-down learning leads to the power law. For instance, if we generalize the differential equation above we obtain a different law:

$$\frac{dT}{dN} = \left(\frac{\alpha}{N^{\beta}}\right) T, \tag{12}$$

where $\beta \neq 1$.

$$T = Be^{-\alpha N^{1-\beta}} \tag{13}$$

A representative curve for β less than 1 is also shown in Fig. 1.13, which produces asymptoting between the exponential and the power law.

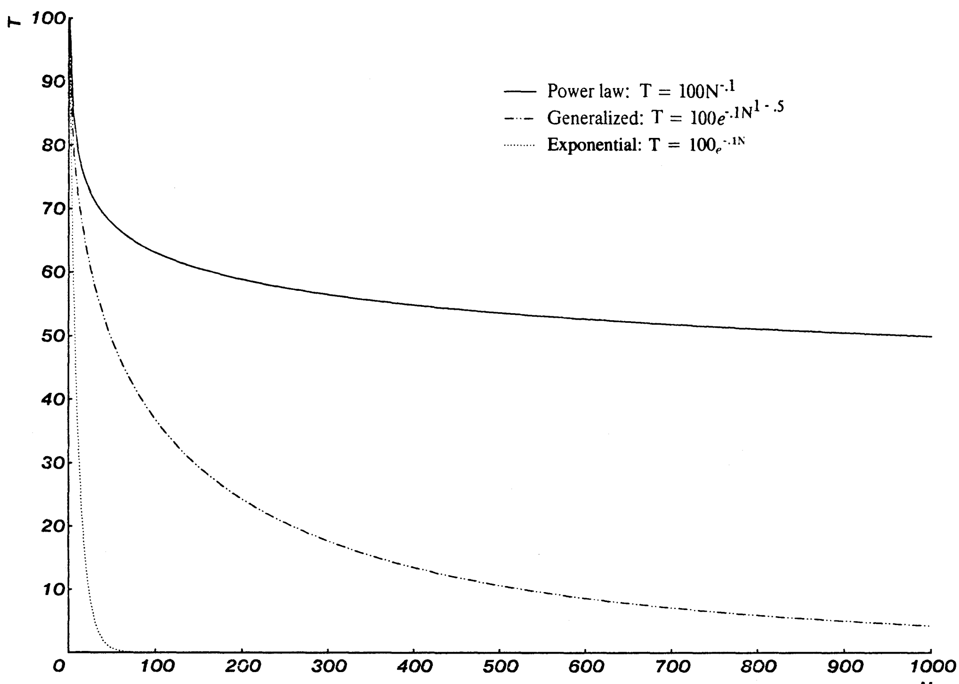

FIG. 1.13. Basic learning curves: power law, exponential, and a generalized curve.

The form of the power law can be appreciated in terms of a simple global rule, as well as in differential form:

Power Law Decay: If T decreases by a factor δ in the first N trials, it will take another $N(N - 1)$ trials to decrease by a factor of δ again.

Comparison with the corresponding global rule for the exponential, shows again how much more slowly the power law drops off:

Exponential Law Decay: If T decreases by a factor of δ in the first N trials, it will take another N trials to decrease by a factor of δ again.

Asymptotes and Prior Experience

As given in Equation 4, the law assumes: (1) the asymptote of the learning is 0 (i.e., the task can be performed in arbitrarily small time after enough learning); and (2) the initial trial of the learning occurs at the first trial of the measured series. Neither of these assumptions need be true.

The more general form of the law is

$$T = A + B(N + E)^{-\alpha} \tag{14}$$

A ($\geqslant 0$) is the *asymptote* of learning as N increases indefinitely. E ($\geqslant 0$) is the number of trials of learning that occurred prior to the first trial as measured (i.e., prior *experience*); it thus identifies the true *starting point* of learning. (Neither $A < 0$ or $E < 0$ make immediate sense, given these interpretations; $A = 0$, $E = 0$ reproduces the basic form of Equation 4.)

Plotting $\log(T - A)$ against $\log(N + E)$ still yields a straight line whose slope is $-\alpha$. The difficulty of course is that A and E are not known in advance, so the curve cannot be plotted as an initial exploratory step in an investigation.

One alternative is just to plot in $\log(T) - \log(N)$ space and understand the deviations:

$$\log(T - A) = \log(B) - \alpha\log(N + E) \tag{15}$$

$$\log(T) = \log(B) - \log(1 - A/T) - \alpha\log(N) - \alpha\log(1 + E/N) \tag{16}$$

There is an error term for each parameter. If T is large with respect to the asymptote, A, then $\log(1 - A/T)$ is close to $\log(1)$, which is 0. This occurs at early values of N. If N is large with respect to E, then $\log(1 + E/N)$ is close to $\log(1)$, which is 0. Thus, the two deviations affect the curve at opposite parts: Non-zero values of E distort the straight line for low N, non-zero values of A distort it for high N.

Figure 1.14 shows a power law with a starting point ($-E$) of -25 and a time asymptote (A) of 5. Figure 1.15 shows the same curve in log–log space. Characteristically, the starting point pulls the initial segment of the curve down toward the horizontal and the finite asymptote pulls the high N tail of the curve up toward the horizontal. A central region of the curve appears as a straight line. It is however less than the true slope ($-\alpha$), as the line shows.

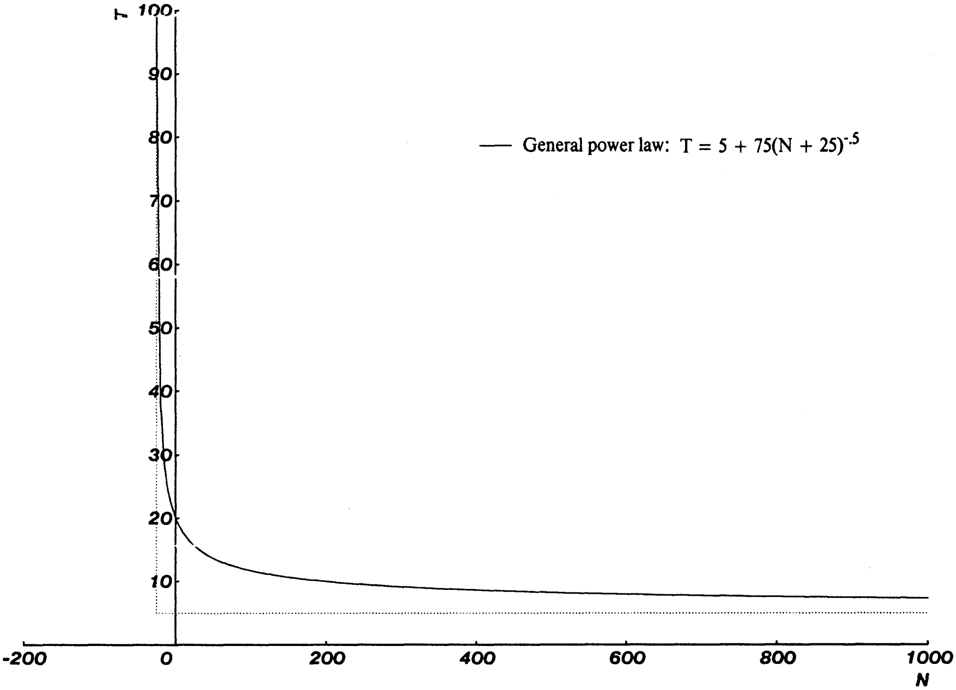

FIG. 1.14. A general power law curve.

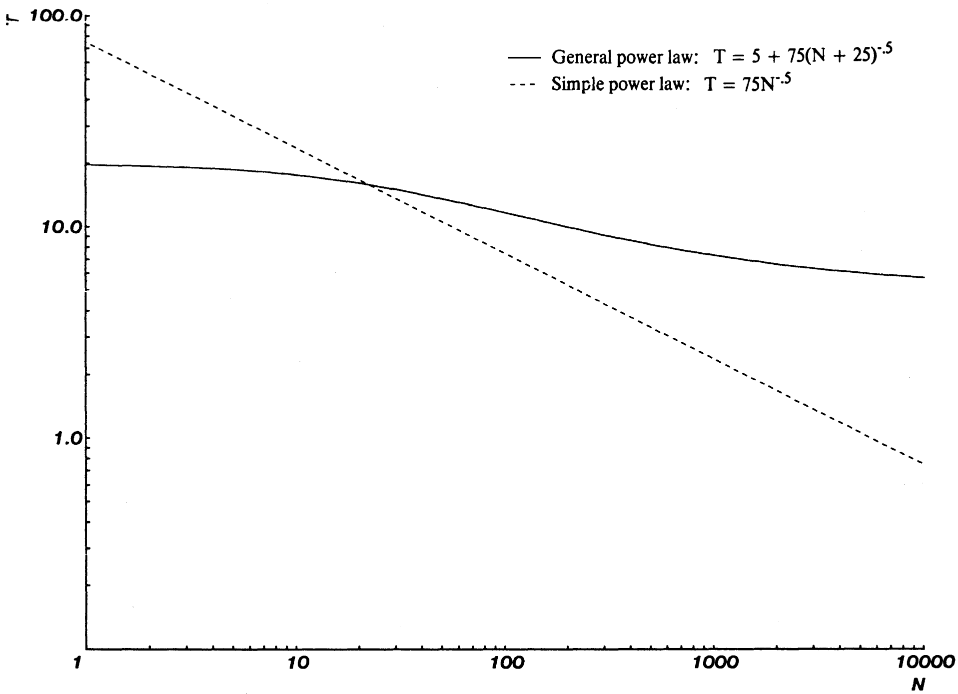

FIG. 1.15. A general power law in log–log coordinates. The simple power law with the same α and B is also shown.

The derivative of the general power function in log–log space is given by

$$\frac{d[\log(T)]}{d[\log(N)]} = -\alpha\left(1 - \frac{A}{T}\right)/\left(1 + \frac{E}{N}\right) \tag{17}$$

It can be seen that the slope is everywhere smaller than α and becomes increasingly so as either A or E increases. A reasonable estimate of the apparent slope as viewed on the graph, α^*, is at the inflection point. It is easy to obtain by setting the derivative of Equation 17 to zero:

$$\frac{d}{dN}\left[\frac{d[\log(T)]}{d[\log(N)]}\right] = -\left(\frac{\alpha}{N}\right)\left(\frac{E}{N} - \frac{\alpha A}{T}\right)\left(1 - \frac{A}{T}\right)\left(1 + \frac{E}{N}\right)^{-2} = 0 \tag{18}$$

$$\alpha^* = \frac{(\alpha N^* - E)}{(N^* + E)} \tag{19}$$

N^* is the point at which the inflection occurs. The exact value of N^* is not expressible in simple terms, but a reasonable approximation is

$$N^* = \left[\frac{BE}{\alpha A}\right]^{1/(1+\alpha)} \tag{20}$$

where $E/N^* << \alpha < 1$.

The structure of Fig. 1.15 suggests that many of the deviations in the empirical curves could be due simply to starting point or asymptote effects. Because the effect of these two phenomena is to bend toward the horizontal at separate ends, it is possible to tell from the curve in log–log space what effect might be operating. The original Snoddy data in Fig. 1.1 provides an example of a clear initial deviation. It cannot possibly be due to an earlier starting point, because the initial curve rises toward the vertical. However, it could be due to the asymptote, because raising the asymptote parameter (A) will pull the right-hand part of the curve down and make its slope steeper. The Seibel data in Fig. 1.6 provides an example where there are deviations from linearity at both ends. Use of a nonzero value for E (previous experience) will steepen the initial portion of the curve, whereas doing likewise for A will steepen the high N portion of the curve. (The results of such a manipulation are seen in Fig. 1.21.)

Trials or Time?

The form of the law of practice is performance time (T) as a function of trials (N). But trials is simply a way of marking the temporal continuum (t) into intervals, each one performance-time long. Since the performance time is itself a monotone decreasing function of trial number, trials (N) becomes a nonlinear compression of time (t). It is important to understand the effect on the law of practice of viewing it in terms of time or in terms of trials.

The fundamental relationship between time and trials is

$$t(N) = T_0 + \sum_{i=1}^{N} T_i = T_0 + \sum_{i=1}^{N} Bi^{-\alpha} = T_0 + B \sum_{i=1}^{N} i^{-\alpha} \tag{21}$$

T_0 is the time from the arbitrary time origin to the start of the first trial. This equation cannot be inverted explicitly to obtain an expression for $N(t)$ that would permit the basic law (Equation 4) to be transformed to yield $T(t)$. Instead, we proceed indirectly by means of the differential forms. From Equation 21 we obtain

$$\frac{dt}{dN} = T \tag{22}$$

Think of the corresponding integral formulation,

$$\frac{d}{dz} \int_a^z f(x)\, dx = f(z)$$

Now, starting with the power law in terms of trials we find:

$$\frac{dT}{dt} = \frac{dT/dN}{dt/dN} = \frac{-\alpha T/N}{T} = \frac{-\alpha}{N} \tag{23}$$

But from the basic Equation (4):

$$N = \left(\frac{T}{B}\right)^{-1/\alpha} \tag{24}$$

Thus, we obtain the trials power law reexpressed in terms of time:

$$\frac{dT}{dt} = -\alpha B^{-1/\alpha} T^{1/\alpha} \tag{25}$$

For $\alpha \neq 1$ this integrates to yield

$$T^{-(1-\alpha)/\alpha} = (1 - \alpha)B^{-1/\alpha}t + C \qquad \text{for } \alpha \neq 1 \tag{26}$$

But C is an arbitrary constant of integration and if the origin and scale of t is adjusted appropriately, we find:

$$T = B' t^{-\alpha/(1-\alpha)} \qquad \text{for } \alpha \neq 1 \tag{27}$$

Thus, a power law in terms of trials is a power law in terms of time, though with a different exponent, reflecting the expansion of time over trials. The results are significantly altered when $\alpha = 1$ (the hyperbolic) however. Equation 25 becomes

$$\frac{dT}{dt} = -B^{-1}T \tag{28}$$

This is no longer the differential form of a power law. Instead it is that of an exponential:

$$T = Ce^{-B^{-1}t} \tag{29}$$

It is left as an exercise for the reader to confirm that an exponential function in trials transforms to a *linear* function in time (hence, Zeno-like, an infinite set of trials can be accomplished in a finite amount of time).

FITTING THE DATA TO A FAMILY OF CURVES

Given empirical curves, such as occur in abundance in the second section, it is important to understand how well they are described by curves of a given family (e.g., power laws) and whether there are alternative general forms that fit them just as well (as noted in the introduction, exponential, hyperbolic, and logistic curves have enjoyed much more favor than power functions). Curve fitting without benefit of a model is notoriously a black art. Nonetheless, we have deliberately chosen not to be model driven initially, because we want to have empirical generalizations as the starting point in the search for theory, not just the raw data.

The basic issue of curve fitting can be introduced from Seibel's own treatment of his data (Fig. 1.6), which appears to be an extremely good fit to the log–log law over an extensive range (40,000 trials). Seibel (1963) fit his points to three curves by least squares: (1) a power law with asymptote only (i.e., E fixed at 0); (2) an exponential with asymptote; and (3) a general power law with both asymptote and starting point.[6] He obtained an r^2 of .991 for the power function with asymptote only. But he also obtained an r^2 of .971 for the exponential with asymptote. His general power law fit was .997. (His parameters for asymptotes and starting points are mostly reasonable but not entirely.) Thus, all the curves give good fits by normal standards. If only differences in the least-squared residual are used, there can hardly be much to choose from. This is an annoying result, in any case; but it is also somewhat unexpected, for the plots that we have shown, though they surely contain noise, are still impressively linear by intuitive standards and involve lots of data.

It is important to recognize that two basic kinds of failure occur in fitting data to a family of smooth curves: (1) failure of the shape of the data curve to fit to the shapes available within the family; and (2) noise in the data, which will not be fit by any of the families under consideration or even noticeably changed by parametric variation within a family. These distinctions are precisely analogous to the frequency spectrum of the noise in the data. However, the analogy probably should not be exploited too literally, because an attempt to filter out the high-frequency noise prior to data fitting simply adds another family of empirical curves (the filters) to confound the issues. What does seem sensible is to attempt to distinguish fits of shape without worrying too much about the jitter.

A simple example of this point of view is the (sensible) rejection of the family of logistic curves from consideration for our data. The logistic provides a sig-

---

[6]The exponential is translation invariant, so a special starting point is not distinguishable for it; that is, $Be^{N+E} = (Be^{E})e^{N} = B'e^{N}$.

moid curve (i.e., a slow but accelerating start with a point of inflection and then asymptoting). No trace of an S-shape appears in any of our data, though it would not be lost to view by any of the various monotone transformations (logs, powers, and exponentials) that we are considering. Hence, independent of how competing the measure of error, the logistic is not to be considered.

The size of the jitter (i.e., the high-frequency noise) will limit the precision of the shape that can be detected and the confidence of the statements that can be made about it. It provides a band through which smooth curves can be threaded, and if that band is wide enough—and it may not have to be very wide—then it may be possible to get suitable members of conceptually distinct curves through it. In all cases, the main contribution to any error measure will be provided by the jitter, so that only relatively small differences will distinguish the different families.

The Data Analysis Procedure

With the elimination of the logistic from consideration, we have focused our efforts on three families of curves: *exponential, hyperbolic,* and *power law*. The analysis procedure that we have ended up using is primarily graphical in nature. We look at what types of deviations remain, once an empirical curve has been fit optimally by a family of theoretical curves. The analysis consists of judgments as to whether the deviations represent actual distortions of shape, or merely jitter. The procedure has the following components:

1. Find spaces where the family of curves should plot as straight lines. Judgments of shape deviation are most easily made and described when the norm is a line. These are the *transformation spaces* of the given family. There may be more than one such space.
2. For each family of curves, find the best linear approximation to the data in the transformation spaces of the family. This will generally involve a combination of search and linear regression.
3. Accept a curve for a family, if the best fit plots as a straight line in the space of that family. Reject it, if it has significant shape distortion.
4. Understand the shape distortion of family X when plotted in the space of family Y. Expect curves of family X to show the characteristic distortion when plotted in the spaces of alternative families.
5. Compute an estimate of fit (r^2) for the best approximation in each transformation space. Expect these values to support the judgments made on the basis of shape distortion.

These criteria contain elements both of acceptance and rejection and provide a mixture of absolute judgments about whether data belong to a given family and relative judgments about the discrimination between families. The parameters for the best fits as well as the estimates of fit (r^2) can be found in Table 1.2.

TABLE 1.2
The General Learning Curves: Parameters from Optimal Fits in the *Log* Transformation Spaces

	Exponential $T = A + Be^{-\alpha N}$				Hyperbolic $T = A + B/(N + E)$				Power Law $T = A + B(N + E)^{-\alpha}$				
Data Set	A	B	α	r^2	A	B	E	r^2	A	B	E	α	r^2
Snoddy (1926)	27.01	38.80	.061	.916	24.49	243.6	1.3	.962	21.74	119.2	0.0	.71	.975
Crossman (1959)	7.19	4.59	1×10^{-7}	.842	7.10	2.4×10^{6}	151000	.983	6.91	20481	31000	.66	.990
Kolers (1975) - Subject HA	1.36	3.82	.018	.849	1.10	94.02	9.8	.915	.18	15.25	0.0	.46	.931
Neisser et al. (1963)													
Ten targets	.06	.83	.13	.905	.00	2.74	.9	.965	.00	2.35	.6	.95	.965
One target	.06	.44	.094	.938	.00	3.16	4.6	.951	.00	2.57	3.9	.94	.951
Card, English & Burr (1978)													
Stepping keys - Subj. 14	2.35	1.99	.011	.335	2.14	171.4	75.2	.338	.02	6.36	9.3	.14	.340
Mouse - Subj. 14	1.46	1.28	.028	.452	1.46	16.70	5.0	.603	.59	4.28	0.0	.33	.729
Seibel (1963) - Subject JK	.371	.461	.000055	.956	.328	3888.1	3042	.993	.324	2439.9	2690	.95	.993
Anderson (1980) - Fan 1	.487	.283	.00055	.774	.466	231.6	319.7	.902	.353	4.322	0.0	.39	.947
Moran (1980)													
Total time	13.80	6.66	.00073	.546	14.77	3335.9	474.6	.637	.03	30.24	0.0	.08	.839
Method time	11.61	3.11	.0010	.652	11.75	1381.8	360.0	.737	.26	19.35	0.0	.06	.882
Neves & Anderson (in press)													
Total time - Subject D	57.5	240.2	.019	.660	45.6	5000.2	7.3	.728	0.0	991.2	0.0	.51	.780
The Game of Stair													
Won games	476	319	.0052	.689	449	29800	40.1	.783	120	1763	0.0	.25	.849
Lost games	152	326	.0016	.634	247	41270	124.1	.751	1	1009	2.5	.19	.841
Hirsch (1952)	2.76	4.35	.070	.819	2.34	37.05	4.9	.897	.00	10.01	0.0	.32	.932
General Power Law													
$T = 5 + 75(N + 25)^{-0.5}$	7.21	6.78	.0037	.983	6.41	1069.6	91.2	.997	5.00	74.85	24.9	.50	1.000
40 Term Additive Mixture	1.60	45.37	.0065	.904	.58	1231.2	10.2	.997	.19	753.1	7.2	.89	.998
Chunking Model													
Combinatorial TE	4.61	4.71	.0046	.957	4.35	365.7	55.3	.992	2.86	17.40	6.6	.33	1.000

The remainder of this section shows how we applied this data analysis procedure. We start by looking at the transformation spaces. This is followed by an examination of the distortions that occur when a theoretical curve is plotted in a space belonging to a different family. We are then in a position to analyze a couple of the empirical curves that appeared in the second section.

The Transformation Spaces

The curves that we are interested in belong to multiparameter families (3 for the exponential and hyperbolic; 4 for the power law). Regression can be used to fit a line to an empirical curve plotted in a multidimensional space. Unfortunately, for the three families that we are interested in, there is no space in which all the parameters (three or four) can be determined by linear regression. The most that we can obtain is two parameters. The remainder must be determined by some other means, such as search. The choice of which parameters are to drop out of the analysis determines the transformation space. We have primarily worked in two different types of transformation spaces. The first type consists of the *log* spaces. These are the most commonly used linearizing spaces for functions with powers. The log transformations that we use are the following:

Exponential: $T' = \log(B) - \alpha N$ for $T' = \log(T - A)$ (30)
Hyperbolic: $T' = \log(B) - N'$ for $T' = \log(T - A)$ and $N' = \log(N + E)$ (31)
Power Law: $T' = \log(B) - \alpha N'$ for $T' = \log(T - A)$ and $N' = \log(N + E)$ (32)

The log spaces for the hyperbolic and the power law turn out to be the standard log–log space, whereas the exponential is in semilog space. Determining fits in these spaces requires a combination of search (over $0 \leqslant A \leqslant T_{\min}$ and $0 \leqslant E$) and regression (for B and α). Because the exponential and hyperbolic families are each missing one of these parameters, the process becomes simpler for them. The exponential only requires a one-dimensional search (over $0 \leqslant A \leqslant T_{\min}$), whereas the hyperbolic can replace the regression (for B and α) with the computation of the average for B.

The log spaces have been used exclusively for the data analyses that are described in the following section (Table 1.2 was computed in the log spaces). It is important to realize though that they are not the only transformation spaces that can be used. We have explored what we call the T–X spaces, though space precludes presenting the analysis. Transforming a curve into its T–X space involves pushing all the nonlinearities into the definition of X as follows:

Exponential: $T = A + BX$ for $X = e^{-\alpha N}$ (33)

Hyperbolic: $T = A + BX$ for $X = \dfrac{1}{(N + E)}$ (34)

Power Law: $T + A + BX$ for $X = (N + E)^{-\alpha}$ (35)

In the T–X spaces, searches are over $\alpha \geqslant 0$ and $E \geqslant 0$, with A and B determined by regression. Only single-dimensional searches are needed for the two three-parameter families. The T-X spaces prove especially useful for estimating the asymptote (A), because it maps into the intercept of the transformed curve.

The Theoretical Curves

When a curve is optimally fit in a space corresponding to its family, it plots as a straight line (by definition). This is not true though when the curve is fit in a space corresponding to some other family. There will be distortions that show up as nonlinearities in the plot. By understanding these characteristic shape distortions, we are able to interpret the deviations that we find when we plot the data in these spaces. This will help us to distinguish between random jitter and distortions that signal a bad fit by the family of curves. Data that plot with the same deviations as one of the theoretical curves have a good chance of belonging to that curve's family.

Figure 1.16 shows the best that a power law can be fit in exponential log space. The power law curve is

$$T = 5 + 75(N + 25)^{-0.5} \tag{36}$$

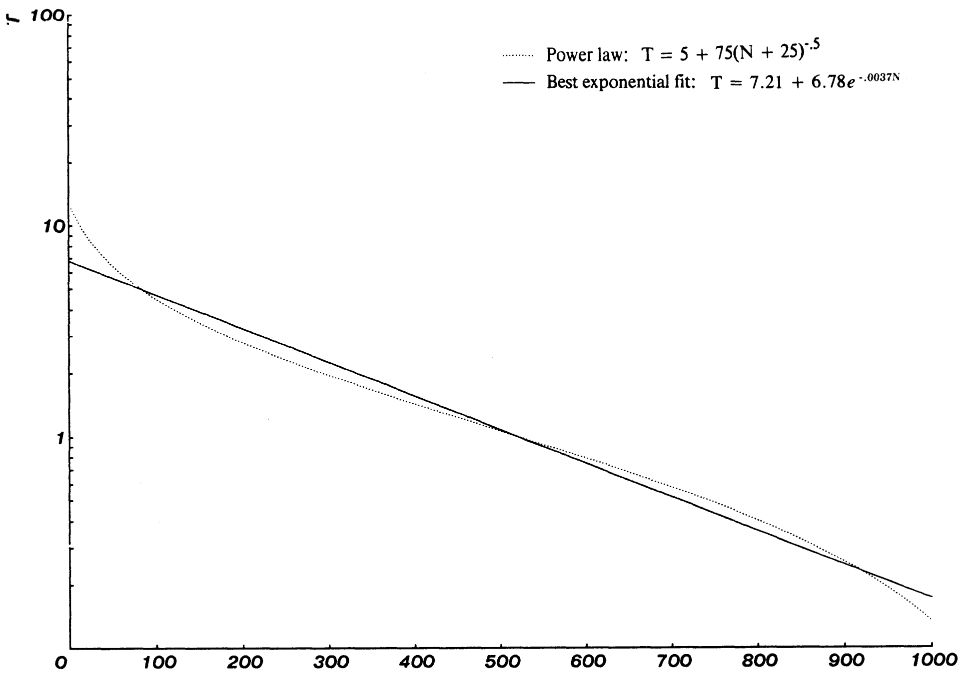

FIG. 1.16. Optimal fit of a power law in the exponential transformation space (semi-log coordinates).

This is the same curve that is plotted in Figs. 1.14 and 1.15. The parameters for the optimal exponential fit can be found in Table 1.2. The r^2 value of .983 is deceptively high, as an examination of Fig. 1.16 shows. There are strong deviations in all portions of the curve. The curve starts out high, goes low, then high again, and finally tails off downward. If we see deviations of this type when a set of data has been optimally fit by an exponential, we can conclude that the exponential family is not a good model for the data and that the power law might be.

Figure 1.17 shows the same curve optimally fit in hyperbolic log space. We see the same sorts of deviations that were found in the exponential case, but they are much attenuated. It will be hard to rule out the hyperbolic family in such a case because the variability of the data is likely to swamp out much of the distortion. At most we can hope to see the slight upturn at low N and the slight downturn for high N.

It is not necessary to look at the theoretical plots for the hyperbolic, as it is a special case of the power law. It will plot with no distortion in the power law log space, and it will have the same type of distortion in the exponential log space as did the power law. This leaves only exponential curves to be examined. We cannot present a plot of the optimal fit of an exponential in the power law log space. All attempts to find such optimal fits have led to at least one of the

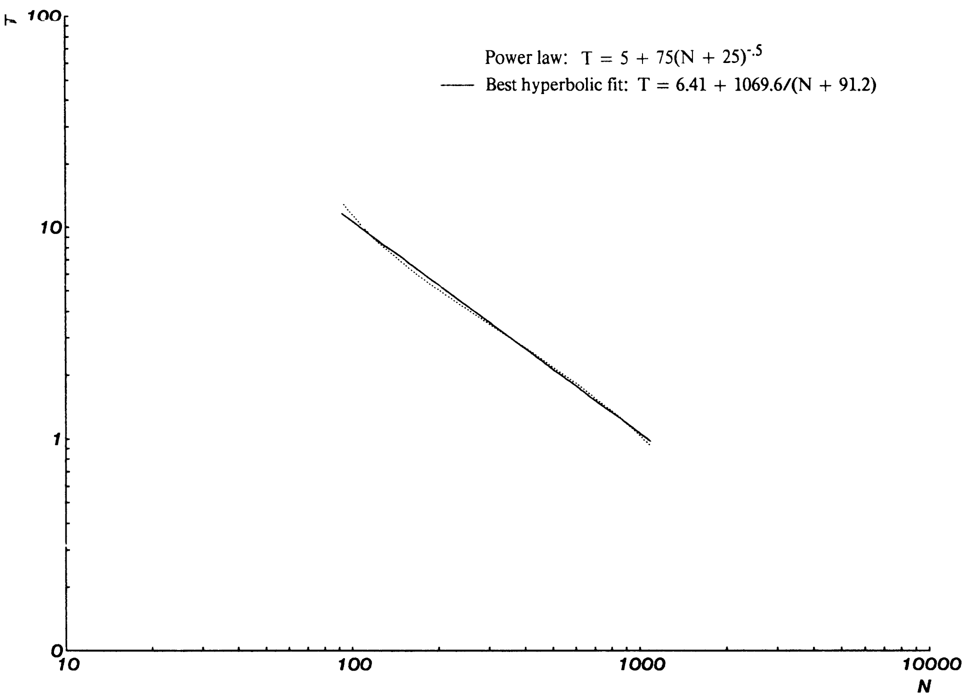

FIG. 1.17. Optimal fit of a power law in the hyperbolic transformation space (log-log coordinates).