Psychology of Learning

Volume 3

SAGE LIBRARY *of*
COGNITIVE AND EXPERIMENTAL PSYCHOLOGY

SAGE LIBRARY IN COGNITIVE AND EXPERIMENTAL PSYCHOLOGY

PSYCHOLOGY OF LEARNING

VOLUME III

EDITED BY
DAVID SHANKS

SAGE

Los Angeles • London • New Delhi • Singapore • Washington DC

First published 2009

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London ECIY ISP

SAGE Publication Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1 Mohan Cooperative Industrial Estate
Mathura Road, Post Bag 7
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
33 Pekin Street #02-01
Far East Square
Singapore 048763

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-1-84787-268-5 (set of five volumes)

Library of Congress Control Number: 2008924785

Typeset by Star Compugraphics Private Limited, Delhi
Printed on paper from sustainable resources
Printed in Great Britain by the MPG Books Group, Bodmin and King's Lynn

Contents

VOLUME III

Part 2:
Learning and Cognition (*Continued*)

~ 27 ~

Toward an Instance Theory of Automatization

Gordon D. Logan

Automaticity is an important phenomenon in everyday mental life. Most of us recognize that we perform routine activities quickly and effortlessly, with little thought and conscious awareness – in short, automatically (James, 1890). As a result, we often perform those activities on "automatic pilot" and turn our minds to other things. For example, we can drive to dinner while conversing in depth with a visiting scholar, or we can make coffee while planning dessert. However, these benefits may be offset by costs. The automatic pilot can lead us astray, causing errors and sometimes catastrophes (Reason & Myceilska, 1982). If the conversation is deep enough, we may find ourselves and the scholar arriving at the office rather than the restaurant, or we may discover that we aren't sure whether we put two or three scoops of coffee into the pot.

Automaticity is also an important phenomenon in skill acquisition (e.g., Bryan & Harter, 1899). Skills are thought to consist largely of collections of automatic processes and procedures (e.g., Chase & Simon, 1973; Logan, 1985b). For example, skilled typewriting involves automatic recognition of words, translation of words into keystrokes, and execution of keystrokes (Salthouse, 1986). Moreover, the rate of automatization is thought to place important limits on the rate of skill acquisition: LaBerge and Samuels (1974) claimed that beginning readers may not be able to learn to read for meaning until they have learned to identify words and letters automatically.

Over the last decade, considerable progress has been made in understanding the nature of automaticity and the conditions under which it may be acquired (for reviews, see Kahneman & Treisman, 1984; LaBerge, 1981; Logan, 1985b; Schneider, Dumais, & Shiffrin, 1984). There is evidence

Source: *Psychological Review,* 95(4) (1988): 492–527.

that automatic processing differs qualitatively from nonautomatic processing in several respects: Automatic processing is fast (Neely, 1977; Posner & Snyder, 1975), effortless (Logan, 1978, 1979; Schneider & Shiffrin, 1977), autonomous (Logan, 1980; Posner & Snyder, 1975; Shiffrin & Schneider, 1977; Zbrodoff & Logan, 1986), stereotypic (McLeod, McLaughlin, & Nimmo-Smith, 1985; Naveh-Benjamin & Jonides, 1984), and unavailable to conscious awareness (Carr, McCauley, Sperber, & Parmalee, 1982; Marcel, 1983). There is also evidence that automaticity is acquired only in consistent task environments, as when stimuli are mapped consistently onto the same responses throughout practice. Most of the properties of automaticity develop through practice in such environments (Logan, 1978, 1979; Schneider & Fisk, 1982; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

Automaticity is commonly viewed as a special topic in the study of attention. The modal view links automaticity with a single-capacity model of attention, such as Kahneman's (1973). It considers automatic processing to occur without attention (e.g., Hasher & Zacks, 1979; Logan, 1979,1980; Posner & Snyder, 1975; Shiffrin & Schneider, 1977), and it interprets the acquisition of automaticity as the gradual withdrawal of attention (e.g., LaBerge & Samuels, 1974; Logan, 1978; Shiffrin & Schneider, 1977). The modal view has considerable power, accounting for most of the properties of automaticity: Automatic processing is fast and effortless because it is not subject to attentional limitations. It is autonomous, obligatory, or uncontrollable because attentional control is exerted by allocating capacity; a process that does not require capacity cannot be controlled by allocating capacity. Finally, it is unavailable to consciousness because attention is the mechanism of consciousness and only those things that are attended are available to consciousness (e.g.,Posner & Snyder, 1975).

However, there are serious problems with the modal view. Some investigators questioned the evidence that automatic processing is free of attentional limitations (e.g., Cheng, 1985; Ryan, 1983). Others found evidence of attentional limitations in tasks that are thought to be performed automatically (e.g., Hoffman, Nelson, & Houck, 1983; Kahneman & Chajzyck, 1983; Paap & Ogden, 1981; Regan, 1981). The single-capacity view of attention, in which the modal view is articulated, has been seriously challenged by multiple-resource theories, which argue that many resources other than attention may limit performance (e.g., Navon & Gopher, 1979; Wickens, 1984).[1] Others argued that performance may not be limited by any resources, attentional or otherwise (e.g., Allport, 1980; Navon, 1984; Neisser, 1976). Moreover, there is growing dissatisfaction with the idea that automatization reflects the gradual withdrawal of attention (e.g., Hirst, Spelke, Reaves, Caharack, & Neisser, 1980; Kolers, 1975; Spelke, Hirst, & Neisser, 1976). Critics argue that the idea is empty unless the learning mechanism can be specified.

The purpose of this article is to propose a theory of automaticity that describes the nature of automatic processing and says how it may be acquired without invoking the single-capacity theory of attention or the idea of resource limitations. The theory is first described generally, then a specific version of the theory is developed to account for the speed-up and reduction in variability that accompany automatization. The theory is then fitted to data from two different tasks – lexical decision and alphabet arithmetic – and experiments that test the learning assumptions of the theory are reported. Finally, the qualitative properties of automaticity are discussed in detail, implications of the theory are developed and discussed, and the theory is contrasted with existing theories of skill acquisition and automatization.

Automaticity as Memory Retrieval

The theory relates automaticity to memorial aspects of attention rather than resource limitations. It construes automaticity as a memory phenomenon, governed by the theoretical and empirical principles that govern memory. Automaticity is memory retrieval: Performance is automatic when it is based on single-step direct-access retrieval of past solutions from memory. The theory assumes that novices begin with a general algorithm that is sufficient to perform the task. As they gain experience, they learn specific solutions to specific problems, which they retrieve when they encounter the same problems again. Then, they can respond with the solution retrieved from memory or the one computed by the algorithm. At some point, they may gain enough experience to respond with a solution from memory on every trial and abandon the algorithm entirely. At that point, their performance is automatic.[2] Automatization reflects a transition from algorithm-based performance to memory-based performance.

The idea behind the theory is well illustrated in children's acquisition of simple arithmetic. Initially, children learn to add single-digit numbers by counting (i.e., incrementing a counter by one for each unit of each addend), a slow and laborious process, but one that guarantees correct answers, if applied properly. With experience, however, children learn by rote the sums of all pairs of single digits, and rely on memory retrieval rather than counting (Ashcraft, 1982; Siegler, 1987; Zbrodoff, 1979). Once memory becomes sufficiently reliable, they rely on memory entirely, reformulating more complex problems so that they can be solved by memory retrieval.

Main Assumptions

The theory makes three main assumptions: First, it assumes that encoding into memory is an obligatory, unavoidable consequence of attention. Attending to a stimulus is sufficient to commit it to memory. It may be remembered well or poorly, depending on the conditions of attention, but it will be encoded. Second, the theory assumes that retrieval from memory is an obligatory, unavoidable consequence of attention. Attending to a stimulus is sufficient to retrieve from memory whatever has been associated with it in the past. Retrieval may not always be successful, but it occurs nevertheless. Encoding and retrieval are linked through attention; the same act of attention that causes encoding also causes retrieval. Third, the theory assumes that each encounter with a stimulus is encoded, stored, and retrieved separately. This makes the theory an *instance* theory and relates it to existing theories of episodic memory (Hintzman, 1976; Jacoby & Brooks, 1984), semantic memory (Landauer, 1975), categorization (Jacoby & Brooks, 1984; Medin & Schaffer, 1978), judgment (Kahneman & Miller, 1986), and problem solving (Ross, 1984).

These assumptions imply a learning mechanism – the accumulation of separate episodic traces with experience – that produces a gradual transition from algorithmic processing to memory-based processing. They also suggest a perspective on theoretical issues that is fundamentally different from the modal perspective, which was derived from assumptions about resource limitations. But are the assumptions valid? Possibly. Each one receives some support.

The assumption of obligatory encoding is supported by studies of incidental learning and comparisons of incidental and intentional learning. The evidence overwhelmingly indicates that people can learn a lot without intending to; incidental learning is usually closer to intentional learning than to chance. The intention to learn seems to have little effect beyond focusing attention on the items to be learned (Hyde & Jenkins, 1969; Mandler, 1967). However, the assumption of obligatory encoding does not imply that all items will be encoded equally well. Attention to an item may be sufficient to encode it into memory, but the quality of the encoding will depend on the quality and quantity of attention. As the levels-of-processing literature has shown, subjects remember the same items better when they attend to their semantic features rather than their physical features (Craik & Tulving, 1975). Dual-task studies show that subjects remember less under dual-task conditions than under single-task conditions (Naveh-Benjamin & Jonides, 1984; Nissen & Bullemen 1987).[3]

The assumption of obligatory retrieval is supported by studies of Stroop and priming effects, in which attention to an item activates associations in memory that facilitate performance in some situations and interfere with it in others (for a review, see Logan, 1980). The most convincing evidence comes from studies of episodic priming that show facilitation from newly learned associates (McKoon & Ratcliff, 1980; Ratcliff & McKoon, 1978,1981), The assumption of obligatory retrieval does not imply that retrieval will always be successful or that it will be easy. Many factors affect retrieval time (Ratcliff, 1978), including practice on the task (Pirolli & Anderson, 1985). The prevailing conditions in studies of automaticity are generally good for retrieval: The same items have been presented many times and so should be easy to retrieve. The algorithm, if used in parallel with retrieval, will screen out any slow or difficult retrievals by finishing first and providing a solution to the task.

The assumption of an instance representation for learning contrasts with the modal view. Many theories assume a *strength* representation (e.g., LaBerge & Samuels, 1974; MacKay, 1982; Schneider, 1985), and others include strength as one of several learning mechanisms (e.g., Anderson, 1982). In instance theories, memory becomes stronger because each experience lays down a separate trace that may be recruited at the time of retrieval; in strength theories, memory becomes stronger by strengthening a connection between a generic representation of a stimulus and a generic representation of its interpretation or its response.

Instance theories have been pitted against strength theories in studies of memory and studies of categorization. In memory, strength is not enough; the evidence is consistent with pure instance theories or strength theories supplemented by instances (for a review, see Hintzman, 1976). In categorization, *abstraction* is the analog of strength. Separate exposures are combined into a single generic, *prototypic* representation, which is compared with incoming stimuli. The evidence suggests that prototypes by themselves are not enough; instances are important in categorization (for a review, see Medin & Smith, 1984). The success of instance theories in these domains suggests that they may succeed as well in explaining automatization. Experiment 5 pits the instance theory against certain strength theories.

The instance representation also implies that automatization is *item-based* rather than process-based. It implies that automatization involves learning specific responses to specific stimuli. The underlying processes need not change at all – subjects are still capable of using the algorithm at any point in practice (e.g., adults can still add by counting), and memory retrieval may operate in the same way regardless of the amount of information to be retrieved. Automaticity is specific to the stimuli and the situation experienced during training. Transfer to novel stimuli and situations should be poor. By contrast, the modal view suggests that automatization is *process-based*, making the underlying process more efficient, reducing the amount of resources required or the number of steps to be executed (e.g., Anderson, 1982; Kolers, 1975; LaBerge & Samuels, 1974; Logan, 1978). Such process-based learning should transfer just as well to novel situations with untrained stimuli as it does to familiar situations with trained stimuli.

There is abundant evidence for the specificity of automatic processing in the literature on consistent versus varied mapping. Practice improves performance on the stimuli and mapping rules that were experienced during training but not on other stimuli or even other rules for mapping the same stimuli onto the same responses (for a review, see Shiffrin & Dumais, 1981). The experiments presented later in the article provide further evidence.

The theory differs from process-based views of automatization in that it assumes that a task is performed differently when it is automatic than when it is not; automatic performance is based on memory retrieval, whereas nonautomatic performance is based on an algorithm. This assumption may account for many of the qualitative properties that distinguish automatic and nonautomatic performance. The properties of the algorithm may be different from the properties of memory retrieval; variables that affect the algorithm may be different from the variables that affect memory retrieval. In particular, variables that affect performance early in practice, when it is dominated by the algorithm, may not affect performance later in practice, when it is dominated by memory retrieval. Thus, dual-task interference and information-load effects may diminish with practice because they reflect difficulties involved in using the initial algorithm that do not arise in memory retrieval.

This theme is developed in detail in a subsequent section of the article.

The assumption that automatic and nonautomatic processing are different does not imply that they have opposite characteristics, as many current treatments of automaticity imply. Automatic processing may be well defined (having the properties of memory retrieval), but nonautomatic processing may not be. The set of algorithms that are possible in the human cognitive system is probably unbounded, and it seems highly unlikely that any single property or set of properties will be common to all algorithms, or even to most of them. Thus, the present theory does not endorse the strategy of defining automaticity by listing dichotomous properties (e.g., serial vs. parallel; effortful vs. effortless) that distinguish it from another specific kind of processing (e.g., *attentional*, Logan, 1980; *controlled*, Shiffrin & Schneider, 1977; *effortful*, Hasher & Zacks, 1979; *strategic*, Posner & Snyder, 1975; and *conscious*, Posner & Klein, 1973).

Quantitative Properties of Automaticity

The theory is primarily intended to account for the major quantitative properties of automatization, the speed-up in processing and reduction in variability that result from practice. The speed-up is the least controversial of the properties of automaticity. It is observed in nearly every task that is subject to practice effects, from cigar rolling to proving geometry theorems (for a review, see Newell & Rosenbloom, 1981). In each case, the speed-up follows a regular function, characterized by substantial gains early in practice that diminish with further experience. More formally, the speed-up follows a power function,

$$RT = a + bN^{-c},$$

where RT is the time required to do the task, N is the number of practice trials, and a, b, and c are constants. a represents the asymptote, which is the limit of learning determined perhaps by the minimum time required to perceive the stimuli and emit a response; b is the difference between initial performance and asymptotic performance, which is the amount to be learned; and c is the rate of learning. The values of these parameters vary between tasks, but virtually all practice effects follow a power function.[4]

The power-function speed-up has been accepted as a nearly universal description of skill acquisition to such an extent that it is treated as a law, a benchmark prediction that theories of skill acquisition must make to be serious contenders (see, e.g., Anderson, 1982; Crossman, 1959; MacKay, 1982; Newell & Rosenbloom, 1981).[5] If they cannot account for the power law, they can be rejected immediately. The instance theory predicts a power-function speed-up.

The reduction in variability that accompanies automatization is not well understood, largely because most theories neglect it. The literature shows that variability decreases with practice (e.g., McLeod, McLaughlin, & Nimmo-Smith, 1985; Naveh-Benjamin & Jonides, 1984), but the form of the function has not been specified; there is nothing akin to the power law. The instance theory predicts that the standard deviation will decrease as a power function of practice. Moreover, it predicts a strong constraint between the power function for the mean and the one for the standard deviation: they must have the same exponent, c.

The predictions for the power law follow naturally from the main assumptions of the instance theory – obligatory encoding, obligatory retrieval, and instance representation. The predictions are developed mathematically in Appendix A. The remainder of this section provides an informal account.

The theory assumes that each encounter with a stimulus is encoded, stored, and retrieved separately. Each encounter with a stimulus is assumed to be represented as a *processing episode*, which consists of the goal the subject was trying to attain, the stimulus encountered in pursuit of the goal, the interpretation given to the stimulus with respect to the goal, and the response made to the stimulus. When the stimulus is encountered again in the context of the same goal, some proportion of the processing episodes it participated in are retrieved. The subject can then choose to respond on the basis of the retrieved information, if it is coherent and consistent with the goals of the current task, or to run off the relevant algorithm and compute an interpretation and a response.

The simplest way to model the choice process is in terms of a race between memory and the algorithm – whichever finishes first controls the response. Over practice, memory comes to dominate the algorithm because more and more instances enter the race, and the more instances there are, the more likely it is that at least one of them will win the race. The power-function speed-up and reduction in variability are consequences of the race.

Memory Retrieval and the Power Law for Means and Standard Deviations

The memory process is itself a race. Each stored episode races against the others, and the subject can respond on the basis of memory as soon as the first episode is retrieved. The race can be modeled by assuming that each episode has the same distribution of finishing times. Thus, the finishing time for a retrieval process involving N episodes will be the minimum of N samples from the same distribution, which is a well-studied problem in the statistics of extremes (e.g., Gumbel, 1958). Intuition suggests that the minimum will decrease as N increases, but the question is, will it decrease as a power function of N?

It would be difficult to prove mathematically that the minimum of N samples from every conceivable distribution decreases as a power function of N, but it is possible to prove it for a broad class of initial distributions (all positive-valued distributions). That proof is presented in Appendix A. The power-function speed-up is a consequence of two counteracting factors: On the one hand, there are more opportunities to observe an extreme value as sample size increases, so the expected value of the minimum will decrease. But, on the other hand, the more extreme the value, the lower the likelihood of sampling a value that is even more extreme, so the reduction in the minimum that results from increasing sample size by m will decrease as sample size increases. The first factor produces the speed-up; the second factor produces the negative acceleration that is characteristic of power functions.

Intuition also suggests that variability will decrease as N increases: The losers of the race restrict the range that the winner can occupy. The more losers, the more severe the restriction, and thus, the smaller the variability. Moreover, the same factors that limit the reduction in the mean limit the reduction in the range that the minimum can occupy, so the reduction in variability should be negatively accelerated like the reduction in the mean. But does it follow a power function? And if so, is the exponent the same as the one for the mean?

The proofs in Appendix A show that the entire distribution of minima decreases as a power function of sample size, not just the mean of the distribution. This implies a power-function reduction in the standard deviation as well as the mean. Because the mean and standard deviation are both functions of the same distribution, the exponent of the power function for the mean will equal the exponent of the power function for the standard deviation.

These predictions are unique to the instance theory. No other theory of skill acquisition or automaticity predicts a power-function reduction in the standard deviation and constrains its exponent to equal the exponent for the reduction in the mean.

The Power Law and the Race between the Algorithm and Memory Retrieval

According to the instance theory, automatization reflects a transition from performance based on an initial algorithm to performance based on memory retrieval. The transition may be explained as a race between the algorithm and the retrieval process, governed by the statistical principles described in the preceding section and in Appendix A. In effect, the algorithm races against the fastest instance retrieved from memory. It is bound to lose as training progresses because its finishing time (distribution) stays the same while the finishing time for the retrieval process decreases. At some point, performance will depend on memory entirely, either as a consequence of statistical properties of the race or because of a strategic decision to trust memory and abandon the algorithm.

Does the transition from the algorithm to memory retrieval compromise the power-law predictions derived in the preceding section and in Appendix A? Strictly speaking, it must. The proofs assume independent samples from n identical distributions, and the distribution for algorithm finishing times is likely to be different from the distribution of retrieval times. But in practice, the deviations from the predicted power law may be small. It is hard to make general analytical predictions because the algorithm and memory distributions may differ in many ways. They may have the same functional form but different parameters (the exponential case is analyzed in Appendix A) or they may have different forms.

Any distortion that does occur will be limited to the initial part of the learning curve. Once performance depends on memory entirely it will be governed by the power law. Before that – during the transition from the algorithm to memory retrieval – the proofs no longer guarantee a power law.

I explored the effects of various transitions on the power-law predictions through Monte Carlo simulation, using truncated normal distributions for the algorithm and the memory process. Earlier simulations showed that the means and standard deviations of the minimum of *n* samples from a truncated normal decreased as a power function of *n*. The current simulations addressed whether a race against another truncated normal with different parameters would distort the power-function fits. The algorithm was represented by nine different distributions, factorially combining three means (350, 400, and 450 ms) and three standard deviations (80, 120, and 160 ms). The memory process was represented by two distributions with different means (400 and 500 ms) and the same standard deviation (100 ms). These parameters represent a reasonably wide range of variation, including cases in which memory is faster and less variable than the algorithm, as well as cases in which it is slower and more variable. This is important because the outcome of the race will depend on the mean and standard deviation of the parent distributions. Other things equal, the distribution with the faster mean will win the race more often. Also, the distribution with the larger standard deviation will win more often because extreme values are more likely the larger the standard deviation.

The effects of 1 to 32 presentations were simulated. The simulations assumed that the algorithm was used on every trial (i.e., the "subject" never chose to abandon it in favor of memory) and that each prior episode was retrieved on every trial. Thus, for a trial on which a stimulus appeared for the *n*th time, reaction time was set equal to the minimum of *n* samples, one from the distribution representing the algorithm and $n-1$ from the distribution representing the memory process. There were 240 simulated trials for each number of presentations (1–32), which approximates the number of observations per data point in the experiments reported in subsequent sections of the article.

The simulations provided three types of data: mean reaction times, standard deviation of reaction times, and the proportion of trials on which the algorithm won the race. Power functions were fitted to the means and standard deviations simultaneously (using STEPIT; Chandler, 1965), such that the exponent was constrained to be the same for means and standard deviations as the instance theory predicts. If the race with the algorithm distorts the relation between means and standard deviations, the constrained power functions will not fit well.

Means and standard deviations. The simulated mean reaction times appear in Figure 1, and the standard deviations appear in Figure 2. The points represent the simulated data and the lines represent fitted power functions, constrained to have the same exponent for means and standard deviations. The exponents appear in Table 1.

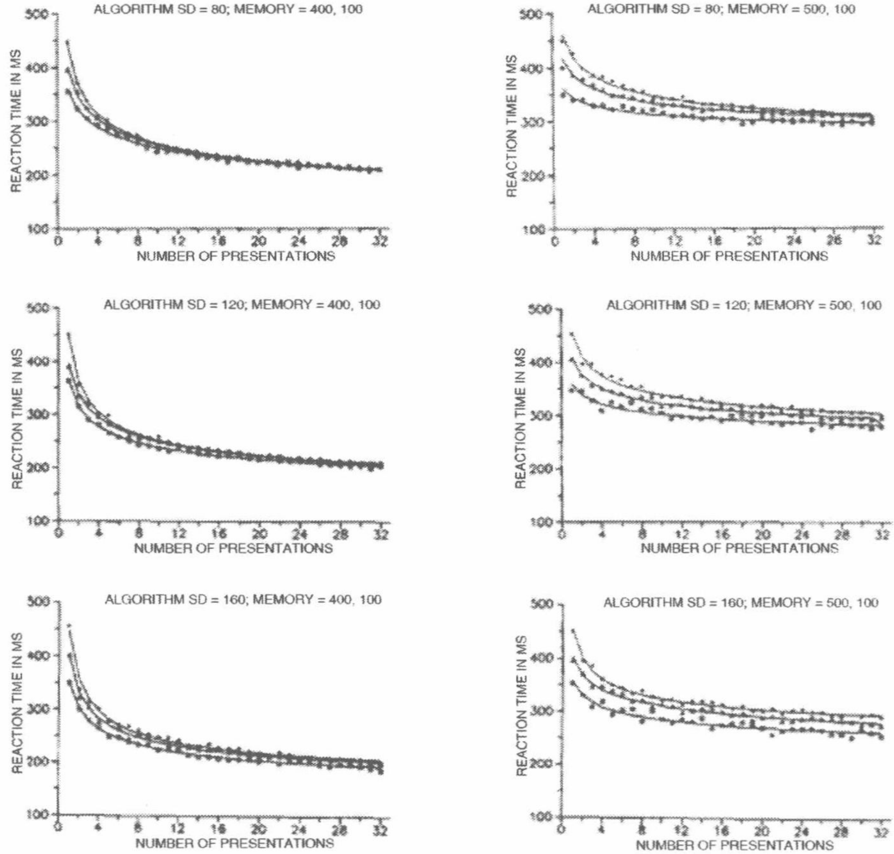

Figure 1. Reaction times from simulations of a race between an algorithm and a memory retrieval process as a function of the number of presentations of an item. (Points represent the simulated data; lines, fitted power functions. Power functions are constrained to have exponents equal to those of power functions fitted to the standard deviations, which are plotted in Figure 2. Each panel portrays three algorithms with different means – 350,400, and 450 from the bottom function to the top – and the same standard deviation – 80 in the top two panels, 120 in the middle two, and 160 in the bottom two – racing against a memory process with a constant mean – 400 in the left-hand panels, 500 in the right – and standard deviation, 100 in all panels.)

Two points are important to note: First, the means and standard deviations both decreased as the number of presentations increased, and the trend was well fit by the constrained power functions (r^2 ranged from .992 to 1.000 with a median of .998; root-mean-squared deviation between predicted and observed values ranged from 2.38 ms to 5.93 ms, with a median of 3.81 ms). Thus, the race does not appear to compromise the power law; the instance theory can predict power functions even when memory retrieval must race against a faster or slower algorithm. Second, the race distorts the form of the power function;

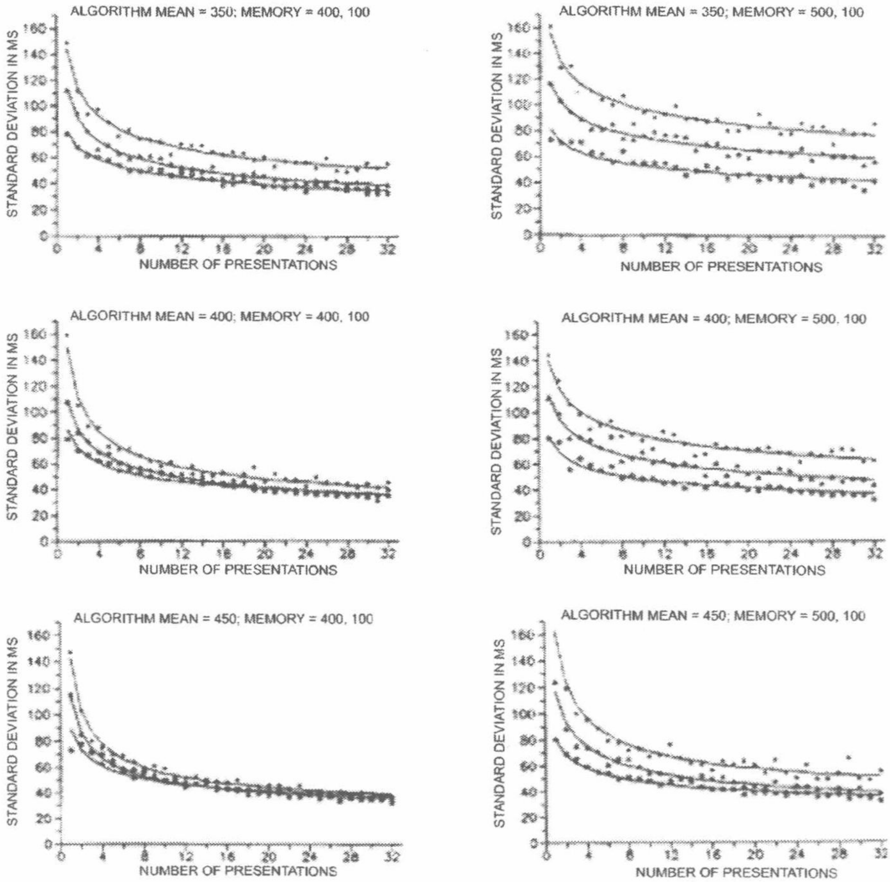

Figure 2. Standard deviations from the simulated race between an algorithm and a memory retrieval process. (Points represent the simulated data; lines, fitted power functions. Power functions are constrained to have exponents equal to those of power functions fitted to the means, which are plotted in Figure 1. Each panel portrays three different algorithms with different standard deviations – 80, 120, and 160 from the bottom function to the top – and the same mean – 350 in the top panels, 400 in the middle, and 450 in the bottom – racing against a memory process with a constant mean – 400 in the left panels, 500 in the right – and standard deviation, 100.)

the exponents from the constrained fits are systematically different from the fits to the memory process by itself. The exponents from the race increase in absolute magnitude as the algorithm mean increases and as the algorithm standard deviation increases.

The simulated data illustrate the effects of "qualitative" differences between the automatic and nonautomatic performance. Each panel has three different versions of the algorithm racing against a single version of the memory retrieval process, and in each case, initial differences due to the algorithm disappear