Psychological Review
1988, Vol. 95, No. 4, 492–527

0033-295X/88/$00.75

Toward an Instance Theory of Automatization

Gordon D. Logan
University of Illinois

This article presents a theory in which automatization is construed as the acquisition of a domain-specific knowledge base, formed of separate representations, *instances*, of each exposure to the task. Processing is considered automatic if it relies on retrieval of stored instances, which will occur only after practice in a consistent environment. Practice is important because it increases the amount retrieved and the speed of retrieval; consistency is important because it ensures that the retrieved instances will be useful. The theory accounts quantitatively for the power-function speed-up and predicts a power-function reduction in the standard deviation that is constrained to have the same exponent as the power function for the speed-up. The theory accounts for qualitative properties as well, explaining how some may disappear and others appear with practice. More generally, it provides an alternative to the modal view of automaticity, arguing that novice performance is limited by a lack of knowledge rather than a scarcity of resources. The focus on learning avoids many problems with the modal view that stem from its focus on resource limitations.

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 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;

This research was supported by National Science Foundation Grant BNS 8510365. It was reported, in part, at the annual meeting of the Psychonomic Society, November 1987, in Seattle, Washington.

Several people contributed to the work reported here through helpful discussions and comments on various drafts of the article. I would like to thank Jerry Busemeyer, Tom Carr, Dick Foster, Rob Kail, Larry Maloney, Doug Medin, Brian Ross, Eliot Smith, and Jane Zbrodoff for their criticisms and suggestions. I am grateful to Walter Schneider and an anonymous reviewer for helpful comments on the article. I would also like to thank Randy Tonks and Leslie Kane for testing the subjects.

Correspondence concerning this article should be addressed to Gordon D. Logan, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, Illinois 61820.

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 in-

[1] The concept of automaticity has not been articulated well in multiple-resource theories (Navon & Gopher, 1979; Wickens, 1984). For the most part, investigators claim that automatic processes use resources efficiently, reiterating the main assumption underlying the modal, single-capacity view. But they don't specify which resources are used more efficiently. The discussion focuses on a single resource, leaving it for the reader to decide why that particular resource should be used more efficiently or whether the other resources come to be used more efficiently as well (also see Allport, 1980; Logan, 1985b).

[2] Strictly speaking, the instance theory considers performance to be automatic when it is based on memory retrieval whether that occurs on the 10th trial or the 10,000th. Early in practice, before subjects rely in memory entirely, performance may be automatic on some trials (i.e., those on which memory provides a solution) but not on others (i.e., those on which the algorithm computes a solution).

tentional 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 & Bullemer, 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.

[3] The assumption of obligatory encoding is similar to Hasher and Zacks's (1979, 1984) notion of automatic encoding of certain stimulus attributed (e.g., frequency of presentation, location). However, Hasher and Zacks assumed that encoding is not influenced by manipulations of attention, intention, or strategy, whereas the instance theory assumes only that it is obligatory. Hasher and Zacks's position was challenged recently by evidence that encoding can be influenced by orienting tasks and dual-task conditions. Greene (1984) and Fisk and Schneider (1984) showed that subjects remembered the frequency of stimuli presented during semantic orienting tasks better than the frequency of stimuli presented during orienting tasks that focused on physical or structural features. Naveh-Benjamin and Jonides (1986) showed that subjects remembered the frequency of stimuli presented under single-task conditions better than the frequency of stimuli presented under dual-task conditions (also see Naveh-Benjamin, 1987). The instance theory can accomodate these findings easily. It assumes that attention to an item will have some impact on memory; it does not assume that all conditions of attention produce the same impact.

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?

[4] Power functions are linear when plotted in logarithmic coordinates. Thus,

$$\log(RT - a) = \log(b) - c\log(N).$$

The power-function speed-up is sometimes called the *log–log linear law of learning* (Newell & Rosenbloom, 1981).

[5] Mazur and Hastie (1978) argued that learning curves were hyperbolic rather than exponential and reanalyzed a great deal of data to demonstrate their point. However, (a) they analyzed accuracy data rather than reaction times—and the theories of the power-function speed-up do not necessarily make predictions about accuracy—and (b) the hyperbolic function is a power function with an exponent of −1, so it is a special case of the power law. Consequently, Mazur and Hastie's arguments and analyses do not contradict the power law.

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 nth 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

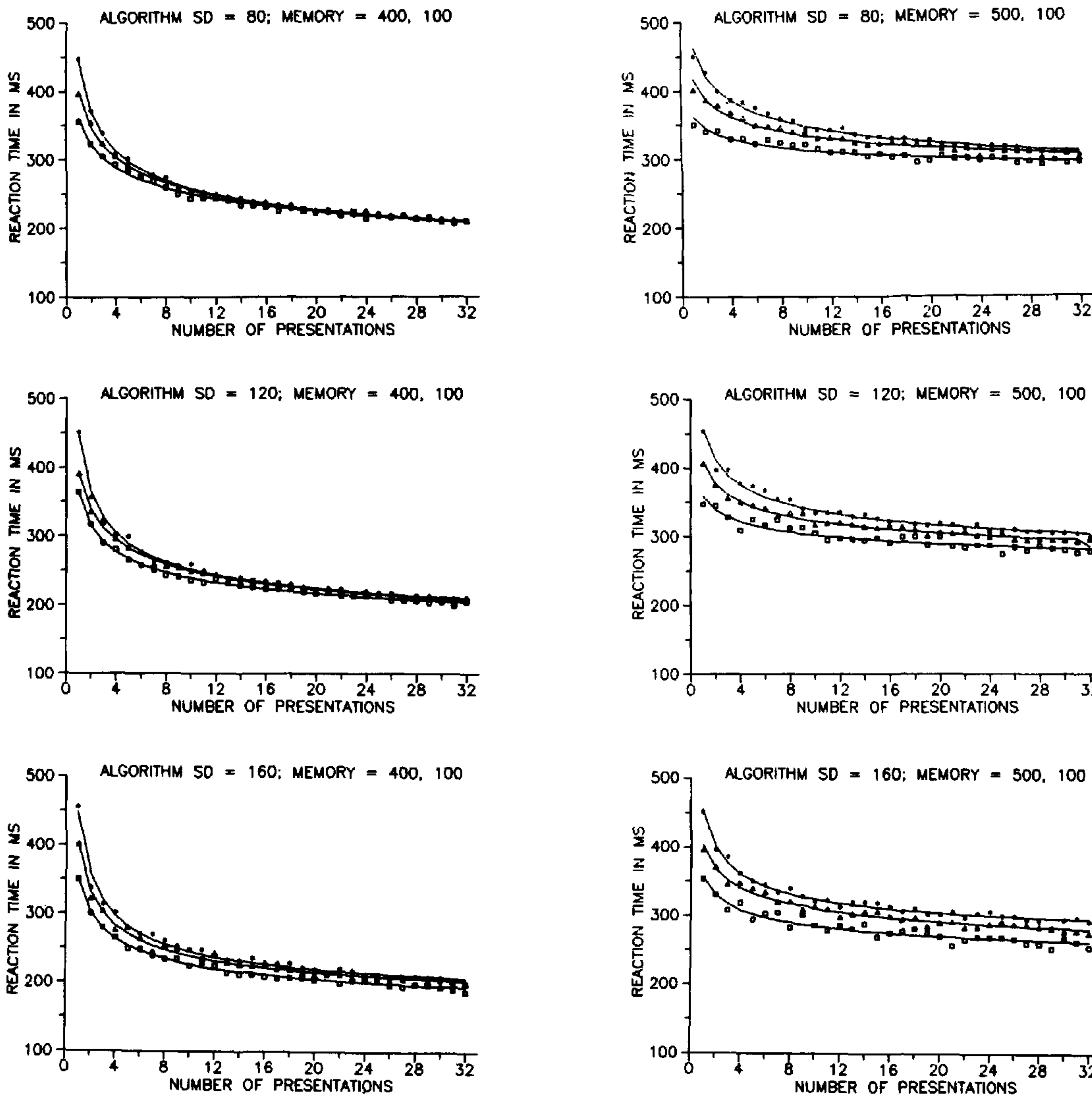

Figure 1. Reaction times from simulations of a race between an algorithm and a memory retrieval process as a function of the number of presentations of an item. (Points represent the simulated data; lines, fitted power functions. Power functions are constrained to have exponents equal to those of power 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.)

the lines represent fitted power functions, constrained to have the same exponent for means and standard deviations. The exponents appear in Table 1.

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

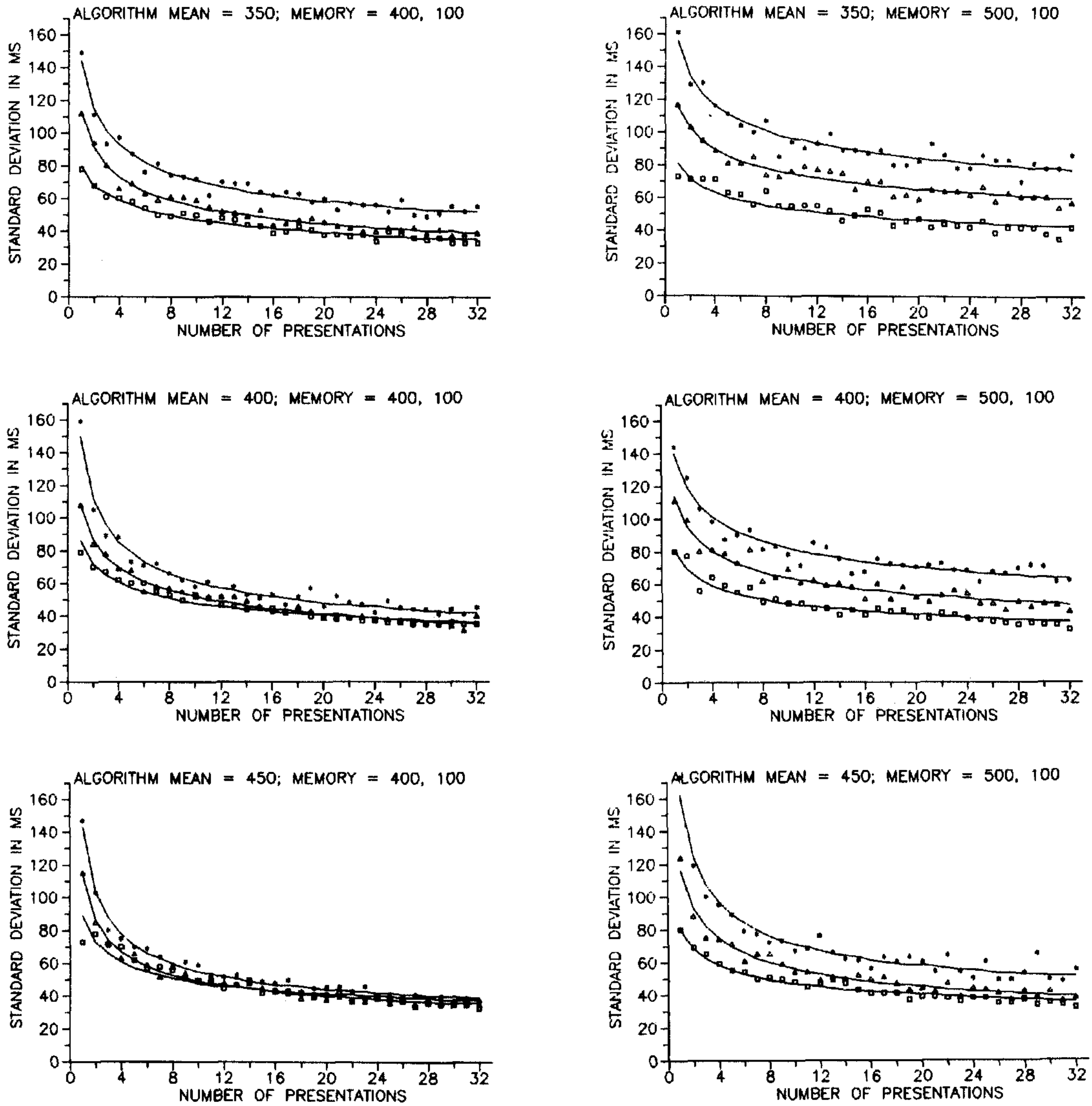

Figure 2. Standard deviations from the simulated race between an algorithm and a memory retrieval process. (Points represent the simulated data; lines, fitted power functions. Power functions are constrained to have exponents equal to those of power functions fitted to the means, which are plotted in 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.)

Table 1
Exponents of Power Functions Fitted to the Simulated Data in Figures 1 and 2, Which Represent Races Between Various Memory Processes and Algorithms

Algorithm	Memory = 400, 100: Algorithm *SD*			Memory = 500, 100: Algorithm *SD*		
M	80	120	160	80	120	160
350	0.156	0.340	0.367	0.185	0.193	0.248
400	0.251	0.301	0.465	0.227	0.252	0.251
450	0.368	0.479	0.550	0.230	0.316	0.452

Note. The exponent for the 400, 100 memory process is 0.302; the exponent for the 500, 100 memory process is 0.309.

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; 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 after a few presentations. Averaged over all six panels, the difference between the 450-ms algorithm and the 350-ms algorithm decreased from 97 ms on the first presentation to 15 ms on the 32nd presentation. If the algorithm were a memory search process and the different conditions corresponded to memory set sizes of 1 and 5, the initial slope of 24 ms/item (i.e., 97/4) would approach zero (i.e., 15/4) after 32 presentations.

Another kind of "qualitative difference" can be seen by comparing the two panels in each row. In this case, differences that were not there initially, emerge with practice. The conditions of the algorithm are the same in the right and in the left panel, but the memory retrieval process is much slower in the right panel. Averaged over the three pairs of panels, the difference between right and left was −1 ms on the first presentation and 84 ms on the 32nd presentation.

The standard deviations show "qualitative differences" similar to those observed for the means. Initial differences due to the algorithm were substantially reduced as the number of presentations increased and memory retrieval came to dominate performance. Averaging over conditions, the initial difference between the 160-ms algorithm and the 80-ms algorithm decreased to 22 ms by the 32nd presentation.

P(algorithm first). Figure 3 presents a different perspective on the outcome of the race, the probabilities that the algorithm finished first. The probabilities were affected by the mean and the variance of the algorithm, increasing as the algorithm became faster and more variable. They were also affected by the mean of the memory process, decreasing as the memory process becomes faster.

The memory process came to dominate the algorithm relatively quickly. Averaged over all conditions, the algorithm won the race only 21% of the time after 16 presentations and 16% of the time after 32 presentations. The speed of memory retrieval had a large effect on the outcome: The 400-ms retrieval process won 90% of the time after 16 presentations and 93% of the time after 32 presentations, whereas the 500-ms retrieval process won only 68% of the time after 16 presentations and 76% of the time after 32 presentations.

The dominance of the memory process after such a small number of trials is important because it suggests that memory retrieval will eventually win the race regardless of the speed and variability of the algorithm. Possibly, memory retrieval will come to win the race even if the algorithm becomes faster with practice. Thus, memory retrieval may provide a back-up mechanism for automatization and skill acquisition even when skill and automaticity are acquired through other mechanisms.

Memorability and the Rate of Learning

The speed with which the memory process comes to dominate the algorithm has important implications for studies of automaticity: It suggests that automatization can occur very quickly (also see Logan, 1988; Naveh-Benjamin & Jonides, 1984; E. Smith & Lerner, 1986). This means that it is feasible to study automatization in a single session, as was done in some of the experiments reported in subsequent sections of this article.

An important question raised by the instance theory is why automatization takes so long in other experiments. In search studies, for example, several thousand trials spanning 10 to 20 sessions are often necessary to produce automaticity (e.g., Shiffrin & Schneider, 1977). The discrepancy may arise for at least three reasons: First, the criterion for automaticity is different in the previous studies. In Shiffrin and Schneider's search studies, automatization was not considered to be complete until the slope of the search function reached zero (but see Cheng, 1985; Ryan, 1983). From the present perspective, automatization may never be complete, in that each additional instance will have some effect on memory, even if its effect does not appear in the primary performance data (also see Logan, 1985b). The present perspective also suggests there may be a shift in the direction of automaticity after only a few trials, and this shift may be a more important phenomenon to study than the zero slope addressed by Shiffrin, Schneider, and others (also see Logan, 1979, 1985b).

Second, differences in the apparent rate of automatization may be artifacts of the way the data are plotted. Instance theory argues that means and standard deviations should be plotted as a function of the number of trials per stimulus because each trial potentially adds a new instance to memory. But data are usually plotted against sessions, disregarding the number of trials per stimulus. Consequently, a task with more stimuli per session (and thus fewer trials per stimulus) will appear to be

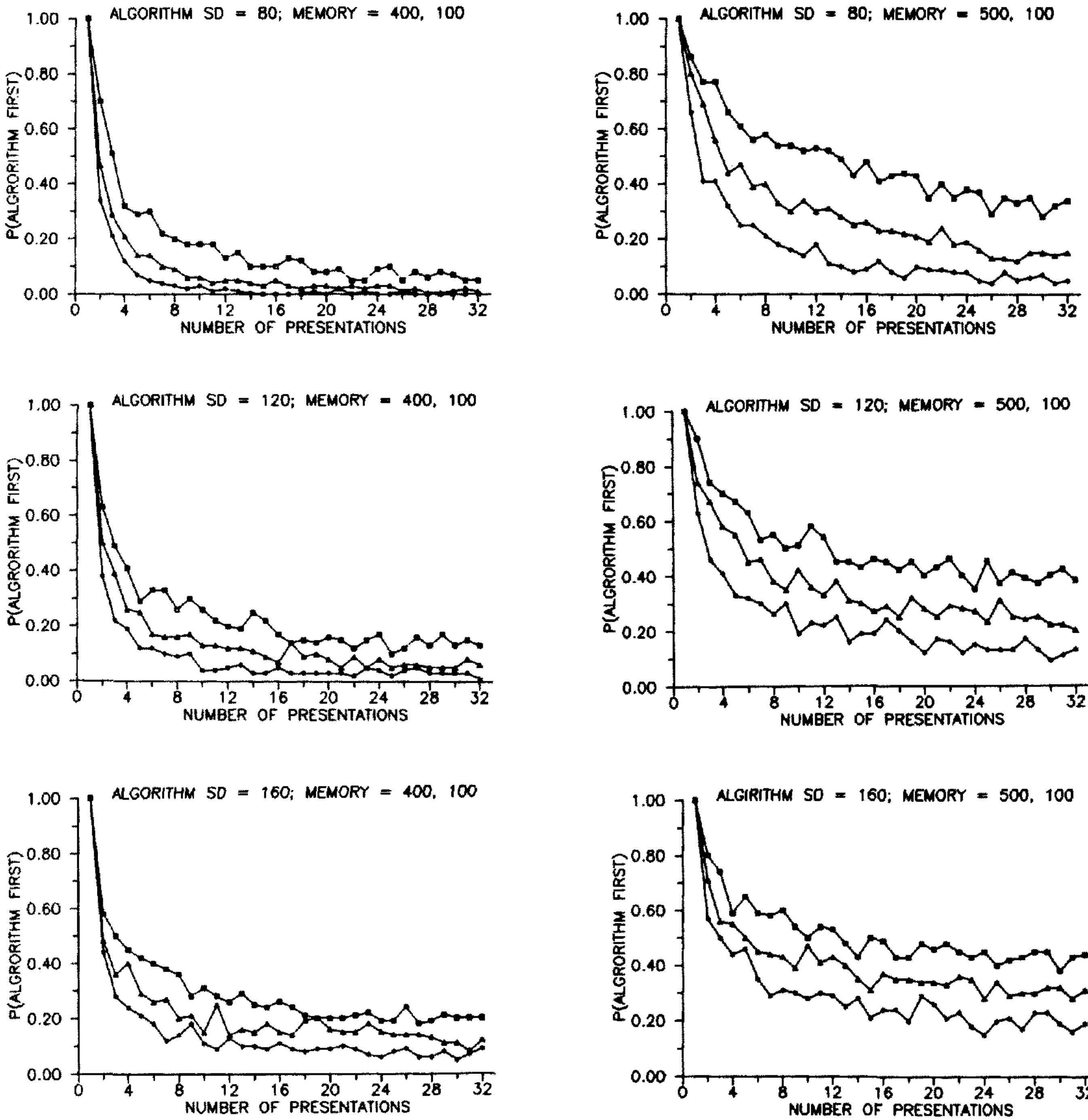

Figure 3. Probability that the algorithm finished first in the simulated race between an algorithm and a memory retrieval process. (Each panel portrays three algorithms with different means—350, 400, and 450—and the same standard deviation—80 in the top two panels, 120 in the middle two, and 160 in the 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.)

learned more slowly than a task with fewer stimuli per session (and thus more trials per stimulus), even if the rate of learning per stimulus is equal.

Third, differences in the rate of automatization may reflect the memorability of the stimuli. According to instance theory, stimuli that are easy to remember will show evidence of automaticity relatively quickly, whereas stimuli that are hard to remember will take a long time to show evidence of automaticity. It may be that the letter arrays studied by Shiffrin, Schneider, and others are hard to remember. By contrast, the simulations so far have assumed that each and every encounter with a stimulus is encoded and retrieved.

The effect of memorability can be modeled by slowing down the retrieval time or by varying the probability that a stimulus will be encoded and retrieved. Most likely, memorability has both effects (Ratcliff, 1978), but it is interesting to consider them separately. The effects of slowing down retrieval can be seen in Table 1: The learning rate, as measured by the power-function exponent was slower for the 500-ms memory process than for the 400-ms memory process when they both raced against the same algorithm.

The effect of varying retrieval probability is to slow the rate of learning by reducing the effective number of traces in the race. Reaction times and standard deviations will still decrease as a power function of n, but with a smaller exponent, reflecting a slower rate of learning.[6]

These analyses suggest that there may be no discrepancy between the rate of automatization predicted by instance theory and the rate observed in typical studies of automaticity. It may be possible to observe automatization in a single session, as was suggested earlier, as long as the number of stimuli is small and the stimuli themselves are easy to remember (also see Logan, 1988; Naveh-Benjamin & Jonides, 1984; Smith & Lerner, 1986).

Conclusions

The theoretical analyses in this section showed that the quantitative properties of automaticity can be accounted for by an instance theory that assumes that subjects store and retrieve representations of each individual encounter with a stimulus. According to the instance theory, automatization reflects a shift from reliance on a general algorithm to reliance on memory for past solutions. Thus, automatization reflects the development of a domain-specific knowledge base; nonautomatic performance is limited by a lack of knowledge rather than by the scarcity of resources.

The power-function speed-up is a statistical consequence of the main assumptions (obligatory encoding, obligatory retrieval, and instance representation), and the theory makes new predictions about the reduction in variability. Standard deviations should decrease as power functions of the number of trials, and the exponents should be the same as the exponents for the means.

What is interesting about the speed-up is that none of the underlying processes change over practice. The algorithm stays the same and so does memory retrieval. Moreover, stimuli are encoded in exactly the same way at every point in practice—each trial results in the encoding and storage of a processing episode. All that changes is the knowledge base that is available to the subject.

It remains to be shown that the instance theory can account for experimental data; that is the purpose of the next section.

Fitting Theory to Data

One of the basic premises of the instance theory is that nonautomatic processes need not have anything in common, except that they are replaced by memory retrieval as practice progresses. In keeping with that premise, the instance theory was fitted to data from two very different tasks, *lexical decision* and *alphabet arithmetic.* The lexical decision task is fast, relatively effortless, and possibly parallel, whereas the alphabet arithmetic task is slow, very effortful, and clearly serial. According to the instance theory, after sufficient practice, both tasks should be performed in the same way—by retrieving past solutions from memory. This prediction was tested indirectly, by examining the fit of power functions to the means and standard deviations of reaction times in both tasks and by looking for evidence of item-specific learning in both tasks. The theoretical analysis in the preceding section makes the strong prediction that both means and standard deviations should decrease as power functions of the number of trials with the same exponent, c. The theoretical analysis also predicts that learning should be item-based; subjects learn specific responses to specific stimuli, and what they learn should not transfer well to different stimuli.

These predictions are tested in Experiments 1–4. Experiment 5 pits the instance theory against certain strength theories.

Experiment 1

Experiment 1 involved a lexical decision task. Subjects were presented with strings of four letters, and their task was to indicate as quickly as possible whether or not the letter string was an English word. The experiment was intended to resemble typical studies of the development of automaticity, in which subjects are exposed to the same items repeatedly throughout practice. Subjects made lexical decisions on the same set of 10 words and 10 nonwords until each word and nonword was presented 16 times. In this paradigm, the average lag between successive presentations is held constant over repetitions, but the degree of nonspecific practice on the task increases with the number of repetitions. To control for nonspecific practice, subjects performed another 16 blocks of lexical decisions, but 10 new words and 10 new nonwords were used in each block. Further details of the procedure and the results of analyses of variance (ANOVAS) on the mean reaction times and standard deviations are presented in Appendix B.

---

[6] If the original power function has an exponent of c and some proportion p of the n episodes are stored and retrieved, then when performance is plotted against n, the data will follow a power function with an exponent $k < c$. That is,

$$n^k = (pn)^c.$$

Taking logs of both sides yields

$$k\log n = c(\log p + \log n),$$

and solving for k yields

$$k = c[(\log p + \log n)/\log n].$$

K must be less than c because $(\log p + \log n)/\log n$ is less than one.

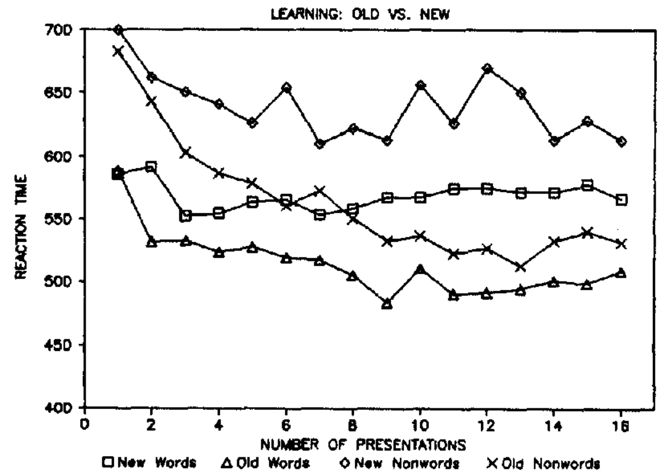

Figure 4. Reaction times to old and new words and nonwords as a function of the number of presentations in the lexical decision task of Experiment 1.

The mean reaction times for new and repeated items are presented in Figure 4. There was some evidence of a general practice effect, which reduced reaction times for new items slightly over blocks. However, the specific practice effect was much stronger: Reaction times to repeated items decreased substantially over blocks, both absolutely and relative to the new-item controls. These effects were apparent for nonwords as well as words.

The means and standard deviations for repeated items appear in Figure 5, the means in the top panel and the standard deviations in the bottom. The points represent the observed data; the lines represent power functions fitted to the data under the constraint that the mean and standard deviation should have the same exponent. Words and nonwords were fitted separately. The estimated parameters for the power functions and measures of goodness of fit (r^2 and root-mean-squared deviation, or *rmsd*) appear in Table 2. Table 2 also contains parameters and goodness-of-fit measures for power functions fitted to means and standard deviations separately, to give some idea of the effect of constraining the exponent.

The data were well fit by power functions. Moreover, the constrained fit was almost as good as the unconstrained fit; *rmsds* were within 2 ms.

The contrast between repeated words and new words suggests that the repetition effect may be instance- or item-based, as the instance theory predicts. Process-based learning predicts no difference, yet a difference was observed. The repetition effect for nonwords is harder to interpret. Nonword reaction times could have decreased because subjects responded to them by remembering what they did when they last saw them, or because subjects responded to them by default (i.e., by failing to find evidence that they were words), and the default response became faster as the word decisions became faster. The present experiment does not allow us to decide between these alternatives. That was a major reason for investigating other schedules of repetition.

Experiment 2

Experiment 2 was intended to resemble typical studies of repetition effects in implicit and explicit memory: Subjects performed several blocks of lexical decision trials. Within each block of trials, some words and nonwords are presented once, some twice, some 4 times, some 6 times, some 8 times, and some 10 times. In this paradigm, the average lag between successive repetitions is confounded with the number of repetitions (being shorter the greater the number of repetitions), but the degree of nonspecific practice is relatively constant over number of repetitions. Also, old nonwords are mixed together with new and old words, so there should be no benefit for nonwords if subjects respond to nonwords by default.

The mean reaction times are plotted in the top panel of Figure 6, and the standard deviations are plotted in the bottom

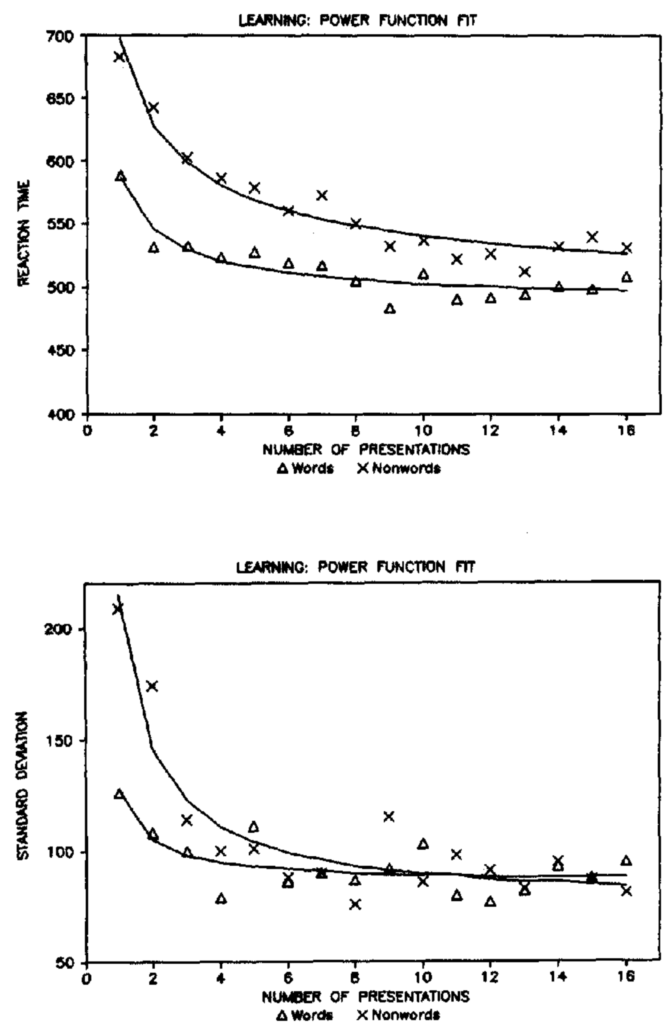

Figure 5. Reaction times (top panel) and standard deviations (bottom panel) as a function of the number of presentations in the lexical decision task of Experiment 1. (Points represent the data; lines represent the best-fitting power function. Power functions for means and standard deviations were constrained to have the same exponent.)

Table 2
Parameter Estimates From Constrained and Separate Fits of Power Functions to Means and Standard Deviations of Reaction Times in the Lexical Decision Tasks of Experiments 1–3

	Experiment 1: Learning		Experiment 2: Repetition		Experiment 3: Mean lag 12		Experiment 3: Mean lag 24	
Parameter	Word	Nonword	Word	Nonword	Word	Nonword	Word	Nonword
				Constrained fits				
ART	485	482	478	518	501	541	518	574
BRT	101	216	106	143	87	141	95	115
C	0.749	0.563	1.033	1.078	1.622	0.702	0.816	0.880
ASD	81	46	84	95	88	89	86	99
BSD	43	157	47	76	54	84	64	75
r^2	0.996	0.994	0.998	0.998	0.996	0.996	0.996	0.996
rmsd	8.9	12.2	5.6	5.6	9.5	10.2	8.6	9.6
				Separate fits				
ART	482	409	479	511	500	540	573	573
BRT	103	280	105	148	88	142	115	115
CRT	0.698	0.327	1.061	0.940	1.571	0.687	1.069	0.859
ASD	85	75	82	102	88	90	61	100
BSD	41	140	48	72	54	84	85	75
CSD	1.063	0.970	0.919	1.601	1.748	0.724	0.387	0.928
r^2	0.996	0.996	0.998	0.998	0.996	0.998	0.998	0.996
rmsd	8.9	10.9	5.6	5.1	9.5	10.2	8.3	9.6

Note. ART = asymptote for mean reaction time; BRT = multiplicative constant for mean reaction time; ASD = asymptote for standard deviation; BSD = multiplicative constant for standard deviation; C = exponent fitted to means and standard deviations simultaneously; CRT = exponent fitted to means separately; CSD = exponent fitted to standard deviations separately; r^2 = squared correlation between observed and predicted values; *rmsd* = root mean squared deviation from prediction.

panel. The points represent the observed data, and the lines represent fitted power functions constrained such that means and standard deviations have the same exponent. As in Experiment 1, the constrained power-function fit was excellent for both the means and the standard deviations. The parameters of the constrained power functions and measures of goodness of fit appear in Table 2. Table 2 also contains parameters and measures of goodness of fit for power functions fitted to the means and standard deviations separately. Again, the constrained fit was nearly identical to the separate fit.

These results confirm the conclusion from Experiment 1: that the repetition effect is specific to individual stimuli because repeated and new stimuli were mixed randomly in each block. Thus, subjects could not have adjusted speed–accuracy criteria, and so forth, in anticipation of repeated stimuli, as they could have in the previous experiment. Significantly, this conclusion applies to the nonwords as well as the words: Subjects responded to repeated nonwords faster than they responded to new words, so they could not have sped up their reaction times to repeated nonwords by default responding. Thus, the data suggest that subjects remembered their previous encounters with specific nonwords and with specific words, as instance theory predicts. It remains possible, however, that the benefit from repeated presentations could be an artifact of the shorter lag between successive repetitions for stimuli that are repeated more often. Experiment 3 was intended to address that issue.

Experiment 3

In Experiment 3, the number of repetitions was manipulated by varying the number of blocks in which a word or nonword appeared. Some words and nonwords were presented in only 1 block, others in 2 consecutive blocks, and others in 4, 8, and 16 consecutive blocks. Thus, the lag between successive presentations was held constant, like the first experiment, but old and new items were mixed randomly, like the second experiment.

In addition, the mean lag between successive presentations was varied between subjects to see whether the confounding of lag with repetitions in Experiment 2 was likely to have affected the results. One half of the subjects had a mean lag of 12 items between successive presentations, and the other half had a mean lag of 24 items (see Appendix B for further details of the procedure).

The data from the mean lag 12 group are presented in Figure 7; the data from the mean lag 24 group are presented in Figure 8. In each case, the top panel contains the mean reaction times and the bottom panel contains the standard deviations. The points represent the observed data, and the lines represent constrained power functions fitted to the means and standard deviations. Mean lag had no significant effects on performance, neither main effect nor interactions (see Appendix B). For both lag conditions, reaction times and standard deviations decreased with repetition, as the instance theory predicts. The constrained power functions fitted the data very well, almost as well as the

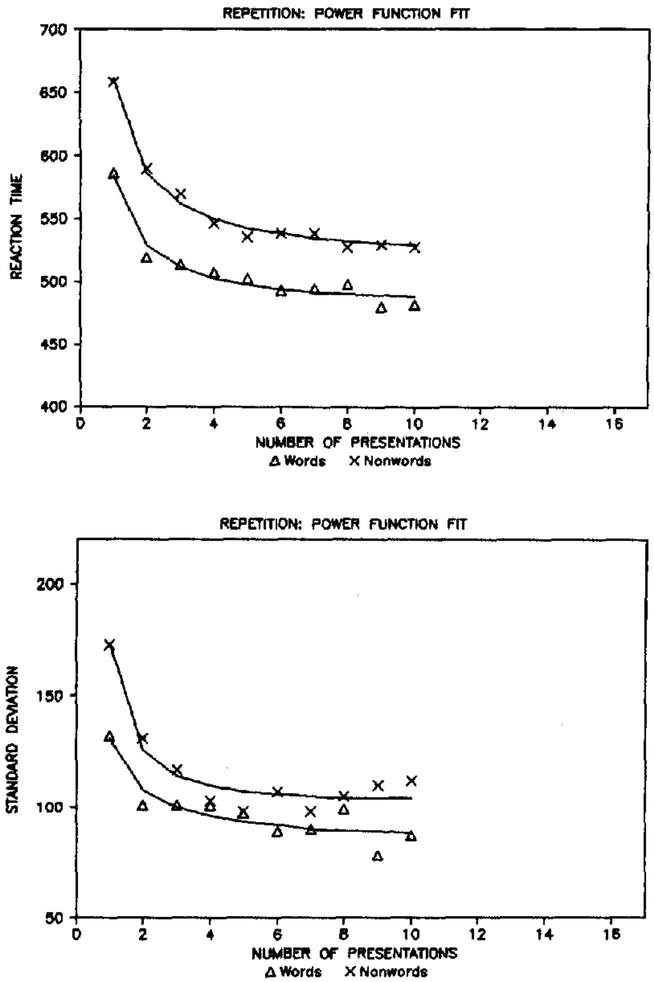

Figure 6. Reaction times (top panel) and standard deviations (bottom panel) as a function of the number of presentations in the lexical decision task of Experiment 2. (Points represent the data; lines represent the best-fitting power function. Power functions for means and standard deviations were constrained to have the same exponent.)

separate fits (see Table 2 for parameter values and measures of goodness of fit).

These results indicate that the benefit from repetition is specific to individual stimuli because the repeated stimuli that showed benefit were mixed randomly with new stimuli. As in Experiment 2, this was true for nonwords as well as words, which suggests that subjects remembered individual encounters with specific nonwords as well as words. Unlike the previous experiment, there was no confound between lag and the number of presentations, so the increased benefit with multiple repetitions was not an artifact of lag. Indeed, the null effect of mean lag suggests that lag is not an important variable when varied within the limits of Experiments 1–3.

Discussion of Experiments 1–3

Experiments 1–3 support several conclusions: First, they demonstrate a constraint between the means and standard deviations of reaction times that was predicted by the instance theory. The means and standard deviations both decreased as power functions of practice, and the functions had the same exponent. This was confirmed in each experiment. It was true for words and for nonwords, and it was true for three different schedules of repetition. The constraint between the means and the standard deviations amounts to predicting co-occurrence of different properties of automaticity, which has been a contentious issue in the recent literature (see Kahneman & Chajzyck, 1983; Paap & Ogden, 1981; Regan, 1981 vs. Logan, 1985b; Jonides, Naveh-Benjamin, & Palmer, 1985); it will be discussed in detail in a subsequent section.

Second, Experiments 1–3 indicate that the repetition effect is specific to individual stimuli, as the instance theory predicts. This was evident in the contrast between repeated items and new-item controls in Experiment 1 and in the contrast between

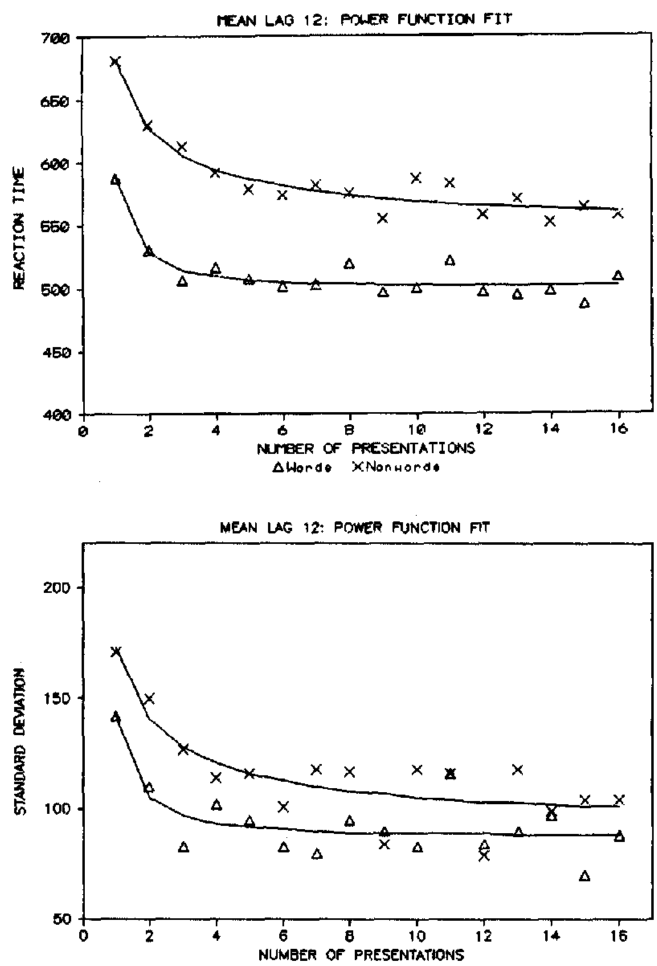

Figure 7. Reaction times (top panel) and standard deviations (bottom panel) as a function of the number of presentations for the mean-lag = 12 condition of the lexical decision task in Experiment 3. (Points represent the data; lines represent the best-fitting power function. Power functions for means and standard deviations were constrained to have the same exponent.)

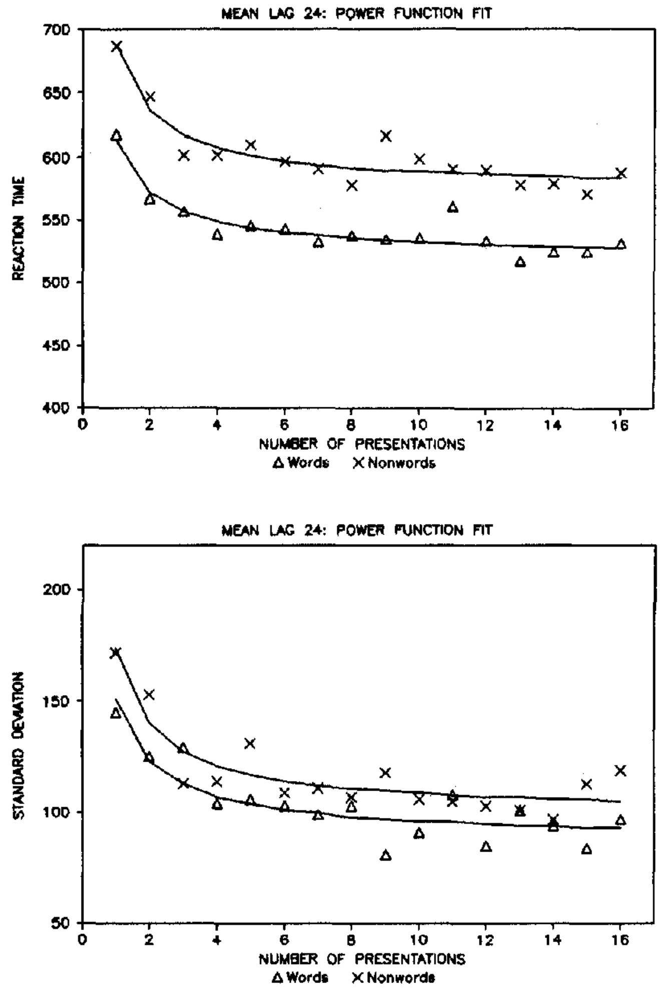

Figure 8. Reaction times (top panel) and standard deviations (bottom panel) as a function of the number of presentations for the mean-lag = 24 condition of the lexical decision task in Experiment 3. (Points represent the data; lines represent the best-fitting power function. Power functions for means and standard deviations were constrained to have the same exponent.)

repeated and new items in Experiments 2 and 3. This specificity of learning rules out process-based theories that explain automatization and skill acquisition in terms of the development of general procedures that deal with new stimuli as effectively as old ones (e.g., Anderson, 1982; Rabbitt, 1981). Such theories may apply in other domains, but they are not appropriate in the present situation.

Third, Experiments 1–3 demonstrate that nonwords as well as words can benefit from repetition. This suggests that the repetition effects are based on memory for specific episodes rather than adjustments to semantic memory (e.g., Feustal, Shiffrin, & Salasoo, 1983; Jacoby & Brooks, 1984; Salasoo, Shiffrin, & Feustal, 1985). It also suggests that repetition effects may derive from the acquisition of associative information rather than item information. If repetition affects item-specific familiarity, subjects would not be able to discriminate words from repeated nonwords; if words showed benefit relative to controls, nonwords should show a cost (e.g., Balota & Chumbley, 1984).

Experiment 4: Alphabet Arithmetic

Experiment 4 examined practice effects in an alphabet arithmetic task. Subjects were asked to verify equations of the form $A + 2 = C$, $B + 3 = E$, $C + 4 = G$, and $D + 5 = I$. Subjects typically reported that they performed the task by counting through the alphabet one letter at a time until the number of counts equaled the digit addend, and then comparing the current letter with the presented answer. For example, $E + 5 = K$ would involve counting five steps through the alphabet (*F, G, H, I, J*), comparing the *J* with the given answer, *K*, and responding *false*. Consistent with subjects' reports, the time to verify alphabet arithmetic equations increases linearly with the digit addend (i.e., with the number of counts). The slope of the function is typically 400–500 ms per count, with an intercept of 1,000 ms (Logan & Klapp, 1988).

This experiment was intended to provide a test of the instance theory that was very different from the lexical decision tasks reported earlier. Reaction times in alphabet arithmetic are nearly an order of magnitude longer than reaction times in lexical decision, and the practice effects extend over several sessions. It was also intended to mimic children's acquisition of addition, which involves a transition from a serial counting algorithm to memory retrieval (e.g., Ashcraft, 1982; Siegler, 1987; Zbrodoff, 1979). Presumably, with enough practice, adult subjects would learn to perform the alphabet arithmetic task by memory retrieval instead of counting.

Subjects were trained on 10 letters from one half of the alphabet (either *A* through *J*, or *K* through *T*). Each letter appeared with four different addends (*2, 3, 4,* and *5*) in a true equation and in a false equation. False equations were true plus one (e.g., $A + 2 = D$) or true minus one (e.g., $A + 2 = B$); the kind of false equation varied between subjects. Thus, each subject experienced a total of 80 different problems during training (10 letters × 4 digit addends × true vs. false). Each problem was presented 72 times; 6 times per session for 12 sessions. There were 480 trials per session. Further details of the procedure can be found in Logan and Klapp (1988).

The means and standard deviations of the reaction times are plotted in Figures 9 and 10. Figure 9 contains data from *true* responses, and Figure 10 contains data from *false* responses. The data are plotted in logarithmic coordinates so that the power functions (solid lines) fitted to the data (points) will appear as straight lines. The instance theory predicts that the line for mean reaction times should be parallel to the line for the standard deviations. The slope of the line is the exponent of the power function in linear coordinates, and the theory predicts that means and standard deviations will have the same exponent. The functions fitted to the data in Figures 9 and 10 were constrained to have the same exponent for the mean and standard deviation, so the lines are parallel. The question is whether the points depart systematically from the parallel lines. Parameters of the fitted power functions and measures of goodness of fit are presented in Table 3.

In evaluating the fits in the figures, it is important to note that

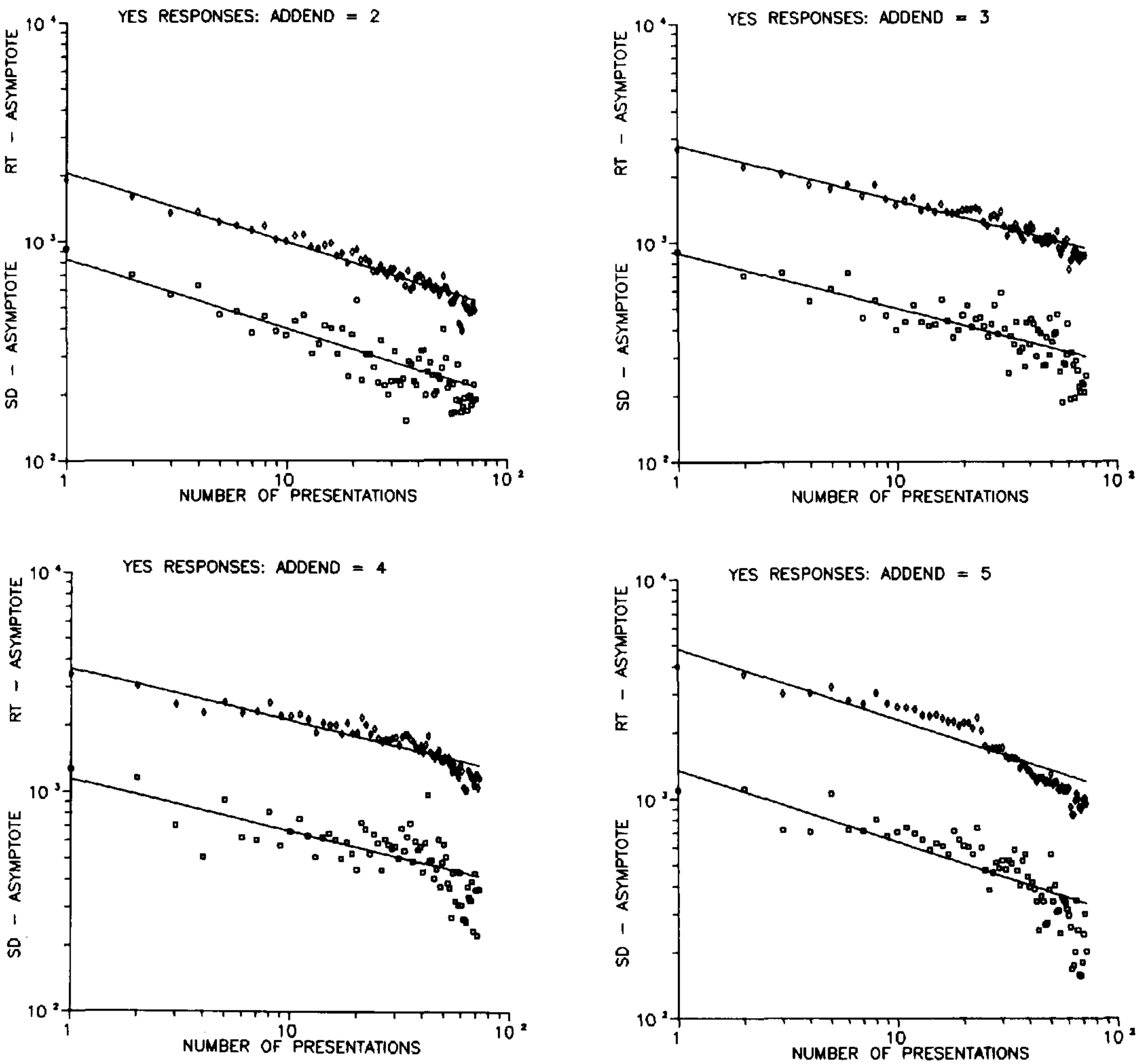

Figure 9. Means and standard deviations of reaction times to verify true alphabet arithmetic equations as a function of the number of presentations and the magnitude of the digit addend. (Points represent observed data; lines represent fitted power functions constrained to have the same exponent for means and standard deviations, parallel lines in log–log plots.)

the log scale introduces systematic distortions. Differences are exaggerated at low values of the variable and compressed at high values. Consequently, low values will appear to fit less well than high values. The standard deviations, which range from 162 ms to 1,408 ms, will appear to fit less well than the means, which range from 816 ms to 4,285 ms. Objectively, the means were fitted better than the standard deviations, but the difference was not as large as it appears in the figures.

The power functions fit the data reasonably well for addends of *2, 3,* and *4.* The functions for means and standard deviations appear parallel, as they should in log–log plots if they both decrease as power functions with the same exponent. Indeed, the constrained fits were as good as separate fits for all addends (see Table 3). Thus, the predictions of the instance theory are confirmed in the arithmetic task as well as in the previous lexical decision tasks.

The fits were not perfect, however. The power functions for addends of *2, 3,* and *4* tended to overestimate the last few pre-

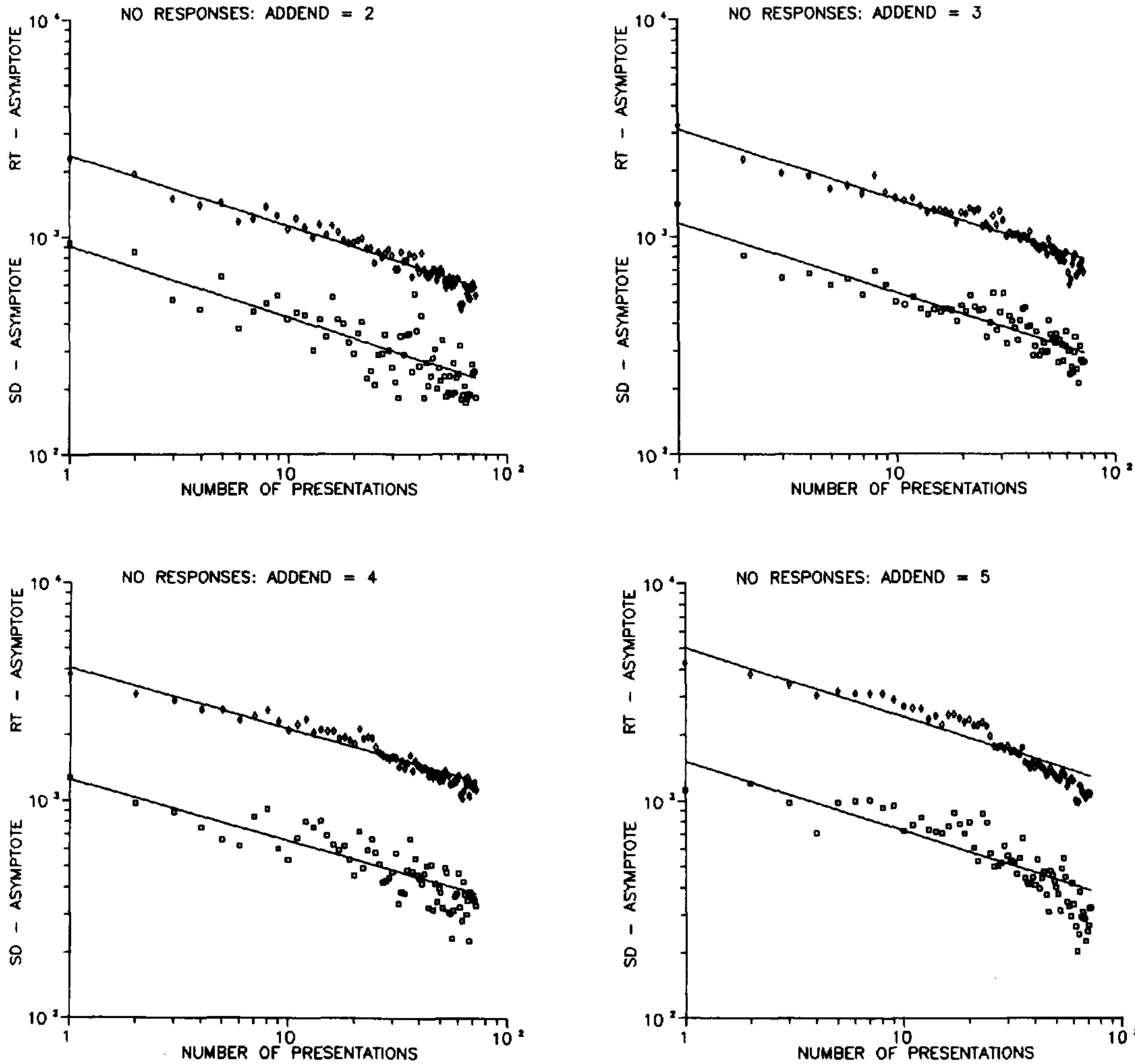

Figure 10. Means and standard deviations of reaction times to verify false alphabet arithmetic equations as a function of the number of presentations and the magnitude of the digit addend. (Points represent observed data; lines represent fitted power functions constrained to have the same exponent for means and standard deviations, parallel lines in log–log plots.)

sentations. The deviation was much worse for addend = 5 equations, where a discontinuity appeared in the learning curves at about 24 presentations; learning seems faster after the discontinuity than before. This discontinuity, which is inconsistent with the instance theory, cannot be accounted for easily by any current theory of skill acquisition. Most theories of skill acquisition predict a power-function learning curve, which is continuous. Thus, the evidence against the instance theory is also evidence against its competitors.

The discontinuity reflects a strategy shift reported by many of the subjects. Several subjects reported deliberately learning the 5s because they were the most difficult problems. Typically, subjects developed mnemonics for the 5s, which allowed them to respond on the basis of memory. For example, when I tried the experiment as a pilot subject, I used psychologists names as mnemonics: I remembered $A + 5 = F$ as A. F. Sanders, $E + 5 = J$ as E. J. Gibson, and $G + 5 = L$ as Gordon Logan. These mnemonics provided anchor

Table 3
Parameter Estimates From Constrained and Separate Fits of Power Functions to Means and Standard Deviations of Reaction Times in the Alphabet Arithmetic Task of Experiment 5

	True equations: Addend				False equations: Addend			
Parameter	2	3	4	5	2	3	4	5
ART	384	195	0	0	357	362	29	0
BRT	2052	2770	3643	4816	2371	3110	4065	5037
C	0.317	0.255	0.238	0.327	0.327	0.324	0.282	0.319
ASD	0	0	0	75	0	0	0	10
BSD	828	893	1144	1346	903	1158	1249	1513
r^2	0.968	0.960	0.908	0.867	0.960	0.960	0.949	0.895
rmsd	59.4	82.0	148.9	206.3	72.3	87.4	116.1	193.9
ART	133	30	5	0	0	98	97	0
BRT	2246	2892	3679	4834	2630	3307	3976	5028
CRT	0.251	0.228	0.241	0.328	0.245	0.274	0.269	0.319
ASD	30	106	112	76	76	0	0	0
BSD	889	817	1099	1334	889	1088	1274	1470
CSD	0.383	0.332	0.303	0.321	0.416	0.547	0.290	0.301
r^2	0.974	0.960	0.908	0.869	0.962	0.964	0.949	0.896
rmsd	54.9	80.9	149.2	206.3	69.9	84.7	115.3	193.5

Note. ART = asymptote for mean reaction time; BRT = multiplicative constant for mean reaction time; ASD = asymptote for standard deviation; BSD = multiplicative constant for standard deviation; C = exponent fitted to means and standard deviations simultaneously; CRT = exponent fitted to means separately; CSD = exponent fitted to standard deviations separately; r^2 = squared correlation between observed and predicted values; *rmsd* = root mean squared deviation from prediction.

points for subjects, which made it easier for them to learn the *4*s.

One might model the addend = *5* data with two learning curves, one reflecting the inefficient mnemonic strategy from Trials 1-24 and another reflecting the more efficient strategy from Trials 25-72. I fitted separate power functions to the data before and after the 24th trial. As before, means and standard deviations were fitted simultaneously with the constraint that they have the same exponent. The fits were much better than the previous ones. For *true* responses, *rmsd* decreased from 206 ms to 90 ms; for *false* responses, it decreased from 194 ms to 88 ms.

Thus, the addend = *5* data may be inconsistent with a strict interpretation of the instance theory, in which neither the algorithm nor the memory process changes over practice, but they are consistent with a more general view of instance theory, which interprets automaticity as a transition from an algorithm to a memory process. There seem to have been two transitions in the addend = *5* condition, one to an inefficient memory process and another to a more efficient one, reminiscent of the plateaus described by Bryan and Harter (1899).

One final piece of evidence deserves comment. Logan and Klapp (1988) reported data from subsequent sessions of the experiment that bear on the instance theory. In particular, they ran a 13th session in which subjects were tested on the untrained half of the alphabet (e.g., subjects trained on A–J were tested on K–T, and vice versa). Reaction times increased substantially, compared with the previous session, and the slope of the function relating reaction time to the magnitude of the addend, which had been close to zero on the previous session, increased to about 400 ms/count. This suggests that subjects learned specific responses to specific stimuli during training, which is consistent with instance- or item-based learning and inconsistent with process-based learning.

To summarize, the predictions of the instance theory were supported in the alphabet arithmetic task as they had been in the lexical decision task of Experiments 1-3. Specifically, the training data showed that the means and standard deviations of reaction times decreased as power functions of practice with the same exponent (appearing parallel in log–log plots) and the transfer data showed that learning was item-based rather than process-based.

Evidence of Separate Instances

The fits in the previous experiments provide evidence that is consistent with the instance theory, but they do not uniquely support it. Most theories of skill acquisition predict a power-function reduction in the mean (e.g., Anderson, 1982; Crossman, 1959; MacKay, 1982; Newell & Rosenbloom, 1981), so the fits to the means are not unique. The fits to the standard deviations, constrained to have the same exponent as the means, were predicted by the instance theory and no other. But the fits do not disconfirm predictions of the other theories; the other theories simply made no prediction. Furthermore, the evidence that automaticity is specific to the stimuli experienced during training may rule out process-based theories of automatization, but it does not uniquely support instance theory. Strength theories, which assume that practice strengthens connections between generic stimuli and generic responses, can also predict item-based learning (e.g., Anderson, 1982; MacKay, 1982; Schneider, 1985).

The instance theory can be distinguished from strength theories by demonstrating the existence of separate memory representations of each encounter with a stimulus (see, e.g., Hintzman, 1976). According to the instance theory, each prior episode is represented in memory, whereas strength theories represent past history by a single strength measure, casting aside the details of individual episodes. The difference between these representational assumptions can be seen clearly in a *varied-mapping* design, in which subjects give different responses or different interpretations on different exposures to the same stimulus. Varied mapping typically produces little or no evidence of learning (e.g., Schneider & Shiffrin, 1977).

According to the instance theory, each exposure would result in a separate trace. At retrieval time, there are two possibilities: All of the separate traces could be retrieved, but they would yield inconsistent or incoherent information about how the current stimulus should be interpreted, so the subject should ignore them and respond on the basis of the algorithm. Alternatively, the current instructional set and the current stimulus could act as a compound retrieval cue and retrieve only those episodes that were encoded in the context of the current instructions. In the first case, there would be no evidence of learning; in the second case, there would be less learning than would be observed in consistent-mapping control conditions.

By contrast, in some versions of strength theory (e.g., Schneider, 1985), the different interpretations in varied mapping cancel one another out, resulting in no net gain in strength, and thus, no evidence of learning. Subjects would have no choice but to rely on the algorithm. Other versions of strength theory might be constructed in which strength accrued separately to each interpretation. As with instance theory, the retrieved interpretations would conflict with each other, so the subjects should ignore memory and rely on the algorithm.

The instance theory can be distinguished from the first version of strength theory by transferring subjects to a *frequency-judgment* task, just as instance theories of memory were distinguished from strength theories by frequency judgment tasks (e.g., Hintzman, 1976). Instance theory predicts that subjects trained under varied interpretation conditions should be just as sensitive to the frequency with which individual stimuli were presented as subjects trained under consistent interpretation conditions, because both groups of subjects would encode representations of each encounter with a stimulus. By contrast, strength theory would predict that subjects trained under varied interpretation conditions would be less sensitive to frequency information than subjects trained under consistent interpretation conditions because there is no separate episodic trace representing each encounter with the stimulus.

In other words, instance theory predicts a dissociation between repetition effects and frequency judgments as measures of memory after varied-interpretation training, whereas certain strength theories predict no such dissociation.

Experiment 5

To test the dissociation, four groups of subjects were trained in the paradigm from Experiment 3 (i.e., some stimuli were presented in 1 block, others in 2, 4, 8, or 16 consecutive blocks), and then transferred to a frequency judgment task. Two groups were trained under consistent interpretation conditions, and two groups were trained under varied interpretation conditions. To manipulate the consistency of interpretation, subjects were shown three kinds of stimuli: words, pronouncible nonwords, and unpronouncible nonwords. Subjects could interpret these stimuli under lexical decision instructions, distinguishing between words and nonwords, or under pronunciation decision instructions, distinguishing between pronouncible and unpronouncible letter strings. Logan (1988) showed that alternating between these interpretations over successive presentations impaired learning, relative to consistent-interpretation controls.

One consistent-interpretation group performed lexical decisions on each training trial, and the other performed pronunciation decisions. The two varied-interpretation groups alternated between lexical decisions and pronunciation decisions over successive presentations of the stimuli. One group began with lexical decisions, and the other began with pronunciation decisions. In the transfer task, all four groups saw new stimuli as well as the stimuli they were trained on, and were asked to estimate the frequency with which each stimulus had appeared during training. Further details of the method can be found in Appendix C.

Training results. The training data are presented in Figure 11 as *benefit scores.* Benefit was calculated by subtracting reaction time for second and subsequent presentations of a stimulus from reaction time for the first presentation of a stimulus, in order to remove differences due to the initial algorithm and facilitate comparisons of learning effects between conditions. Benefit scores for consistent-interpretation lexical decision subjects and for varied-interpretation subjects who began with lexical decision are presented in the top panel of Figure 11. Benefit scores for consistent-interpretation pronunciation subjects and for varied-interpretation subjects who began with pronunciation decisions are presented in the bottom panel of Figure 11. The error rates are presented in Appendix C.

The lexical decision subjects showed more benefit for consistent interpretations than for varied interpretations. The difference was largest for pronouncible nonwords, intermediate for words, and smallest for unpronouncible nonwords. The pronunciation subjects showed less clear-cut results: Consistent interpretation produced an advantage over varied interpretation only for pronouncible nonwords, and even that difference diminished after 10 or 11 presentations. These conclusions were confirmed by ANOVA, reported in Appendix C.

The results of the lexical decision group are sufficient for the present purpose, which is to determine whether there is a dissociation between repetition effects and frequency judgments as measures of memory following training with varied interpretation. If there is a dissociation, then frequency judgments following lexical decisions ought to be just as accurate for varied-interpretation subjects as for consistent-interpretation subjects; if there is no dissociation, then frequency judgments should be less accurate for varied-interpretation subjects than for consistent-interpretation subjects.

Transfer results. The average frequency estimates for each group are presented in Figure 12. The left-hand panels present the data from consistent- and varied-interpretation lexical deci-

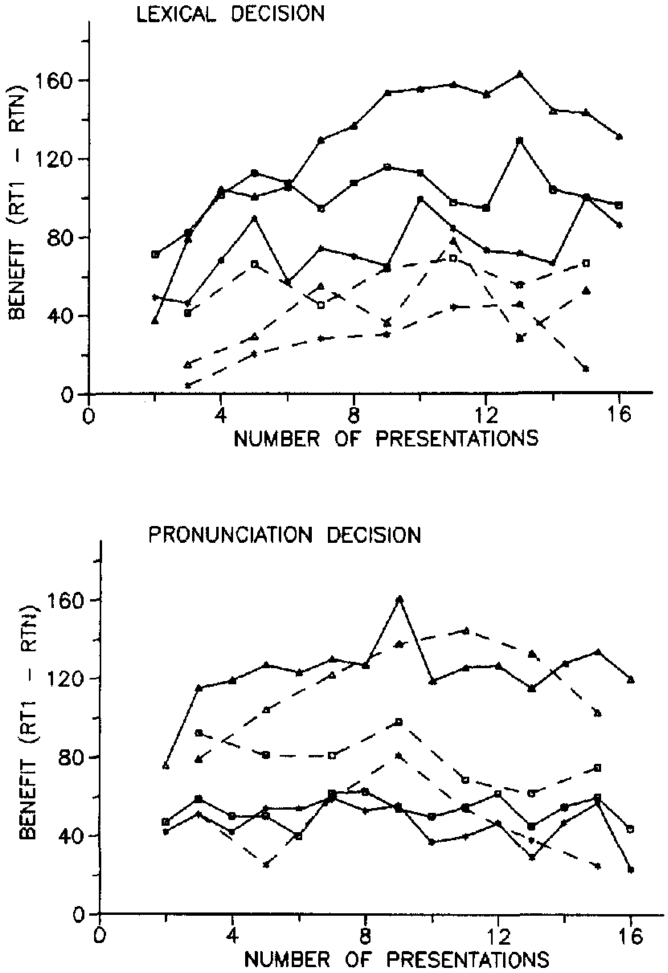

Figure 11. Benefit scores as a function of the number of presentations for consistent (solid lines) and varied (broken lines) interpretations of words (boxes), pronouncible nonwords (triangles), and unpronouncible nonwords (stars) in the lexical decision task (top panel) and pronunciation task (bottom panel) of Experiment 5.

sion subjects; the right-hand panels present the data from consistent- and varied-interpretation pronunciation subjects. The top panels present frequency estimates for words, the middle panels present estimates for pronouncible nonwords, and the bottom panels present estimates for unpronouncible nonwords.

Frequency estimates from subjects trained with varied interpretations were very close to the estimates from subjects trained with consistent interpretations. If anything, they were a little more accurate. Thus, in the lexical decision subjects, at least, there was a dissociation between frequency estimates and repetition effects as measures of memory. These conclusions were confirmed by ANOVAs, reported in detail in Appendix C.

Conclusions

The results from the lexical decision subjects showed a dissociation between repetition effects and frequency judgments as measures of memory: Consistency of stimulus-to-interpretation mapping had strong effects on lexical decision performance, but had negligible effects on frequency judgments. This dissociation was predicted by the instance theory, on the assumption that subjects encode separate representations of each encounter with each stimulus. It would not be predicted in a straightforward manner by strength theories in which strengthening one interpretation weakens the other, resulting in no net gain in strength (e.g., Schneider, 1985).

It is possible, however, to have a theory in which multiple exposures to a stimulus are represented both as a strength value and as separate episodic traces. That sort of theory could account for the present results, with the strength values underlying the repetition effects and the separate traces underlying the frequency judgments. However, the theoretical development of the instance theory showed that separate traces can produce repetition effects, so it would not be clear whether the strength values or the separate traces were responsible for the repetition effects. A more parsimonious interpretation would be that there are only separate traces in memory, which produce both the repetition effect and the frequency judgments. The onus would be on the strength theorists to show that strength values did in fact underlie repetition effects.

General Discussion

The experiments tested the instance theory in three ways: First, they tested the power-law predictions and found that, as predicted, means and standard deviations of reaction times decreased as power functions of the number of trials with the same exponent (Experiments 1–4). Second, the experiments determined whether learning was item-based, as the instance theory predicts, or process-based, as the modal view predicts. The data supported the instance theory: Subjects learned specific responses to specific stimuli and did not transfer well to new stimuli (Experiments 1–4) or to new approaches to old stimuli (Experiment 5). Third, Experiment 5 tested the instance theory against certain strength theories, asking whether subjects retain separate representations of each stimulus even if they don't use them to support performance. The data showed evidence of separate representations, as the instance theory predicts.

The success of the theory in these three tests suggests that it should be taken seriously in studies of skill acquisition and automaticity. The remainder of this section discusses implications of the instance theory for current issues and controversies.

Instance Theory and the Properties of Automaticity

The instance theory provides a new perspective on many of the qualitative properties that distinguish automatic and nonautomatic processing. In some cases, the properties derive from the assumptions about representation and retrieval from memory. In other cases, the differences occur because automatic and nonautomatic processing are based on different processes. Factors that affect the initial algorithm need not affect the memory retrieval process, and vice versa. In the remainder of this section, these principles are applied to several of properties of automaticity.

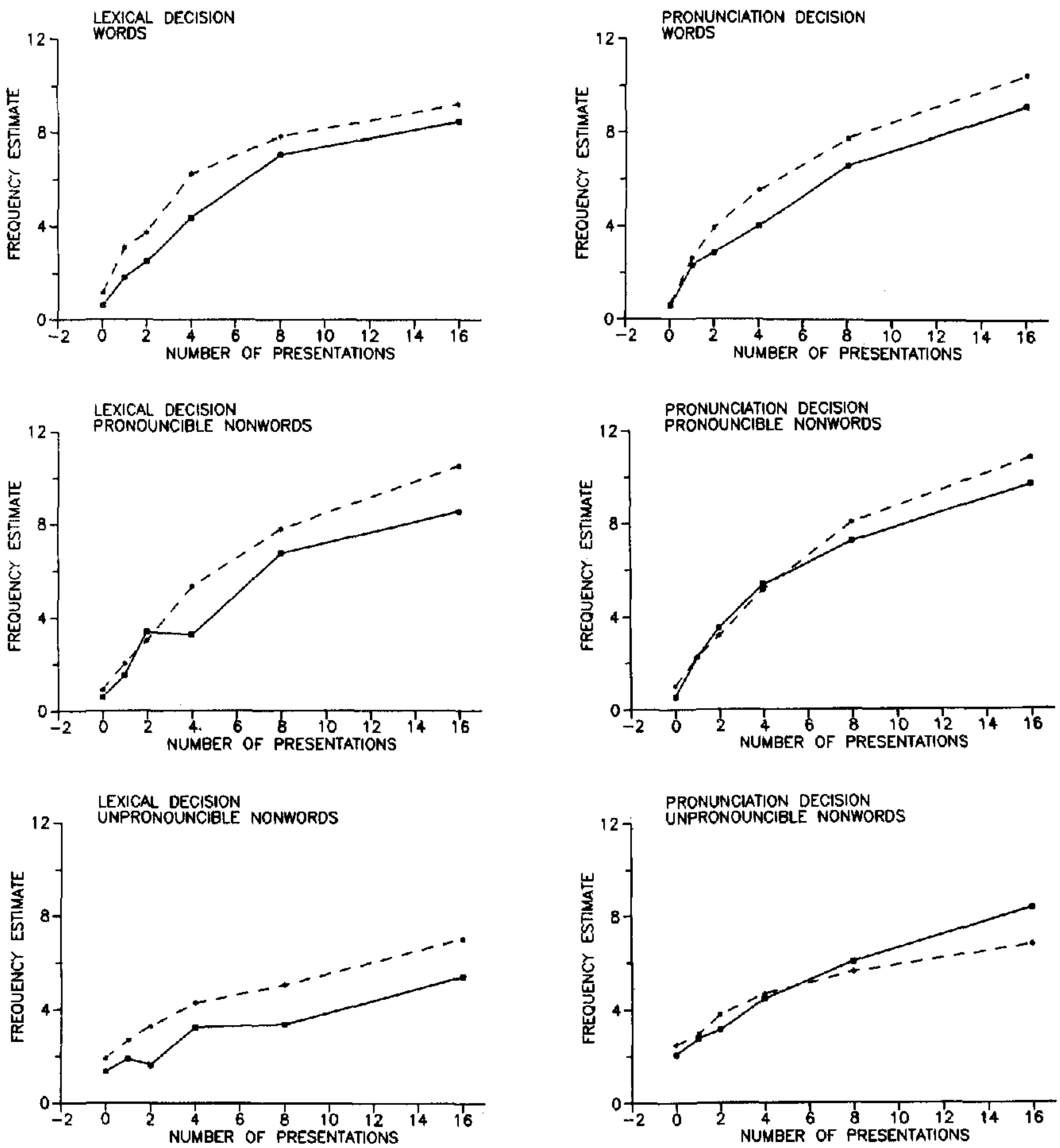

Figure 12. Estimated frequencies of occurrence as a function of actual frequencies of occurrency for consistent (solid lines) and inconsistent (broken lines) interpretations of words (top panels), pronouncible nonwords (middle panels), and unpronouncible nonwords (bottom panels) after training on the lexical decision task (left panels) and the pronunciation task (right panels) of Experiment 5.

Autonomy

Phenomenal experience and experimental data suggest that automatic processing is autonomous, in that it can begin and run on to completion without intention (for reviews, see Kahneman & Treisman, 1984; Logan, 1985b; Zbrodoff & Logan, 1986). Instance theory accounts for the autonomy by assuming that memory retrieval is obligatory; attention to a stimulus is sufficient to cause the retrieval of all of the information associated with the stimulus, whether or not the subject intends to retrieve it.

The major evidence for the autonomy of automatic processing comes from Stroop and priming studies, in which an irrelevant stimulus influences the processing of a relevant stimulus (but see Zbrodoff & Logan, 1986). For example, subjects are slower to name colors when the colors form words that represent irrelevant color names (BLUE in red ink; Stroop, 1935), and subjects make lexical decisions faster if the target word (e.g., DOCTOR) is preceded by a related word (e.g., NURSE) than if it is preceded by an unrelated word (e.g., BUTTER; Meyer & Schvaneveldt, 1971). The modal interpretation of such effects is that the irrelevant stimulus activates its memory representation and that activation speeds or slows responses to related stimuli—in other words, the irrelevant stimulus retrieves response-related information from memory. The modal interpretation also assumes that activation is not only obligatory but also is independent of intention and attention; it occurs whether or not the subject attends to the stimulus or intends to process it (e.g., Logan, 1980; Neely, 1977; Posner & Snyder, 1975).

Two recent findings are difficult to reconcile with the modal view. First, Francolini and Egeth (1980) and Kahneman and Henik (1981) showed that Stroop interference was stronger when subjects attended to the irrelevant stimulus. Kahneman and Henik showed subjects two words (e.g., MOST and BLUE), one of which was colored (e.g., red) and one of which was black. The task was to name the color of the colored word (in this case, red). The irrelevant color word (BLUE) interfered with color naming much more when it was colored, and hence was attended, than when it was black, and hence unattended. Francolini and Egeth (1980) found similar results in a numerical Stroop task in which subjects counted red letters or digits and ignored black ones; irrelevant digits interfered more when they were red, hence attended, than when they were black, and hence ignored.

Second, M. Smith (1979; M. Smith, Theodor, & Franklin, 1983) and Henik, Friedrich, and Kellogg (1983) showed that Stroop and priming effects can be completely eliminated by manipulating the way in which the subject processes the irrelevant stimulus. M. Smith and her colleagues showed that priming effects could be eliminated if subjects treated the priming stimulus as a letter string and not as a word (i.e., by searching for a letter within it vs. making a lexical decision). Henik et al. showed that Stroop effects could be eliminated in the same way.

Both sets of results are difficult to deal with in the modal interpretation of Stroop and priming effects because an irrelevant stimulus is supposed to activate its memory representation whether or not the subject attends to it or intends to process it. Instance theory can account for the attention effects (i.e., Francolini & Egeth, 1980; Kahneman & Henik, 1981) with its assumption that attention is sufficient to cause associated information to be retrieved from memory. Attention may not be necessary, but it is sufficient, which means that (a) information at the focus of attention may be more strongly activated than information outside the focus, and therefore may provide stronger retrieval cues, and (b) information at the focus will be more likely to activate its memory representation than information outside the focus.

Instance theory would interpret the intention effects (i.e., Henik et al., 1983; M. Smith, 1979; M. Smith et al., 1983) as retrieval effects: The instruction to process the irrelevant stimulus in a different way leads subjects to use different retrieval cues in conjunction with the stimulus, to retrieve the required information from memory. The effects on subsequent processing will depend on what was retrieved and how it is related to the subsequent task. If subjects make a lexical decision about the prime, then information about the meaning of the prime will be retrieved and available to influence a subsequent lexical decision. But if subjects search for a letter in the prime, information about letters will be retrieved, which may not affect a subsequent lexical decision.

Stroop and priming studies are interesting because they focus on a different sense of automaticity than the practice studies that were addressed by the instance theory. Stroop and priming studies are concerned with the activation of memory representations, whereas practice studies are concerned with the use of activated memory representations in the performance of tasks. Stroop and priming studies consider a stimulus to be processed automatically if it retrieves *anything* from memory, whereas practice studies consider a stimulus to be processed automatically only if it retrieves *something useful* from memory. The different senses are apparent in interpreting control conditions in the different paradigms: In Stroop tasks, neutral (noncolor) words are assumed to activate their memory representations automatically even though the activation has no effect on color naming. But in practice studies, varied-mapping control stimuli are not considered to be processed automatically, even though they may retrieve information from memory (cf. Experiment 5).

It is tempting to think of Stroop-type automaticity as one component of the automaticity addressed in practice studies because memory retrieval is the first step in performing a task automatically. One could imagine a progression from one type of automaticity to the other, as a small number of stimulus exposures may cause enough memory activation to produce Stroop and priming effects but not enough to provide a reliable basis for performing the task. However, Stroop and priming effects do not provide a pure measure of memory retrieval. Memory retrieval affects performance by interacting with a subsequent decision process, just as it does in practice studies (e.g., Logan, 1980). The intentional and attentional effects described earlier suggest that the appearance of Stroop and priming effects depends on the relation between retrieval cues and decision processes, just as practice effects do. Thus, one need not expect Stroop and priming effects to appear sooner than practice effects. It may be possible to relate the two senses of automaticity theoretically, but that does not mean that their empirical manifestations will be related in a straightforward manner.

Control

Phenomenal experience suggests that automatic processing is closely controlled. Behavior on "automatic pilot" is usually coherent and goal-directed (Reason & Myceilska, 1982); skilled practitioners are better than novices even though they perform automatically (Logan, 1985b). Experimental data also suggest that automatic processing is closely controlled. Skilled activities such as speaking and typing can be inhibited quickly in response to an error or a signal to stop, often within a syllable or a keystroke after the error or stop signal (Ladefoged, Silverstein, & Papcun, 1973; Levelt, 1983; Long, 1976; Logan, 1982; Rabbitt, 1978). Thoughts may be harder to inhibit than overt actions (Logan, 1983, 1985a), but even they can be controlled by deliberately thinking of other things (Wenger, Schneider, Carter, & White, 1987).

By contrast, the modal view is that automatic processing is not well controlled. Shiffrin and Schneider (e.g., 1977) explicitly distinguish between automatic and *controlled* processing, and the idea is implicit in other approaches (e.g., LaBerge & Samuels, 1974; Posner & Snyder, 1975). The main motivation for the modal view on control is the evidence that automatic processes are autonomous. But autonomy does not imply lack of control. A process is autonomous if it can begin and run on to completion without intention (Zbrodoff & Logan, 1986); that does not mean it cannot be initiated and guided to completion intentionally.

According to the instance theory, automatic processes are used intentionally. Automaticity exploits the autonomy of the retrieval process, harnessing it so that it provides information that is relevant the person's goals. Thus, automatic processes are controlled. The retrieval process can be controlled by manipulating retrieval cues or stimulus input, or both, and the subsequent decision process can be inhibited before it results in an overt response. Automatic processing may be a little harder to control than algorithm-based processing, but only because it tends to be faster and allows less time for an act of control to take effect. It may be controlled differently, but it is controlled nonetheless (for related views, see Logan, 1985b; Neumann, 1984).

Effortlessness

Phenomenal experience and experimental data also suggest that automatic processing is effortless (see Jonides, 1981; Logan, 1978, 1979; Posner & Snyder, 1975; Shiffrin & Schneider, 1977). Typical treatments of automaticity assume that processing capacity or cognitive resources are severely limited and that automatization is a way around the capacity limitations. By contrast, the instance theory does not assume any capacity limitations; novice performance is limited by a lack of knowledge rather than by scarcity of processing capacity or resources. Subjects perform slowly at first because they have no other way to perform the task. As their knowledge base builds up with practice, they have the choice of responding on the basis of the algorithm or on the basis of memory retrieval. Presumably, they do not switch to memory retrieval until it is faster than the algorithm and at least as reliable.

The major experimental evidence for effortless automatic processing comes from dual-task experiments that show that practiced subjects suffer less interference from a concurrent task than do unpracticed subjects (e.g., Bahrick & Shelley, 1958; Logan, 1978, 1979). The usual interpretation is that the practiced task requires fewer resources than the unpracticed task. That may be the case, but instance theory would suggest that the reduction in dual-task interference occurs because of the shift from the algorithm to memory retrieval: Because memory retrieval and the algorithm are different processes, tasks that interfere with the algorithm will not necessarily interfere with memory retrieval. The reduction in interference may occur only because experimenters choose concurrent tasks that interfere with the initial algorithm; if the concurrent task does not interfere with the algorithm, another concurrent task is chosen. So the reduction in interference may be an artifact of the experimenter's choice of procedures rather than a general reduction in resource demands. Indeed, instance theory suggests that it may be possible to find an *increase* in dual-task interference with practice, in cases in which the concurrent task interferes with memory retrieval but not with the initial algorithm. Thus, instance theory predicts a *shift* in dual-task interference rather than a reduction in dual-task interference (also see Logan, 1985b).

Alternatively, instance theory would predict a reduction in dual-task interference with automatization because automatization provides subjects with more ways to do the task. Initially, they have only one way to perform the task—following the instructed algorithm—so a concurrent task that interferes with the algorithm must affect their performance. After automatization, however, they can use the algorithm or rely on memory retrieval. If the concurrent task interferes with the algorithm, they can use memory retrieval; if it interferes with memory retrieval, they can use the algorithm. In either case, their performance need not suffer. The general point is that automatic performance can be more flexible than nonautomatic performance, providing the subject with ways to avoid dual-task interference (also see MacKay, 1982).

Another major line of evidence that automatic processing is effortless comes from search studies that manipulate the number of items in the memory set, the number of items in the display, or both. Initially, reaction time increases linearly with the number of memory set or display items, but after considerable practice, the slope approaches zero (for a review, see Schneider & Shiffrin, 1977). The initial linear increase is interpreted as evidence that unpracticed search is effortful—search is assumed to be serial in order to minimize the momentary load on capacity—and the zero slope at the end of practice is interpreted as evidence that search has become effortless or capacity free.

There are many criticisms of this interpretation (e.g., Cheng, 1985; Ryan, 1983), but the basic findings can be handled easily by instance theory: Initially, performance depends on a search algorithm in which subjects try to find the probe item in the memory set or the memory item in the display. Several different algorithms could be used for the task, most of which would produce the linear increase in reaction time with memory set size or display size (see Townsend & Ashby, 1983). After practice, subjects retrieve the appropriate response directly from memory, without searching, when given a probe or a multi-item display as a retrieval cue (cf. Ryan, 1983; Schneider, 1985). This scheme provides a natural account of memory search; whether it can work for visual search is not immediately clear (i.e., visual search may train an automatic attention response, which is not part of the instance theory; Shiffrin & Schneider, 1977). The principle here is the same as in dual-task studies: Factors that affect the algorithm, such as display size or memory set size, do not necessarily affect memory retrieval.

Instance theory accounts for the phenomenal experience of effortlessness during automatic performance by suggesting that memory retrieval is often easier than the algorithm. Indeed, subjects would not be expected to switch from the algorithm to memory retrieval until memory retrieval was quick and effortless.

Unconsciousness

The evidence that automatic processes are unconscious is primarily phenomenal—we cannot easily introspect on the things we do automatically. Traditional views have some difficulty dealing with this aspect of automaticity because consciousness is not an easy concept to express in the information-processing language that dominates current theorizing (e.g., Carr et al., 1982; Marcel, 1983; Posner & Snyder, 1975). I suspect that the attribution of unconsciousness to automatic processing stems from the belief that automatic processing is processing without thinking. In everyday language, for example, we say that we solved a problem automatically when the solution springs to mind without much thought or when it is the first thing that occurs to us. Instance theory can capture this intuition by specifying precisely what it means to think about a solution (i.e., to compute a solution by applying an algorithm) and what it means to find a solution without thinking (i.e., to retrieve a solution from memory), and both of them can be expressed easily in information-processing language.

Another reason for believing that automatic processing may be unconscious is that algorithms may involve a series of steps or stages on the way to a solution, each of which may be introspected upon, whereas memory retrieval is a single-step process (e.g., the resonance metaphor of Hintzman, 1986, and Ratcliff, 1978, or the holographic retrieval process described by Eich, 1982, and Murdock, 1982, 1983). Thus, automatic processes may not be available to the "mind's eye" long enough to provide conscious evidence of their inner workings.

Poor Memory

Phenomenal experience suggests that automatic processing results in poor memory for what was processed. When distracted, we may start up from a stop light, shift gears, and attain cruising speed on "automatic pilot," and then find we have no memory of having done so (also see Reason, 1984). There have not been many experimental tests of memory for things processed automatically, but the few that exist support the poor-memory hypothesis.

For example, Kolers (1975) investigated acquisition of skill at reading spatially transformed text (rof elpmaxe) and found that memory for transformed texts was better than memory for normal texts. He attributed the difference to the automaticity of normal reading. He showed as well that memory for transformed text declined as subjects acquired skill. Thus, his results suggested that subjects can remember stimuli they process automatically, although not as well as stimuli they processed algorithmically.

Fisk and Schneider (1984) had subjects search for exemplars of target categories in a series of words presented one after the other. One experiment examined memory early in practice, during controlled processing; the other examined memory following automatic processing. The results were clear: Subjects remembered much more accurately in Experiment 1 than in Experiment 2. In fact, there was little evidence of memory for some stimuli (novel distractors) in Experiment 2, which led Fisk and Schneider (1984) to conclude that automatic processing left no traces in memory.[7] However, automaticity was confounded with dual-task conditions: Experiment 1 used only single-task conditions, and Experiment 2 used only dual-task conditions (i.e., the category search task performed concurrently with a difficult digit search task). Thus, either automaticity or the dual-task conditions (or both) could have impaired memory in Experiment 2. One cannot tell from their experiments which was the important factor. The literature suggests that dual-task conditions may have been responsible: Recent evidence indicates that dual-task conditions at encoding severely impair memory (Naveh-Benjamin & Jonides, 1986). Nissen and Bullemer (1987) presented data suggesting that subjects cannot remember stimuli processed in dual-task conditions. And Kolers's (1975) data suggested that subjects can remember stimuli processed automatically. They may not remember well, but they do remember.

Typically, poor memory for stimuli processed automatically is interpreted as an encoding deficit—encoding is either cursory or not done at all. By contrast, the instance theory interprets it as a retrieval effect. The theory assumes that events are encoded in the same way on each trial, whether it be the 10th or the 10,000th. In each case, a separate trace is created to represent the event. However, the traces may have a different impact on retrieval, depending on how many other traces there are in memory and depending on the retrieval task. The impact of trace $i + 1$ relative to trace i will decrease as i increases, following a kind of Weber function. Adding one trace to zero makes more of a difference than adding one trace to 10 or one trace to 1,000.

The nature of the retrieval task is also important. Some retrieval tasks, like recognition and recall, require subjects to identify one trace out of many. Other retrieval tasks, like the kind used in studies of automaticity and categorization (e.g., Hintzman, 1986; Medin & Schaffer, 1978), allow subjects to respond to the whole set of retrieved traces without focusing on any one in particular. In the former, *competitive* retrieval tasks, the other traces can interfere with retrieval of the one desired trace, for example, by adding noise to the decision process. The task is like finding a particular tree in a forest; the more dense the forest, the harder it is to find. Performance should get worse as the task becomes automatic and more traces are added to memory. By contrast, in the latter, *cooperative* retrieval tasks, the different traces serve the same purpose, working together for a common goal. The task is like finding a forest; the more trees there are, the easier it is to find. Performance should get better as the task becomes more automatic and more traces are added to memory.

The distinction between competitive and cooperative retrieval tasks is well illustrated in the *frequency paradox* in recognition versus lexical decision:[8] Low-frequency words are recognized better than high-frequency words, but low-frequency words produce slower lexical decisions than high-frequency

[7] The poor memory for novel distractors may be a result of shallow processing and incongruity rather than automaticity. The targets were very familiar and the distractors were new, so a simple familiarity judgment would allow accurate performance. Subjects may not interpret the distractors beyond deciding they are unfamiliar, and that would not produce good memory. Also, memory is typically worse for *no* items than for *yes* items, possibly because *no* items are not congruent (Craik & Tulving, 1975).

[8] I would like to thank Tom Carr for providing this example.

words. This paradox occurs because the tasks require different retrieval processes: Recognition requires subjects to discriminate one particular exposure to a word from all other exposures to the word. High-frequency words have more exposures than low-frequency words, so a single exposure is harder to discriminate. By contrast, lexical decision requires subjects to discriminate *any* exposure to the word from *no* exposure. High-frequency words have more exposures, so more of them are likely to be retrieved, making the discrimination easier.

Instance Theory and Issues in Automaticity

The instance theory provides new perspectives on the diagnosis of automaticity and the relation between automaticity and categorization. The perspectives are described in this section.

Diagnosing Automaticity

Much of the interest in automaticity comes from studies that attempt to determine whether specific processes, such as letter encoding (Posner & Boies, 1971), lexical access (Becker, 1976), and abstraction of superordinate categories (Barsalou & Ross, 1986), are automatic. The processes are often acquired through experience outside the laboratory, so the assessment of automaticity is an absolute judgment; the process is either automatic or not automatic. Researchers often assess the properties of automaticity and consider a process to be automatic if it possesses some or all of the properties.

However, the diagnosis of automaticity is fraught with problems: Different researchers use different lists of properties and disagree on the necessity and sufficiency of the properties they list. For example, Posner and Snyder (1975) list only 3 properties, whereas Hasher and Zacks (1979) list 5, and Schneider, Dumais, and Shiffrin (1984) list 12. Hasher and Zacks (1979) have argued that all properties must be present in a truly automatic process, but others are less restrictive. To make matters worse, some researchers have questioned the agreement between properties, showing for example, that an obligatory process may be effortful (Kahneman & Chajzyck, 1983; Paap & Ogden, 1981; Regan, 1981).

Automaticity is also difficult to diagnose absolutely from the perspective of instance theory, because instance theory does not assume that any of the properties are necessary or sufficient. Automaticity is defined in terms of the underlying process—automatic performance is based on memory retrieval—and not in terms of necessary and sufficient properties. As described earlier, many of the properties may be characteristic of automaticity, but no property or set of properties define it. To determine whether a process is automatic, one must determine whether it is based on memory retrieval, and that is difficult to do because there are no accepted criteria for deciding whether something is based on memory retrieval.[9]

Instance theory suggests that automaticity is both absolute and relative. It is absolute in that performance may sometimes be based entirely on memory retrieval and sometimes entirely on the algorithm. It is relative in that performance may be based on memory retrieval on some proportion of the trials. It may be possible to estimate the proportion without knowing which trials were memory-based and which were algorithm-based. Automaticity is also relative in that memory strengthens progressively throughout practice, and it is appropriate to say that performance is more automatic after 10,000 trials than after 1,000 trials, even if both performances are entirely memory-based. Each trial has the same impact on memory regardless of the number of trials that went before it. Thus, instance theory suggests there are no limits on the degree of automaticity that may be attained; automaticity may never be complete.

It is easier to judge automaticity relatively than absolutely: The more automatic the performance, the faster it should be, the less effortful, the more autonomous, and so on (also see Logan, 1985b). Assessments of relative automaticity can be made most confidently when two performances of the same task are compared, preferably at two different levels of practice. One would expect practice to make a task more automatic, and the more practiced task is likely to be more automatic. It is more difficult to assess the relative automaticity of two different tasks. For example, one task may appear less effortful than another because its algorithm is easier, not because its performance is based more on memory retrieval. However, many of these problems may be minimized and sometimes avoided by a careful task analysis (e.g., Jonides et al., 1985).

Automaticity and Categorization

The instance theory of automaticity bears a strong resemblance to instance theories of categorization (e.g., Hintzman, 1986; Jacoby & Brooks, 1984; Medin & Schaffer, 1978). Instance theories of categorization argue that subjects decide the category membership of a test stimulus by comparing it with the instances stored in memory and assigning it to the category containing the most similar instances. These theories are similar to the instance theory of automaticity in that both assume that (a) performance depends on retrieval of specific instances from memory and (b) the retrieval process is cooperative, comparing all of the available traces with the test stimulus. The theories are also similar in that they focus on learning; both automaticity and categorization are acquired abilities.

Studies of automatization differ from studies of category learning in that there is an initial algorithm that allows subjects to perform accurately until memory retrieval can take over. In category learning, there is no algorithm to "bootstrap" performance. Performance is inaccurate until the category is well learned. Studies of category learning are also different in that the stimuli presented to subjects are usually very similar to each other and subjects are expected to exploit the similarities in forming categories and in generalizing to new instances. Studies of automatization, by contrast, often use easily discriminable stimuli (e.g., words or letters) and rarely test for generalization to new instances (but see Salasoo et al., 1985; Schneider & Fisk, 1984). Finally, studies of category learning usually involve sub-

[9] It may be possible to diagnose instance-based automaticity by showing better transfer to trained instances than to untrained ones. That requires knowing something about the subject's history with the task and materials, which is often not easy in practice.

stantially less practice than do studies of automatization—one session versus 10 or 20.

None of these differences seem very fundamental. Category learning studies could use orienting tasks to allow accurate initial performance, and studies of automatization could vary stimulus similarity and test for generalization and transfer. Category learning studies could investigate practice effects over the range used in studies of automatization, and studies of automatization could focus on the effects of initial practice (e.g., Experiments 1–3). The differences could be viewed as guidelines for future research to bring the different areas closer together. As a result, automaticity might come to be viewed more as a general cognitive phenomenon and less as a special topic in the psychology of attention.

The current version of the instance theory of automaticity would need to be changed in order to deal with the effects of stimulus similarity. The current version assumes that each repetition of the same stimulus produces exactly the same trace and that each different stimulus produces an entirely different trace. There is no cross talk between nonidentical traces to produce the effects seen in the studies of category learning. To account for those effects, I would have to make more detailed assumptions about how the trace is represented and how the retrieval process works (cf. Hintzman, 1986). That is an important direction for future work, but it is beyond the scope of this article.

The close relations between categorization and automatization envisioned by instance theory contrast sharply with the views of Shiffrin and Schneider (1977; also see Schneider & Shiffrin, 1985), who argued that automatization was more than categorization. The main evidence for their claim is a visual search experiment in which subjects were trained to detect letters from one set in arrays made from another set of letters. After several sessions of practice, during which performance improved dramatically, Shiffrin and Schneider switched sets: Subjects now had to detect the former distractor letters in arrays made from former targets. If automaticity depended only on categorization, there should be good transfer because the discrimination between targets and distractors is essentially the same. However, transfer was abysmal. Performance was as bad as it was on the initial practice session and took the same number of training sessions to reach pretransfer levels.

Shiffrin and Schneider (1977) argued that automatization affected the targets' ability to attract attention. Well-practiced targets elicit an *automatic attention response,* which pulls attention away from its current focus. In the transfer task, the former targets would pull attention away from the current targets (former distractors), causing failures of detection. Extensive post-transfer practice would be necessary to build up the automatic attention responses to the current targets. Other studies provided further evidence for automatic attention responses, so Shiffrin and Schneider concluded that there was more to automaticity than categorization; automaticity involved categorization and the automatic attention response.

I do not deny the automatic attention response or its importance in some cases of automaticity. However, it is not the only mechanism of automaticity; I argue that memory for instances is another. The automatic attention response may be specific to visual search tasks, in which the subject must focus spatially on a target item. There is a cost associated with filtering out the distractors (Kahneman, Treisman, & Burkell, 1983), and subjects may learn to overcome that cost. I suspect that the automatic attention response may not be important in tasks that do not require spatial filtering, such as the lexical decision and alphabet arithmetic tasks studied earlier. For those tasks, categorization (and hence, retrieval of instances) may be enough. Karlin and Bower (1976) and Jones and Anderson (1982) presented data suggesting that categorization may be sufficient for automatization in memory search.

Categorization has many properties that are important in automaticity. In particular, categorizing an item makes available a host of facts associated with category membership, permitting inferences that go beyond the given perceptual information (Murphy & Medin, 1985). When this happens quickly and effortlessly by a single act of memory retrieval, it captures the essence of automaticity.

Perhaps the main motivation for distinguishing between automaticity and categorization is the fear that categorization is somehow more fundamental; if automaticity is like categorization, then it is epiphenomenal (e.g., Cheng, 1985). I believe this fear is ungrounded. Automaticity and categorization are both fundamental constructs, reflecting different aspects of the same learning process, namely, the storage and retrieval of instances.

Process-Based Versus Instance-Based Learning

The instance theory assumes that there is no change in the initial algorithm or in the memory process as practice progresses. All that changes with practice is the data base on which the memory process operates. This assumption makes the theory easy to analyze and simulate, but it is unlikely to be true in general. The memory process may change through forgetting or through changes in attention (e.g., the addend = 5 condition in Experiment 4). Or the algorithm may change through process-based learning. There is some evidence of process-based learning in the literature (Pirolli & Anderson, 1985; E. Smith & Lerner, 1986) and even in the present experiments. For example, the new-item control condition in Experiment 1 showed some improvement with practice even though none of the stimuli were repeated.

Possibly, what appears to be process-based learning may actually reflect a different sort of instance-based learning. Subjects may be reminded of a better way to approach the task by retrieving an example of an approach that was successful on a similar problem in the past (see Ross, 1984). Thus, process changes may reflect a discrete shift to a different strategy rather than a gradual evolution or refinement of a single process. Alternatively, subjects may parse their experience into instances in a different way than the experimenter imagines. In a category judgment task, for example, a subject who is asked to decide whether a trout is a fish may encode the trial as another instance of *fish* rather than the first instance of *trout* (cf. Barsalou & Ross, 1986), and show learning that depends on the number of presentations of *fish* rather than *trout.*[10] It may be difficult to sepa-

[10] I would like to thank Eliot Smith for suggesting this hedge. It may account for two aspects of Schneider and Fisk's (1984) data that seem troublesome for instance theory: (a) the null effect of number of in-

rate these variations of instance-based learning from true process-based learning, but the issue is important and the results may be worth the effort. Whatever the outcome, we would learn something important about the strategies people use in acquiring and utilizing knowledge.

Retrieval of instances may play an important role in process-based learning. Even when subjects learn an algorithm, the conditions necessary for instance-based learning are satisfied—subjects may have no choice but to encode and retrieve specific instances. One could imagine interactions between instance-based and process-based learning in which retrieval of instances helps eliminate ineffective variations of the process (they would be less likely to finish before memory retrieval) and execution of the process helps eliminate ineffective memory retrievals.

Relations to Other Theories of Skill Acquisition and Automaticity

Several existing theories predict a power-function speed-up from assumptions that differ substantially from the instance theory. Many of the theories addressed different tasks and processes than the instance theory addresses, and few of them deal with automaticity directly. In this section, I compare those theories with instance theory, but I do not intend to argue that instance theory is more correct or more accurate than the other theories. Instead, I view the theories as addressing different situations, and some theories may fit some situations better than others; it seems likely to me that humans can learn in more than one way. Choice among theories is something like choice among statistical tests. One considers the assumptions of the statistical model and determines whether they can be satisfied in a given situation. If so, one uses the test; if not, one chooses another test. Analysis of variance, for example, is not wrong or inaccurate because it cannot deal with data from Bernouli trials; it is simply inappropriate. When two different tests can be used on the same data, one can ask which is more powerful and more accurate, as in comparisons between parametric and nonparametric statistics; but first, one must be sure that the assumptions fit the situation.

Crossman

Crossman (1959) proposed a theory of skill acquisition in which subjects sampled methods for performing a task until they found the fastest one. He assumed there was a pool of possible methods and that subjects would select one at random for each performance of the task. Afterward, they compared the speed of the method they selected with their average speed over several trials, and if the method was faster, they increased the probability of selecting it for the next trial. In the long run, the fastest method would have the highest probability of being selected. The theory predicts a power-function speed-up because it is easier to find a faster method early in practice than later on.

Crossman's (1959) theory applies to situations in which there are several different methods for performing a task and subjects know or have available all of the different methods when they first begin the task. There is no provision for learning new methods or improving old ones as practice progresses. In this respect, it differs from the instance theory, which assumes that subjects acquire new methods (i.e., memory for past solutions) that strengthened over practice (i.e., by accumulating similar instances). Crossman's theory may account well for the acquisition of motor skills, like cigar rolling or typing, but it would not account for the lexical decision and alphabet arithmetic tasks described earlier, which involve developing and strengthening new methods.

Newell and Rosenbloom

Newell and Rosenbloom (1981; Rosenbloom & Newell, 1986) proposed a theory based on the idea of chunking. They argued that subjects acquire responses to stimulus patterns or chunks, which they can execute the next time the pattern occurs. They assumed that subjects learned patterns at several different levels, some encompassing single elements, some encompassing several elements, and some encompassing the whole stimulus. They argued that the smaller patterns recur more frequently than the larger patterns (e.g., moving a single pawn vs. opening a chess game), so subjects will have more opportunities to manifest their learning the smaller the pattern. This principle accounts for the power-function speed-up: Subjects benefit from having learned the smaller patterns early in practice because they recur often. Later on, they will have learned most of the smaller patterns and will benefit no more from subsequent occurrences. Subjects will tend to benefit from larger patterns later in practice because they recur less often and because there are more of them to be learned. Thus, early learning will be rapid, as the smaller patterns are acquired and utilized, and later learning will be relatively slow, as the larger patterns are gradually learned and gradually utilized.

Newell and Rosenbloom's theory differs from the instance theory in that it assumes that there is no strengthening of the response to an individual chunk once it is learned. Their theory applies best to situations in which the stimuli are highly patterned, allowing chunks to be formed at many different levels. It would not apply well to the lexical decision and alphabet arithmetic experiments reported earlier because the tasks could not be performed by breaking the stimuli down into chunks and responding to chunks at different levels. None of the component letters predicted which response to make; subjects had to respond to the stimuli as wholes.

MacKay

MacKay (1982) proposed a theory in which learning occurs by strengthening connections between nodes in a network that describes the representation underlying perception and action. His theory produces the power-function speed-up in two ways: First, the connections are strengthened in proportion to the difference between the current strength and the maximum possible strength. The proportion is constant over learning, which means that changes in strength will be greater early in learning

stances (4 vs. 8 per category) on the rate of learning and (b) the successful transfer to new instances (ranging from 60% to 90%).

than they will be later on. Second, MacKay assumed that the representation was hierarchical and that the higher nodes were smaller in number and farther from their maximum strength than the lower nodes. Thus, early learning would be dominated by large changes in the strength of a few higher nodes, and later learning would be dominated by small changes in the strength of all nodes.

MacKay's (1982) theory differs primarily from the instance theory in that it is a strength theory. It assumes that practice strengthens connections between generic stimuli and generic responses, whereas the instance theory accounts for strengthening by encoding and retrieving separate representations of each encounter with a stimulus. MacKay intended his theory to apply to behavior that is more complex than the behavior to which the instance theory has been applied (i.e., generating sentences vs. performing lexical decisions), but that is not a major difference. It should be possible to decide between his theory and mine empirically, with experiments like the present Experiment 5 (also see Hintzman, 1976).

Anderson

Anderson (1982, 1987) proposed a theory of skill acquisition that has several different learning mechanisms. Some involved translating verbal instructions into procedures for performing a task, others involved generalizing and differentiating existing procedures, and one involved simple strengthening. The most important mechanisms in the present context are *composition* and *strengthening.*

Composition involves collapsing a series of steps into one by combining adjacent procedures. The amount of practice necessary to reduce a complex procedure to a single step depends on the number of steps and on the probability of combining adjacent steps. The more steps and the lower the probability of combining, the longer it will take. Composition reduces the number of steps by a constant proportion on each iteration of the procedure, producing rapid learning early in practice and slower learning later on.

Strengthening involves increasing the speed with which productions are executed. It operates mainly on composed productions, increasing the strength on each exposure by a constant proportion. Strengthening and composition work together to produce the power-function speed-up. The other learning mechanisms may contribute something to the speed-up, but composition and strengthening are the major contributors.

Anderson's (1982, 1987) theory differs from the instance theory in being much more detailed and embedded in a very powerful theory of general cognition. Some of its learning mechanisms are stimulus-specific, like those of instance theory, but others are more general, providing process-based learning. Thus, Anderson's theory would apply better than the instance theory to situations in which people learn general procedures rather than specific responses to specific stimuli. Anderson's composition process will work only when the structure of the task allows adjacent steps to be collapsed. It is unlikely to work in the lexical decision experiments reported above, in which one single-step process was replaced by another. Nor is it likely to work in alphabet arithmetic or arithmetic in general. Subjects did not learn to count by twos, then fours, and so on; instead, they switched from counting by ones to remembering.

Schneider

Walter Schneider has been developing a theory of automatization for several years (Schneider, 1985; Schneider & Detweiler, 1987). Schneider's theory involves two kinds of learning, *priority* learning, which attracts attention to display positions that are likely to contain targets, and *association* learning, which connects stimuli directly to responses. The mechanism underlying both kinds of learning is proportional strengthening; after each successful trial, priority and associative strength are both increased by an amount proportional to the difference between their current strength and the maximum possible strength. The power-function speed-up is a consequence of proportional strengthening.

Schneider's theory differs from the instance theory in specific details. Schneider was concerned with the microstructure of skill acquisition and made what he considered to be physiologically plausible assumptions about the underlying representations and the processes that operate on them. So far, he has addressed automatization only in tasks that combine visual and memory search, developing a detailed model of initial performance on those tasks. It is not obvious how his theory would deal with other tasks at the same level of detail (e.g., lexical decision and alphabet arithmetic).

Schneider also interpreted the properties of automaticity differently than I do, having taken a more conventional position. For example, he characterized nonautomatic processing (which he called *controlled* processing) as "slow, generally serial, effortful, capacity-limited, [and] subject-controlled" (Schneider, 1985, p. 476), whereas I argue that there may be no characteristics common to all or even to most instances of nonautomatic processing. He assumed that controlled processing is necessary for learning; there is no further learning once processing is automatic. By contrast, I assume that learning occurs on each trial, whether processing is automatic or not. It may be harder to find evidence of learning once processing is automatic, for reasons I described earlier, but each trial continues to lay down a separate trace.

Finally, Schneider's theory differs from the instance theory in assuming two learning mechanisms instead of one. His association-learning mechanism is similar to the learning mechanism in the instance theory, but there is nothing in the instance theory corresponding to his priority learning mechanism. I believe priority learning is important primarily in visual search, in which targets must be discriminated from simultaneously presented distractors. Association learning (or instance learning) should be sufficient to account for situations that do not require such discrimination (e.g., memory search, lexical decision, and alphabet arithmetic).

Despite these differences, Schneider's theory is similar to the instance theory in assuming that automatization reflects a transition from algorithm-based processing ("controlled" processing) to memory-based processing. The language may be very different, but the underlying idea is basically the same. The most fundamental differences lie in the assumptions about strengthening; his is a strength theory, and mine is not. One could distinguish between the theories empirically, with experiments that distinguish strength theories from instance theories (e.g., Experiment 5; also see Hintzman, 1976).

Concluding Remarks

The purpose of this article was to present an alternative to the modal view of automatization as the gradual withdrawal of attention. The instance theory accounts for many of the facts addressed by the modal view without assuming any resource limitations, attentional or otherwise. Novice performance is limited by a lack of knowledge rather than by a lack of resources. The theory accounts for the power-function speed-up that is perhaps the most stable and least controversial property of automatization. In doing so, it predicted a power-function reduction in the standard deviation and a constraint between the mean and standard deviation (same exponent), which was confirmed in each experiment.

An important feature of the theory as developed so far is that it accounts for the facts of automaticity by assuming that only the knowledge base changes with practice. This assumption may strike many readers as implausible, but it accounts for a surprising amount of variance in a number of learning tasks. The theory may be viewed as a null hypothesis against which to evaluate competing theories that assume changes in the underlying processes with practice. The main point of the fits to experimental data is that there may not be much variance left for competing theories to explain.

References

Allport, D. A. (1980). Attention and performance. In G. Claxton (Ed.), *Cognitive psychology* (pp. 112–153). London: Routledge & Kegan Paul.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369–406.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94,* 192–210.

Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review, 2,* 213–236.

Bahrick, H. P., & Shelley, C. (1958). Time-sharing as an index of automatization. *Journal of Experimental Psychology, 56,* 288–293.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357.

Barsalou, L. W., & Ross, B. H. (1986). The roles of automatic and strategic processing in sensitivity to superordinate and property frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12,* 116–134.

Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 556–566.

Bryan, W. L., & Harter, N. (1899). Studies of the telegraphic language. The acquisition of a hierarchy of habits. *Psychological Review, 6,* 345–375.

Carr, T. H., McCauley, C., Sperber, R. D., & Parmalee, C. M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 757–777.

Chandler, P. J. (1965). *Subroutine STEPIT: An algorithm that finds the values of the parameters which minimize a given continuous function* [Computer program]. Bloomington: Indiana University, Quantum Chemistry Program Exchange.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55–81.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological Review, 92,* 414–423.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

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

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Feustal, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112,* 309–346.

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proceedings of the Cambridge Philosophical Society, 24,* 180–190.

Fisk, A. D., & Schneider, W. (1984). Memory as a function of attention, level of processing, and automatization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 181–197.

Francolini, C. M., & Egeth, H. (1980). On the nonautomaticity of "automatic" activation: Evidence of selective seeing. *Perception and Psychophysics, 27,* 331–342.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum [On the law of probability of the maximum deviation]. *Annales de la Société Polonaise de Mathématique, Cracovie, 6,* 93–116.

Greene, R. L. (1984). Incidental learning of event frequency. *Memory and Cognition, 12,* 90–95.

Gumbel, E. J. (1958). *Statistics of extremes.* New York: Columbia University Press.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39,* 1372–1388.

Henik, A., Friedrich, F. J., & Kellogg, W. A. (1983). The dependence of semantic relatedness effects on prime processing. *Memory and Cognition, 11,* 366–373.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–91). New York: Academic Press.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace model. *Psychological Review, 93,* 411–428.

Hirst, W., Spelke, E. S., Reaves, C. C., Caharack, G., & Neisser, U. (1980). Dividing attention without alternation or automaticity. *Journal of Experimental Psychology: General, 109,* 98–117.

Hoffman, J. E., Nelson, B., & Houck, M. R. (1983). The role of attentional resources in automatic detection. *Cognitive Psychology, 15,* 379–410.

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization and recall of highly associated words. *Journal of Experimental Psychology, 82,* 472–481.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 1–47). New York: Academic Press.

James, W. (1890). *Principles of psychology.* New York: Holt.

Jones, W. P., & Anderson, J. R. (1982). Semantic categorization and high-speed scanning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 237–242.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye movement. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum.

Jonides, J., Naveh-Benjamin, M., & Palmer, J. (1985). Assessing automaticity. *Acta Psychologica, 60,* 157–171.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Chajzyck, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 497–509.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Erlbaum.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

Kahneman, D., & Treisman, A. M. (1984). Changing views of attention and automaticity. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.

Kahneman, D., Treisman, A., & Burkell, J. (1983). The cost of visual filtering. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 510–522.

Karlin, M. B., & Bower, G. H. (1976). Semantic category effects in visual word search. *Perception and Psychophysics, 19,* 417–424.

Kolers, P. A. (1975). Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 689–701.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

LaBerge, D. (1981). Automatic information processing: A review. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 173–186). Hillsdale, NJ: Erlbaum.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

Ladefoged, P., Silverstein, R., & Papcun, G. (1973). Interruptibility of speech. *Journal of the Acoustical Society of America, 54,* 1105–1108.

Landauer, T. K. (1975). Memory without organization: Properties of a model with random storage and undirected retrieval. *Cognitive Psychology, 7,* 495–531.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41–104.

Logan, G. D. (1978). Attention in character classification: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General, 107,* 32–63.

Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 189–207.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12,* 523–553.

Logan, G. D. (1982). On the ability to inhibit complex movements: A stop-signal study of typewriting. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 778–792.

Logan, G. D. (1983). On the ability to inhibit simple thoughts and actions: 1. Stop-signal studies of decision and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 585–606.

Logan, G. D. (1985a). On the ability to inhibit simple thoughts and actions: 2. Stop-signal studies of repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 675–691.

Logan, G. D. (1985b). Skill and automaticity: Relations, implications and future directions. *Canadian Journal of Psychology, 39,* 367–386.

Logan, G. D. (1988). *Repetition priming and automaticity: Common underlying mechanisms.* Manuscript submitted for publication.

Logan, G. D., & Klapp, S. T. (1988). *Automatizing alphabet arithmetic: Evidence for an instance theory of automatization.* Manuscript submitted for publication.

Long, J. (1976). Visual feedback and skilled keying: Differential effects of masking the printed copy and the keyboard. *Ergonomics, 19,* 93–110.

MacKay, D. G. (1982). The problem of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review, 89,* 483–506.

Mandler, G. (1967). Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1, pp. 327–372). New York: Academic Press.

Marcel, A. T. (1983). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology, 15,* 238–300.

Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. *Psychological Bulletin, 85,* 1256–1274.

McLeod, P., McLaughlin, C., & Nimmo-Smith, I. (1985). Information encapsulation and automaticity: Evidence from the visual control of finely timed actions. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 391–406). Hillsdale, NJ: Erlbaum.

McKoon, G., & Ratcliff, R. (1980). Priming in item recognition: The organization of propositions in memory for text. *Journal of Verbal Learning and Verbal Behavior, 19,* 369–386.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35,* 113–138.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Murdock, B. B. (1974). *Human memory: Theory and data.* Hillsdale, NJ: Erlbaum.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B. (1983). A distributed memory model for serial-order information. *Psychological Review, 90,* 316–338.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Naveh-Benjamin, M. (1987). Coding of spatial location information: An automatic process? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 595–605.

Naveh-Benjamin, M., & Jonides, J. (1984). Maintenance rehearsal: A two-component analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 369–385.

Naveh-Benjamin, M., & Jonides, J. (1986). On the automaticity of frequency coding: Effects of competing task load, encoding strategy, and intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 378–386.

Navon, D. (1984). Resources—A theoretical soup stone? *Psychological Review, 91,* 216–234.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86,* 214–255.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Neisser, U. (1976). *Cognition and reality.* San Francisco: Freeman.

Neumann, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz & A. F. Sanders (Eds.), *Cognition and motor processes* (pp. 255–293). Berlin: Springer-Verlag.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning. *Cognitive Psychology, 19,* 1–32.

Paap, K. R., & Ogden, W. C. (1981). Letter encoding is an obligatory but capacity-demanding operation. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 518–527.

Pirolli, P. L., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 136–153.

Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review, 78,* 391–408.

Posner, M. I., & Klein, R. (1973). On the functions of consciousness. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 21–35). New York: Academic Press.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Rabbitt, P. M. A. (1978). Detection of errors by skilled typists. *Ergonomics, 21,* 945–958.

Rabbitt, P. M. A. (1981). Sequential reactions. In D. H. Holding (Ed.), *Human skills* (pp. 153–175). New York: Wiley.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior, 17,* 403–417.

Ratcliff, R., & McKoon, G. (1981). Automatic and strategic components of priming in recognition. *Journal of Verbal Learning and Verbal Behavior, 20,* 204–215.

Reason, J. T. (1984). Lapses of attention in everyday life. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 515–549). New York: Academic Press.

Reason, J. T., & Myceilska, K. (1982). *Absent minded: The psychology of mental lapses and everyday errors.* Englewood Cliffs, NJ: Prentice-Hall.

Regan, J. E. (1981). Automaticity and learning: Effects of familiarity on naming letters. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 180–195.

Rosenbloom, P. S., & Newell, A. (1986). The chunking of goal hierarchies: A generalized model of practice. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2, pp. 247–288). Los Altos, CA: Morgan Kaufmann.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16,* 371–416.

Ryan, C. (1983). Reassessing the automaticity-control distinction: Item recognition as a paradigm case. *Psychological Review, 90,* 171–178.

Salasoo, A., Shiffrin, R. M., & Feustal, T. C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General, 114,* 50–77.

Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin, 99,* 303–319.

Schneider, W. (1985). Toward a model of attention and the development of automatic processing. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 475–492). Hillsdale, NJ: Erlbaum.

Schneider, W., & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21, pp. 53–119). New York: Academic Press.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1984). Automatic and control processing and attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 1–27). New York: Academic Press.

Schneider, W., & Fisk, A. D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. *Perception and Psychophysics, 31,* 160–168.

Schneider, W., & Fisk, A. D. (1984). Automatic category search and its transfer. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 1–15.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: 1. Detection, search and attention. *Psychological Review, 84,* 1–66.

Schneider, W., & Shiffrin, R. M. (1985). Categorization (restructuring) and automatization: Two separable factors. *Psychological Review, 92,* 424–428.

Shiffrin, R. M., & Dumais, S. T. (1981). The development of automatism. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 111–140). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116,* 250–264.

Smith, E. R., & Lerner, M. (1986). Development of automatism of social judgments. *Journal of Personality and Social Psychology, 50,* 246–259.

Smith, M. C. (1979). Contextual facilitation in a letter search task depends on how the prime is processed. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 239–251.

Smith, M. C., Theodor, L., & Franklin, P. E. (1983). The relationship between contextual facilitation and depth of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 697–712.

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition, 4,* 215–230.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes.* Cambridge, MA: Cambridge University Press.

Wenger, D. M., Schneider, D. J., Carter, S., & White, T. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology, 53,* 1–9.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63–102). New York: Academic Press.

Zbrodoff, N. J. (1979). *Development of counting and remembering as strategies for performing simple arithmetic in elementary school children.* Unpublished master's thesis, University of Toronto, Toronto.

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General, 115,* 118–130.

(*Appendixes follow on next page*)

Appendix A

Formal Analysis of the Power-Function Speed-Up

The instance theory assumes that (a) each encounter with a stimulus is encoded into memory, (b) all prior encounters with a stimulus are retrieved when the stimulus is encountered again, (c) each encounter is stored and retrieved independently, and (d) the subject can respond as soon as the first trace is retrieved from memory. Assuming further that the distribution of retrieval times is the same for each of the n traces, reaction times can be modeled as the minimum of n independent samples from the same distribution. The purpose of Appendix A is to show that reaction times will decrease as a power function of sample size, which leads to the prediction that means and standard deviations decrease as power functions of practice with the same exponent. Also addressed is the race between the algorithm and memory retrieval, examining the effect of the algorithm on the power-function speed-up and the number of trials required for memory to dominate the algorithm.

The distribution of minima from an arbitrary distribution can be written as a function of the initial distribution. The cumulative distribution function is

$$F_1(t) = 1 - [1 - F(t)]^n, \tag{A1}$$

and the probability density function is

$$f_1(t) = n[1 - F(t)]^{n-1}f(t) \tag{A2}$$

(Gumbel, 1958, p. 76). However, it is difficult to derive general predictions for changes in the mean and standard deviation with sample size, n, that would be true for every initial distribution. Instead, we derived specific predictions for two initial distributions, the exponential and the Weibull, and we derived general predictions for the class of positive-valued distributions by working backward from the asymptotic distribution of minima.

The exponential was chosen for the initial derivation because it is easy to work with; the entire model can be expressed with exponential distributions. However, predictions from the exponential model deviate systematically from empirical data, so aspects of the model were explored in a particular generalization of the exponential, the Weibull distribution. The Weibull was chosen for three reasons: First, with appropriate parameterization, it resembles reaction-time distributions; second, it avoids the deviation from the data found with the exponential model; and third, as discussed in the final section, it turns out to be an important asymptotic distribution of minima from positive-valued distributions.

The Exponential Distribution

The exponential distribution is commonly used in models of memory retrieval because of its empirical success and analytical tractability (Murdock, 1974; Ratcliff, 1978). Its distribution function is

$$F(t) = 1 - \exp[-wt],$$

and its density function is

$$f(t) = w\exp[-wt].$$

From Equation A1, the distribution function for the minimum of n samples from the same exponential distribution is

$$F_1(t) = 1 - \exp[-nwt],$$

and from Equation A2, the density function is

$$f_1(t) = n\exp[-(n-1)wt]w\exp[-wt]$$
$$= (nw)\exp[-nwt].$$

Thus, the distribution of minima from an exponential distribution is itself an exponential distribution, with a rate constant n times larger than the rate constant for the initial distribution.

The mean of the distribution decreases as a power function of n with an exponent of -1:

$$\bar{T}_1 = (nw)^{-1} = (n^{-1})(w^{-1}). \tag{A3}$$

Since the standard deviation of an exponential distribution equals the mean, Equation A3 implies that the standard deviation also decreases as a power function of n with the same exponent, -1. This was the prediction tested in Experiments 1–4. The exponential model makes a stronger prediction, that the mean equals the standard deviation. Later, the results are generalized to other distributions so that the exponent need not equal -1, and the mean need not equal the standard deviation.

The exponential distribution permits an easy analysis of the race between the algorithm and memory retrieval. In general, the distribution of the minima from two distributions, $f_a(t)$ and $f_m(t)$, is

$$f_r(t) = f_a(t)[1 - F_m(t)] + f_m(t)[1 - F_a(t)].$$

If the distribution of finishing times for the algorithm, $f_a(t)$, and the memory process, $f_m(t)$, are exponential, then

$$f_r(t) = w_a\exp[-w_at]\exp[-nw_mt] + nw_m\exp[-nw_mt]\exp[-w_at]$$
$$= (w_a + nw_m)\exp[-(w_a + nw_m)t].$$

The parameters of the resulting distribution are the sums of the parameters of the parent distributions. The mean (and the standard deviation) of the resulting distribution,

$$\bar{T}_r = (w_a + nw_m)^{-1}, \tag{A4}$$

decreases as n increases. If the mean for the algorithm is the same as the mean for memory, then the mean and the standard deviation of the resulting distribution decrease as a power function of $n + 1$ with an exponent of -1. To the extent that the algorithm mean differs from the memory mean, this relation will be distorted. However, the distortion will decrease as n increases and memory wins the race more often. At some point, memory will win virtually all the time.

The probability that the algorithm will win the race can be derived easily if the underlying distributions are exponential. If so, memory retrieval and the algorithm can be viewed as simultaneous Poisson processes with rates w_a and nw_m, respectively. Then the probability that the algorithm finishes first is

$$P(\text{algorithm first}) = w_a/(w_a + nw_m),$$

which decreases rapidly as n increases. If the mean for the algorithm equals the mean for the memory process, then the probability that the algorithm finishes first equals $1/(n + 1)$; it decreases as a power function of $n + 1$ with an exponent of -1.

In summary, the instance theory can be modeled as a race between two exponential distributions, one representing finishing times for the algorithm and one representing finishing times from a memory process with n independent traces. That model predicts (a) a power-function reduction in mean reaction time, (b) a power-function reduction in the standard deviation of reaction times, and (c) a common exponent for means and standard deviations. These predictions are not compromised much by the race with the algorithm, provided that the mean for the algorithm is reasonably close to the mean for memory retrieval.

The exponential distribution imposes severe restrictions on the predictions, namely, that the exponent of the power function equals -1 and that the mean equals the standard deviation. These restrictions create problems for the model. Most exponents from empirical power functions are less than 1 in absolute magnitude (Newell & Rosenbloom, 1981), and the reduction in the mean rarely equals the reduction in the standard deviation (see Experiments 1–4). How can the exponential model deal with these problems?

One possibility is to relax (and make more realistic) the assumption that each and every encounter with a stimulus is stored and retrieved. If instead, each encounter was stored and retrieved with probability p, the observed power function would be

$$n^k = (pn)^{-1}.$$

Taking logs of both sides yields

$$k\log n = -(\log p + \log n).$$

Solving for k yields

$$k = -(\log p + \log n)/\log n,$$

which is less than 1 if p is less than 1 (since logs of proportions are negative), and it decreases as p decreases (since logs of proportions increase in absolute magnitude as the proportion decreases); it predicts slower learning, the lower the probability of storage and retrieval, which is reasonable.

Thus, the exponential model, supplemented by the (reasonable) assumption of imperfect storage and retrieval, predicts power functions with exponents less than 1 in absolute magnitude, which is commonly observed (Newell & Rosenbloom, 1981). However, the exponential model still predicts an identical reduction in the mean and standard deviation, which is not generally observed (see, e.g., Experiments 1–4).

Weibull Distribution

The Weibull distribution is a generalization of the exponential distribution. Its distribution function is

$$F(t) = 1 - \exp[-(t/a)^c],$$

and its probability density function is

$$f(t) = (c/a)(t/a)^{c-1}\exp[-(t/a)^c].$$

If $c = 1$, the Weibull reduces to an exponential distribution with rate parameter $1/a$. The Weibull is a flexible distribution. The parameter c determines its shape, ranging from exponential when $c = 1$ to normal looking when $c = 5$. Intermediate values produce density functions that resemble reaction time distributions—truncated on the left, with a long tail on the right. Thus, the Weibull may provide a reasonable approximation for retrieval times from a variety of processes.

From Equation A1, the distribution function for the minimum of n samples from the same Weibull distribution is

$$\begin{aligned} F_1(t) &= 1 - \{\exp[-(t/a)^c]\}^n \\ &= 1 - \exp[-n(t/a)^c] \\ &= 1 - \exp[-((n^{1/c}t)/a^c] \\ &= F(n^{1/c}t). \end{aligned}$$

Thus,

$$\bar{T}_1 = n^{-1/c}\bar{T}$$

and

$$\sigma_1 = n^{-1/c}\sigma.$$

The distribution of minima from a Weibull distribution is itself a Weibull distribution. As sample size increases, the distribution of minima contracts as a power function of sample size with an exponent of $-1/c$. Thus, the mean and the standard deviation and all of the quantiles of the distribution should decrease as power functions of n with the same exponent, $-1/c$. Experiments 1–4 tested the equality of the exponents for means and standard deviations.

The Weibull imposes less severe restrictions on the predicted power functions than does the exponential. The exponent must be the same for the mean and standard deviation, but it is not fixed at any particular value. Since c is 1 or larger in most applications, the exponents of power functions, $-1/c$, should be 1 or less, as is commonly observed (Newell & Rosenbloom, 1981). Moreover, the mean of a Weibull is not equal to the standard deviation.

So far, a power-function speed-up and reduction in standard deviation has been predicted from two specific initial distributions. The next section attempts to generalize these results to the class of positive-valued distributions, of which the exponential and Weibull are members.

Stability Postulate

The power law can be generalized further by working backward from the *asymptotic* distribution of the minimum. Most readers will be familiar with the normal distribution as the asymptotic distribution of sums or averages; by the central limit theorem, the distribution of sums or averages from an arbitrary initial distribution will converge on the normal distribution as sample size increases, regardless of the form of the initial distribution. As it turns out, there are three distributions that are asymptotic in that sense for minima. The distribution of minima from an arbitrary initial distribution will converge on one of the three asymptotic distributions as sample size increases (Gumbel, 1958). Which of the three it converges on depends on rather general properties of the initial distribution (whether the extremes are limited or unlimited). Only one asymptotic distribution—the third—is relevant to reaction time data because it applies to random variables with only positive values. Interestingly, the third asymptotic distribution is the Weibull.

Proving that the asymptotic distribution of minima follows the power law is important because it implies that minima from any positive-valued distribution will eventually conform to the power law as sample size increases. At some point, the initial distribution will converge on the asymptote and what is true of the asymptotic distribution will be true of samples from the initial distribution. Before the distribution converges, the power law may be approximately correct.

The power-law predictions derive from the *stability postulate,* which was used initially by Fréchet (1927) and Fisher and Tippett (1928) to derive the three asymptotic distributions (Gumbel, 1958). The following proof is a condensed version of one given by Gumbel (pp. 157–162). It deals with the maximum rather than the minimum to simplify calculation, but the results apply to the minimum as well as the maximum; what is true of t for the maximum is true of $-t$ for the minimum.

According to the stability postulate, a distribution is stable with respect to its maximum (or minimum) if the distribution of maxima (minima) sampled from it retains the same form as the initial distribution as sample size increases, changing only in its scale or in translation of its origin. As shown earlier, the exponential and Weibull are stable in this sense. Analogous to Equation A2, the distribution function for the maximum of n independent samples from the same distribution is

$$F_n(t) = F^n(t).$$

The stability postulate states that the probability that the largest value is t or less after n samples is equal to the probability of a linear function of t derived from the initial distribution. That is,

$$F^n(t) = F(a_n t + b_n), \qquad \text{(A5)}$$

where a_n and b_n are both functions of n, the sample size. For the third asymptotic distribution, t is negative with a maximum of zero (for the minimum, t is positive with a minimum of zero). The additive constant, b_n, drops out, and Equation A5 reduces to

$$F^n(t) = F(a_n t). \tag{A6}$$

Equation A6 implies that raising $F(t)$ to the mth power is the same as multiplying t by a_m; that is,

$$F^{mn}(t) = F(a_m a_n t).$$

But from Equation A6,

$$F^{mn}(t) = F(a_{mn} t).$$

This implies the functional equation:

$$a_{mn} = a_m a_n.$$

The solution to Equation A7 is a power function,

$$a_n = n^c,$$

because

$$(mn)^c = (m)^c(n)^c.$$

Consequently,

$$F(t)^n = F(n^c t),$$

$$\bar{T}_1 = n^{-c}\bar{T},$$

and

$$\sigma_1 = n^{-c}\sigma.$$

For the third asymptotic distribution, c is positive. Thus, the scale of the distribution contracts as a power function of n, the sample size, whether we are dealing with the maximum (in which case, t is negative and bounded above at zero) or the minimum (in which case, t is positive and bounded below at zero). This means that all quantiles of the distribution of minima should decrease as a power function of n, and all the power functions should have the same exponent, $-c$. In particular, the mean and the standard deviation should both decrease as power functions of n with the same exponent, $-c$.

Gumbel (1958) then derived the form of the three asymptotic distributions, showing that the third is a Weibull distribution. But that is beyond the requirements of this article. The important point for now is that an arbitrary distribution that is bounded on the left (i.e., has only positive values) will converge on the Weibull distribution as sample size increases. There may be departures from the power law for small sample sizes, before the distribution of minima reaches asymptote, but afterward the distribution will contract following the power law.

Appendix B

Method and Results From Experiments 1–3

The details of the method and results from the lexical decision experiments are reported in this section.

Experiment 1

Method

Subjects. In all, 24 subjects were tested. Some subjects came from introductory psychology courses and received course credit for participating; some were volunteers and were paid $4 for participating.

Apparatus and stimuli. The stimuli were four-letter words and nonwords. The words were common nouns selected from the Kucera and Francis (1967) frequency norms. The words ranged in absolute frequency from 7 to 923 per million, with a mean of 56.05. The nonwords were constructed by replacing one letter of each word. In most cases, the nonwords were pronouncible.

The stimuli were displayed on a point–plot CRT (Tektronix Model 604, equipped with P31 phosphor), controlled by a PDP 11/03 computer. They were displayed as uppercase letters, formed by illuminating about 20 points in a 5 × 7 matrix. They were displayed at the center of the screen. Viewed at a distance of 60 cm (maintained by a headrest), each word and nonword subtended 2.67° × .57° of visual angle.

Each trial began with a fixation point exposed in the center of the screen and a 900-Hz warning tone. The tone and fixation point were presented for 500 ms, followed immediately by the word or nonword for that trial, which was also presented for 500 ms. After the word or nonword was extinguished, a 1,500-ms intertrial interval began.

Subjects responded by pressing the two outermost telegraph keys in a panel of eight mounted on a moveable board that sat between the headrest and the CRT.

Procedure. Subjects were told that on each trial they would see a word or a nonword and that their task was to indicate whether a word or a nonword had appeared. They were told to respond as quickly as possible without making more than 10% errors. One half of the subjects pressed the right-most key of the panel of eight to indicate that they saw a word and pressed the left-most key to indicate that they saw a nonword, whereas the other half did the opposite. Subjects were not told anything about the schedule of stimulus repetitions or even that some stimuli would be repeated.

There were two main conditions, the experimental condition, in which the same set of words and nonwords were exposed repeatedly, and the control condition, in which new words and nonwords were shown on each trial. Each condition involved blocks of 20 trials, 10 with words and 10 with nonwords. In the experimental condition, a set of 10 four-letter words and 10 four-letter nonwords were chosen randomly from the population of 340 stimuli. Each word and nonword was shown once per block in a different random order in each block, for a total of 16 blocks. Thus, the lag between successive repetitions varied from 1 to 40, with a mean of 20.

In the control condition, a different set of 10 four-letter words and 10 four-letter nonwords was chosen randomly without replacement from the set of 340 stimuli for each of the 16 blocks. Each control stimulus appeared in only one block and never appeared in the experimental blocks.

A different random sampling of stimuli was prepared for each subject, and the order of trials within blocks was randomized separately for each subject.

Results

An ANOVA on the reaction-time data revealed the following effects: The main effect of number of presentations was significant, $F(15,$

Table B1
Error Rates From Lexical Decision Task in Experiment 1

	Condition			
Presentation	Old word	Old nonword	New word	New nonword
1	.05	.11	.08	.06
2	.05	.07	.07	.06
3	.02	.07	.08	.04
4	.02	.07	.10	.09
5	.03	.05	.04	.05
6	.02	.05	.04	.05
7	.02	.05	.05	.08
8	.04	.04	.09	.07
9	.03	.06	.07	.08
10	.03	.02	.10	.05
11	.04	.06	.10	.05
12	.03	.04	.08	.05
13	.04	.05	.06	.04
14	.04	.05	.05	.03
15	.03	.03	.07	.04
16	.03	.04	.08	.05

345) = 7.52, $p < .01$, $MS_e = 6{,}809.85$, as was the main effect of control versus experimental conditions, $F(1, 23) = 31.15$, $p < .01$, $MS_e = 52{,}687.35$, and the main effect of word versus nonword, $F(1, 23) = 43.11$, $p < .01$, $MS_e = 31{,}748.23$. Presentations and conditions interacted significantly, $F(15, 345) = 2.92$, $p < .01$, $MS_e = 7{,}186.63$, reflecting the extra benefit from repeating specific stimuli in the experimental condition. Presentations and word versus nonword interacted significantly, $F(15, 345) = 4.23$, $p < .01$, $MS_e = 2{,}177.17$, reflecting the greater effect of repetition with nonwords than with words. In addition, there were significant interactions between conditions and word versus nonword, $F(1, 23) = 5.59$, $p < .01$, $MS_e = 7{,}904.85$, and between presentations, conditions, and word versus nonword, $F(15, 345) = 1.84$, $p < .05$, $MS_e = 2{,}306.41$.

An ANOVA on the standard deviations revealed a significant main effect of the number of presentations, $F(15, 345) = 3.44$, $p < .01$, $MS_e = 5{,}011.76$, and a significant main effect of control versus experimental conditions, $F(1, 23) = 29.30$, $p < .01$, $MS_e = 19{,}509.57$. In addition, there were significant interactions between presentations and conditions, $F(15, 345) = 2.32$, $p < .01$, $MS_e = 5{,}407.23$, and between presentations and word versus nonword, $F(15, 345) = 2.46$, $p < .01$, $MS_e = 3{,}656.31$.

The error rates are presented in Table B1. Several subjects had error rates of zero in many cells of the design, so no statistical analysis was attempted. Note, however, that error rates tended to be lower for repeated items than for new items and lower for words than for nonwords.

Experiment 2

Method

Subjects. In all, 26 subjects were recruited from the population sampled in Experiment 1.

Apparatus and stimuli. Apparatus and stimuli were the same as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, with the following exceptions: In each block of the experiment, one word and one nonword were presented once, one was presented twice, one 4 times, one 6 times, one 8 times, and one 10 times, for a total of 62 trials. The lag between successive repetitions varied randomly. The mean and range of lags decreased as the number of repetitions increased.

Table B2
Error Rates for the Lexical Decision Task in Experiment 2

	Condition	
Presentation	Word	Nonword
1	.06	.05
2	.02	.04
3	.02	.03
4	.01	.03
5	.02	.03
6	.02	.02
7	.01	.02
8	.02	.03
9	.00	.03
10	.01	.02

A different sample of words and nonwords was selected for each block, for a total of 10 blocks. A different random sample of stimuli was used for each subject, and the order of trials within blocks was randomized separately for each subject.

Subjects were instructed as in Experiment 1.

Results

An ANOVA on the reaction times revealed a main effect of number of presentations, $F(9, 225) = 65.41$, $p < .01$, $MS_e = 961.51$; a main effect of word versus nonword, $F(1, 25) = 60.77$, $p < .01$, $MS_e = 5{,}276.15$; and an interaction between presentations and word versus nonword, $F(9, 225) = 5.28$, $p < .01$, $MS_e = 650.99$.

An ANOVA on the standard deviations revealed a main effect of number of presentations, $F(9, 225) = 12.37$, $p < .01$, $MS_e = 1{,}279.93$, a main effect of word versus nonword, $F(1, 25) = 14.65$, $p < .01$, $MS_e = 2{,}955.69$, and a significant interaction between them, $F(9, 225) = 2.11$, $p < .05$, $MS_e = 1{,}118.63$.

The error rates are presented in Table B2. Again, no statistical analyses were attempted, but the error rates tended to decrease with repetition and tended to be lower for words than for nonwords.

Table B3
Error Rates From Lexical Decision Task in Experiment 3

	Lag = 12		Lag = 24	
Presentation	Word	Nonword	Word	Nonword
1	.06	.07	.06	.04
2	.02	.04	.02	.03
3	.03	.05	.02	.04
4	.03	.04	.02	.04
5	.01	.03	.01	.04
6	.02	.04	.04	.04
7	.03	.03	.03	.03
8	.04	.05	.02	.04
9	.01	.02	.02	.02
10	.02	.05	.02	.02
11	.02	.05	.02	.02
12	.01	.02	.03	.02
13	.04	.05	.03	.01
14	.02	.02	.02	.02
15	.02	.07	.03	.03
16	.01	.02	.04	.01

Experiment 3

Method

Subjects. In all, 32 subjects from the population sampled in the previous experiments served in Experiment 3. A total of 16 subjects served in the mean-lag-12 condition and 16 served in the mean-lag-24 condition.

Apparatus and stimuli. Apparatus and stimuli were the same as in the previous experiments.

Procedure. The procedure was the same as in the preceding experiments with the following exceptions: Each word and nonword appeared once in each block of trials. Some words and nonwords appeared in only 1 block, others appeared in 2 successive blocks, others appeared in 4, 8, and 16 successive blocks. The mean lag and the range of lags was the same for each number of repetitions. The mean lag was varied between subjects by manipulating the number of stimuli in each repetition condition presented in a single block of trials. One half of the subjects (16) had a mean lag of 12; only one stimulus from each repetition condition occurred in each block, except for the 16-repetition condition, in which two words and two nonwords occurred in each block. Each subject experienced four sets of 16 blocks of 12 trials. The other half of the subjects (16) had a mean lag of 24; two stimuli from each repetition condition were presented each block, except for the 16-repetition condition, in which four words and four nonwords were presented each block. Each subject experienced two sets of 16 blocks of 24 trials.

Each subject received a different random sample of the 340 stimuli, and the order of stimuli within blocks was randomized separately for each subject.

Results

An ANOVA on the reaction times revealed a significant main effect of repetition, $F(15, 450) = 17.70$, $p < .01$, $MS_e = 2{,}333.58$; for word versus nonword, $F(1, 30) = 78.29$, $p < .01$, $MS_e = 13{,}943.47$; and for the interaction between them, $F(15, 450) = 2.44$, $p < .01$, $MS_e = 1{,}086.01$. There was no significant effect of mean lag, $F(1, 30) < 1$, $MS_e = 218{,}538.63$, and no interaction between lag and number of presentations, $F(15, 450) < 1$, $MS_e = 2{,}333.58$, or between lag and word versus nonword, $F(1, 30) = 1.27$, $MS_e = 13{,}943.47$.

An ANOVA on the standard deviations revealed a significant main effect of repetition, $F(15, 450) = 5.47$, $p < .01$, $MS_e = 3{,}003.22$, and a significant effect for word versus nonword, $F(1, 30) = 11.46$, $p < .01$, $MS_e = 5{,}307.10$. No other effects were significant.

The error rates appear in Table B3. No statistical analyses were attempted. Error rates tended to decrease with repetition and to be lower for words than for nonwords.

Appendix C

Details of Method and Results for Experiment 5

Method

Subjects. A total of 48 subjects from an introductory psychology class served as subjects to fulfill course requirements.

Apparatus and stimuli. The apparatus was the same as that used in Experiments 1–3. The stimuli were five-letter words, pronouncible nonwords, and unpronouncible nonwords. The words were nouns selected from the Kucera–Francis (1967) norms to match exactly the distribution of log frequencies of the four-letter words used in Experiments 1–3. The average absolute frequency was 75.27 per million, with a range of 8 to 787. There were 340 words in total. The nonwords were made by substituting letters in the words, making a total of 340 pronouncible and 340 unpronouncible nonwords. Pronouncibility was determined by consensus of three native speakers of English.

Procedure. In the training phase, subjects saw words, pronouncible nonwords, and unpronouncible nonwords. Some of them were presented in only 1 block, others in 2 consecutive blocks, others in 4 consecutive blocks, others in 8 consecutive blocks, and others in 16 consecutive blocks. Altogether, there were 16 stimuli of each type presented once and 8 stimuli of each type presented 2, 4, 8, and 16 times. Each block involved a total of 48 trials, and altogether, there were 16 training blocks.

Subjects in the consistent interpretation groups ($n = 12$ per group) made lexical decisions or pronunciation decisions throughout the training phase. Subjects in the varied interpretation groups alternated between lexical decisions and pronunciation decisions each block throughout training, one half beginning with lexical decisions and one half beginning with pronunciation decisions. Because of the way the blocks were structured, subjects interpreted each stimulus in one way on odd-numbered presentations and the other way on even-numbered presentations, regardless of the total number of presentations each stimulus received.

In the transfer phase, subjects saw the stimuli they were presented with in the training phase randomly intermixed with 16 new words, 16 new pronouncible nonwords, and 16 new unpronouncible nonwords. The stimuli were presented one at a time in the center of the CRT for 500 ms, and subjects gave verbal estimates of the frequency with which they were presented in the training phase. The experimenter sat in the room with the subject and typed each frequency estimate into the computer. Subjects were told that some stimuli were new and some had been presented 16 times, so their frequency estimates should range between 0 and 16.

Results

Training phase. ANOVAs were performed on the benefit scores. Two separate analyses were performed, one for lexical decision subjects (i.e., the consistent lexical decision group and the varied interpretation group that began with lexical decisions) and one for the pronunciation decision subjects (i.e., the consistent pronunciation group and the varied interpretation group that began with pronunciation decisions). Each analysis involved consistent versus varied interpretation as a between-subjects factor, and stimulus type (word vs. pronouncible nonword vs. unpronouncible nonword) and number of presentations (3, 5, 7, 9, 11, 13, and 15) as within-subjects factors. The number-of-presentations factor included only the odd-numbered presentations, for which the same decision was made in each group.

For lexical decision subjects, the main effect of consistent versus varied interpretation was significant, $F(1, 22) = 10.79$, $p < .01$, $MS_e = 44{,}573.09$, as was the main effect of stimulus type, $F(2, 44) = 5.71$, $p < .01$, $MS_e = 11{,}068.67$, and the main effect of number of presentations, $F(6, 132) = 4.28$, $p < .01$, $MS_e = 3{,}297.74$. There were significant interactions between presentations and stimulus type, $F(12, 264) = 1.80$, $p < .05$, $MS_e = 1{,}590.09$, and between presentations, stimulus type, and consistent versus varied interpretation, $F(12, 264) = 1.85$, $p < .05$, $MS_e = 1{,}590.09$.

Table C1
Error Rates for Lexical Decision and Pronunciation Tasks in Experiment 5

	Lexical decision						Pronunciation					
	Consistent			Varied			Consistent			Varied		
Presentation	WORD	PRNW	UPNW	WORD	PRNW	UPNW	WORD	PRNW	UPNW	WORD	PRNW	UPNW
1	.11	.15	.02	.04	.12	.00	.10	.16	.02	.03	.12	.02
2	.06	.16	.01				.06	.18	.01			
3	.04	.16	.01	.00	.20	.00	.04	.17	.01	.00	.09	.02
4	.04	.14	.02				.05	.15	.02			
5	.04	.15	.02	.00	.15	.00	.03	.16	.03	.01	.06	.03
6	.04	.13	.02				.03	.15	.02			
7	.03	.13	.01	.02	.18	.01	.05	.15	.01	.01	.07	.04
8	.04	.13	.00				.04	.14	.00			
9	.01	.16	.01	.01	.04	.05	.01	.18	.01	.00	.07	.04
10	.02	.18	.00				.02	.19	.00			
11	.08	.16	.00	.01	.02	.03	.06	.18	.00	.03	.07	.03
12	.03	.16	.00				.02	.18	.01			
13	.05	.16	.00	.00	.18	.03	.03	.18	.00	.01	.05	.04
14	.04	.16	.01				.03	.18	.01			
15	.01	.15	.00	.05	.19	.00	.01	.14	.00	.01	.07	.04
16	.05	.12	.00				.05	.09	.00			

Note. PRNW = pronouncible nonword; UPNW = unpronouncible nonword.

For pronunciation subjects, the main effect of consistent versus varied interpretation was not significant, $F(1, 22) < 1$, $MS_e = 61{,}977.54$, but the main effect of stimulus type was significant, $F(2, 44) = 13.91$, $p < .01$, $MS_e = 18{,}579.01$. No other effects were significant.

The error rates are presented in Table C1. No statistical analyses were performed on the error rates.

Transfer phase. ANOVAs were performed on the frequency estimates. As before, one analysis was performed on the lexical decision groups and one on the pronunciation decision groups. For lexical decision subjects, there were no effects of consistent versus varied interpretation; neither the main effect, $F(1, 22) = 2.99$, $p < .10$, $MS_e = 41.608$, nor the interactions were significant. However, there were significant effects of number of presentations, $F(5, 110) = 109.65$, $p < .01$, $MS_e = 4.696$; stimulus type, $F(2, 44) = 10.96$, $p < .01$, $MS_e = 5.861$; and the interaction between presentations and stimulus type, $F(10, 220) = 15.93$, $p < .01$, $MS_e = 1.418$.

For pronunciation subjects, there were no significant effects of consistent versus varied interpretation; neither main effect, $F(1, 22) < 1$, $MS_e = 26.433$, nor interactions. There were significant effects of presentations, $F(5, 110) = 220.20$, $p < .01$, $MS_e = 2.817$; Presentations × Stimulus Type, $F(10, 220) = 14.62$, $p < .01$, $MS_e = 1.189$; and Group × Presentations × Stimulus Type, $F(10, 220) = 3.25$, $p < .01$, $MS_e = 1.189$.

Received February 3, 1987
Revision received November 10, 1987
Accepted December 16, 1987 ■