The Nature of Cognition

edited by Robert J. Sternberg

The MIT Press

From The MIT Press

MITCogNet

This book was set in Sabon by Achorn Graphic Services, Inc. and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

The nature of cognition / edited by Robert J. Sternberg.
p. cm.
"A Bradford book."
Includes bibliographical references and indexes.
ISBN 0-262-19405-8 (hardcover : alk. paper).—ISBN 0-262-69212-0 (pbk. : alk. paper)
1. Cognition. 2. Mental representation. I. Sternberg, Robert J.
BF311.N37 1998 97-47622
CIP

10

Response Time versus Accuracy in Human Memory

Michael Kahana and Geoffrey Loftus

One of the first decisions confronting a behavioral scientist is the choice of a measurement instrument that appropriately captures some relevant aspect of human behavior. Consider a typical memory experiment. A subject is presented with a list of words to remember. Immediately after studying the list, a word is presented, and the subject's task is to judge, as quickly and accurately as possible, whether the word was shown in the studied list. Data in this experiment consist of both the subject's response ("yes" or "no") and the time it took the subject to make the given response (called *response latency, response time,* or *reaction time;* hereafter, RT). Many scientists seem to religiously adhere to the study of either response accuracy or response time; rarely are both investigated simultaneously in a given experimental design. Is this a mistake, or are accuracy and response time perhaps just two sides of the same coin—two measures that can be used interchangeably, depending on which is more convenient in a given experimental design? The goal of this chapter is to attempt to answer this question through a selected review and analysis of some of the basic experimental results and theoretical issues in the area of human memory.

Although interest in RTs has been around for a long time (e.g., Donders, 1868/1969; Helmholtz, 1850), until recently research in human memory has been almost exclusively concerned with measures of response accuracy. In a survey of memory texts published during the 1970s (Baddeley, 1976; Crowder, 1976; Hall, 1971; Kausler, 1974; Murdock, 1974), fewer than 4% of the experiments cited reported data on RTs. Beginning in the late 1960s, however, a whole host of new problems emerged that required the use of RT as the measure of interest: semantic-priming effects (e.g., Meyer & Schvaneveldt, 1971); perceptual priming (Neely, 1981); implicit serial, or sequence, learning (Jimenez, Mendez, & Cleeremans, 1996; Reber, 1967); and short-term memory (Sternberg, 1966)—each of which will be addressed in this chapter. However, it was not until the mid-1970s, when real-time personal computers became standard tools in the psychological laboratory, that the study of RTs became standard in the field. A recent text on human memory (Anderson, 1995) contains a healthy mix of accuracy and RT data.

Not only is there now a heightened interest in RT within cognitive research, but new experimental techniques that combine measurement of processing time and response accuracy have emerged.[1] Later in this chapter, we examine some of these techniques in detail. In discussing the relation between RT and accuracy in human cognition, we will focus primarily on data and theory within the domain of human memory.

Accuracy and Interresponse Times in Free Recall

A first analysis of memory tasks reveals that making a task harder increases error rates and RTs. As a case study, consider the correlation between measures of accuracy and measures of RT in one of the classic verbal memory tasks—*free recall.* In free recall, subjects are presented, one by one, with a set of to-be-remembered items and are then asked to recall as many items as they can remember in any order. The task is "free" because unlike most other memory tasks, the experimenter exerts minimal control over the retrieval process; all cues other than the general cue to recall the list items are internally generated by the subject.

The free recall task is deceptively simple. The experimenter asks the subject a very simple question, but the subject is free to do a great many things. Consider first the nature of the responses. How many list items did the subject recall? In what order were the items recalled? Were any nonlist items recalled? What was the relation between these items and the items in the studied list? How much time elapsed between successive responses? Do these interresponse times vary as a function of the number of items recalled or the length of the list? These questions just begin to point out the wealth of data obtained using this task.

An initial examination of recall accuracy reveals several regularities. Early and late list items are remembered better than items from the middle of the list. The advantage for the first and last few items are referred to as a *primacy* effect and a *recency* effect, respectively. The curve that describes the relation between the position of the items in the list and the probability of recall is termed a *serial position curve.* These results are remarkably stable across subjects, stimulus materials, and many incidental characteristics of the experimental design (Greene, 1992; Murdock, 1962).[2]

It is not sufficient to merely describe these data; the goal of cognitive science is to characterize the memory processes that produce the observed results. Much research has been devoted to understanding the process of free recall, and one successful model of this task is the *Search of Associative Memory* (SAM) model (Raaijmakers & Shiffrin, 1980, 1981; Shiffrin & Raaijmakers, 1992). A central notion in memory research, which is captured in SAM, is that items, processed sequentially or in a common temporal context, become *associated* or linked with one another. In terms of the data, an association between two items, *A* and *B,* simply means that the likelihood of recalling *B* is increased in the presence of *A* (either as an externally or internally provided cue). Is this true in our free recall task? Kahana (1996) reanalyzed data from a number of free recall studies and found that after recall of a given list item, the probability of recalling one of its neighbors (in terms of its position in the studied list) is greatly enhanced. A *conditional response probability* (CRP) function relates the probability of recalling a given item to its distance (in the study list) from the last item recalled. Figure 10.1 (left panel) shows the CRP function for data obtained by Murdock and Okada (1970).

Two aspects of the CRP functions are consistently obtained in studies of free recall: contiguity and asymmetry. Contiguity refers to the finding that items tend to be recalled after other items that were studied in adjacent list positions. For example, item 6 is more likely to be recalled immediately after item 5 than immediately after item 3. Asymmetry refers to the finding that among successively recalled items that were adjacent in the study list, forward transitions (item 5 then item 6) are about twice as likely as backward transitions (item 6 then item 5).

So far, we have just considered accuracy. What can be said of the *interresponse times* (IRTs) between successively recalled items? Like CRP

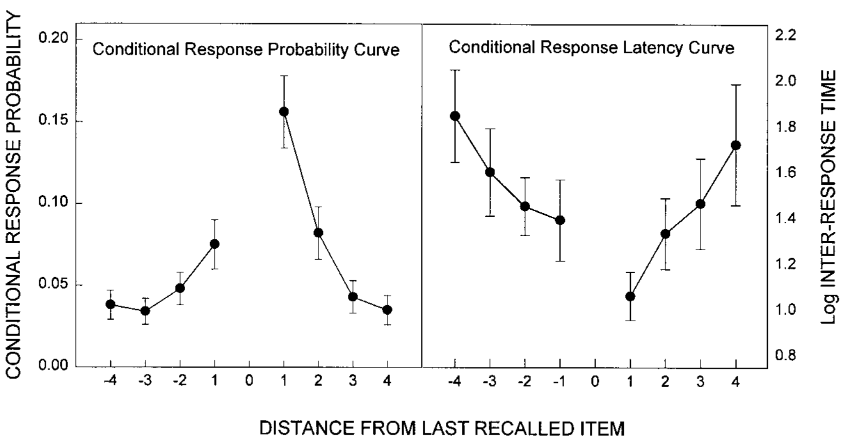


Figure 10.1
Conditional response probability curve (left panel) and conditional response latency curve (right panel) for Murdock & Okada's (1970) study of free recall. Log interresponse time (IRT) is computed as ln $(1 + IRT)$. Error bars reflect 95% confidence intervals around each mean. Confidence intervals were calculated using the Loftus & Masson (1994, appendix B) procedure for within-subject designs.

curves, *conditional response latency* (CRL) functions relate IRTs between successively recalled items to their proximity in the original study list. CRL data from Murdock and Okada (1970) are shown in figure 10.1 (right panel). IRTs are short when neighboring list items are recalled successively. IRTs increase as the separation between the items' positions in the study list increases.

The IRT functions mimic the basic result portrayed in the CRP functions—namely, the more likely the transition, the faster the transition. It is tempting to say that both CRP and CRL functions reflect the operation of a single latent construct—associative strength.[3] Nearby items are more strongly associated with each other than are distant items. The stronger the association, the higher the probability and the shorter the IRT between successively recalled items. This is one version of a strength theory of memory—accuracy and IRTs are just two measures of the strength of information stored in memory.

When average accuracy and RT data show similar patterns, we are tempted to hypothesize that these commonalities are indicative of a single underlying process. However, this is not necessarily the case. To take a common example, height and weight show highly similar patterns and yet it is unlikely that they reflect a single underlying variable. Different eating behaviors can affect weight without having any effect on height.

Semantic Clustering

Preexperimental semantic relations among list items also exert a powerful influence on recall order and on recall accuracy.[4] This is often investigated using a *categorized free-recall* task. Subjects study a list of words drawn from a number of different categories (e.g., *airplane, ruby, dog, celery, diamond, car, truck, elephant, tomato, mule, cabbage, boat*). These words are presented in a random order, and subjects are asked to recall the items in any order they like (standard free-recall instructions). The relevant data are the order of recall and the IRTs between successively recalled items. When subjects recall a categorized word list, items belonging to the same category are usually recalled successively and in rapid succession (short within-category IRTs). These categorically related word clusters are separated by long between-category IRTs (Patterson, Meltzer, & Mandler, 1970; Pollio, Kasschau, & DeNise, 1968; Pollio, Richards & Lucas, 1969; Wingfield, Lindfield, & Kahana, 1998).

Although there is substantial data on free recall of categorized lists, there is a paucity of data on the effects of interitem similarity on free recall of random word lists. Unfortunately, studies that have carefully measured interitem similarity and output order in free recall (e.g., Cooke, Durso, & Schvaneveldt, 1986; Romney, Brewer, & Batchelder, 1993) have not simultaneously collected data on IRTs.

Like the data on conditional response probability and latency, semantic cluster effects can be interpreted as reflecting differences in associative strength. Items that are similar in meaning, or members of a common category, are more strongly associated. These items will tend to be recalled together, and the IRTs will be very short. To get from recalling items within a given category to recalling the items in the next category requires subjects to rely on the weaker associations that link all of the experimental items together—associations that stem from the common experimental context in which the items were studied.

Exponential Increase in IRTs

Another feature of the categorized recall is that within-category and between-category IRTs start out fast and slow down with each transition. This finding mirrors a basic result observed in free recall of random word lists: IRTs increase *exponentially* with output position. Figure 10.2 shows the increase in IRTs with output position reported by Murdock and Okada (1970). Rohrer and Wixted (1994) have shown that this finding holds up under variations in list length, presentation rate (of the list items), and a number of other variables.

What causes this increase in IRTs? According to strength theory, items with the strongest representations are recalled first and fastest. The remaining items, being necessarily weaker, take longer to recall, thus producing the accelerating interresponse times with output position. Note

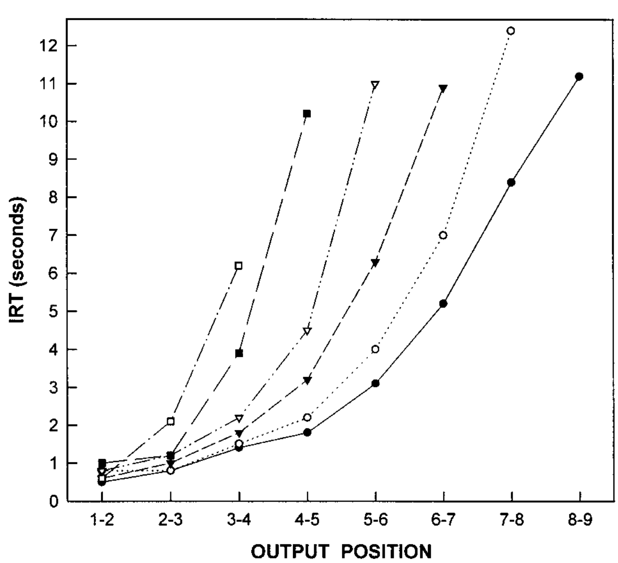


Figure 10.2
This figure shows data from Murdock & Okada (1970) illustrating the exponential increase in IRTs with output position. Subjects in this experiment studied lists of 20 common words presented visually. Vocal responses were tape recorded and IRTs were measured. Each of the six curves in this graph represents a different total number of words recalled (4–9).

that this view does not assume that recalling some items has an effect on recall of subsequent items.

A more cognitively oriented model might propose that another process causes the IRTs to increase. One such model, the *random search-with-replacement* model, has been advocated by Rohrer and Wixted (1994; see also Wixted & Rohrer, 1994). In its simplest form, this is a pure retrieval model. According to this account, recall involves two stages. First, subjects search through recently activated items in memory (termed the *search set*) and sample an item for possible recall. If the item has already been recalled, it is rejected. If it has not already been recalled, the item is recalled. Assuming that it takes time to sample each item, as the ratio of recalled to non-recalled items increases in the search set, the time to recall the remaining items will increase as a consequence of the resampling and rejection of items already recalled. The process of random search with replacement mathematically predicts exponential growth of IRTs (McGill, 1963).

Wixted and Rohrer's (1996) random search with replacement account of IRTs in free recall can be seen as a specific instantiation of the general notion of output interference: that the act of recalling list items impairs access to other list items (cf. Tulving & Arbuckle, 1963). Such interference could be due to the resampling of the recalled items, or to a direct effect of the recalled list items on the accessibility of the not-yet-recalled memories.

Rohrer and Wixted take the generality of the exponential growth of IRTs in free recall as support for the notion of random search with replacement. Although they acknowledge that retrieval in free recall is influenced by many factors not captured in the oversimplistic random search model, they believe that a process akin to resampling is the likely cause of the rapid growth in IRTs with output position. This view is very different from strength theory in that random search with replacement argues that changes in accuracy and RT may be caused by primarily different, though interdependent, memory processes. Nevertheless, without independent evidence for the role of resampling in free recall, it is hard to reject the position that accuracy and RT in free recall are simply two sides of the same memory-strength coin.

Analysis of Memory under Conditions of High Accuracy

We began this chapter by raising the question of whether RT and accuracy are two sides of the same coin. If they are, why not let those who study accuracy live on in blissful ignorance of those who study RT, and vice versa? One reason not to do this revolves around the investigation of well-learned tasks such as reading, speech, naming objects, or performing a practiced motor sequence. In these tasks, people rarely make errors, yet speed may be of the essence. Therefore, to study tasks that are performed essentially without errors, we must consider RTs.

It is probably fair to say that almost all RT research is concerned with tasks where error rates are negligible. Entire areas of cognitive science rely on RTs as their exclusive source of data. For example, one major area of memory research is concerned with the structure of preexperimental semantic representations. These researchers use a variety of techniques including *lexical decision tasks* (LDT; i.e., deciding whether a letter string, such as VOLVAP, is a word) and *sentence verification tasks* (answering yes or no to questions such as "Is a canary a bird?"). RT data from these tasks provide insights into mental processes without making reference to response accuracy.

Even in the case of memory for newly acquired information, there are situations in which performance is relatively error free. Consider the learning of words in a foreign language. Initially subjects will make many errors, but after sufficient repetition, errors will be negligible. Considerable research has shown that RTs speed up dramatically even after accuracy reaches 100%. The reduction in RT with practice is characterized by what is called a *power law* (Newell & Rosenbloom, 1981). In almost any cognitive task, RT varies with practice according to an equation of the form

$$RT = aP^{-b} \tag{10.1}$$

where P is the number of practice trials, and a and b are positive constants that depend on the details of the material, the kind of practice, and the type of learning task. Such regularity summarizes a great deal of data across a variety of domains of cognition and serves as a benchmark that theories must meet.

A nice illustration of the power law can be found in a study by Woltz, Bell, Kyllonen, and Gardner (1996, experiment 3). Woltz et al. examined the contributions of *instance memory* and *rule memory* to the acquisition of a cognitive task. Subjects were given four-digit strings that could be transformed into a single digit by the sequential application of some combination of different rules. Each rule transforms two adjacent digits to a single digit (for example, if two adjacent digits are successive they are transformed into the next item in the sequence: 56 becomes 7; 65 becomes 4). Figure 10.3 shows RT and accuracy as a function of training. Over the two training blocks, RT decreased dramatically and in accord with the power law. Accuracy, on the other hand, remained essentially constant.

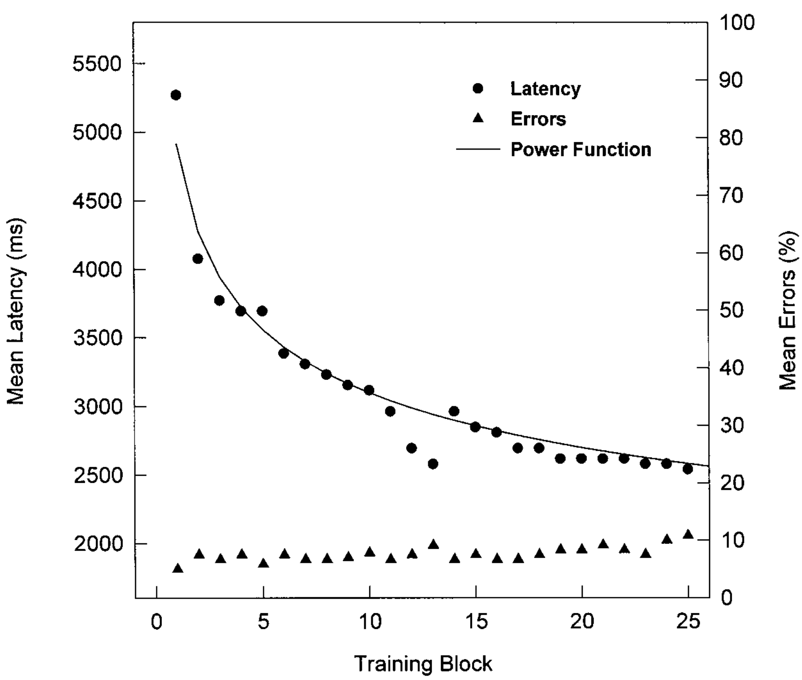


Figure 10.3
Accuracy and latency data in a digit-recoding task (Woltz, Bell, Kyllonen, & Gardner, 1996, experiment 3). The smooth line through the latency data represents the best-fitting power function. These data illustrate what is often called the *power law of learning.*

With practice, subjects got faster at transforming the four-digit strings into single digits. Is this because they were better at using the rules or because they had memory for instances of digit pairs, triples, or quads that they could simply recall? In a final phase of the experiment, subjects were given three types of multidigit strings to recode: strings that were identical with those recoded (i.e., transformed into a single digit) in the earlier training phases (old strings/old rules), new strings based on previously practiced recoding steps (new strings/old rules), and new strings that required new sequences of recoding operations (new strings/new rules). Woltz et al. (1996) found the most improvement for old strings/old rules, an intermediate amount for new strings/old rules, and the least transfer to new strings/new rules. These results were interpreted as evidence for both *instance-based learning* (i.e., learning of particular examples) and *rule-based learning*.

RT and Accuracy in Implicit Sequence Learning and Explicit Sequence Prediction

Sometimes RT (or accuracy) can reveal memory in the absence of intention to retrieve information from a learned episode. This type of memory is referred to as *implicit memory* and distinguished from intentional retrieval, which is termed *explicit memory* (Tulving & Schacter, 1991). Implicit memory has been examined for single items (Jacoby & Dallas, 1981; Tulving, Schacter & Stark, 1982), associations between items (Goshen-Gottstein & Moscovitch, 1995a, 1995b; Graf & Schacter, 1995), and sequences (Reber, 1967).

In studying implicit sequence learning, strings of letters or digits are generated through the use of a finite state grammar (Cleeremans & McClelland, 1991; Reber, 1967). To create the sequence of stimuli, one starts at a given node (figure 10.4) and probabilistically chooses a path to another node. Note that the first node is identical to the last node, so this generative process may be repeated indefinitely. The label of the chosen path is the current stimulus. To introduce some additional noise into the task, on some proportion of trials the stimuli are chosen randomly (i.e., without using the finite state grammar).

In one version of this approach (Jimenez et al., 1996), a simple RT task was employed; letters were shown one by one on the screen, and

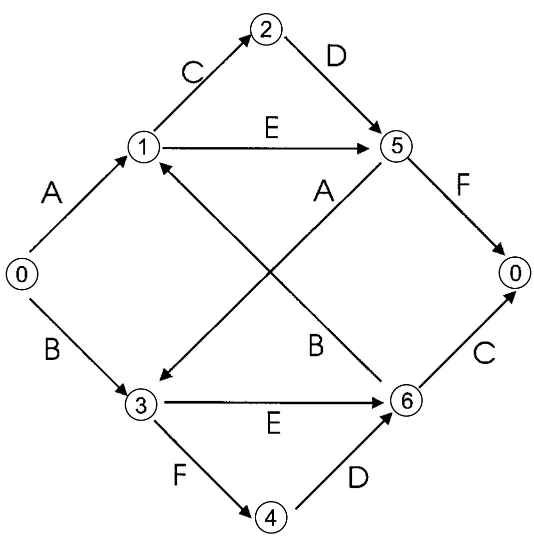


Figure 10.4
Depiction of the artificial grammar used by Jiménez, Méndez, & Cleeremans (1996). An artificial grammar is used to create a sequence of stimuli based on probabilistic rules. To create a sequence, begin with node zero and choose a random path. For simplicity, let us suppose that each path is chosen with an equal probability. If the path from 0 to 1 is chosen, the sequence begins with *A*. From node 1 there are two possible transitions. If node 5 is chosen, the second element in the sequence is *E*. This process may repeat indefinitely. Note that the sequence AEFBFDBC is generated by traversing the nodes 015034612. Not all sequences are valid. For example, there is no way to generate the sequence ABCD from this artificial grammar.

subjects pressed the key corresponding to the displayed letter. Unbeknownst to the subjects, there was a probabilistic pattern to the sequence of stimuli (letters). The basic result was that RTs consistently improved over trials. If a new grammar was switched to, RTs slowed down—this shows that the facilitation in performance was not simply due to a learning-to-learn effect. As is generally the case, the facilitation in RTs followed a power law. This improvement has typically been taken as evidence for implicit sequence learning.

Jiménez et al. (1996) introduced a second task in which subjects were instructed to press the key of the letter they thought would follow the probe letter. In this task, accuracy was the primary variable of interest and retrieval was explicit. Interestingly, subjects were able to predict the next letter at a rate substantially better than chance.

Because of the introduction of random letters every so often in the sequence, Jimenez et al. (1996) were able to assess the degree to which prior items in the sequence facilitated subsequent performance. They found that in both the explicit prediction task and the implicit reaction time task, two prior items added significantly to a single prior item, but a third prior item did not significantly improve performance. Contrary to expectations, implicit memory showed the same general pattern as explicit memory. Also, accuracy measures exhibited the same basic pattern of results as RT measures.

Accuracy and RT analysis of the Ranschburg Effect

Sequences of items that contain a repeated element are harder to reproduce than sequences consisting of all unique elements. For example, the sequence of digits 723856391 is harder to recall in order than the sequence 723856491. This finding is known as the *Ranschburg effect.*[5] At first glance, one would expect repetition of a list item to improve rather than worsen memory for ordered lists. A list with a repeated element has fewer different elements to be learned. This is especially evident in the case of words where the pool of possible elements is very large. In addition, we might expect processing the first of the repeated elements to facilitate, or prime, the processing of its repetition.

Rather than just considering subjects' overall ability to reproduce the list, Crowder (1968) and Jahnke (1969, 1970, 1972) examined error rates for individual list elements. They found that repeating elements at separated list positions resulted in impaired memory *only* for the second instance of the repeated element. In a sequence such as 723856391, subjects made more errors on the second of the repeated 3s than on an item from the same position in a control list (containing all unique elements). If, however, an element was repeated successively, subjects were *better* at recalling both repeated elements, but showed no facilitation or impairment in recalling the rest of the list. In a sequence such as 723385691, subjects performed better on the repeated 3s than on items from the same positions in a control list.

Greene (1991) suggested that a guessing strategy might account for the Ranschburg effect. Most studies of the Ranschburg effect employ lists of between 8 and 10 digits with only a single repeated element. When the set of elements is determined (i.e., the digits 0–9), the task only requires that subjects remember which elements belong in which positions. Even with lists of 8 or 9 digits, subjects have most of the information about the list elements, and the task depends primarily on remembering the order. At the end of the list, where performance is generally poor,[6] subjects are most likely to guess. Because only a single element is repeated, it makes sense to guess from among the elements (digits) that they have not already recalled. This will boost performance for all but the repeated items. Because the second repeated item is usually near the end of the list, where poor memory encourages guessing, recall of that element will show greatest impairment relative to the control list. Greene (1991) tested this guessing hypothesis by either encouraging subjects to guess liberally or telling subjects not to guess at all. When encouraged to guess liberally, the Ranschburg effect was enhanced. When guessing was strongly discouraged, the Ranschburg effect was eliminated.

Kahana and Jacobs (1998) wondered if a Ranschburg effect would be obtained using latency (IRTs) rather than accuracy as the variable of interest. Subjects studied lists of nine consonants with or without a single repeated element. The process of studying and recalling elements was repeated until each sequence was reproduced perfectly on three successive trials. On these final three perfect trials, the computer recorded the subjects' IRTs between successive recalls. Relative to a control list with no repeated elements, subjects had longer IRTs to the second repeated element if the repeated elements were spaced apart in the list. In contrast, subjects had shorter IRTs to the second repeated element if the repetitions were in nearby list positions. IRTs to the first of the repeated elements were unaffected by the repetition (as compared with control lists of nonrepeated elements). These results are shown in figure 10.5.

This study demonstrates that the Ranschburg effect, previously only known from accuracy data, can also be revealed using latency data (IRTs). But the latency data make Greene's (1991) guessing account far less appealing. Because the latency data are examined only after accuracy

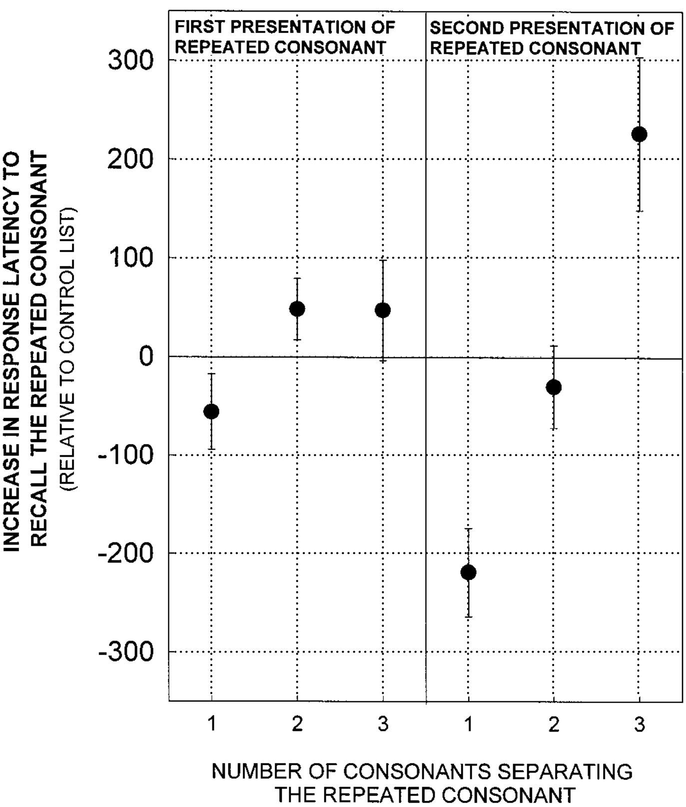


Figure 10.5
Data illustrating the effects of within-list item repetition on reaction times in a sequence recall task (Kahana & Jacobs, 1998). Error bars denote 95% confidence intervals. See text for details.

has reached 100%, it is unlikely that these data are contaminated by guesses. In addition, using lists of nine *consonants* as stimuli makes guessing relatively ineffective. As subjects pass the halfway point in the list, there are 16 possible items for only four remaining positions—guessing is not very helpful under these circumstances. Based on the Kahana & Jacobs study, it seems that the Ranschburg effect is a reliable memory phenomenon that can be revealed using both accuracy and latency measures. Although the parallel finding of Ranschburg interference in both accuracy and RT data may suggest that accuracy and RT are "two sides of the same coin," these two measures provide crucially different kinds of information with respect to the theories of the task.

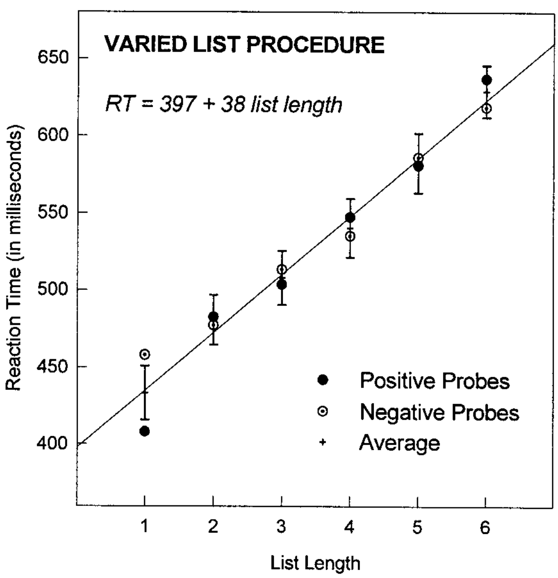
The Subspan Item-Recognition Task

For a sufficiently easy task, one needn't engage in extensive practice to achieve near perfect accuracy. For example, if you ask someone to remember a five-digit number over a period of time during which there is no distracting information presented, error rates will be negligible. *Memory span* is a term used to denote the number of items that a person can reproduce in the correct order without errors. Lists of items that are shorter than an average person's memory span (about seven digits, six letters, or five words; Crannell & Parish, 1957) are called *subspan* lists, and lists that exceed memory span are termed *supraspan* lists.

Sternberg (1966) examined recognition memory for subspan lists. In the subspan item-recognition task (also called the Sternberg task or the memory-scanning task), subjects are presented with a short list of items (digits, words, letters, etc.). Following a brief delay (typically 1–2 seconds), a probe item is shown, and subjects indicate whether the item was one of the elements of the original list. In Sternberg's original experiments, subjects were well practiced at this task.

Because the list is subspan, there are few errors (less than 5%) and consequently the dependent variable of interest is RT. The effects of numerous experimental manipulations on RT have been investigated. These include variations in list length, probe type (list items versus nonlist items), serial position of the probe item (if it is in the list), or recency of the probe item (if it is not in the list), and the kind of materials used (e.g., letters, digits, words, random polygons, etc.). When a probe item is one of the list items, it is called a *positive* probe (because it warrants a positive, yes response). Similarly, nonlist items are called *negative* probes.[7]

In Sternberg's 1966 paper, two procedures were introduced. In the *varied list* procedure, lists are randomly chosen for each trial and list length varies from trial to trial. In the *fixed list* procedure, a given list of items is prememorized and then repeatedly tested. This process is repeated for prememorized lists of various lengths. Sternberg's results are shown in figure 10.6. Panel A presents data obtained using the varied list procedure, and panel B presents data obtained using the fixed list procedure.



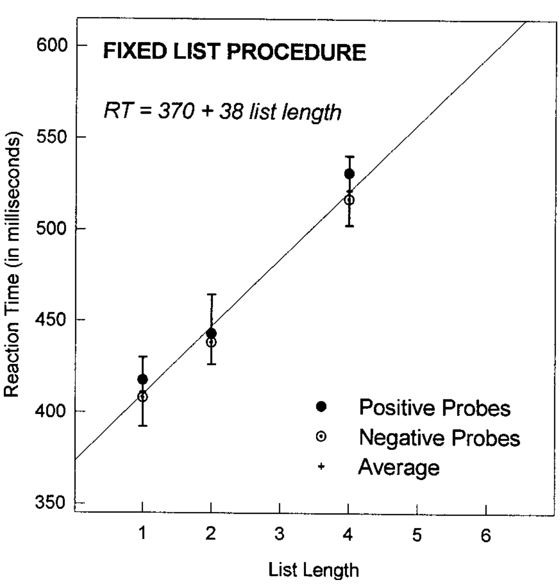


Figure 10.6
Reaction time as a function of list length for Sternberg's (1966) subspan item-recognition experiments. Panel A shows data obtained using a varied list procedure; panel B shows data obtained with a fixed list procedure. In the varied list procedure, items (typically digit, letters, or words) change from trial to trial. In the fixed list procedure, the items are the same for each trial; only the test cue changes. The equations given in each figure characterize the best-fitting line through the average data for positive and negative probes.

In both cases, Sternberg found that mean RT increases linearly with list length. Two features of these data are particularly striking. First, the slopes of positive and negative probes are indistinguishable. Second, the slopes of the linear list-length functions are equivalent for both the fixed list and the varied list procedures.

Sternberg (1966) proposed a simple model to account for these data. He assumed that the probe item is serially compared with each member of the set of items that are activated in memory (the search set). In a *serial* comparison process, a new comparison does not begin until the previous comparison has been completed. This explains why RTs increase linearly with list length (each additional item requires one additional comparison), but why are the slopes identical for positive and negative probes?

Consider what happens if the memory comparison process is *self-terminating*. The probe item is compared with each list item until a match is detected or the list is exhausted. This is called a self-terminating search because the comparison process terminates as soon as a match is detected. Consider a list of three items. If given a positive probe (randomly chosen from among the three list items), there is an equal probability of finding a match after one, two, or three comparisons. On average, two comparisons are required. If a negative probe is given, three comparisons are always required (all three list items must be rejected). What happens if the list length is increased from three to four? A positive probe now requires either one, two, three, or four comparisons, resulting in an average of 2.5 comparisons. A negative probe requires all four comparisons to be made. Consequently, increasing the list length by one item results in an increase of 1 comparison for negative items, but only 0.5 comparisons (on average) for positive items. Thus, the slope for negative probes should be twice as great as the slope for positive probes. This is clearly not the case (see figure 10.6).

To explain the equivalence of slopes for positive and negative probes, Sternberg suggested that the serial comparison process is *exhaustive*. Exhaustive search means that the probe item is compared with every item in the search set, and a decision is not made until all comparisons have been made. This idea may seem unrealistic, but if the comparison process is extremely fast and the decision process is noisy and slow, it makes sense to do all of the comparisons prior to making a decision rather than making a separate decision after each comparison (Sternberg, 1969b).

Sternberg (1969a) presented a more complete description of the basic scanning model. The model has four stages: *stimulus encoding, memory comparison, decision,* and *response.* The following claim is critical to the analysis of this model: a given process or stage is not initiated until the previous stage is completed. This claim is reasonable if a stage acts on information produced by a preceding stage that must be available in a fairly complete form (Sternberg, 1998a). Much debate has centered on the validity of this claim (e.g., Hockley & Murdock, 1987; McClelland, 1979). We will return to this issue later in this chapter. A final important detail of the model is that scanning times needn't be fixed. It is often assumed that the time to scan a given item comes from a distribution of possible values. The shape of the distribution (e.g., normal vs. exponential) and its mean and variance are important in generating model predictions (e.g., Luce, 1986, chapter 11).

As discussed previously, recall performance depends crucially on the recency of the items being tested (see "Accuracy and Interresponse Times in Free Recall," p. 324). One may ask how the RT to recognize an item depends on the item's position in the study list. In particular, are subjects faster at recognizing recently presented items? In Sternberg's short-term item-recognition task significant recency effects are often, but not always, obtained (see McElree & Dosher, 1989, and Sternberg, 1975, for reviews). In the clearest case, Monsell (1978) found dramatic recency (i.e., facilitation of positive responses to recent list items) using either letters (experiment 1) or words (experiment 2) as stimuli (figure 10.7). In Monsell's study, the test probe followed the last list item either immediately (right panel) or after a brief delay (left panel). The delay condition required subjects to name a vowel presented immediately after the last list item (this took an average of 500 ms). This step was performed to prevent subjects from rehearsing the list items during the delay period. In an earlier study, Forrin & Cunningham (1973) showed that increasing the length of an unfilled delay between study and test eliminates the recency effect in short-term item recognition. In general, experimental conditions that reduce or eliminate rehearsal tend to produce large recency effects, and those that allow for rehearsal (e.g., Sternberg, 1966) typically have flat serial-position curves.

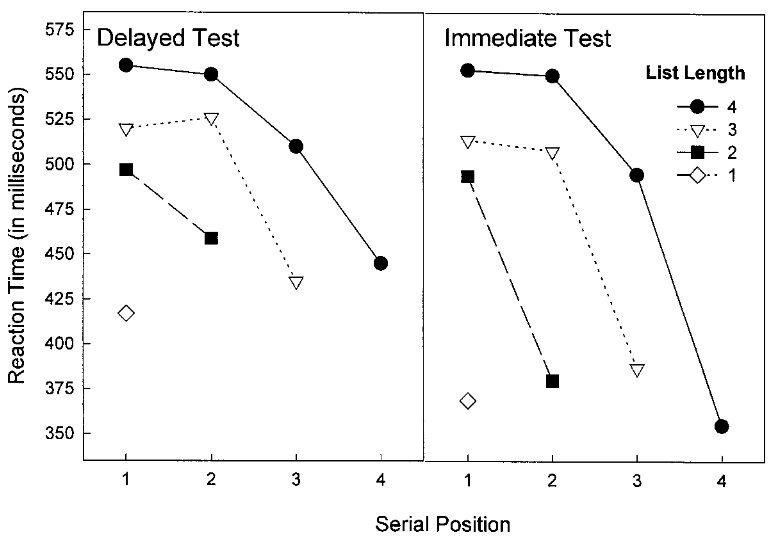


Figure 10.7
Serial position data from Monsell (1978, experiment 1). A fast presentation rate (500 ms/item) was designed to minimize rehearsal. In the immediate test, the probe item immediately followed the presentation of the last list item; in the delayed test a vowel had to be named after the offset of the last list item. As soon as a response was detected, the probe item was presented.

According to Sternberg's serial exhaustive-scanning (SES) model, a response cannot be made until every comparison has been performed. Consequently, the time required to perform the memory scan should be independent of serial position. In light of this, the marked serial position effects obtained by Monsell (1978) and others present a challenge to the Sternberg model. In fact, some authors have rejected the Sternberg model because of this evidence alone. In response, two points need to be made. First, the Sternberg model was designed to explain data obtained under conditions where subjects could freely rehearse highly familiar items (e.g., Sternberg, 1966). Under these conditions, significant serial position effects are consistently absent (Ferrin & Cunningham, 1973). Second, facilitation in performance may be occurring in other stages of the model (Sternberg, 1975). For example, recent items may speed, the encoding of the probe item or the execution of the response—thereby resulting in faster RTs (for a similar argument in the literature on same-different comparisons, see Proctor, 1981). This priming account of the recency effect is difficult to reconcile with the problem of recent negatives. If a probe that was not on the current list was presented as a target on a recent prior list, RTs to respond "no" to the negative probe are significantly increased (e.g., McElree & Dosher, 1989, experiment 2) this finding has proven difficult to reconcile with Sternberg's SES model.

Another challenge to Sternberg's SES model comes from studies that examine list length effects beyond the span of immediate memory. As mentioned earlier, near perfect accuracy is attained either when lists are short (subspan) or through practice (for longer, supraspan, lists). Burrows and Okada (1975) used a prememorized list technique to study RTs in an item recognition task with list lengths far beyond the limits of immediate memory. Their results are shown in figure 10.8. For subspan list lengths (two–six), the slope of the best-fitting line is 37 ms—replicating the classic Sternberg effect. However, a separate line fit to the supraspan lists (lengths greater than eight) yielded a much shallower slope of 13 ms/item. Burrows and Okada also fit a single logarithmic function to their subspan and supraspan data. They found that this function fit all of the data points as well as both the bilinear subspan and supraspan functions, but with fewer parameters. According to the serial exhaustive scanning model, each additional item in the memory set should result in a constant increase in mean RT for both positive and negative probes. These data indicate that the increase in mean RT is not a constant, but varies with list length. This finding is not easily reconciled with the SES model.

So far we have discussed serial position effects (e.g., Monsell, 1978) and list length effects (e.g., Burrows & Okada, 1975) in the context of RT studies of short-term memory. Another major variable that is studied in human memory is repetition. Baddeley and Ecob (1973) wondered what would happen if a single list element in a Sternberg task were repeated. Under these conditions, mean RT is significantly faster for responding to the repeated element than to nonrepeated elements. Like the serial position effects reported by Monsell (1978), these data seem inconsistent with the SES model. If each element must be scanned before a response can be made, it should not matter how many times an element is presented. This critique of the Sternberg SES model assumes that other

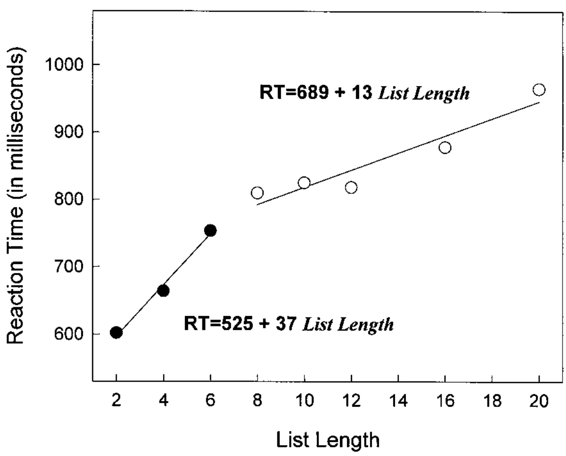


Figure 10.8
Data illustrating the effect of a large range of list lengths on mean reaction time in a probe recognition task (Burrows & Okada, 1975, experiment 2). To achieve nearly errorless performance, lists were "prememorized"; that is, before testing a given list length, the list was already well learned. In the figure, a bilinear function is fit to the data. For short list lengths, the slope of the RT-list length function is similar to results obtained by Sternberg (1966). For longer list lengths, mean RT rises slowly as list length increases (the slope of the best-fitting line is only about 13 ms). As noted by Burrows and Okada (1975), a simple log function fits the data as well as the two linear functions shown here. This raises questions about the linearity of the list length–RT functions reported for short lists.

stages are not influenced by repetition. It is not unreasonable to suppose that the decision stage is executed more quickly when two matches have been registered than when only a single match has been registered.

In the years since the publication of Sternberg's original paper, Sternberg's basic experimental findings have been replicated and extended hundreds of times in studies that manipulated dozens of different experimental variables (see Sternberg, 1975, for a review). Although the data are solid, there has been a long debate about the meaning and interpretation of these findings. Many models have been proposed to account for the basic data, yet none of these models has succeeded in capturing most of the benchmark effects (Sternberg, 1975). Although the simplest version of any model can easily be rejected, the model's creators can often patch things up to correct for the most serious problems. By the 1970s it was already becoming clear that many different types of models can produce identical predictions for data on mean RTs (e.g., Anderson, 1973, Townsend, 1976, 1984).

More recently, attention has shifted from looking at mean RTs to examining the actual shape of the RT distribution. It turns out that although very different types of models can explain the same pattern in the mean RTs, explaining the exact shape of the distribution and how it changes with manipulation of experimental variables is more difficult. Memory theorists have begun to tackle this issue with promising results (Ashby, Tein, & Balakrishnan, 1993; Hockley & Murdock, 1987; Ratcliff, 1978).

The three findings just reviewed, list length, serial position, and repetition effects, all show parallel effects on accuracy and latency. Longer lists are harder to remember than shorter lists. Recently presented items are easier to remember than items presented earlier in the list. Repeated items are easier to remember than nonrepeated items. It may be argued that both the time it takes to recognize an item and the likelihood of successful recognition are reflections of a single construct—the strength of the memory trace.[8] Appealing as this idea may seem, we will later see that the precise nature of the relationship between accuracy and latency may provide important information for testing models of memory.

Task Analysis Using Accuracy and RT Data

If the goal of information-processing research is to break down a complex task into logically distinguishable mental operations and then characterize and model those operations, how do we go about breaking down the complexity of real-world tasks? Among researchers who are concerned with accuracy, the standard method of task analysis is to look for experimental factors (e.g., word frequency or the spacing of repeated elements) that have different effects on different memory tasks or on different aspects of subjects' performance in a given task. Consider the serial position curve in free recall. Presenting the list auditorally results in a larger recency effect (better memory for the last few list items) than does presenting the list visually (this phenomenon is referred to as a *modality effect*). However, the mode of presentation (auditory vs. visual) has no effect on the rest of the serial position curve.[9] This finding is called a *functional* *dissociation* between the recency and the prerecency part of the serial position curve. Another experimental variable, list length, has no effect on the recency part of the serial position curve but has a substantial effect on the prerecency items. With this second, complementary dissociation, the tasks are said to be *doubly dissociated*. Such double-dissociations are sometimes taken to support the view that recency and prerecency items represent the operation of distinct *short-term* and *long-term memory systems*.[10]

If we are willing to make the assumption that some sets of mental operations are arranged (at least approximately) in nonoverlapping stages (i.e., one stage begins only after the prior stage is done with its processing), we can perform a more sophisticated task analysis using mean RT data. This approach is called the *additive factors* method (Sternberg, 1969a; Sternberg, 1998a). The key to this approach is the factorial experimental design. Two or more factors that are known to affect overall RT are manipulated factorially. If each of the factors selectively affects a different processing stage, then total RT should be given by the sum of the separate effects of each factor on RT. If however, the two factors influence a common stage, total RT will deviate from the sum of the separate effects, and the factors can be said to interact, in a statistical sense.

As an example of the additive factors approach, consider an experiment in which the RT-list length relationship is examined as a function of some other variable: in this case the variable of whether or not the test probe is degraded (made to look blurry) via reduction of contrast or randomizing pixels. Sternberg (1967b) conducted this experiment and his results are shown in figure 10.9. The nearly identical slopes of the two functions indicate a lack of interaction between list length and whether the probe item is degraded: that is, the RT difference between the clean and degraded conditions is approximately the same for each value of list length. In this experiment two factors are varied: list length and probe degradation. Probe degradation simply lowers the RT-list length function by a fixed amount. Statistically, it is said that these factors do not interact (see figure 10.9). Such additivity is predicted by a discrete stage model in which stimulus degradation affects one stage (perhaps encoding) and set size affects another stage (perhaps comparison). If two factors affected the same stage, one would expect to find a statistical interaction (i.e., the

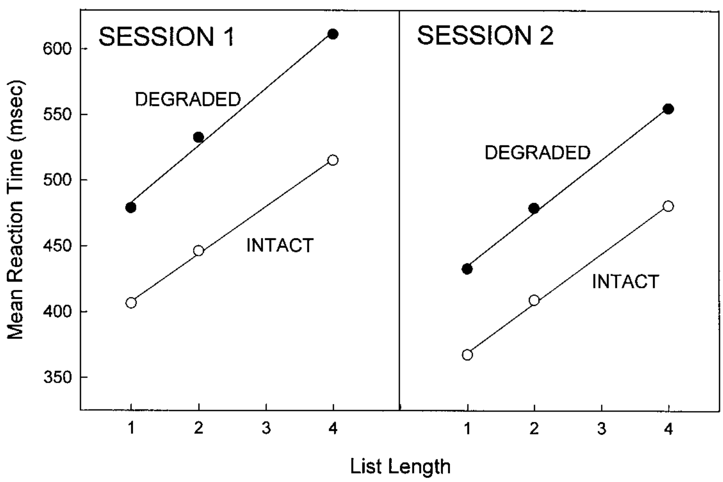


Figure 10.9
Data from Sternberg (1967), illustrating the effect of visually degrading the probe on the RT–list length relation. For both degraded and clean probes, RT is linearly related to list length. There is no interaction, as indicated by the nearly parallel lines for the two conditions. The left panel shows data for the first session of doing the task, and the right panel shows data for the second session. Degree of practice (session 2 versus session 1) does not seem to affect either the slopes of the function or the difference between degraded and intact performance. Rather, practice just speeds everything up (as indicated by the lower intercepts for session 2).

slope of the list length-RT function would be different for degraded and nondegraded stimuli). Sanders (1980) and Sternberg (1998a) reviewed a great deal of evidence from a broad range of factorial RT studies. They found that many variables that would logically be expected to influence different processing stages do have additive effects on RT.

The additive factors method is not without its detractors. An early criticism of the method is that it relies entirely on RT measurements. These measurements may not be comparable across experimental conditions that differ even slightly in accuracy. Pachella (1974) presented a cogent review and critique of the research on RTs during the prior 10 years. As will be described in detail, Pachella pointed out that when you correct for the differences in error rates across conditions, some of the additive effects observed in RT data disappear. This type of error rate correction assumes something called a speed-accuracy trade-off, which will be discussed in the next section.

A subsequent challenge to the additive factors approach came from demonstrations that models that assume continuous transmission of information (i.e., the products of a given stage are constantly available to the next processing stage) can often produce additive effects on mean RT (Ashby, 1982; McClelland, 1979). Roberts and Sternberg (1993) performed a detailed analysis of the McClelland-Ashby model. They found that although the model could predict additive effects on mean RT for some parameter values, the model did not provide a reasonable fit to additivity at the level of the entire RT distribution. Roberts and Sternberg's work exemplifies the recent trend toward fitting RT distributions rather than simply mean RT. Distributional tests provide investigators with significantly greater resolution in distinguishing theories.

In an interesting development in this area, Schweikert (1985) and Roberts (1987) have each expanded the additive factors approach to deal with accuracy and response rate data respectively. Consider a model in which a correct response relies on the completion of two independent processes, A and B. Further, assume that process B must act on the completed output of process A. If processes A and B provide adequate information for a correct response with probabilities $p(A)$ and $p(B)$ respectively, then the probability of a correct response is given by $p(A) \times p(B)$. Converting to logarithms, we can write

$$\log p(A \text{ and } B) = \log p(A) + \log p(B). \qquad (10.2)$$

If each of two factors selectively influence each of the two hypothetical processes, one would expect additive effects of the factors on the logarithm of recall probability. This finding has been observed by a number of investigators in a number of different experimental paradigms (see Schweikert, 1985, for a review).

Complications Introduced by the Possibility of Speed-Accuracy Tradeoffs

As we have noted, several individuals, most notably Wickelgren and Pachella, wrote of serious difficulties that are entailed when RT studies do not consider variation in error rates across experimental conditions (e.g., Corbett & Wickelgren, 1978; Pachella, 1974; Wickelgren, 1977).

Consider the curves shown in figure 10.10A. This figure introduces the concept of a speed-accuracy trade-off. The general idea is simple: the more time you allot to some task, the better you will do at that task. For instance, if you are typing, you could type slowly and make relatively few errors or, alternatively, you could type more quickly and make more errors. In this example, you, the typist, decide what you will do in terms of trading off additional speed for less accuracy.

In figure 10.10A, *condition 1* and *condition 2* refer to two conditions in some RT experiment. Condition 2 is assumed to be more difficult than condition 1; thus conditions 1 and 2 could, for example, be three- and five-item lists in our familiar item-recognition task. For each condition, probability correct is plotted as a function of what is termed *processing time*. For the moment, processing time, which is measured from the beginning of stimulus onset, is an unobservable construct. The idea here is that the onset of the to-be-processed stimulus (e.g., the probe word in the recognition test) triggers appropriate perceptual and cognitive processing. The more such processing occurs, the more information about the stimulus is obtained. At any given processing time, some specific amount of information has been obtained. Probability correct corresponding to that particular processing time is the probability that with *only the information obtained thus far,* a correct response would be made.

The curves that relate probability correct to processing time are called *speed-accuracy trade-off* (SAT) curves. The greater difficulty of condition 2 compared to condition 1 is embodied in the observation that in order to obtain some fixed level of response probability, more processing time is required for condition 2 than for condition 1. Notice that SAT curves are like typical RT curves (e.g., as in a Sternberg paradigm) rotated by 90°. Whereas in a typical RT function, processing time is plotted as a function of amount of required processing (e.g., of memory list length), in an SAT curve amount of processing (measured in terms of probability correct) is plotted as a function of allotted processing time.

Implications of Only Observing RTs

In a typical RT task, SAT curves are not directly measured. Rather, in a given experimental condition, subjects adopt (implicitly or explicitly) some *criterion* point on the SAT curve (just as when you are typing you

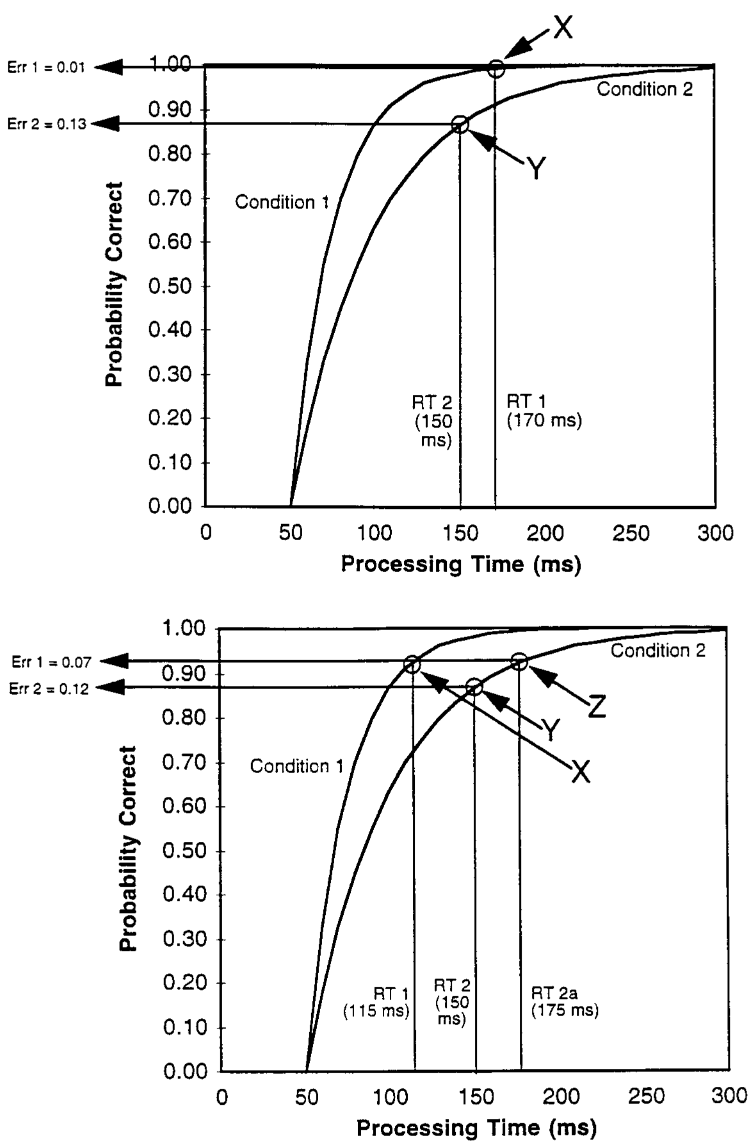


Figure 10.10
Hypothetical speed-accuracy tradeoff (SAT) curves for two conditions. In (A), *X* and *Y* indicate where the speed-accuracy criteria are placed. In (B), *X, Y,* and *Z* indicate where the speed-accuracy criteria are placed. In both panels, *RT1* and *RT2* are the RTs for conditions 1 and 2; *Err 1* and *Err 2* are error rates for conditions 1 and 2. In panel B, *RT 2a* shows the RT for condition 2, assuming an error rate equal to the error rate of condition 1.

must decide how fast you are going to type, which, in turn, will produce some corresponding degree of accuracy). The processing time corresponding to this criterion is the observed RT in the experiment, and the probability correct corresponding to this criterion is 1.00 minus the observed error probability in the experiment.[11]

To illustrate the complexities of doing standard RT experiments, consider the curves in figure 10.10A. In condition 1, the criterion point is labeled *X*. This corresponds to an observed RT of 170 ms, and an observed error rate of .01 (1.00 − .99 probability correct value). In condition 2, the observed RT is 150 ms, and the observed error rate is .13. Clearly, something is amiss. The presumably more difficult condition 2 has a shorter observed RT (150 ms) than does the presumably less difficult condition 1 (170 ms). Thus with RT information only, the experimenter would, incorrectly, conclude that condition 2 is *less* difficult than condition 1.

Fortunately, experimenters would not be quite so naive. Rather, the experimenter would quickly note that the observed error rate in condition 2 (.13) is greater than the observed error rate in condition 1 (.01) and would become suspicious that the shorter observed RT in condition 2 may be due to a speed-accuracy trade-off—that is, in condition 2, observers are (for whatever reason) sacrificing accuracy for increased speed—and it is for this reason that RT is shorter in condition 2 than in condition 1. This would lead the experimenter to suspend judgment about the relative difficulty of the two conditions and rerun the experiment, changing the subjects instructions so as to eliminate this speed-accuracy confounding.

Necessary Conditions for Safe Ordinal Conclusions

Let us imagine that this rerunning produces the data of figure 10.10B. Again, the two speed-accuracy criterion points are labeled *X* and *Y* for condition 1 and 2 (ignore point *Z* for the moment). Now the observed RTs are 115 ms and 150 ms for condition 1 and 2 respectively. Thus, these observed RTs now correctly reflect the greater difficulty of condition 2 compared to condition 1. In addition, the observed error rates are .07 and .12 for conditions 1 and 2, respectively. In short, condition 2 now has both a longer observed RT and a higher observed error rate than does condition 1. Thus, the experimenter could correctly conclude that condition 2 is more difficult than condition 1. In addition, because of the higher observed error rate in condition 2 compared to condition 1, the experimenter would be confident that the longer RT of condition 1 could not have come about because of a speed-accuracy trade-off. To summarize: when one condition (condition 2) produces *both* longer RTs and higher errors than another condition (condition 1), the experimenter can safely conclude that condition 2 is intrinsically harder than condition 1.

Quantitative Interpretational Difficulties

However, the speed-accuracy trade-off problem has not been completely solved even when the data emerge as in figure 10.10B. Suppose the experimenter were interested in the *magnitude by which* the condition-2 processing time exceeded the condition-1 processing time. The best estimate from the figure 10.10B data would be that this value is the difference of the observed RTs, that is, 150 ms − 115 ms = 35 ms. But would this be accurate? No, it wouldn't, because the two conditions differ in terms of error rate as well as in terms of RT.

Suppose the experimenter had been lucky enough that the error rates were identical—say, .07—in both conditions. Now the two speed-accuracy criterion points would be *X* and *Z* on figure 10.10B. Note that the RTs corresponding to condition 1 and 2 would be 115 and 175 ms (the latter is labeled as *RT 2a* in the figure). Therefore, the *real* difference between the two conditions—that is, the RT difference with error rates held constant at .07—would be 175 ms − 115 ms = 60 ms. Quantitatively, this is quite a different conclusion from the 35-ms figure that we would have arrived at from the actual data that entailed the different error rates. This means that many patterns of RT data are difficult to interpret when error rates differ among the conditions.

Consider Sternberg's (1966) finding that RT increases linearly with list length in a subspan item recognition paradigm (see figure 10.6). It was this result that led Sternberg to postulate a serial comparison process (i.e., when the probe item is compared with each element of the memory set, one comparison does not begin until the prior comparison has been completed). But, if error rates vary systematically as a function of list length, it is unlikely that the observed linear RT functions would be obtained under conditions in which error rates did not vary with list length (see Pachella, 1974).

Equalizing Error Rates Is Still Not Enough

Suppose one could carefully control error rates so that they were identical in the various conditions. Going back to figure 10.10B, suppose that the speed-accuracy criteria for conditions 1 and 2 were at points *X* and *Z*. Note there that the associated RTs of 115 and 175 ms are those associated with a particular error rate—specifically, .07. However, this error rate is *arbitrary;* that is, there is no reason why we should be interested in the RTs associated with this error rate as opposed to any other error rate. It is easy to see that if we observed RTs associated with some other error rate—say .50—then both the RTs associated with the individual conditions and the difference between the two RTs would be different. It is for this reason that more and more investigators have adopted the somewhat time-consuming but more informative strategy of mapping out entire SAT curves for various conditions. The means by which this mapping is done are described in the next section.

An Important Caveat

The foregoing analysis indicates *potential* problems with the RT approach. Suppose, however, that every mental process is characterized by two independent variables: the time it takes (resulting in measured RT) and the information it provides (resulting in measured accuracy). Forcing subjects to respond quickly may still increase errors because a response may be required before the process completes. Yet, left to their own devices, subjects may respond as soon as the necessary processing is complete, and the measured RT may then be directly interpreted as reflecting the time required to perform a given task. There is still considerable debate as to how serious the problems are with the RT approach (see Sternberg, 1998b, appendix A, for a detailed discussion of this point).

SAT Curves in the Study of Human Memory

Schouten and Bekker (1967) introduced an experimental technique to study the SAT function. In this technique, called the *response signal pro-*

cedure, (RSP) subjects are trained to make their response as soon as a signal is given. An SAT curve is constructed by varying the onset time of the response signal. At the very shortest delays, the subjects response is essentially a "guess"; the information processing needed to make a correct response has not begun to become available. As the signal delay increases, the subject has more time, and presumably more information becomes available. Performance increases with the time of the response signal until it reaches some asymptotic value.

As you might imagine, subjects find this task to be quite difficult. To ensure that subjects respond almost immediately after the onset of the response signal requires considerable practice. One of the hardest aspects of this task is withholding a correct response until the signal appears. Some subjects simply cannot do this, and consequently they are excluded from participating.

Before describing the mathematical form of the SAT curve, it is necessary to introduce a special index of performance that is often used in studying detection, discrimination, and recognition. Consider our familiar recognition memory task. This task can be seen as a discrimination task between two sets of items—studied items and nonstudied items. Performance is then characterized in terms of subjects' ability to discriminate studied from nonstudied items. One way of measuring discriminability is by taking the probability of a correct yes response (called a hit) and correcting for the probability of an incorrect yes response (called a false alarm). The way this is done is by first transforming hit rate and false alarm rate to *z scores* (i.e., convert the raw scores into standard scores). The difference between the z-transformed hit rate and the z-transformed false alarm rate is termed *d prime* (written as d').

Using d prime as our measure of performance, it has been shown that SAT curves for individual subjects are well fit by an exponential growth to a limit, given by the equation

$$d' = \lambda(1 - e^{-\beta(t-\delta)}),\ t > \delta. \tag{10.3}$$

The three parameters in equation 10.3, λ (lambda), β (beta), and δ (delta), characterize three phases of information processing. In phase one, $t < \delta$, no information is available. After $t = \delta$, the information rises with rate β (phase 2) until it reaches an asymptotic level of performance (phase 3).

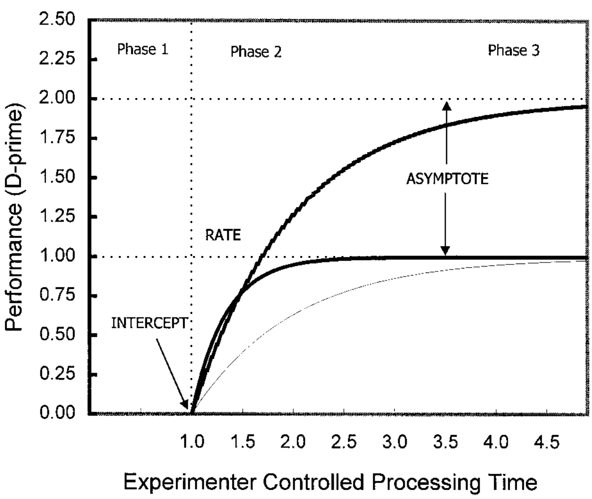


Figure 10.11
Hypothetical speed-accuracy tradeoff (SAT) curve generated by an exponential rise to an asymptote (see equation 9.3). In this figure, all of the curves assume an intercept, δ, set to 1. For the lowest curve (long dash), the rate parameter, β, and the asymptote, λ, are also set to 1. Above this curve, the dark solid curve has a rate parameter of 2. The uppermost curve is defined by a rate parameter of 1 and an asymptote of 2.

Note that we must choose a certain point on the SAT curve to characterize the asymptotic phase. The form of equation 10.3 for different parameter values is shown in figure 10.11.

Reed (1973, 1976) applied this response signal procedure to studying short-term item recognition. McElree and Dosher (1989), following up on Reed's (1976) work, examined SAT curves in the Sternberg task. They conducted two experiments that replicated the standard effects in the literature: asymptotic accuracy varied linearly with list length, and pronounced serial position effects were observed. SAT curves for lists of four and six items are shown in figure 10.12. Analyzing these curves separately for different serial positions and list lengths revealed a surprising result: neither the rate nor the intercept of the SAT curves varied with list length.

According to the classic Sternberg (serial exhaustive-scanning) model, what predictions can one make about the shape of the SAT curves? If each comparison has an equal duration, increasing list length should require more comparisons. Consequently, each added item should cause

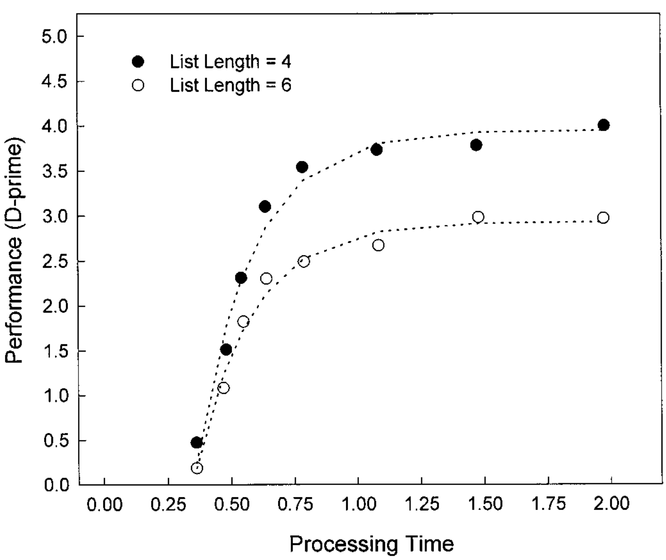


Figure 10.12
Observed average performance (as measured by d-prime) as a function of total processing time for list lengths of four and six. Smooth functions are based on the fits of an exponential rise to an asymptotic function (equation 9.3).

the minimum processing time to increase by the comparison time. This would produce a difference only in the intercept of the SAT curve. Reed's (1976) data ruled out this hypothesis.

Consider a more sophisticated version of the Sternberg model in which the comparison durations vary from trial to trial and from item to item. Because a response cannot be made until all comparisons are complete, longer list lengths will still require more processing time. For the case of a single list, the comparison will be completed very fast for some items and very slowly for other items (with a range of comparison times in between); this variability will result in a gradually increasing SAT curve (assume that subjects guess if they haven't completed all memory comparisons). If there are more items in the list, the likelihood of all of the items having fast comparison rates is quite low, so the SAT curve should rise more gradually as list length increases. As illustrated in McElree and Dosher's (1989) study (see figure 10.12), the rate of increase in the SAT curve does not vary with list length. This finding cannot be reconciled with any known variant of the serial exhaustive-scanning model. However, Ratcliff's (1978) diffusion model, a *parallel* model of RTs, does provide a reasonable account for the basic SAT data (McElree & Dosher, 1989; Ratcliff, 1978). The diffusion model will be discussed in more detail later in this chapter.

Criticisms of the SAT Approach and the Response Signal Procedure

In our earlier discussion, we pointed out some of the potential dangers involved in comparing RTs for conditions in which accuracy varies. It was assumed that variation in accuracy could result from a speed-accuracy trade-off that would disguise true RT differences. As Pachella (1974) pointed out, conditions that yield short RTs often result in lower error rates than conditions that yield long RTs. But the correlation is not 100%. In some cases, error rates vary independently of RT even when subjects are under considerable time pressure to respond (e.g., Sternberg, 1969b).

Given the availability of the response signal procedure as a means of mapping out the effect of experimental variables on the complete SAT curve, it may seem surprising that the field has not completely adopted this approach. Aside from the added complexity of this experimental technique, there have been several potentially serious problems with the RSP that should be pointed out.

The first and most serious problem is that the response signal may alter the way in which information is processed in a given task. Essentially the response signal procedure turns a single task into a dual task. While subjects are busy trying to derive the information needed to make a response, they must be constantly attentive to the response signal. Then, even if they are ready to respond, they must wait until the response signal arrives. This turns a fairly simple task into a relatively complex one, making the task of interest much more difficult to model.

Another important criticism of the response signal procedure is that it cannot distinguish all-or-none processing from continuous accrual of information. If all of the relevant information for a cognitive judgment becomes available at some variable instant in time, SAT curves will still appear to increase smoothly. How, then, can one distinguish between these fundamentally different views of cognition—all-or-none versus continuous processing? Meyer, Irwin, Osman, & Kounios (1988) proposed a variant of the response signal technique, called speed-accuracy decomposition (SAD), as a means of resolving this ambiguity. In the SAD technique, subjects are given regular (no signal) trials randomly interspersed with signal trials. Because subjects don't know if a trial is going to have a signal until the signal arrives, subjects are free to respond as soon as they are ready. On response signal trials, subjects may be responding on the basis of complete information (prior to the onset of the signal) or on the basis of partial information (after the signal is given). By obtaining RT distributions for both response signal trials and regular (no signal) trials, it is possible to determine the separate contributions of complete and partial information to the RTs obtained on the response signal trials. To do this, one must be willing to make certain assumptions about the relationship between responses based on complete and partial information (see Meyer et al., 1988, for details). Although there has been some debate as to the validity of these underlying assumptions (De Jong, 1991, but see Smith, Kounios, & Osterhout, 1997), in the worst case these assumptions leave the investigator with no less information than would be available using the more traditional SAT technique. Evidence obtained using the SAD procedure has shown that under many conditions information is accumulated continuously (e.g., Kounios, 1993; Kounios, Montgomery, & Smith, 1994; Meyer et. al., 1988). However, a recent study of problem solving revealed evidence for all-or-none processing using a SAD procedure (Smith & Kounios, 1996).

Performance Curves to Investigate Visual Information Acquisition

Speed-accuracy trade-off curves, of the sort described in the previous section, can also be used to study relatively low-level processes such as attention and visual information acquisition. When low-level processes are under investigation a simple paradigm can be used in which, on each of a series of trials, the following sequence of events occurs:

1. Usually a trial begins with a warning tone and a fixation point.
2. A stimulus (e.g., a picture) is presented for a variable but short duration (e.g., a duration ranging from 20 to 250 ms).

3. The stimulus is followed by a visual mask that prevents information acquisition from the iconic image that typically follows the visual stimulus. Thus, the time the observer has available to process the stimulus is carefully controlled.
4. Eventually, memory for the presented stimulus is tested. For instance, if pictures were shown as stimuli, memory for the pictures might be tested in a later recognition test.

In this paradigm, memory performance can be plotted as a function of stimulus duration. This form of a speed-accuracy curve is known as a *performance curve*. Performance curves have been generated by numerous researchers to investigate a variety of issues.[12]

An Example: Using Performance Curves to Investigate Effects of Priming

To illustrate how generation of such curves can be instrumental in formulating precise conclusions about the mechanisms by which some variable exerts its effect, consider the phenomenon of *priming*. In general, priming refers to the effect of some *priming stimulus* on the perceptual and cognitive processing of some related *target stimulus* that occurs near in time to the priming stimulus. A classic example is that of a lexical decision task (e.g., Meyer & Schvaneveldt, 1971). In an LDT, an observer is presented with a target letter string that is either a word or a nonword (e.g., NURSE or NIRSE) and must decide, as quickly as possible, whether the letter string is a word or a nonword.

To see the effect of priming in this paradigm, consider the letter string NURSE, to which, of course, the response "word" should be given. A universally reported result is that the RT for correctly responding "word" to the target NURSE is faster when NURSE is preceded by a related word (such as DOCTOR) than by an unrelated word (such as LION). Thus the word DOCTOR is said to *prime* the related word, NURSE. One way or another it shortens the time to correctly respond.

How does priming work? Consider two possibilities:

Possibility 1. The prime acts as if the observer has been given a brief "advance peek" (e.g., a 50-ms advance peek) at the target word. In this case, of course, RT should be 50 ms faster in the primed than in the unprimed condition.

Possibility 2. The prime acts to speed up processing of the target word. In this case, RT should still be faster in the primed compared to the unprimed condition, but by how much is not clear.

In a typical RT experiment, these two possibilities cannot be distinguished, because simply observing a shorter RT to the primed compared to the unprimed condition is consistent with either one. However, Reinitz et al. (1989) investigated priming by observing performance curves. Briefly, their experiment was as follows. On each of a series of trials, a target picture of an object was presented for varying durations and was followed by a visual mask. For instance, the target on one trial might be a guitar. In a *primed condition* the target was preceded by a related word (the word *guitar* in this example), whereas in an *unprimed condition* the target was preceded by an unrelated stimulus (which was either an unrelated word, such as *lamp,* or just a row of *X*s; these two unprimed conditions produced identical performances, so we will lump them together and just call them both the *unprimed condition.* Later, memory for the target stimuli was tested in a recognition test.

The results of this experiment took the form of two performance curves: performance as a function of original target stimulus duration for both the primed and the unprimed conditions. Now the two possibilities just sketched make different predictions, which are shown in figures 10.13 A, B.

Possibility-1 Prediction: Horizontally Parallel Curves Figure 10.13A shows the quite straightforward prediction corresponding to possibility 1: If having a prime is like having an "advanced peek" of, say, *X* ms at the target stimulus, then the two performance curves corresponding to primed and unprimed conditions should be *horizontally parallel;* that is, the horizontal difference between them should be some constant. The magnitude of the horizontal difference corresponds to the duration that the "advanced peek" is worth. If the data in figure 10.13A were obtained, the experimenter would conclude that possibility 1 is correct and that having a prime is like having an advance peek at the target picture of duration 50 ms (the magnitude of the horizontal difference between the curves). For any performance level achieved in the unprimed condition,

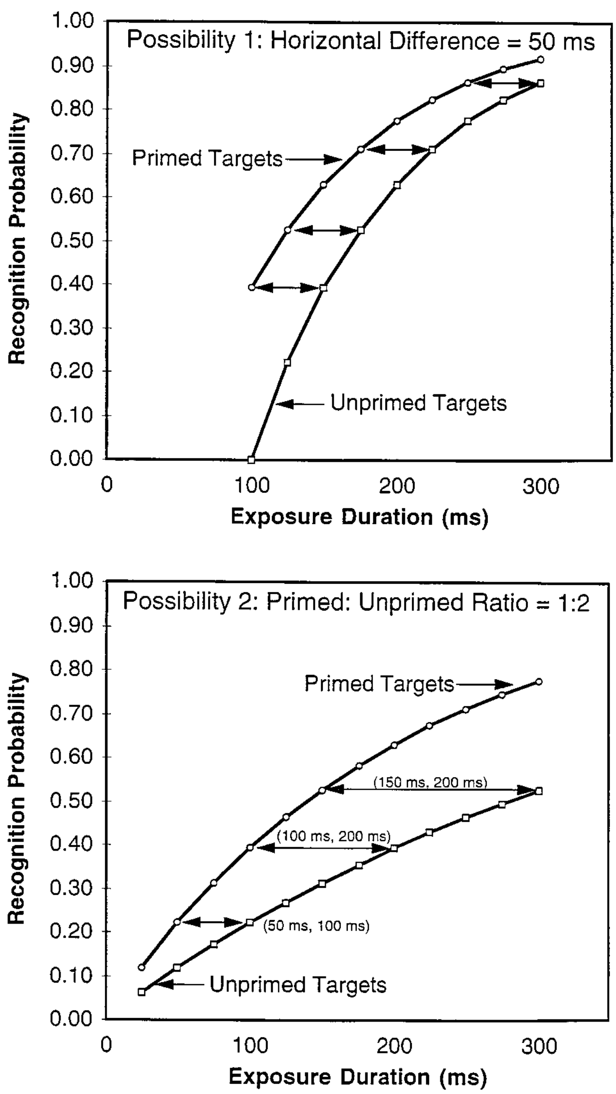


Figure 10.13
Panel (A) illustrates the predictions for possibility 1 (the "advanced peek" possibility). Note that the curves are horizontally parallel, separated by 50 ms. Panel (B) illustrates the predictions for possibility 2 (the "speedup" possibility). Note that the curves are horizontally diverging at a ratio of 1:2. Panel (C) illustrates possibility 2, with duration plotted on a log axis. In this case, the curves are once again horizontally parallel.

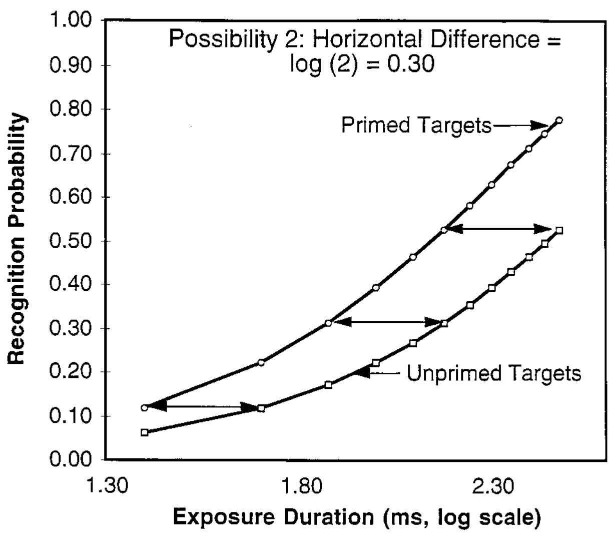


Figure 10.13
continued

the subject needs 50 ms less in the primed condition because of the 50-ms advanced peek provided by the prime.

Possibility-2 Prediction: Constant-Ratio Diverging Curves The prediction for possibility 2 is a bit more complicated, and it is illustrated in figure 10.13B. The idea here is that if the prime speeds up processing of the target, by some ratio, *r,* then it should take *r* times as long to achieve any given performance level for unprimed compared to primed targets. In the illustration of figure 10.13B, $r = 2$; that is, the prime speeds up target processing by a factor of 2. Thus, for instance, to achieve a performance of about .23 requires 50 ms for the primed targets but 100 ms for the unprimed targets. To achieve a performance level of about .40 requires 100 ms for the primed targets, but 200 ms for the unprimed targets, and so on.

One methodological note is of some interest here. Suppose you have a data set corresponding to the primed and unprimed performance curves, and you wish to see whether the data correspond to the prediction of possibility 1 (figure 10.13A) or to the prediction of possibility 2 (figure

10.13B). Testing the possibility-1 prediction is relatively straightforward: you just "slide" the two curves horizontally relative to one another (either physically, using transparencies, or electronically) and see if you can get them to exactly overlap.

Testing the possibility-2 prediction shown in figure 10.13B is not so straightforward. However, there is a trick that allows one to test the possibility-2 prediction in a similarly simple way. This is to plot the curves on a *log duration* axis rather than on a linear duration axis (a linear duration axis, as in figure 10.13B, is the normal way of plotting). Because equal linear *ratios* correspond to equal log *distances,* the possibility-2 prediction is that when plotted on a log duration axis, the performance curves should again be horizontally parallel. This possibility is illustrated in figure 10.13C. How did the data actually come out? The answer to this question is a bit complicated, and we will not describe it in detail here. Suffice it to say that initially possibility 2 was confirmed: that is, at least during the very early stages of perception, priming has the effect of speeding up the rate at which processing takes place.

Models of RT Data in Human Memory

Earlier we asked the following question; Are accuracy and RT two sides of the same coin? We then went on to show that under some circumstances, error rates are negligible and an analysis of RT reveals many interesting features of human behavior. As an example of one particularly well-explored domain, we considered RT data in the Sternberg subspan item-recognition task. Three basic empirical findings emerged from these studies. First, longer lists are associated with longer RTs—we called this a list length effect. Second, in conditions designed to eliminate rehearsal, recent items are recognized more quickly than earlier list items—we called this a recency effect. Finally, repeated items are remembered better than once presented items. Not surprisingly, all of these effects have perfect analogues in the literature on accuracy in recognition memory tasks involving longer lists. The list length effect (in recognition memory) has been known since Strong's 1912 study. Although there is still much debate as to the cause of this effect (See Murdock & Kahana, 1993a, 1993b; Shiffrin et al. 1993), it is found in every type of memory test regardless of whether accuracy or RT is the dependent variable. The beneficial effects of recency are equally ubiquitous in the memory literature. Rubin and Wenzel (1996) and Wixted and Ebbesen (1991) have shown that across a wide range of tasks and materials response accuracy declines as a power function of time, or the number of items intervening between study and test (termed *study-test lag*). In short-term item recognition, Monsell (1978) found a similar type of recency effect in RT data (see figure 10.7). In a continuous recognition task, Hockley (1982) also found dramatic recency effects in RT data.[13] Clearly, recency, repetition, and list length effects are fundamental properties of human memory. It is with this in mind, that we can entertain the possibility that memories vary in the strength of the cue-target match. The amount of information is then a single dimension that has a single SAT function. More information results in faster and more accurate responses.

Unidimensional Strength Theory of Recognition (or Signal Detection Theory)

Consider the familiar item-recognition task (e.g., Sternberg, 1966) as a discrimination between two categories: items that were shown in the list and those that were not. According to *strength theory* (Norman & Wickelgren, 1969) items vary along the dimension of information that distinguishes these two categories (this dimension is sometimes called memory strength). A crucial assumption of the theory is the idea of a noisy system: items within each category may vary greatly in their values along the "strength" dimension. This results in two strength distributions: a distribution for list items and a distribution for nonlist items. Responses are made based on the value of an item along this informational dimension. If an item's strength exceeds some *criterion* value, a positive response is made; otherwise a negative response is made. This theory is considered because it can provide a simple, unified account for both RT and accuracy data in a broad range of recognition memory tasks (Murdock, 1985). Although the model is overly simplistic, the basic ideas it encompasses have become part of almost every current model of human memory (e.g., Hintzman, 1986; Metcalfe, 1982; Murdock, 1982). *Strength theory* is the term used for the application of *signal detection theory* (SDT) to recognition memory tasks (Egan, 1958). In order to see how this theoretical framework can provide an account for both accuracy and RT data, the basic elements of signal detection theory will be briefly introduced (for a more thorough treatment, the reader is encouraged to consult Murdock, 1985).

In a categorization or recognition task, there may be different payoffs associated with incorrectly classifying a nonlist item as a list item (a false alarm) and for correctly classifying a list item as such (a hit). Such differential payoffs are easily modeled by assuming that the criterion can be adjusted to meet the task demands. In a case where we want to avoid false alarms at all costs, we set a high criterion. In a case where we want to maximize hits, but where false alarms aren't too bad, we set a low criterion. Moving the criterion should not affect the discriminability of the two distributions; only the relative numbers of hits and false alarms will change.

A graph that plots hit rate against false alarm rate for different criteria is called a receiver operating characteristic (ROC) curve. According to strength theory, plotting the z-transformed hit rate versus the z-transformed false alarm rate should result in a linear function. If the two distributions have equal variance, the slope of the z-ROC curve should be 1. In the recognition memory task we can vary the criterion by collecting data on *judgements of confidence*. In this technique subjects are asked, "how confident are you that X was on the list?" A typical scale used for confidence judgements is as follows:

−3	−2	−1	1	2	3
High		Low	Low		High
	"No"			"Yes"	

We can now mimic a subject with high criterion by grouping confidence judgments that are less than 3 into the "no" category, leaving only confidence judgements of 3 in the "yes" category. Based on this grouping, our imaginary conservative subject only responds "yes" when our real subject responds "yes" with high confidence. Similarly, we can mimic a subject who is slightly less conservative by grouping confidence judgments that are less than 2 into the "no" category, leaving confidence judgments of 2 and 3 in the "yes" category. Moving the criterion further down, we reach the criterion of our real subject, with positive confidence judgments reflecting "yes" responses and negative confidence judgments reflecting "no" responses. Finally, we can move our criterion yet further down, all the way to the point were only a confidence judgement of -3 is in the "no" category, and all other confidence judgments are grouped in the "yes" category. This hypothetical liberal subject will only withhold a "yes" response when he/she is certain the item was not on the list. If we plot hits against false alarms for each of these hypothetical subjects, we can construct an ROC curve. At the most conservative end of the spectrum, both the hit rate and the false alarm rate will be low because the subject rarely makes "yes" responses. At the most liberal end of the spectrum, both the hit and false alarm rates will be high because the subject almost always makes "yes" responses. The points representing hit rate and false alarm rate for each criteria level (liberal to conservative) traces out the bow-shaped ROC curve (See Swets, 1998 for background information on ROC curves and their applications to some real world problems). Studies of recognition memory (Koppell, 1977; Ratcliff, McKoon, & Tindall, 1994; Yonelinas, 1997) have found that the z-ROC curves are nearly linear but have slopes that are consistently less than 1 (around 0.8 in most studies). The linearity of the z-ROC functions is consistent with the view that strengths of list and nonlist items are distributed normally. The finding that the slope of the z-ROC curve is less than 1 indicates that the variability in the strength of items' representations in memory is greater for list items than for non nonlist items (see Ratcliff, McKoon, & Tindall, 1994)

Strength theory may be extended to deal with RT data in a fairly straightforward manner. If we plot RT as a function of confidence judgment values (which maps directly onto the distance from our yes-no criterion), we find that as we approach the criterion from either direction, RTs increase quite dramatically. According to the *RT-distance* hypothesis, RT increases monotonically as the criterion is approached from either direction (Koppell, 1977; Murdock & Anderson, 1975). Murdock (1985) proposed an extension of strength theory to handle RT data. This model has been shown to fit data on list length effects and recency effects, as well as RT distributions, in both the Sternberg (1966) subspan item-recognition task and in the supraspan study-test recognition paradigm. The power of the RT-distance hypothesis is that it can be applied to any domain of signal detection theory. Maddox and Ashby (1996) incorporated the RT-distance hypothesis into the *generalized recognition theory* of multidimensional categorization tasks (Ashby & Perrin, 1988). In this manner they were able to simultaneously fit both accuracy and RT data in a variety of categorization tasks.

The Diffusion Model

The diffusion model (Ratcliff, 1978; Ratcliff & Van-Zandt, submitted) is an abstract mathematical model of any cognitive task that involves choosing from among a number of sources of information. These tasks include recognition memory as well as multidimensional perceptual discrimination tasks. Consider an application of this model to the basic Sternberg item recognition paradigm. A probe is compared in parallel with a defined (but potentially large) set of items in memory (see figure 10.14). Each memory trace begins with a base level of activation. As time progresses, the activation drifts, or diffuses, with a variable rate toward either a lower or upper boundary. The model is self-terminating on a match (i.e., if an item reaches its upper bound, the model produces a positive response) and exhaustive on nonmatches (i.e., all items must reach the lower bound before a negative response can be made). For appropriately chosen parameter values, this model can produce many of the major findings in the Sternberg paradigm—including both asymptotic accuracy and RT distributions. In addition, it provides a reasonable account of SAT functions (McElree & Dosher, 1991). The diffusion model has also been successfully applied to data on multielement comparisons (Ratcliff, 1981) and choice reaction time[14] (Ratcliff & Van-Zandt, submitted). One criticism of the diffusion model is that it does not explain the basis of the processes it postulates. The model does not explain how items are represented, how they are compared, what causes the variability in drift rates, or even how the upper and lower boundaries are instantiated.

Nonetheless, a diffusion-type mechanism may be incorporated into models that *do* make explicit assumptions about item comparison and representation. Nosofsky and Palmeri (1997) extended Nosofsky's (1986) exemplar-based model of categorization to account for RTs in speeded categorization tasks. In their model, exemplars of items and their associated categories are stored as separate memory traces. A to-be-

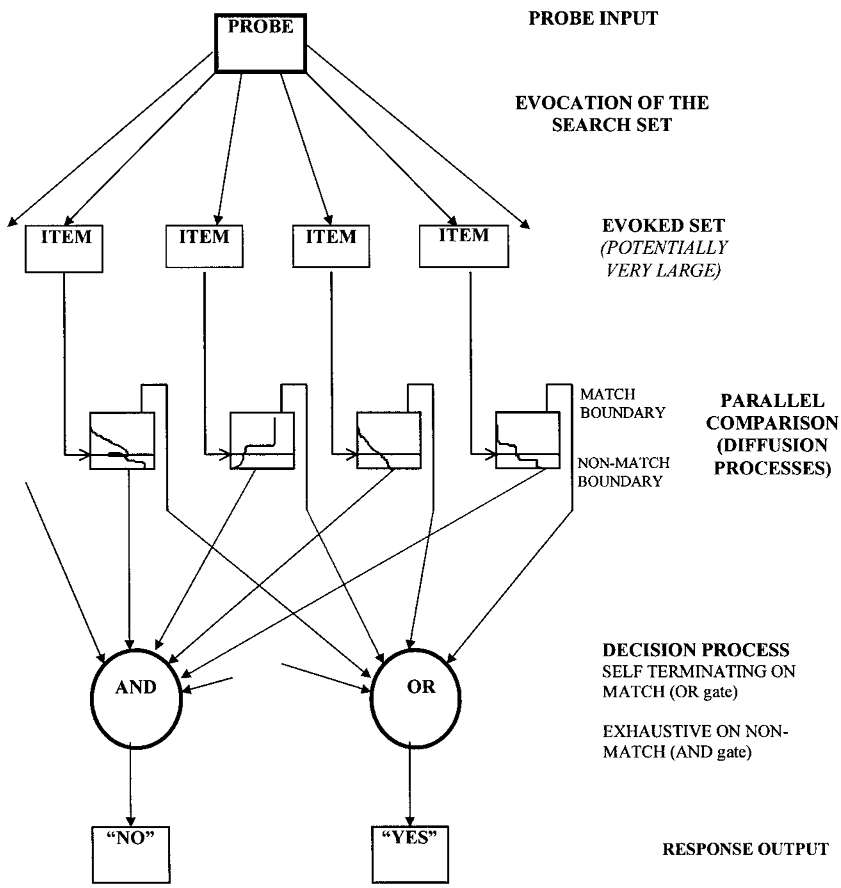


Figure 10.14
An illustration of the diffusion model applied to an item recognition task. The process begins at the top of this figure with the encoding of the probe item. The probe item is then compared, in parallel (i.e., all comparisons begin at the same time), with each of the items in the memory set. Each comparison results in a matching strength value that begins at a *baseline level* and then continuously increases or decreases at a *variable rate.* A positive yes, response is made if any of the comparisons reaches a *match boundary;* a negative, no, response is made if all of the comparisons reach a *nonmatch boundary.* Model parameters include the values of the match and nonmatch boundaries and the mean and variance of the matching strength for each item in the memory set. (Adapted from figure 3, Ratcliff, 1978.)

classified probe item is simultaneously compared with each stored exemplar. The likelihood of successful retrieval is determined jointly by the strength of the exemplar in memory and its similarity to the probe item. Each retrieved exemplar adds evidence in favor of the category with which it has been associated. When a criterion of evidence is reached in favor of a particular category, a response is made. Nosofsky and Palmeri (1997) obtained good fits to both mean RT and to RT distributions.

Connectionist Models

During the last 15 years there has been a surge of interest in connectionist models of memory. These models typically assume that a unit of memory is represented by a pattern of activation across a large number of processing units. The set of activation values across these units defines a vector in a multidimensional space. Given a sufficiently large number of units, the same population of units can be used to store a multitude of items. Interactions among processing units determine the storage of new memories and the dynamics by which the model can reconstruct an entire pattern given a partial input.

Connectionist models of memory and cognition can be subdivided into three major classes: multiple-layer, feedforward models of recognition and categorization (e.g., McClelland, 1979; Usher & McClelland, 1996); autoassociative models of recognition and pattern completion (e.g., Chappell & Humphreys, 1994; Masson, 1995; Metcalfe, 1990); and recurrent, heteroassociative models of sequence memory (e.g., Cleeremans & McClelland, 1991). (See also chapter 8, this volume.)

These models provide mechanisms that give rise to the constructs that are characterized abstractly within models such as the diffusion model or strength theory. In one of the earliest applications of a connectionist model to accuracy and RT data in human memory, Anderson (1973) showed how a simple distributed memory model could account for the basic linear RT functions obtained by Sternberg (1966). Further efforts to model the Sternberg task employed nonlinear models with multiple layers (e.g., McClelland, 1979).

Usher and McClelland (1996) propose a two-layer network model for choice reaction time tasks. The first layer represents the stimulus as a pattern of activation across a set of units. These units send their activation through weighted paths to a second, decision, layer with *N* units (one for each possible choice). The Usher and McClelland model proposes both recurrent excitatory connections and mutual inhibitory connections between the units in the decision layer. The interplay between the excitatory and inhibitory mechanisms results in a generalized diffusion toward one of two decision bounds. This model encompasses the diffusion model as a special case while providing an even better account of SAT data.

Although the Usher and McClelland model is appropriate for choice reaction time tasks, it cannot do pattern reconstruction or serial recall tasks. To do these tasks, a class of connectionist models known as recurrent, or autoassociative, neural networks have been developed. These models allow for associations between an item and itself (autoassociation) as well as associations among different items (heteroassociations). They follow a simple learning principle called the Hebb rule (after Hebb's, 1949, hypothesis about synaptic plasticity). When two units are coactive the connection between them is strengthened, and when two units have uncorrelated activities the connection between them is weakened. These principles are related to the biological mechanisms of long-term potentiation and long-term depression (Brown & Chattarji, 1995; McNaughton & Morris, 1987; Treves & Rolls, 1994). The network evolves dynamically according to a simple model of neural function (usually a derivative of the classic McCulloch & Pitts, 1943, model). Typically, the activity of a unit is a monotonic function of the weighted sum of the input to that unit.

The Hopfield (1982) model is an example of a simple attractor neural network capable of mimicking human RT and accuracy data in priming experiments (Masson, 1995). Chappell and Humphreys (1994) expanded this approach to explain a number of phenomena in recognition and recall memory tasks. Although there has been some criticism of neural network models of RTs (Ratcliff & Van-Zandt, 1996), these models provide a natural account of both accuracy and RT data across a broad range of cognitive tasks.

Models of accuracy and RT data often assume that a common dimension of information underlies both accuracy and RT judgments. Few models have tackled the difficult problem of fitting SAT functions in a wide range of tasks. The two models that have been largely successful in accounting for SAT data assume variability in the rate with which information continuously accrues (e.g., Ratcliff, 1978; Usher & McClelland, 1996).[15]

Conclusions: Are Accuracy and RT Data Two Sides of the Same Coin?

Superficially, it appears that our review of theory and data concerning accuracy and RT in human memory supports the view that these two measures may reflect a single underlying dimension of information. However, this conclusion leaves us somewhat uneasy. To further examine this question, we have set out to find a few examples of cases in the literature where accuracy and RT have not provided comparable answers.

Sometimes variables that have significant effects on accuracy do not affect RTs (MacLeod & Nelson, 1984). Sternberg (1969b) reported an experiment in which subjects studied a list of items presented either once, twice, or three times. After the list presentation, a single item was presented as a cue to recall the next item in the list. As with the standard item-recognition task, RTs in this task increased linearly with list length, but interestingly, RT was not affected by the number of times the list was presented. In contrast, error rates for the longest list (six items) were quite high (23%) when the list was only presented once, but less than 5% when the list was presented three times. In this study, accuracy differences were not reflected in RT data.

There are fewer cases in the literature where a variable has a significant effect on RT data but no discernible effect on response accuracy (when accuracy is far from ceiling). Sanders, Whitaker, & Cofer (1974) found that in a recognition task, subjects did not suffer from associative interference when measured using accuracy but showed substantial interference when RT was examined. Subjects took as many trials to learn *C-D* word pairs after learning *A-B* pairs as they did after learning *A-D* pairs. In contrast, RTs were significantly slower when tested on the *A-D* list, presumably because of interference from the *A-B* pairs learned in the first list.

Santee & Egeth (1982; see also Mordkoff & Egeth, 1993) found that accuracy and latency were affected quite differently from each other in a letter recognition task. Perceptual interference caused by displaying targets very briefly affected accuracy at detection but not latency. In contrast, response competition caused by having to respond to a given target in the face of competing information from another target affected latency of responses but not accuracy.

In a recent study examining accuracy and RT in various types of associative recall tasks, Kahana (1998) found that the order in which a pair of items was studied has no significant effect on accuracy, yet forward recall was significantly faster than backward recall (ART $>$ 400 ms). Accuracy for forward recall was 87.7%, and for backward recall it was 85.1% ($p > .10$). This result makes sense if one assumes that an association is a single integrated unit of information that is unpacked in the order in which it was encoded.

Perhaps the most striking example comes from a *judgment of recency* (JOR) task. In this task, subjects are presented with a short list of items (usually words or letters). Immediately after list presentation, two items are presented and the subject must select the more recent list item. For example, suppose the list consists of items XTLVDGBNW and the probe items are *V* and *N*. In this case, the subject might correctly select *N* as the more recent list item. Muter (1979) and Hacker (1980) independently discovered that RTs in this task are dependent only on the position of the more recent item and not on the relative recency of the two items. From a strength-type theory, we would expect that the difference in the recency of the two items would affect both accuracy and RT data. Data from Hacker's study are shown in figure 10.15. The peculiar finding that the position of the less recent item does not affect RT led Hacker (1980) to propose a self-terminating, backward serial-scanning model of this task. If we scan backward from the end of the list, it will take the same time to find the more recent item regardless of the position of the less recent item (cf. Murdock's, 1974, conveyer belt model of recognition memory).

McElree and Dosher (1993) performed a SAT analysis of the Muter-Hacker JOR task. They succeeded in replicating the Muter-Hacker finding that mean RTs are only affected by the recency of the more recent probe (and not the distance between the two items). The SAT study of the same task showed that there *is* an effect of the relative recency of the two probe items. Specifically, the rate of approach to asymptote was more rapid as the less recent probe was more distant. In contrast, the more recent probe had the expected effects on both the intercept and the

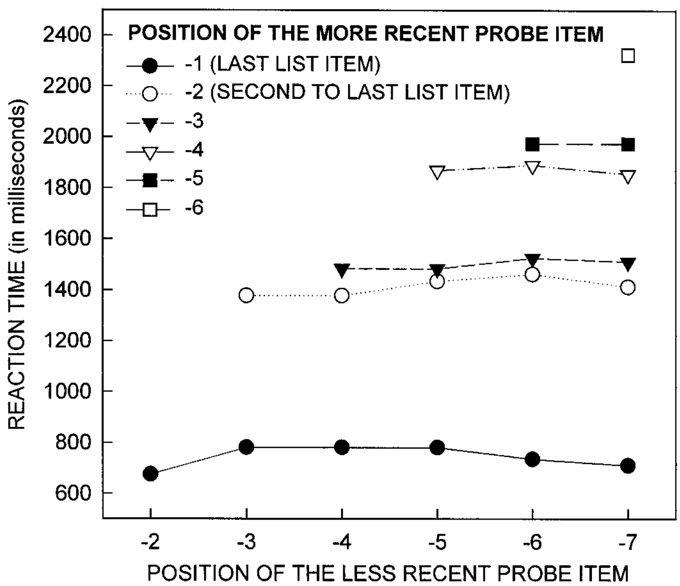


Figure 10.15
Response accuracy and latency in a *judgment of recency* task (Hacker, 1980). Mean correct RT is strongly influenced by the recency of the more recent probe item, but is unaffected by the recency of the less recent probe item. These data conflict with the reasonable prediction that the *relative* recency of the two items influences mean RT. The same basic pattern of data has also been obtained by Muter (1979), Hockley (1984), and McElree & Dosher (1993).

asymptote of the SAT functions. This SAT approach clearly demonstrated that the relative lag of the first list item, which did not affect mean RTs, did have a significant effect on processing rate.

Additional studies using SAT techniques have begun to provide convergent evidence against the idea that memories vary along a single dimension. Rather, SAT studies of human memory have lent support to the view that different types of information are represented in memory (e.g., Murdock, 1974; Underwood, 1983). Gronlund and Ratcliff (1989) compared single-item recognition with associative recognition (recognition that two items were paired together in a list). They found that item information became available before associative information. Ratcliff and McKoon (1989) and Dosher (1984) found that preexperimental relations among items influenced recognition of word pairs or sentences early in processing and that the necessary contextual information did not become available until later stages of processing. Hintzman and Curran (1994) found evidence that item recognition judgments are influenced by a fast-acting familiarity mechanism followed by a slower recall-like retrieval process (c.f., Atkinson & Juola, 1973).

All of these results support the cognitive idea that multiple kinds of information provide us with our "memory strength." In particular, the range of studies reported converge on the need to distinguish between information on item familiarity (the closest idea to the traditional notion of strength), experimentally formed associations between items, and contextual information that binds items and associations to a particular time and place. These different kinds of information are often characterized by different SAT functions. If several different types of information, or memory processes, mediate performance in a task, the accuracy-RT relation would have to be identical for each process in order for accuracy and RT to be measuring the same thing. If each process has a different accuracy-RT relation, it is good cause for studying the effects of experimental variables on both accuracy and RT data.

In recent years the evidence for the involvement multiple processes and types of information in memory tasks has been accumulating. More information implies better accuracy and shorter RT, making accuracy and RT measures highly correlated. But the results of SAT studies have shown that the precise pattern of accuracy-RT effects may teach us a great deal about memory processes. In addressing the question posed in the beginning of this chapter, accuracy and RT cannot be two sides of the same coin unless the cognitive process of interest is a single operation acting on a single type of information. Consequently, consideration of both accuracy and RT data is often critical in distinguishing theories of cognition, and the use of only one of these measures may provide a skewed interpretation of the phenomena of interest.

Notes

1. See, for example, Gronlund & Ratcliff (1989); Hintzman & Curran (1994); Kounios, Osman, & Meyer (1987); McElree (1996); McElree & Dosher (1989, 1993); Meyer, Irwin, Osman & Kounios (1988); Ratcliff & McKoon (1989); Ratcliff & Van Zandt (1996); and Rohrer & Wixted (1994).

2. In these experiments lists are usually made up of between 15 and 40 randomly chosen words. The advantage in recall is for the first 3 to 4 words and the last 6 to 8 words. The size of the recency effect does not depend on the length of the list, the presentation rate, or other variables that generally effect overall memory (Murdock, 1962).

3. Latent constructs are variables (often representing mental processes) that are not directly observed but whose existence is inferred from the data. The idea of association is a latent construct, as is intelligence or morale.

4. See, for example, Brown, Conover, Flores, & Goodman (1991); Cooke, Durso, & Shvaneveldt (1986); and Romney, Brewer, & Batchelder (1993). See Shuell (1969) for a review of the earlier literature.

5. In rapid serial visual presentation (RSVP) of sentences, subjects are impaired at recalling the second presentation of a repeated element. This is known as *repetition blindness* (Kanwisher, 1987). There is some debate as to how repetition blindness is related to the Ranschburg effect: Kanwisher (1987) maintains that the two are distinct phenomena; however, Fagot and Pashler (1995) suggest that the two phenomena may be closely related (see also Whittlesea, et al., 1996).

6. The reader may note that in free recall, performance is *best* at the end of the list. However, in serial or ordered recall, performance is best at the beginning of the list. This makes sense because subjects must start recalling at the beginning in serial recall but are free to recall from the end in free recall.

7. In the recognition memory literature, investigators often call positive probes *old items* and negative probes *new items*.

8. see Murdock (1985) for an attempt to fit such a model to data from the Sternberg task.

9. See Murdock & Walker (1969).

10. For opposing views see Baddeley & Hitch (1974, 1977); Crowder (1982) and Greene (1986, 1992).

11. For an interesting alternative view see Sternberg (1998b).

12. See, for example, Loftus (1985); Loftus & Bell (1975); Loftus, Busey, & Senders (1993); Loftus, Duncan, & Gehrig (1992); Loftus, Johnson, & Shimamura (1985); Reinitz (1990); Reinitz, Wright, & Loftus (1989); Rumelhart (1970); Shibuya & Bundeson (1988); and Townsend (1981).

13. In a continuous recognition task, there is no differentiation between the study and test phases. Stimuli are presented one by one, and as each stimulus appears, subjects respond yes if they think they have seen it before and no if they think it is a new word.

14. In a choice reaction time task a stimulus is presented (e.g., a row of asterisks on a computer screen) and subjects are supposed to make one of several discrete responses according to the qualities of the stimulus (e.g., many or few asterisks). Although a recognition memory task is a kind of two-choice RT task (was the presented word on the list, yes or no?), the term *choice reaction time* is used to

refer to judgments concerning a stimulus that is present rather than judgments concerning one's memory for a stimulus.

15. Hanes and Schall (1996) have found an interesting parallel to the variable rate assumption in single-cell studies of the rhesus monkey.

References

Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review, 80,* 417–438.

Anderson, J. R. (1995). *Learning and memory: An integrated approach.* New York: Wiley.

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review, 89,* 599–607.

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review, 95,* 124–150.

Ashby, F. G., Tein, J. Y., & Balakrishan, J. D. (1993). Response time distributions in memory scanning. *Journal of Mathematical Psychology, 37,* 526–555.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1, pp. 242–293) San Francisco: Freeman.

Baddeley, A. D. (1976). The psychology of memory. New York: Basic Books.

Baddeley, A. D., & Ecob, J. R. (1973). Reaction time and short-term memory: Implications of repetition effects for the high-speed exhaustive scan hypothesis. *Quarterly Journal of Experimental Psychology, 25,* 229–240.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–90). New York: Academic Press.

Baddeley, A. D., & Hitch, G. (1977). Recency Re-examined. In S. Dornic (Ed.), *Attention and Performance* (Vol. 6, pp. 647–667). Hillsdale, NJ: Erlbaum.

Brown, S. C., Conover, J. N., Flores, L. M., & Goodman, K. M. (1991). Clustering and recall: Do high clusterers recall more than low clusterers because of clustering? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 710–721.

Brown, T. H., & Chattarji, S. (1995). Hebbian synaptic plasticity. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 454–459). Cambridge, MA: MIT Press.

Burrows, D., & Okada, R. (1975). Memory retrieval from long and short lists. *Science, 188,* 1031–1033.

Chappell M., & Humphreys, M. S. (1994). An autoassociative neural network for sparse representations: Analysis and application to models of recognition and cued recall. *Psychological Review, 101,* 103–128.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120,* 235–253.

Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 538–549.

Corbett, A., & Wickelgren, W. (1978). Semantic memory retrieval: Analysis by speed-accuracy tradeoff functions. *Quarterly Journal of Experimental Psychology, 30,* 1–15.

Crannell, C. W., & Parish, J. M. (1957). A comparison of immediate memory span for digits, letters and words. *Journal of Psychology, 44,* 319–327.

Crowder, R. G. (1968). Intraserial repetition effects in immediate memory. *Journal of Verbal Learning and Verbal Behavior, 7,* 446–451, 1968.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Crowder, R. G. (1982). The demise of short-term memory. *Acta Psychologica, 50,* 291–323.

De Jong, R. (1991). Partial information or facilitation? Different interpretations of results from speed-accuracy decomposition. *Perception & Psychophysics, 50,* 333–350.

Donders, F. C. (1969). On the speed of mental processes. In W. G. Koster (Ed. & Trans.), *Attention and performance* (Vol. 2, pp. 412–431). Amsterdam: North Holland. (Original work published 1868.)

Dosher, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology, 16,* 519–555.

Egan, J. P. (1958). Recognition memory and the operating characteristic. Technical Note AFCRC-TN-58-51. Indiana University, Hearing and communication laboratory. See Green, D. M., and Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley.

Fagot, C. & Pashler, H. (1995), Repetition blindness: perception or memory failure? *Journal of Experimental Psychology: Human Perception & Performance, 21,* 275–92.

Forrin, B., & Cunningham, K. (1973). Recognition time and serial position of probed item in short-term memory. *Journal of Experimental Psychology, 99,* 272–279.

Goshen-Gottstein, Y., & Moscovitch, M. (1995a). Repetition priming for newly formed and preexisting associations: Perceptual and conceptual influences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1229–1248.

Goshen-Gottstein, Y., & Moscovitch, M. (1995b). Repetition priming effects for newly formed associations are perceptually based: Evidence from shallow encoding and format specificity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1249–1262.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 501–518.

Greene, R. L. (1986). Sources of recency effects in free recall. *Psychological Bulletin, 99,* 221–228.

Greene, R. L. (1991). The Ranschburg effect: The role of guessing strategies. *Memory and Cognition, 19,* 313–317.

Greene, R. L. (1992). *Human memory: Paradigms and paradoxes,* Hillsdale, NJ: Erlbaum.

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 846–858.

Hacker, M. J. (1980). Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 651–675.

Hall, J. F. (1971). Verbal learning and retention. Philadelphia: Lippincott.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science, 274,* 427–430.

Hebb, D. O. (1949). *Organization of behavior.* New York: Wiley.

Helmholtz, H. von (1853). *Philosophical Magazine, 4,* 313–325. (Original work published 1850.)

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

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33,* 1–18.

Hockley, W. E. (1982). Retrieval processes in continuous recognition. *Journal of Experimental Psychology:Learning, Memory, and Cognition, 8,* 497–512.

Hockley, W. E. (1984). The analysis of reaction time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 10,* 598–615.

Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review, 94,* 341–358.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, U.S.A., 84,* 8429–8433.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Jahnke, J. C. (1969). The Ranschburg effect. *Psychological Review, 76,* 592–605.

Jahnke, J. C. (1970). Probed recall of strings that contain repeated elements. *Journal of Verbal Learning and Verbal Behavior, 9,* 450–455.

Jahnke, J. C. (1972). The effects of intraserial and interserial repetition on recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 706–716.

Jahnke, J. C. (1974). Restrictions on the Ranschburg effect. *Journal of Experimental Psychology, 103,* 183–185.

Jimenez, L., Mendez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 948–969.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24,* 103–109.

Kahana, M. J. (1998). *An analysis of distributed memory models of ordered recall: Effects of compound cueing, target ambiguity, and recall direction.* Manuscript submitted for publication.

Kahana, M. J., & Jacobs, J. (1998). *A response time analysis of the Ranschburg effect: Implications for distributed memory models of serial recall.* Manuscript in preparation.

Kanwisher, N. G. (1987). Repetition blindness: type recognition without token individuation. *Cognition, 27,* 117–43.

Kausler, D. H. (1974). Psychology of verbal learning and memory. New York: Academic Press.

Koppell, S. (1977). Decision latencies in recognition memory: A signal detection theory analysis. *Journal of Experimental Psychology, 3,* 445–457.

Kounios, J. (1993). Process complexity in semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 338–351.

Kounios, J. (1996). On the continuity of thought and the representation of knowledge: Electrophysiological and behavioral time-course measures reveal levels of structure in human memory. *Psychonomic Bulletin & Review, 3,* 265–286.

Kounios, J., Montgomery, E. C., & Smith R. W. (1994). Semantic memory and the granularity of semantic relations: Evidence from speed-accuracy decomposition. *Memory & Cognition, 22,* 729–741.

Kounios, J., Osman, A. M., & Meyer, D. E. (1987). Structure and process in semantic memory: New evidence based on speed-accuracy decomposition. *Journal of Experimental Psychology: General, 116,* 3–25.

Loftus, G. R. (1985). Picture perception: Effects of luminance level on available information and information-extraction rate. *Journal of Experimental Psychology: General, 114,* 342–356.

Loftus, G. R., & Bell, S. M. (1975). Two types of information in picture memory. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 103–113.

Loftus, G. R., Busey, T. A., & Senders, J. W. (1993). Providing a sensory basis for models of visual information acquisition. *Perception & Psychophysics, 54,* 535–554.

Loftus, G. R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 530–549.

Loftus, G. R., Johnson, C. A., & Shimamura, A. P. (1985). How much is an icon worth? *Journal of Experimental Psychology: Human Perception and Performance, 11,* 1–13.

Luce, R. D. (1986). *Response times.* New York: Oxford University Press.

MacLeod, C. M., & Nelson, T. O. (1984). Response latency and response accuracy as measures of memory. *Acta Psychologica, 57,* 215–235.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception & Performance, 22,* 795–817.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 3–23.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

McCulloch, W. S., & Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5,* 115–133.

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory & Cognition, 24,* 173–187.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General, 118,* 346–373.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General, 122,* 291–315.

McNaughton, B. L., & Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neuroscience, 10,* 408–415.

Metcalfe, J. (1990). Composite holographic associative recall model (CHARM) and blended memories in eyewitness testimony. *Journal of Experimental Psychology: General, 119,* 145–160.

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

Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. *Psychological Review, 95,* 183–237.

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.

Miller, J. (1993). A queue-series model for reaction time, with discrete-stage and continuous-flow models as special cases. *Psychological Review, 100,* 702–715.

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology, 10,* 465–501.

Mordkoff, J. T., & Egeth, H. E. (1993). Response time and accuracy revisited: Converging support for the interactive race model. *Journal of Experimental Psychology: Human Perception & Performance, 19,* 981–991.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482–488.

Murdock, B. B. (1974). *Human memory: Theory and data.* Potomac, MD: 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. (1985). An analysis of the strength-latency relationship. *Memory and Cognition, 13,* 511–521.

Murdock, B. B., & Anderson, R. E. (1975). In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium.* (pp. 145–194) Hillsdale, NJ: Erlbaum.

Murdock, B. B., & Kahana, M. J. (1993a). Analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 689–697.

Murdock, B. B. & Kahana, M. J. (1993b). List-strength and list-length effects: Reply to Shiffrin, Ratcliff, Murnane, and Nobel (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1450–1453.

Murdock, B. B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior, 86,* 263–267.

Murdock, B. B., & Walker, K. D. (1969). Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior, 8,* 665–676.

Muter, P. (1979). Response latencies in discriminations of recency. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 160–169.

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.* Hillsdale, NJ: Erlbaum.

Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in short-term memory. *Journal of Mathematical Psychology, 6,* 192–208.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental psychology: General, 115,* 39–57.

Nosofsky, R. M., & Palmeri, T. J. (in 1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.) *Human information processing: Tutorials in performance and cognition.* New York: Halstead Press.

Pachella, R. G., & Fisher, D. F. (1969). Effect of stimulus degradation and similarity on the trade-off between speed and accuracy in absolute judgments. *Journal of Experimental Psychology, 81,* 7–9.

Patterson, K. E., Meltzer, R. H., & Mandler, G. (1971). Inter-response times in categorized free recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 417–426.

Pollio, H. R., Kasschau, R. A., & DeNise, H. E. (1968). Associative structure and the temporal characteristics of free recall. *Journal of Experimental Psychology, 76,* 190–197.

Pollio, H. R., Richards, S., & Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior, 8,* 529–536.

Proctor, R. W. (1981). A unified theory for matching-task phenomena. *Psychological Review, 88,* 291–326.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory,* (Vol. 14, pp. 207–262). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

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

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review, 88,* 552–572.

Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology, 21,* 139–155.

Ratcliff, R., McKoon, G., & Tindall, M. H. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 763–785.

Ratcliff, R., & Van-Zandt, T. (1996). *Connectionist and diffusion models of reaction time.* Manuscript submitted for publication.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855–863.

Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science, 181,* 574–576.

Reed, A. V. (1976). List length and the time-course of recognition in immediate memory. *Memory & Cognition, 4,* 16–30.

Reinitz, M. T. (1990). Effects of spatially directed attention on visual encoding. *Perception and Psychophysics, 47,* 497–505.

Reinitz, M. T., Wright, E., & Loftus, G. R. (1989). The effects of semantic priming on visual encoding of pictures. *Journal of Experimental Psychology: General, 118,* 280–297.

Roberts, S. (1987). Evidence for distinct serial processes in animals: The multiplicative-factors method. *Animal learning & Behavior, 15,* 135–173.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.) *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience,* pp. 611–653. Cambridge, MA: MIT Press.

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory and Cognition, 22,* 511–524.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science, 4,* 28–34.

Rubin, D., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 4,* 734–760.

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology, 7,* 191–218.

Sanders, A. F. (1980). Stage analysis of reaction processes. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 331–354). Amsterdam: North-Holland.

Sanders, A. F., Whitaker, L., & Cofer, C. N. (1974). Evidence for retroactive interference in recognition from reaction time. *Journal of Experimental Psychology, 102,* 1126–1129.

Santee, J. L., & Egeth, H. E. (1982). Do reaction time and accuracy measure the same aspects of letter recognition? *Journal of Experimental Psychology: Human Perception & Performance, 8,* 489–501.

Schouten, J. F. & Bekker, J. A. M. (1967) Reaction time and accuracy. *Acta Psychologica, 27,* 143–153.

Schweickert, R. (1985). Separable effects of factors on speed and accuracy: Memory scanning, lexical decision, and choice tasks. *Psychological Bulletin, 97,* 530–546.

Shibuya, H., & Bundsen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 591–600.

Shiffrin, R. M., & Raaijmakers, J. (1992). The SAM retrieval model: A retrospective and prospective. In A. F. Healy, S. M. Kosslyn, and R. M. Shiffrin (Eds.). *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 1), Potomac, MD: Erlbaum.

Shiffrin, R. M., Ratcliff, R., Murnane, K., & Nobel, P. (1993). TODAM and the list-strength and list-length effects: A reply to Murdock and Kahana. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1445–1449.

Shuell, T. J. (1969). Clustering and organization in free recall. *Psychological Bulletin, 72,* 353–374.

Smith, R. W., & Kounios, J. (1996). Sudden insight: All-or-none processing revealed by speed-accuracy decomposition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1443–1462.

Smith, R. W., Kounious, J., & Osterhout, L. (1997). The robustness and applicability of speed and accuracy decomposition: A technique for measuring partial information. *Psychological Methods, 2,* 95–120.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654.

Sternberg, S. (1967a). Retrieval of contextual information from memory. *Psychonomic Science, 8,* 55–56.

Sternberg, S. (1967b). Two operations in character-recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics, 2,* 45–53.

Sternberg, S. (1969a). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30,* 276–315.

Sternberg, S. (1969b). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57,* 421–457.

Sternberg, S. (1975). Memory Scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27,* 1–32.

Sternberg, S. (1998a). Discovering mental processing stages: The method of additive factors. In D. Osherson (Series Ed.), D. Scarborough & S. Sternberg (Vol. Eds.), *An invitation to cognitive science. Vol. 4: Methods, models, and conceptual issues* (2nd ed., pp. 703–861). Cambridge, MA: MIT Press.

Sternberg, S. (1998b). Inferring mental operations from reaction-time data: How we compare objects. In D. Osherson (Series Ed.), D. Scarborough, & S. Sternberg, (Vol. Eds.), *An invitation to cognitive science. Vol. 4: Methods, models, and conceptual issues* (2nd ed., pp. 365–454). Cambridge, MA: MIT Press.

Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological Review, 19,* 447–462.

Swets, J. A. (1998). Separating discrimination and decision in detection, recognition, and matters of life and death. In D. Osherson (Series Ed.), D. Scarborough, & S. Sternberg, (Vol. Eds.), *An invitation to cognitive science. Vol. 4: Methods, models, and conceptual issues* (2nd ed., pp. 635–702). Cambridge, MA: MIT Press.

Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology, 14,* 219–238

Townsend, J. T. (1981). Some characteristics of visual-whole report behavior. *Acta Psychologica, 47,* 149–173.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology, 28,* 363–400.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus, 4,* 374–391.

Tulving, E., & Schacter, D. L. (1991). Priming and human memory systems. *Science, 247,* 301–305.

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 336–342.

Underwood, B. J. (1983). *Attributes of memory.* Glenview, IL: Scott, Foresman.

Usher, M. & McClelland, J. L. (1996). *On the time course of perceptual choice: A model based on principles of neural computation.* Manuscript submitted for publication.

Whittlesea, B. W. A., & Podrouzek, K. W. (1995). Repeated events in rapid lists: part 2. Remembering repetitions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1689–1697.

Whittlesea, B. W. A., Dorken, M. D., & Podrouzek, K. W. (1995). Repeated events in rapid lists: part 1. Encoding and representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1670–1688.

Wickelgren, W. (1977). Speed-accuracy tradeoff and information-processing dynamics. *Acta Psychologica, 41, 67–85.*

Wingfield, A., Lindfield, K., & Kahana, M. J. (1998). Adult age differences in temporal characteristics of category free recall. *Psychology & Aging, 13,* 256–266.

Wixted, J. T. & Ebbesen (1991). On the form of forgetting. *Psychological Science, 2,* 409–415.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review, 1,* 89–106.

Woltz, D. J., Bell, B. G., Kyllonen, P. C., & Gardner, M. K. (1996). Memory for order of operations in the acquisition and transfer of sequential cognitive skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 438–457.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: the contribution of recollection and familiarity. *Memory & Cognition, 25,* 747–63.