Remembering and Forgetting

John T. Wixted

In its most common usage, the word memory refers to an assemblage of mental representations of past experience. To study memory from this point of view is to study the structures and processes that have evolved to store and manipulate these representations. In its behavior-analytic sense, by contrast, memory refers not to static mental entities but to the potential to manifest in behavior the effects of past experience. To study memory from this point of view is to study behavior that reflects a previously presented stimulus (i.e., remembering) or the loss of that kind of stimulus control (i.e., forgetting). The vast majority of memory research with humans has been performed with the former interpretation in mind, but many of the techniques developed in that literature are well-suited to the behavioral analysis of remembering and forgetting as well.

Why would anyone be interested in studying memory from a behavior-analytic point of view? One compelling answer to this question was given by Watkins (1990). This article was recently reprinted in The Behavior Analyst because of its obvious appeal to behaviorally oriented psychologists. In essence, Watkins argued that the experimental analysis of memory has reached a dead end because of the field’s heavy reliance on mediationism, according to which the act of remembering is best explained by the existence of a mental representation (or trace) that bridges the gap between the occurrence of an event and the remembering of that event. This practice, according to Watkins, has shifted attention away from behavior and toward an ever-growing and increasingly complicated collection of hypothetical constructs and processes that are impervious to empirical disconfirmation. As an alternative, Watkins suggests focusing on empirical laws relating the stimulus conditions that prevail during learning to the behavior we call remembering. This chapter reviews some of the techniques and procedures that might be used to advance such an endeavor. Before describing what those methods are, a brief review of traditional memory terminology is in order. Although some of these terms are associated with theoretical constructs and hypothetical mental pro-

John T. Wixted - Department of Psychology, University of California at San Diego, La Jolla, California 92093-0109.
cesses, their importance from a behavior-analytic point of view lies in the proce-
dural and conceptual distinctions to which they refer (Wixted, 1989).

THE LANGUAGE OF MEMORY

All learned behavior, including classically conditioned eye blinks and simple
operants such as a keypeck, can be construed as involving memory. Several com-
prehensive memory classification schemes have been proposed in recent years that
attempt to organize all learned behavior according to the different types of mem-
ory they reflect (e.g., Squire, Knowlton, & Musen, 1993). Rather than focus on the
entire gamut of learned behavior, however, the present chapter will focus on the
analysis of behavior that (most would agree) qualifies as remembering. This sec-
tion outlines some of the language used by researchers involved in the study of
remembering in humans. Behavior analysts might prefer a different terminology, but
it is important for anyone planning to conduct research in this area to be aware of
some of the conceptual distinctions that have guided research in this area for most
of the twentieth century.

Short-Term versus Long-Term Memory

Perhaps the most fundamental and well-known distinction is that between
short-term and long-term memory. In their simplest usage, these terms refer to
nothing more than the approximate time interval since the discriminative stimu-
lus was last presented. Tests of short-term memory typically involve delay inter-
vals that range up to 30 seconds or so, with tests of long-term memory involving
delays of days, weeks, or even years.

A distinction based only on temporal parameters is not really very helpful be-
cause it is easy to imagine situations in which the time since stimulus presenta-
tion is irrelevant. For example, if the stimulus in question is a phone number, a
simple and effective mnemonic strategy is to repeat the phone number continu-
ously until the call is made. Under these conditions, it does not matter if the de-
lay interval is less than 10 or greater than 60 seconds. No forgetting is likely to oc-
cur because the effective delay interval is continuously reset to zero by means of
behavioral strategy (rehearsal). Thus, an alternative and more useful interpretation
of the difference between short-term and long-term memory is that the former in-
volves the continuous, uninterrupted control of responding by a stimulus where-
as the latter involves the noncontinuous, delayed control by a stimulus. Although
both cases constitute valid examples of remembering, they are quite distinct. In-
deed, the well-known amnesic H. M. can easily remember the name of a new ac-
quaintance for an indefinite period of time so long as the name is continuously re-
peated. By contrast, if H. M.'s behavior is momentarily controlled by another
stimulus (e.g., if he engages in a brief conversation with another person), the name
is lost forever. Unlike normal subjects, H. M. cannot learn a new name (Ogden &

Note that this chapter will be primarily concerned with delayed stimulus con-
trol, which is what others would refer to as long-term memory (even though the
amount of time between stimulus presentation and the behavior it controls may be

The critical feature of an experiment concerned with delayed stimulus control is an intervening period of time during which behavior is brought under the control of another stimulus. In experiments concerned with the study of long-term memory, this is usually accomplished by means of a distractor task. That is, while subjects may be given a list of pictures or a story to memorize, a short irrelevant task intervenes between study and test. A common distractor task used in verbal learning studies is counting backwards by 3s, but many other tasks may be used.

Semantic versus Episodic Memory

A second dichotomy, advanced most recently by Endel Tulving (1972), distinguishes between responding based on cumulative learning across time and situations (semantic memory) versus responding based on a discrete prior learning experience (episodic memory). Asking a subject to name as many U.S. cities as possible is an example of a semantic memory task because the subject need not (and usually cannot) describe the situations under which each piece of information was acquired. Instead, the information was learned over time under many different conditions. In an episodic memory task, by contrast, the subject is required to identify a situationally specific stimulus. Thus, for example, asking a subject what he or she ate for breakfast that morning or to name the words on a recently presented list would represent tests of episodic memory.

As with the short-term/long-term memory distinction, cognitive psychologists regard episodic and semantic memory as different memory systems (perhaps subserved by different areas of the brain). Whether or not that is the case, the distinction also serves a useful purpose at the level of procedure, which is why it is considered here. Most of the studies discussed below would be classified as episodic (because the experimenter supplies the to-be-remembered material), but a few semantic memory studies will be considered as well.

Implicit versus Explicit Memory

A relatively recent distinction that currently commands a great deal of attention is that between explicit and implicit memory. Explicit memory tasks are those in which the subject is explicitly asked to remember a previously presented stimulus (whether presented during the course of the experiment or not), which is what is done in most human memory experiments. Implicit memory tasks, on the other hand, are indirect tests of memory because the subject is asked to perform some task that does not require, but may nevertheless reveal, the influence of a previous episode. Interest in this distinction is high, in part because amnesic subjects who are severely impaired on explicit memory tasks (be definition) are often unimpaired on implicit tasks (e.g., Warrington & Weiskrantz, 1970). An example of an implicit memory task is “mirror reading.” At first, reading textual information in a mirror is a slow and difficult process because the image is reversed. With practice, however, the task becomes much easier. Amnesic subjects often show improvement on this task that rivals that of normal subjects even though the amnesic subjects have no recollection of having performed the task before.

Although the study of implicit memory is currently in high gear, the present
chapter will mainly be concerned with tests of explicit memory. Only explicit memory tests set the occasion for what can properly be described as an act of remembering. The most commonly used tests in this regard involve either recall or recognition. In a recall test, the subject is asked to reproduce some facsimile of the stimulus in question. For example, the subject might be asked to say aloud or write down all of the words that were presented on a list 10 minutes ago, or to draw as accurately as possible a series of geometric forms that were presented earlier in a session, or to describe a crime enacted on a video that he or she watched the day before. In each case, a reproduction (in one form or another) of the original stimulus is required.

Unlike recall tests, recognition tests do not require the subject to generate previously presented stimuli. Instead, the experimenter presents those stimuli again and asks the subject to decide whether or not the item was seen before. Thus, for example, after studying a list of words, subjects may be presented with those words again and asked to decide whether or not they appeared on the list. Typically, half of the items are old (i.e., they did appear on the list) and half are new (i.e., they did not appear on the list), but other arrangements are certainly possible. In a line-up study, for example, the subject must decide whether or not a “criminal” seen earlier in a video clip is included in a set of six photographs. Alternatively, in a continuous recognition procedure, study and test are intermingled. In this procedure, subjects are asked to decide whether or not each item in a long list of items is being presented for the first time or whether it was presented earlier in the list.

The following sections review specific procedures used to investigate remembering in humans. Although the discussion to follow necessarily makes use of the terms presented above, the main emphasis is placed on how one might investigate remembering and forgetting from a behavior-analytic perspective. The first section considers a variety of procedures used to study the effect of reinforcement on recall and recognition. Subsequent sections consider procedures designed to investigate the rate and pattern of remembering and the role of interference in forgetting.

**REINFORCEMENT AND HUMAN MEMORY**

The last comprehensive review of the effects of reinforcement on human memory was provided by Nelson (1976), and anyone interested in this subject would be well-advised to consult that source. This section reviews some of the methods discussed by Nelson as well as a variety of additional methods a behavior analyst may wish to employ.

**Reinforcement Manipulations during Study**

A number of studies conducted in the early 1970s were concerned with the effect of reinforcement on learning that was later explicitly tested by means of recall or recognition. Loftus (1972), for example, used a recognition memory procedure in which subjects were presented with pairs of pictures to memorize. During study, subjects were informed that one picture of each pair would be worth nine
points for being correctly recognized on a later test whereas the other would be worth only one point. As might be expected, the number of eye fixations to the high-value pictures was greater than to the low-value pictures during list presentation (although that behavior was not itself reinforced). On the subsequent recognition test, subjects were also more likely to correctly recognize high-value pictures than low-value pictures. However, when the data were plotted according to the number of eye fixations a picture received, no effect of reinforcer magnitude was observed. That is, regardless of its assigned value, if a picture happened to receive few fixations during study it was unlikely to be recognized on a later test, but if it happened to receive many fixations during study the probability of correct recognition was high. Thus, the reinforcement manipulation influenced study time, which in turn influenced the accuracy of recognition.

Nelson (1976) reported an interesting study along the same lines that varied reinforcement magnitude either between-subjects or within-subjects. In the between-subject design, subjects studied a list of 40 words and were later asked to recall the words in any order they wished. Half of the subjects (the low-value group) were told they would receive 1 cent for each word correctly recalled. The other half (the high-value group) were told they would receive 10 cents for each correctly recalled word. In agreement with many other studies of this kind, the manipulation was essentially ineffective (probably because subjects were already sufficiently motivated to attend to the task even without a financial incentive). In the within-subject design, by contrast, some of the words on the list were designated as low-value words (1 cent for each one correctly recalled) and others were designated a high-value words (10 cents for each one correctly recalled). When reinforcement magnitude was manipulated in this manner, the effect on subsequent performance was quite large, with the advantage going to the high-value words. Note that his within-subject manipulation is like the one reported by Loftus (1972) pictures, and the result was basically the same. Thus, although adult subjects are generally sufficiently motivated to perform to the best of their ability (which is why the between-subject manipulation was ineffective), it is possible to differentially manipulate attention within a list by varying reinforcer magnitude associated with the remembering response.

Cuvo (1974) reported findings in direct support of the idea that monetary incentive can be used to influence which words subjects study and rehearse (and, therefore, which words they are most likely to subsequently recall). This study also illustrates a procedural detail that will be of particular interest to behavioral psychologists involved in the study of remembering and forgetting. As in the within-subject condition of Nelson's study, subjects were asked to learn a list of words, some of which would earn the subject 10 cents if correctly recalled (and these words were designated as such) and some of which earn 1 cent. While learning the list, subjects were instructed to repeat aloud any words they happened to be thinking about. This overt rehearsal procedure renders behavior that is ordinarily covert observable (cf. Fischler, Rundus, & Atkinson, 1970; Rundus, 1971). Cuvo found that the effect of reinforcer magnitude on later recall was mediated by differential rehearsal. That is, high-value words were rehearsed to a greater extent and were more likely to be recalled than low-value words. When rehearsal was prevented by means of a distractor task during list presentation, on the other hand, the effect
of monetary incentive was not significant. Similar findings have been reported by a number of research groups (Eysenck & Eysenck, 1982; Kuzinger & Witryol, 1984).

Note that the results discussed above suggest that, under certain conditions, reinforcement manipulations can affect the amount of rehearsal applied to different items in a list (i.e., a higher reinforcer magnitude induces subjects to selectively rehearse the high-value words). Although the issue has not yet been extensively investigated, it seems likely that reinforcement manipulations could also be used to affect the kind of rehearsal an item receives. Wixted and McDowell (1989), for example, showed that when the amount of overt rehearsal was held constant for different thirds of a 15-item list, the effects on delayed recall varied considerably. In this study, subjects were instructed to overtly rehearse 5 items from the list for 15 seconds. The remaining 10 items were presented very rapidly and therefore received very little rehearsal. Sometimes the 5 rehearsed items were the first 5 items of the list, sometimes the second 5, and sometimes the last 5. Each list was followed by free recall. At the end of the session an unexpected delayed recall test was administered for all of the previously studied lists (cf. Craik, 1970). Even though the original amount of overt rehearsal was the same for different thirds of the list, the effect of rehearsal on the 5 items at the beginning of each list was much more pronounced than the effect of rehearsal on the last 5 items of each list. Indeed, the latter items were no more likely to be recalled than items receiving no rehearsal.

Effects like these are usually explained by differing kinds of rehearsal. The general conclusion is that elaborative rehearsal (e.g., forming associations, creating mental images) facilitates later recall whereas maintenance rehearsal (rote repetition) does not. In the study discussed above, subjects presumably relied on elaborative rehearsal for items at the beginning of each list (because they knew more items were on the way) and maintenance rehearsal for items at the end of each list (because they knew those items could be rehearsed right up to the point of recall). In one of the few studies of its kind, Bauer and Peller-Porth (1990) found evidence that reinforcement manipulations can influence the kind of rehearsal an item receives. More specifically, they found that the use of monetary reinforcement for correct recall increased the number of items recalled from early list positions in children (including some learning-disabled children) but had no effect on recall for words occupying later positions in the list. They interpreted their findings in terms of incentive-induced elaborative rehearsal applied to the initial items of a list.

Research on differing types of rehearsal may seem foreign to many behavior analysts. However, it is important to keep in mind that “rehearsal” is nothing more than behavior governed by prevailing reinforcement contingencies (whether those contingencies are specified by the experimenter or not). The fact that this behavior is ordinarily covert creates special complications but does not impose an insurmountable obstacle. Moreover, the fact that elaborative rehearsal affects later recall, whereas rote repetition does not, is a simple principle of behavior, not a theory. Cognitive theories of the kind Watkins (1990) criticized seek to explain why elaborative rehearsal has the effects it does (e.g., elaborative rehearsal creates multiple retrieval routes to the memory trace), but, thus far, those theories add little to the principle of behavior they seek to explain. Although cognitive theories of that
Reinforcement Manipulations during a Recall Test

Can reinforcement be used to influence the likelihood that something will be correctly recalled after the study phase is completed? To investigate this issue, researchers do not explain the reinforcement contingencies to the subject until they are about to be tested. Thus, for example, subjects might be asked to study a list of words for later recall. Following a distractor task, one group of subjects might be informed that each correctly recalled word will be reinforced with a relatively large amount of money and another group would be informed that each correctly recalled word will be reinforced with a relatively small amount of money. Typically, this manipulation has little or no effect on performance (e.g., Heinrich, 1968), although an interesting exception is discussed below.

Loftus and Wickens (1970) showed that, under certain conditions, reinforcement contingencies manipulated at test can affect the accuracy of recall. In this study, subjects studied nonsense syllables paired with letters (e.g., DAX-P). On a later test, the nonsense syllable was presented as a cue and the subject was asked to supply the letter with which it has been paired. The critical feature of this experiment was that the trials were self-paced. That is, subjects were free to spend as much time as they wished trying to recall the associated letter before moving on to the next item. Note that this is unlike the typical study in which recall time is fixed. Loftus and Wickens found that recall performance was enhanced at test when the stimulus (e.g., DAX-) was designated as a high-value item (such that correct recall would yield a relatively large reinforcer) compared with when the stimulus was designated as a low-value item. Reaction-time data suggested that this effect occurred because subjects spent more time in the presence of a high-value stimulus (before moving on to the next item) relative to a low-value stimulus. Because recall is not an instantaneous process (a point considered in much more detail later), the more time spent in the presence of the stimulus, the more likely the subject was to eventually produce the correct response. Thus, whenever time to recall is allowed to freely vary, reinforcement is likely to influence performance.

Note that in free recall (in which subjects are asked to recall all previously presented items in any order), it is not possible to vary reinforcement magnitude on an item-by-item basis. In cued recall, however, it is possible. The study reported by Loftus and Wickens represents one example of how this can be done, but many other possibilities exist. For example, imagine that subjects were given a long list of words (say, 60) to study, followed by a free recall test. A subject who could successfully recall 20 words from the list would probably require about 15 minutes or more to do so. Unfortunately for the experimenter, recalling words is only one of several activities the subject might engage in for those 15 minutes. That is to say, the rate of reinforcement for extraneous activity (r_e) unrelated to recall may not be zero even in a bleak running room equipped with nothing more than a computer. Thus, if the payoff for correct recall is low, other stimuli in the environment may begin to control the subject’s behavior after just a few minutes (such that the recall
total would never reach 20 words). If the payoff is high relative to \( r_e \), on the other hand, time devoted to remembering would probably increase considerably (and correspondingly more words would be recalled). Students of the matching law will recognize this argument as a variant of Herrnstein’s (1970) interpretation of behavior (for additional information, see Mazur, Chapter 5). According to this account, the effect of a given rate of reinforcement on an operant response (e.g., a pigeon pecking a key) will vary inversely with the amount of extraneous reinforcement \( (r_e) \). A similar idea may apply in the case of free recall. The details of how one might quantify the flow of remembering in a free recall task (and thereby more precisely measure the effects of a reinforcement manipulation, for example) is covered in more detail in a later section concerned with the rate and pattern of remembering.

The examples discussed above illustrate the effect of reinforcement on recall. The results essentially show that reinforcement manipulations can affect performance by influencing time on task. If time to recall is fixed by the experimenter, reinforcement manipulations have little or no effect. The next section considers the effects of reinforcement manipulations on recognition.

**Reinforcement Manipulations during a Recognition Test**

In a yes/no recognition experiment, test items are presented one at a time during the test phase and the subject must indicate whether or not the item appeared on the list. One cannot study the effect of reinforcement on recognition performance without first considering the dependent measure used to assess performance. At first glance, the most natural dependent measure to use in a situation such as this is the same one used for recall, namely, the percentage of correct responses. However, the use of percent correct as the dependent measure can easily lead to incorrect conclusions about the role of a reinforcement manipulation (or any other manipulation for that matter).

The subject’s responses on a yes/no recognition test can be classified according to a 2-by-2 matrix formed by the combination of the response (yes or no) and the item’s status (old or new). If the item is old (i.e., if it did appear on the list), a yes response is termed a *hit* whereas a no response is termed a *miss*. If the item is new (i.e., if it did not appear on the list), a yes response is termed a *false alarm* whereas a no response is termed a *correct rejection*. Thus, the hit *rate* is the proportion of old test trials in which the subject correctly reported yes and the false alarm *rate* is the proportion of new test trials in which the subject incorrectly reported yes. The hit and false alarm rates provide the information needed to assess a subject’s performance.

Note that if a subject is asked to respond in a liberal manner, the percentage of yes responses will increase (thereby increasing both the hit and the false alarm rate). If that same subject is then asked to respond in a conservative manner, the percentage of yes responses will decrease (thereby decreasing both hit and false alarm rates). Although the percentage of correct responses will usually change considerably as a result of this manipulation (because the change in hits will not exactly offset the change in false alarms), other measures designed to capture dis-
criminability independent of the biasing manipulation will (ideally, at least) remain essentially constant.

**Signal Detection Theory**

As discussed by Irwin and McCarthy (Chapter 10), the most commonly used dependent measure on a task such as this is \(d'\) of signal detection theory (which also yields a measure of bias, \(\beta\)). Figure 9.1 illustrates the signal detection analysis of yes/no recognition data. The abscissa represents strength of evidence (a subjective variable) that an individual test item was seen before. The analysis assumes that the evidence variable associated with new test items (i.e., lures) varies from trial to trial according to a Gaussian distribution. The evidence variable associated with old items (i.e., targets) is also normally distributed and is, on average, stronger than that associated with new items. The vertical line \(c\) represents the subject's decision criterion. On trials in which evidence exceeds \(c\), the subject responds positively, otherwise the response is negative.

The placement of the criterion \(c\) and the distance between the two distributions (\(d'\)) determine the pattern of hits, misses, correct rejections, and false alarms. Whereas \(d'\) is determined by trial-specific factors (e.g., study time), the position of \(c\) is determined by the biasing condition. In the liberal biasing condition, the criterion would be placed more to the left, whereas in the conservative biasing condition, it would be placed more to the right. The hit rate and false alarm rate from an experimental condition can be taken to a reference table and used to determine \(d'\) and \(\beta\) (e.g., McNicol, 1972). Similar measures can be derived from a less theoretical detection account described by Davison and Tustin (1978).

![Signal Detection Theory](image)

**Figure 9.1.** A graphical illustration of signal detection theory. The target and lure distributions correspond to previously seen and previously unseen test items, respectively. The decision criterion is represented by \(c\). When strength of evidence exceeds \(c\), the subject responds "yes," otherwise the response is "no."
ROC Analysis. One advantage of using the yes/no recognition procedure is that a receiver operating characteristic (ROC) analysis can be performed on the data to answer interesting theoretical questions. An ROC plot depicts the hit rate versus false alarm rate obtained from several (at least three) biasing conditions in which memory variables like study time and delay interval are held constant. Figure 9.2 depicts a hypothetical ROC curve (top panel) and corresponding signal detection interpretation (bottom panel). The point higher and to the right in the ROC reflects a liberal criterion for giving a "yes" response (corresponding to $c_1$ in the bottom panel). The point lower and to the left reflects a conservative criterion for giving a "yes" response (corresponding to $c_3$). Some older and intuitively appealing theories predict that the form of this function will be linear. Indeed, "high threshold theory," which once prevailed as an account of recognition memory,
predicts a linear relationship between the hit rate and false alarm rate (Green & Swets, 1966). The Gaussian distributions of signal detection theory are consistent with the curvilinear shape shown in the figure. The shape of this function also serves as a test of the assumption that the new and old distributions have equal variances. If so, the curve will be symmetrical about the diagonal, otherwise a more serpentine path will be traced out by the data points.

ROC plots are best analyzed using data from individual subjects rather than using hit rates and false alarm rates averaged over subjects. This generally means that subjects should be run for multiple sessions during which they learn and attempt to recognize multiple lists of words. Ratcliff, Sheu, and Gronlund (1992), for example, tested four subjects for 20 recognition sessions during which they learned 10 lists per session. Bias was manipulated by varying the probability that a test item was a target or a lure. For example, when 80% of the test items were lures (and subjects were informed of that fact), response bias was tilted in favor of a "no" response. When 80% of the test items were targets, bias was tilted in favor of a "yes" response. Over the 20 sessions, a sufficient amount of data was collected to permit an ROC analysis for each subject. The ROC data were not perfectly symmetrical, suggesting that the most accurate detection representation would involve a target distribution with slightly greater variance than the lure distribution.

Effect of Reinforcement on Bias in Recognition. Instead of varying the mix of target and lures on the recognition test to influence bias, reinforcement contingencies can be manipulated. Thus, for example, when a correct "yes" response yields a much larger reinforcer (or a higher probability of reinforcement) than a correct "no" response, subjects will be much more inclined to give "yes" responses (and give "no" responses only when they are certain the test item did not appear on the list). Although such effects have not been thoroughly investigated, the general assumption is that monetary incentives affect criterion placement (i.e., bias) without affecting $d'$ (e.g., Zimmerman & Kimble, 1973). However, this is an issue that probably needs to be examined in greater detail. In the animal literature, for example, reinforcement manipulations in discrimination tasks can affect both bias (the general inclination to choose one alternative) and discriminability (the ability to distinguish between two previously presented sample stimuli). The literature on the differential outcomes effect in animals shows this quite clearly (e.g., Santi & Roberts, 1985). That may be true in humans as well, in which case signal detection theory would not unambiguously disentangle these two measures of performance.

Baron and Surdy (1990), in one of the few human memory studies reported in the Journal of the Experimental Analysis of Behavior, used a continuous recognition procedure to study the effect of reinforcement on bias and discriminability (i.e., remembering) in older and younger adults. As indicated earlier, a continuous recognition procedure intermingles the study and test phases of the experiment. During a session, a long string of items is presented and some of the items are occasionally repeated and for every item, the subject must decide (yes or no) whether or not it was presented earlier in the series. Baron and Surdy manipulated the type of stimuli used (alphanumeric strings, words, or sentence) and manipulated the
payoff matrix to influence bias as well. Their results showed that accuracy (measured by $A'$, which is conceptually similar to $d'$) was lower for older subjects, and the reinforcement manipulation affected bias to a greater extent in younger subjects. This latter result suggested that older subjects tended to adopt a more rigid response style than younger subjects. That is, when reinforcement contingencies change in a recognition procedure, older subjects do not adapt to the changed contingencies as readily as younger subjects. This result may or may not be restricted to the domain of remembering, but this study nevertheless illustrates the potential value of a behavioral analysis of human recognition.

Detection Theory and Behavior Analysis. Classical detection theory obviously entails slightly more theoretical assumptions than most accounts in behavior analysis. However, it is worth noting that the classical model has stood the test of time in a way that other theoretical accounts have not. Indeed, its developmental history is quite unlike that associated with the personalized cognitive theories described by Watkins (1990). Unlike most models, signal detection theory has not been relentlessly embellished with a collection of trace features, time tags, mental conveyor belts, and the like. Instead, as applied to a simple yes/no recognition task, it is exactly the same theory that was introduced to psychology more than 30 years ago. The theory avoids gratuitous embellishment because it is mathematically precise and self-contained (constrained as it is by the shape of an ROC). Indeed, the theory can be viewed as merely one way to represent behavioral data (as much a tool as a theory). Unlike unconstrained cognitive models that become increasingly divorced from the empirical world over time, important components of signal detection theory (e.g., the location of the decision criterion) are directly translatable into behavior. The next section considers some issues concerning criterion placement that may be of particular interest to behavior analysts.

Reinforcement and Confidence in Recognition Decisions

Yes/no recognition decisions are sometimes made with little or no confidence and sometimes made with great confidence. Asking a subject to supply a confidence rating for each response introduces a new and interesting angle to consider. How are those ratings to be interpreted and why would anyone, especially a behavior analyst, be interested in them? On the surface, the analysis of subjective ratings may seem to be better suited to cognitive psychology than behavior analysis. However, despite appearances to the contrary, this is one area of research where the information processing perspective encounters an uncomfortable dilemma that is rather easily explained in terms of a subject’s reinforcement history.

The Confidence-Based ROC Analysis

The signal detection framework discussed above can be readily extended to the analysis of confidence judgments. As a first step, ROC plots can be produced by computing hit and false alarm rates separately for each confidence rating (instead of for different biasing conditions). Consider, for example, an experiment in
which each yes or no judgment is accompanied by a confidence rating on a 1 to 3 scale (ranging from guessing to absolute certainty). With this scheme, the possible responses to any test item are no-3, no-2, no-1, yes-1, yes-2, or yes-3 in increasing order of confidence that the test item appeared on the list. Note that this provides a richer set of data to analyze than is provided by simple yes or no responses.

For the target items (i.e., those that were on the list), a hit rate can be determined separately for each confidence rating by simply computing the proportion of targets that receive a rating at least as high as the confidence rating in question. Every target receives a rating of at least no-3 (because no-3 is the lowest possible rating), so a meaningful hit rate for this confidence rating cannot be computed (i.e., the hit rate is always 1.0 for the no-3 rating). However, a smaller proportion of targets (perhaps .95) receives a confidence rating of at least no-2 (i.e., no-2, no-1, yes-1, yes-2, or yes-3). That proportion is the "hit" rate associated with the no-2 confidence criterion even though, technically, any targets receiving a response of no-2 or no-1 represent incorrect responses. In a similar way, one can compute the hit rate associated with the no-1 criterion by calculating the proportion of target responses that receive a rating of at least no-1. Continuing in this manner yields five meaningful hit rates, one each for no-2, no-1, yes-1, yes-2, and yes-3. False alarm rates for each of these five confidence criteria can be computed in exactly the same way. The false alarm rate for the no-2 confidence criterion, for example, is the proportion of lures receiving a rating of at least no-2, and the false alarm rate for the no-1 confidence criterion is the proportion of lures receiving a rating of at least no-1. The five hit and false alarm rates produced in this manner can be used to construct an ROC (which, again, is simply a plot of hit rate versus false alarm rate). Almost invariably, the procedure yields an orderly (and quite typical) ROC plot (MacMillan & Creelman, 1991).

Figure 9.3 shows the signal detection interpretation of the confidence-based ROC. The locations of the five confidence criteria are nothing more than direct translations of the five false alarm rates. The criteria are placed such that the proportion of the lure distribution exceeding each criterion corresponds to the false alarm rate associated with that criterion. For example, if 2.5% of the lures receive

![Figure 9.3. A graphical illustration of the signal detection interpretation of confidence judgments. Each vertical line represents a different confidence criterion. When strength of evidence exceeds y3, a high-confident yes response is given. When it exceeds y2 but no y3, a less confident yes response is given (i.e., yes-2). Similarly, if strength of evidence falls above n2 but below n1, a response of no-2 if given. Note that if evidence falls below n2, a high-confident no response is given (i.e., no-3).](image-url)
a high-confident yes response (i.e., if the false alarm rate for yes-3 is 2.5%), then
the yes-3 criterion is assumed to be located 2 standard deviations above the mean
of the lure distribution (such that only 2.5% of the lure distribution falls to the
right of yes-3). If 50% of the lures receive a response of no-1 or greater, then the
no-1 criterion is located at the mean of the lure distribution (such that half of that
distribution falls to the right of no-1). Similarly, the target distribution is posi-
tioned in such a way that the proportion of the signal distribution to the right of
each confidence criterion is equal to the hit rate associated with the level of con-
fidence. For the sake of simplicity, equal variance distributions are assumed in this
example (although that assumption need not be made). The smooth curve drawn
through the ROC data represents the locus of points that would be produced by an
infinite number of confidence criteria.

One question of interest is how the ROC plots (and corresponding signal de-
tection representations) change as a function of discriminability (i.e., $d'$). To an-
swer this question, one merely manipulates discriminability (e.g., by manipu-
late ing study time) and plots the confidence-based ROCs for each condition. To avoid
carry-over effects in the use of the confidence scale, the two discriminability con-
ditions should probably be run in separate sessions. Figure 9.4 shows a repre-
sentative finding from my laboratory. In this study, 14 subjects studied two lists of 48
words in each of two sessions. Each of the four lists (two in each session) was
followed by a yes/no recognition test involving the 48 targets randomly intermixed
with 48 lures. During the recognition tests, subjects were asked to decide whether
or not the item appeared on the previous list (yes of no) and to supply a confidence
rating (1 to 5). In one session (the weak condition), the words comprising each list
were presented at a rate of one per second. In another session (the strong
condition), the words were presented three times each during list presentation. The left
panel of Figure 9.4 shows the confidence-based ROCs using group data, and the
right panel shows the signal detection interpretation of these results (the identical
analysis performed on individual subject data showed these findings to be repre-
sentative of individual subjects). The finding of interest here is that the confidence
criteria fan out as $d'$ decrease (which is tantamount to saying that the points in beh-
avioral ROC data spread out).

This fan effect is just what one venerable information-processing theory of dis-
criminability predicts. That theory states that subjects place their criteria in such a
way as to maintain constant likelihood ratios (e.g., Glanzer, Adams, Iverson, &
Kim, 1993). Thus, for example, the yes-3 criterion might always be placed at the
point where (say) the odds are 10 to 1 in favor of a yes response being correct (re-
gardless of what $d'$ is). That is, no matter what $d'$ might be, the yes-3 criterion will
be placed at the point where the height of the target distribution is 10 times that
of the noise distribution. Although not intuitively obvious, this theory requires the
fan effect shown in Figure 9.4.

This theory makes the right prediction, but even many cognitively oriented
psychologists are dubious of the likelihood ratio account because of the extensive
knowledge and considerable computational ability the theory assumes that sub-
jects have in their possession. Specifically, in order to maintain a constant likeli-
hood ratio of 10 to 1 for yes-3, subjects would need to know the mathematical forms
of the target and lure distributions (Gaussian) and be able to compute the ratio of
the heights of those two distributions at a given point on the evidence axis. Assuming such abilities, subjects could, no matter what \( d' \) is, locate the point on the evidence axis where the height of the target distribution is 10 times that of the lure distribution and place the yes-3 criterion at that point. Doing so would ensure that the odds that a high-confident yes response is correct is always better than 10 to 1 (and the fan effect shown in Figure 9.4 would be observed).

The problem with this account is that it seems to throw the concept of parsimony to the winds. Even if subjects are aware of the Gaussian shapes of the target and lure distributions (which seems like a rather strong assumption), the further assumption that subjects can compute the ratio of two Gaussian distributions at a given point on the evidence axis stretches the imagination to the breaking point. Nevertheless, it is hard to imagine an information-processing explanation for the results shown in Figure 9.4 without appealing to this computationally intensive idea.

On the other hand, the result shown in Figure 9.4 is almost obvious when the subject's probable reinforcement history is taken into account. Indeed, when considered from this point of view, the likelihood ratio computation required by an information-processing account is readily seen as a surrogate for the subject's reinforcement history. Subjects do not step out of a vacuum into the experimental situation, but instead arrive after long training in the use of the English language. As part of that training, subjects have presumably encountered extensive feedback concerning expressions of confidence in their recognition decisions. Sometimes,
perhaps, they have expressed a great deal of confidence in decisions that turned out to be wrong (and in those cases there may have been considerable consequences to pay). Other times, the subject may have expressed low confidence in recognition decisions that turned out to be correct, in which case the subject may have missed out on rewards that the verbal community may otherwise have offered (given that confidence is a highly regarded trait).

What would the effect of that kind of training be? Unless they are immune to the consequences of their behavior, subjects would learn that when conditions are unfavorable (e.g., when learning time was brief), one should be very conservative before giving a high-confident yes or not response. In terms of signal detection theory, this translates into a more conservative placement of the extreme no-2 and yes-3 criteria when \( d' \) is low. If a more conservative strategy were not adopted when conditions were unfavorable (e.g., if the criteria did not fan out), subjects would soon find themselves making many high-confident errors under those conditions.

The effect of consequences on confidence ratings has not been extensively investigated, but the area seems ripe for behavior analysis. Some general strategies that have been used in this regard are reviewed next. Note that, as evidenced by the discussion presented below, one need not adopt the signal detection model to investigate this issue in a productive way. Nevertheless, interpreting the results in terms of that relatively simple model offers a way to connect to the cognitive literature and directly contrast the interpretation of findings from a behavioral point of view versus an information-processing point of view.

**The Experimental Analysis of the Effects of Feedback on Confidence**

The developmental literature shows that, on a variety of tasks, young children exhibit confidence judgments that are poorly calibrated to accuracy compared with those of older children. In fact, young children have been referred to as “eternal optimists” because of their tendency to express high confidence in all of their decisions (Newman & Wick, 1987). With feedback, however, the calibration exhibited by both younger and older children improves. An example of the effect of feedback on confidence in a nonmemory task is provided by Newman and Wick (1987). They exposed children to a task in which they were required to estimate the number of random dots in a display and to report their confidence in each response. Some displays were relatively easy (involving relatively few dots) and some were hard (involving many dots). In the absence of feedback about whether or not the response was correct, subjects’ confidence ratings were poorly calibrated to accuracy (i.e., confidence did not decrease as difficulty increased and performance worsened). Following feedback, however, the calibration exhibited by older children improved significantly (as it did for higher skilled younger children as well). That is, as accuracy decreased with increasing task difficulty, confidence decreased as well.

Dot counting is not the same as remembering, but this experiment serves to underscore the point that feedback can affect the validity of confidence judgments. Using a continuous recognition procedure, Berch and Evans (1973) found that even children in kindergarten were capable, to some extent, of monitoring the accuracy of their own performance. In this experiment, subjects were given a series
of digits and asked to recognize each one as being new (i.e., as being presented for the first time) or old (i.e., as having appeared before in the series). Confidence judgments were obtained by having the child point to one of two pictures representing differing degrees of confidence. One picture showed a child with a puzzled expression (low confidence) and another showed a child with a self-satisfied smile (high confidence). Thus, a two-point confidence rating scale was used. The older children exhibited a clear awareness of their own abilities. That is, when accuracy was low, confidence was low as well (as it should be). The younger children show the same effect, but it was much less pronounced. Presumably, the younger children exhibited lower calibration because they had not yet experienced sufficient training in expressions of confidence. In terms of signal detection theory, the younger children would probably be less likely to exhibit the fan effect shown in Figure 9.4.

Very little research using adults has been directed at the effects of feedback on confidence in remembering, but a study by Stock, Kulhavy, Pridemore, and Krug (1992) is consistent with the ideas presented earlier. Subjects in this experiment completed self-paced general-knowledge multiple-choice questions (i.e., this was a semantic memory task) and provided a confidence rating for each response. The experimenter-supplied feedback (correct versus incorrect) after every response. The finding of interest was that subjects spent more time studying the feedback (before moving on to the next question) for high-confident errors than for low-confident errors. Similarly, they spent less time studying feedback for high-confident correct responses than for low-confident correct responses. The authors interpreted their findings in the following way: "These results were explained by the proposition that people try to reduce discrepancies between what they think they know and what feedback indicates they know" (p. 654). One can reasonably assume that the same happens in everyday life and that this, perhaps, accounts for what looks to be rather extraordinary likelihood ratio computation when a subject’s learning history is ignored.

The limited amount of work reviewed above on the relationship between confidence and accuracy on the one hand and corrective feedback on the other shows that this area of research is wide open. The behavior-analytic community seems particularly well-suited to investigate these issues because it is unlikely to follow the information-processing path that ends with the necessary assumption that subjects are capable of extraordinary computational feats.

**RATE OF AND PATTERN OF REMEMBERING**

One of the defining features of early behavior analysis was a detailed inquiry into the rate and pattern of a continuous stream of behavior. That same approach applied to remembering may eventually prove to be equally fruitful. The act of remembering, at least by means of free recall, is generally not an instantaneous process, but instead occurs over an extended period of time. When a subject is asked to list as many foreign capitals as possible (a semantic memory task), or to recall a recently presented list of words (an episodic memory task), the result is almost always the same: a rapid burst of responding followed by a gradual decline
to zero. Performance in this situation is perhaps best thought of as being under the control of multiple temporally remote stimuli, each competing for the control of behavior in a given instant (which is the behavioral interpretation of what a cognitive psychologist would call a search set). Almost all of the previous research on free recall has been concerned with the number or proportion of items correctly recalled. More interesting behavioral properties, such as the rate and pattern of free recall, have received scant attention. These are properties that seem to be of natural interest to the behavioral community, which has long been concerned with the rate of pattern of behavior maintained by schedules of reinforcement.

As indicated above, the time course of free recall can be studied using either semantic or episodic memory procedures. In a semantic memory procedure, the subject is simply asked to list as many items in a category as possible (e.g., “Name as many cities as you can think of”). The responses are timed, literally producing a cumulative record. Originally, timing methods were quite crude, but adequate. Subjects were asked to write down their responses (instead of saying them aloud) and to draw a line under the most recently recalled word every minute. In that way, the experimenter could later reconstruct cumulative progress (i.e., the number of words recalled up to each minute of the recall period).

Another timing method that more closely approximates the operations of a cumulative recorder was first used by Bousfield and Sedgewick (1944) and, apparently, next used by Wixted and Rohrer (1993). In this method, subjects are asked to recall items aloud and the experimenter taps a lever attached to a recording device. That device might be a rolling drum that traces out a record (Bousfield & Sedgewick, 1944) or a computer, which records the time of each response and permits the later plotting of progress in either cumulative or noncumulative form (Wixted & Rohrer, 1993). A voice-activated relay attached to a computer provides an even more accurate way to time recall. Figure 9.5 shows one recall record produced by a single subject studied by Bousfield and Sedgewick (1944). The subject was asked to name as many U.S. cities as possible and the resulting “kymographic” record is shown in the figure (estimated from their Figure 4). This figure clearly shows the negatively accelerated time course that typifies free recall.

The record presented in Figure 9.5 does not include repetitions. Thus, for example, if the subject said “Tucson” 10 seconds into the recall period and repeated that city 5 minutes later, only the first response would be included. Also, if the subject mistakenly said “Arkansas,” the response key would not be depressed. Fortunately, such errors occur very rarely and do not significantly affect the cumulative record whether they are included or not.

Exactly the same procedures can be followed to study free recall performance on an episodic memory procedure. In this case, the subject is first exposed to a list of words and is then asked to remember as many of the words as possible (with each response timed as before). In a typical case, a subject might be asked to read 10 words presented one at a time for 1 second each on a computer screen. Following a short distractor task designed to prevent rehearsal, the signal to begin recalling the words is given. The duration of the recall period should be long enough to allow performance to approach asymptotic levels. For lists involving 5 or fewer words, 30 seconds is probably sufficient. However, for lists as long as 60 words,
a recall period of 15 to 20 minutes may be needed. This fact suggests an interesting point about how contingencies of reinforcement might affect free recall. Specifically, reinforcement for correct responding may not improve memory per se, but, as indicated earlier, it may nevertheless improve performance by motivating subjects to continue on task long after they mistakenly believe they have exhausted the supply of recallable words (Nelson, 1976).

Generally, too few words are recalled on a single trial to produce a smooth record. Therefore, recall totals over multiple trials (and over subjects) are often summed together. Bousfield and Sedgewick (1944) observed that their individual and group cumulative recall functions were reasonably well described by an exponential of the form

$$F(t) = N(1 - e^{-t/\tau})$$

(1)

where $F(t)$ represents the cumulative number of items recalled by time $t$, $N$ represents the number of items recalled given unlimited time (i.e., asymptotic recall), and $\tau$ represents the average latency to recall associated with the $N$ items that are ultimately recalled. The smaller $\tau$ is, the faster the rate of approach to asymptote and vice versa.

Figure 9.6 presents a group cumulative free recall function taken from one of the conditions reported by Bousfield and Sedgewick (pleasant activities) along with the best-fitting exponential. This figure, which shows the cumulative number of items recalled up to each point in the recall period, clearly illustrates the two properties that characterize the time course of free recall: asymptotic recall (indicated by the dashed line) and rate of approach to asymptote, which in this case is rather gradual. Experimental manipulations that affect one property of recall may or may not affect the other. Most free recall experiments report the num-

![Figure 9.5](image-url)
FIGURE 9.6 The cumulative number of pleasant activities generated as a function of time. The data were taken from Table 1 of Bousfield and Sedgewick (1944), Journal of General Psychology, Vol. 30, p. 151. The points represent the average values of 18 subjects and the smooth curve represents the best fit of Equation 1. The estimated asymptotic level of recall is indicated by the dashed line. Reprinted with permission of the Helen Dwight Reid Educational Foundation. Published by Heldref Publications, 1319 18th St. NW, Washington, DC 20036-1802. Copyright 1944.

ber of items recalled by the end of an arbitrarily defined recall period (which represents a single point on the cumulative recall curve). This measure fails to indicate whether or not subjects were still making progress when the recall period ended and, if not, whether the final level of performance was achieved rapidly or slowly. A much better alternative is to track the entire time course of recall and analyze the data by fitting a two-parameter growth function (or, equivalently, by analyzing the noncumulative recall latency distributions).

What variables affect the pattern of behavior shown in Figure 9.6? The first question that might occur to an operant psychologist in this regard is how reinforcement for correct responses affects $N$ and $\tau$. The answer is unknown because the experiment has never been performed. Based on the review presented earlier, one might surmise that a simple schedule of reinforcement (e.g., 25 cents for each correct response) would induce subjects to remain engaged in the task for a longer period of time. If so, asymptotic recall ($N$) might increase. The effect on $\tau$ is harder to predict. Rohrer and Wixted (1994) and Wixted and Rohrer (1993) showed that some procedural variables (e.g., list length) affect $\tau$ and some (e.g., study time) do not, but much more work is needed to elucidate the principles governing this basic property of remembering.

The considerations discussed above do not apply to situations in which accurate recall consists of a single response (e.g., “Who was the lead actor in Taxi Driver?”). However, much remembering in the real world consists of a stream of behavior, as when a witness is asked to describe a crime scene, or a student is asked to recall the details of a lecture. Under these conditions, questions about the factors that govern the rate and pattern of responding become relevant. With some notable exceptions (e.g., Roediger & Thorpe, 1978), such questions have been all but ignored by cognitive researchers and, it seems fair to say, by behavior analysts as well.
INTERFERENCE AND FORGETTING

The ability of a stimulus to exert delayed control over behavior is not independent of the stimuli that precede or follow it. Thus, for example, it is relatively easy to remember the name of one new student introduced to a class. Remembering that student’s name after hearing 20 more students introduced is quite a bit more difficult. This example illustrates the obvious point that much forgetting is related to the learning of other material (i.e., interference). This section reviews some of the procedures that can be used to study this phenomenon. Research conducted along these lines may be of particular interest to behavior analysts because, as we shall see, the contrast between cognitive theories advanced to explain interference effects and an empirical law proposed by Watkins (1990) is stark indeed.

Paired-Associates

In this procedure, subjects learn one list of item pairs, which can be represented as the A-B list, and then learn a second list of item pairs in which the first item remains the same but the second is changed. This list can be represented as the A-C list. The individual items may consist of digits, nonsense syllables, words, or any combination of these. During the recall test, subjects are presented with the A items and asked to recall either the corresponding B items or the corresponding C items. An impairment in the ability to recall the targeted items relative to a control group that learns only one list can be attributed to interference caused by the learning of the other list.

Retroactive Interference

To study retroactive interference (i.e., interference by subsequent learning), investigators typically used the A-B, A-C, A-B paradigm, in which subjects study one list of A-B item pairs, followed by a list of A-C item pairs, followed by a cued recall test in which the A terms are supplied and subjects are asked to supply the B terms. A control group might receive no A-C list or be given a list consisting of all new items. A decrement in recall performance for the experimental group relative to the control group (the typical finding) can be attributed to interference of A-B associations caused by the subsequent learning of A-C associations. The typical finding with this procedure is that performance is quite impaired immediately after the A-C list is learned, but the degree of impairment relative to the control group lessens as the retention interval increases (see Crowder, 1976, for a review).

Although the analysis discussed above relied on lists of words, many other conceptually similar procedures can be used. Prominent among these are studies of event memory (Loftus, Feldman, & Dashiel, 1995). In a typical experiment of this kind, subjects view an event (such as a videotaped robbery or car accident) and later answer questions about it. One prominent issue in this literature concerns the effect that misleading postevent information has on what a subject reports as having occurred. For example, if the event involved a car accident at an intersection controlled by stoplights, the subject might be asked the following misleading question: “Did the red car or the blue car run the stop sign?” Later, when
asked to describe the accident scene, subjects will often report having seen a stop sign. Note that this can be construed as a retroactive interference design because the interfering material is presented after the to-be-remembered even is observed.

**Proactive Interference**

To study proactive interference (i.e., interference by prior learning), investigators often used a similar procedure represented by A - B, A - C, A [C]. In this case, recall is cued by the A terms (as before) but subjects are asked to supply the C terms. A control group either receives no A-B list or studies a list of word pairs that share no items with the subsequent A-C list. A decrement in recall performance here can be attributed to interference of A-C associations caused by the prior learning of A-B associations. The typical finding with this procedure is that performance is relatively unimpaired immediately after the A-C list is learned but drops off rapidly as the retention interval increases.

Variations of this basic procedure were developed to answer interesting questions about the status of associations that appeared to be lost via interference. Did those associations fully extinguish or were they merely overshadowed by the interfering associations? For example, in the modified free recall (MFR) procedure, subjects were asked to supply the first item that the A stimulus brought to mind (rather than instructing them to recall the B term or the C term) to determine which association still exists. This technique, however, still does not reveal whether both associations (A-B and A-C) are simultaneously intact. Thus, in the modified modified free recall (MMFR) procedure, subjects were asked to supply both the B and C terms in response to A. In the proactive interference procedure, this technique helps to reveal if a subject’s inability to recall a particular C item is related to the continued survival (or perhaps spontaneous recovery) of the B item previously associated with A. Curiously, results from studies like this often reveal that the especially rapid forgetting of C items as a function of time occurs despite the fact that B items cannot be remembered either (Crowder, 1976).

**The Brown-Peterson Task**

Another popular procedure designed to study interference and forgetting is the Brown-Peterson task. In this procedure, subjects are presented with a to-be-remembered item (such as a word or a nonsense syllable), followed by a distractor task of varying duration (e.g., requiring the subject to count backwards by 3s), followed by a free recall test (i.e., the subject is asked to supply the word). The typical finding is that performance declines in curvilinear fashion as a function of the duration of the retention interval (i.e., the duration of the distractor task). Note that if no distractor task were used, performance would almost surely remain close to 100% correct (because the subject would simply rehearse the item continuously).

**Retroactive Interference**

As with the paired-associates procedure, variants of the Brown-Peterson task can be used to study the effects of retroactive and proactive interference. For ex-
ample, the similarity of the distractor task material to the to-be-remembered item can be varied to study retroactive interference. Such effects are most easily observed when the to-be-remembered item consists of items such as digits or nonsense syllables rather than words. For example, if the to-be-remembered item is “hkj,” recall performance will be worse if subjects are asked to shadow (i.e., repeat) consonants during the retention interval than if they are asked to shadow digits.

As an aside, it might be noted here that the standard Brown–Peterson procedure described above can be used to study the time course of forgetting over the short term. Wixted and Ebbesen (1991), for example, presented subjects with lists of six words to study. Each list was followed by a demanding distractor task (repeating additional words) for varying lengths of time ranging from 2.5 to 40 s. If subjects are run for several sessions, forgetting functions (i.e., performance plotted as a function of delay) can be plotted for each subject individually. The advantage of this procedure is that it allows for a rather precise quantification of the behavior of interest. Wixted and Ebbesen (1991) found that the time course of forgetting was well-described by a power function and poorly described by the exponential. The quantification of behavior has not been a main focus of cognitively oriented memory researchers, but it seems to be a natural focus for behavior analysts. As yet, little or no work has investigated what variables affect the parameters of the power function.

**Proactive Interference**

The Brown–Peterson procedure turned out to be an especially useful way to study the powerful effects of proactive interference as well. Initially, it was observed that the rate of forgetting on the first trial of a session was much less than that on later trials. More specifically, Keppel and Underwood (1962) found that whereas performance following a 3-s retention interval was relatively unaffected by the number of previous trials, performance following an 18-s retention interval was profoundly (and negatively) affected with each passing trial.

Wickens (1972) developed an illuminating procedure that showed rather convincingly that the proactive interference effects described above stemmed from the similarity of the to-be-remembered item and previously learned items. In this procedure, the to-be-remembered items were drawn from a single category for several trials in succession. For example, the items for Trials 1 through 4 might be *ruby, diamond, emerald, and sapphire*. On the first trial, the probability of recalling *ruby* after a 15-s retention interval might be quite high. With each ensuing trial, however, the probability of recalling the precious stone presented on that list would decline. Wickens showed that the original level of performance could be restored by switching to a new category. For example, if the to-be-remembered item on the fifth trial were *Spain*, the probability of recall would increase to a level close to that observed on Trial 1 (but would decline with subsequent trials involving countries). This phenomenon is known as *release from proactive interference*. That phenomena such as this might be of interest to behavior analysts is perhaps best illustrated by considering the radically different ways such data can be explained.
Cognitive versus Behavioral Accounts of Interference

Most accounts of the interference phenomena discussed above are (not surprisingly) cognitive in nature. However, one notable exception can be found in the literature. Watkins (1990) forcefully argued that such mediationist accounts of remembering tend to remove attention from basic principles of behavior that accommodate the data in a more direct and much simpler way. Specifically, all of the interference phenomena discussed above are consistent with the cue overload principle (a principle that involves no hypothetical constructs). This principle simply states that the more items associated with a cue, the less effective that cue will be. The power of this principle is easy to underestimate because of its sheer simplicity.

The cue overload principle easily accounts for both retroactive and proactive interference effects. In both cases, associating more items with a cue results in the loss of that cue's effectiveness. The main difference between the two procedures concerns when the interfering items are associated with the cue. In the retroactive interference procedure described above, the interfering item is associated with the cue after the to-be-remembered item is associated with the cue. In the proactive interference procedure (both the paired-associates and release from PI paradigms), the interfering items are associated with the cue before the to-be-remembered item is.

This principle can also accommodate a number of other robust findings. Prominent among these is the list-length effect. This simply refers to the fact that the longer a list is, the worse performance tends to be (and this is true whether recall or recognition is used). A number of cognitive models designed to explain this result (for recognition) assume that, when faced with a test item on a yes/no recognition test, subjects mentally compare the test item to each memorized item from the list. Each comparison yields at least some feeling of familiarity (even if the test item does not match the memorized item). The closer the match between the memorized item and the test item, the higher the familiarity is. Thus, for targets, the comparison process yields a series of low familiarity values and one high familiarity value (which occurs when the target is compared with the matching item from the list stored in memory). For lures, the comparison process yields a series of low familiarity values only (because the lure does not match any of the memorized items). Finally, the familiarity values from all such comparisons are summed to yield the actual familiarity associated with the test item. If that value exceeds a decision criterion, a yes response is given, otherwise the response is no. This kind of model is known as a global matching model (e.g., Clark & Gronlund, 1996). It can accommodate list length and interference because each new item in the list (or each interfering item) contributes noise to the process, thereby decreasing discriminability.

Although global matching models can explain the list-length effect, its theoretical intricacies are plain to see. By contrast, the cue overload principle accounts for the effect in a much more efficient way. The more items on the list, the more burdened the cue becomes (the cue in this case being nothing more than the experimental context). The claim by Watkins (1990) that his is a simpler and more empirically testable approach is especially evident when the cue overload principle is compared with a global matching model.
Some of the procedures devised to investigate remembering and forgetting in humans have a cognitive flavor to them, but that should not necessarily be taken to imply that the interesting phenomena they generate are immune to a behavioral analysis. Indeed, the selective review presented here suggests that many important topics in this field are more profitably studied from a behavioral perspective than from a purely information-processing point of view. Although it is perhaps true that the behavioral analysis of human memory to date does not have an especially Skinnerian look and feel, it is nevertheless clear that much can be gained by pursuing the subject according to principles most closely identified with the field of behavior analysis.

One important aspect of remembering that is conspicuously missing from nearly all cognitive models of memory is the effect of the subject’s reinforcement history. By excluding any consideration of the consequences of remembering, purely cognitive theories are sometimes forced to rely on seemingly implausible assumptions about a subject’s computational abilities. This was especially true of theories concerned with the confidence a subject expresses in his or her recognition decisions. Because subjects appear to behave in a more-or-less optimal way, the implication appears to be that they are capable of computing likelihood ratios on a moments’ notice. How else could it be that subjects know when to appropriately express high confidence in a recognition decision and when to appropriately concede that they are merely guessing? It is a question a behavior analyst is likely to ask, and the answer, undoubtedly, involves some consideration of the subject’s learning history. The details of this learning process are mostly unknown, but only because behavior analysts have not yet taken a close look at this interesting issue.

Another important issue that may be of particular interest to behavior analysts concerns the search for empirical laws of remembering. Over the last few decades, the search for empirical laws has given way to the construction of comprehensive theories about the inner working of memory. Watkins (1990) bemoaned that disturbing trend and offered a concrete example of what simple empirical laws have to offer. His cue overload principle is surprisingly powerful, but it is surely not the only one of its kind. Other, equally powerful laws are presumably waiting to be discovered by those who are willing to search for them. Behavior analysts have always been adept at finding general empirical laws of behavior, and it is hard to imagine why they would not be equally successful within the domain of remembering and forgetting.

REFERENCES


