

Forgetting functions

K. GEOFFREY WHITE

University of Otago, Dunedin, New Zealand

Forgetting functions describe how accuracy or discriminability declines as the temporal distance from the event to be remembered increases. Several two-parameter mathematical functions are equally successful in fitting data from a range of studies. The choice of function must therefore be made on theoretical grounds. Consistent with the proposal that remembering follows the principles of discrimination, remembering is specific to the time of retrieval and is independent of the level of performance at earlier times in the retention interval. Empirical evidence for this temporal independence supports constant-rate forgetting. Exponential functions are therefore favored descriptions of forgetting functions because they describe how the influence of the temporally distant event declines at a constant rate. A combination of exponential generalization with an exponential decrease in the effect of the temporally distant stimulus holds promise as a theoretical account of forgetting functions.

Forgetting and remembering pose interesting scientific challenges, in that they raise the question of action at a temporal distance. In the present paper, I argue that remembering can be treated as discrimination at the time of retrieval and that temporal distance is a dimension of the stimulus complex. Accounting for the effect of temporal distance is an issue of psychophysical scaling. The mathematical form of the forgetting function cannot be decided solely on the basis of goodness of fit, however, because several functions describe the data equally well. The property of temporal independence suggests that exponential functions should be favored.

The Evolutionary Context

Remembering and its complement, forgetting, involve the influence of earlier events on later behavior. In the first experimental demonstration of remembering in nonhuman animals, Thorndike (1898) observed that, after cats had a month or two of holiday from the experimental apparatus, their latency to escape from puzzle boxes clearly benefited from their earlier experience. His demonstration placed memory in an evolutionary context, especially in terms of the cross-species continuity of mental function that was proposed by Charles Darwin (1871/1899) in *The Descent of Man*. In an evolutionary context, remembering is adaptive, as illustrated by studies of species

differences in spatial memory. For example, Clark's nutcracker (*Nucifraga columbiana*) lives at high altitudes and stores over 30,000 pine seeds each summer for retrieval in the winter months (Kamil & Balda, 1990). In laboratory procedures for measuring the accuracy of spatial memory (over short delays), the performance of Clark's nutcracker is superior to low-altitude corvids and pigeons (Olson, 1991). In studies in which the memory performance of different species is similar, however, the similarities are more likely to reflect the structure of the task or the procedures (White, Juhasz, & Wilson, 1973), just as the differences may reflect the impact of procedural factors or task requirements (Macphail, 1996). In an evolutionary context, remembering is an adaptive reflection of environmental demands.

Forgetting too is adaptive. Kraemer and Golding (1997) have argued convincingly that forgetting should be seen not as involving process failure but instead as enhancing behavioral plasticity. For example, forgetting minimizes proactive interference from prior memories and facilitates the discrimination of recency. Normally, its adaptive utility stems from its primary characteristic—namely, diminution in accuracy with increasing temporal distance. The hallmark of experimental studies of remembering is thus manipulation of the temporal distance variable—that is, the retention interval.

Forgetting Functions

The first experimental studies of forgetting in humans date from well over a century ago. Using percent savings as a measure of retention, Ebbinghaus (1885/1964) reported a systematic reduction in amount retained over several weeks since original learning. Hundreds of studies have been completed since then, using a variety of performance measures. Woodworth and Schlosberg (1954) summarized some of the earlier studies, which showed that, with the exception of the savings measure, the general form of the forgetting function was similar for different

I thank my graduate students for questioning my ideas, and Angela Ruske and Barry Dingwall and his team for assistance and support in the conduct of the experiment reported here. I thank Peter Killeen for challenging some of the notions floated here during my leave at Arizona State University in 1998. I thank Camilla Gobbo and Angelo Bisazza for hosting my visit to the University of Padua, where some of the ideas were developed. I thank Glenn Brown, Peter Killeen, and John Wixted for their excellent comments on the manuscript. Correspondence should be addressed to K. G. White, Department of Psychology, University of Otago, Dunedin, New Zealand (e-mail: kgwhite@otago.ac.nz).

—This article was invited by the editors.

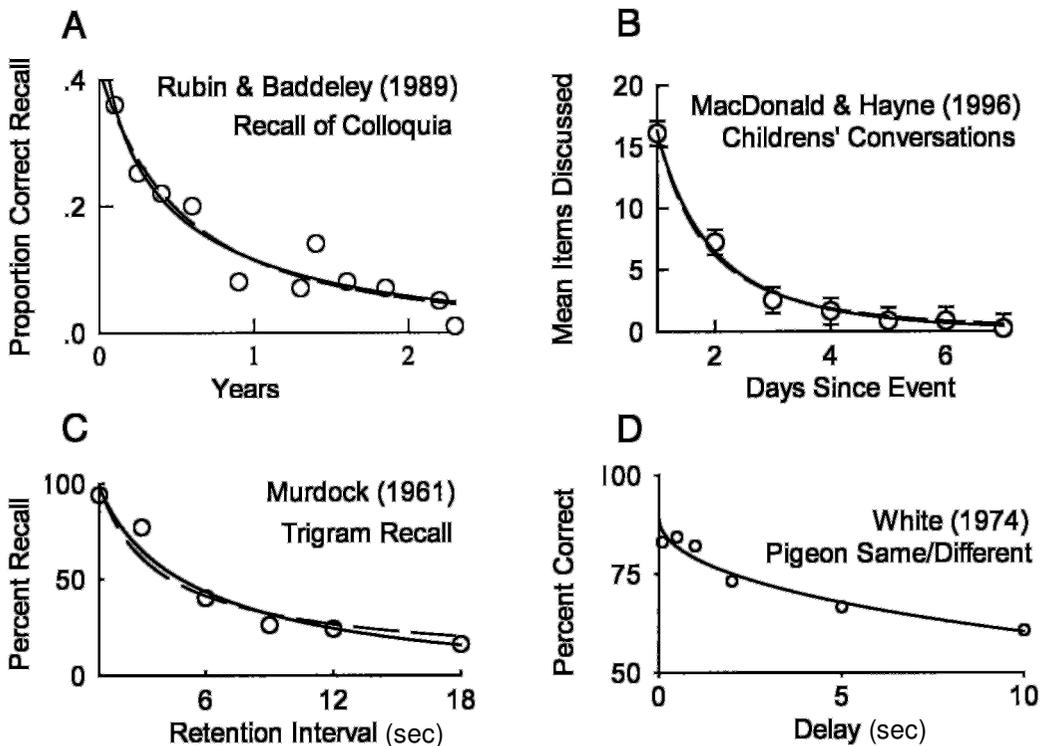


Figure 1. Examples of forgetting functions over long delays. Solid lines are best-fitting exponential to \sqrt{t} functions, and dashed lines are best-fitting power functions. (A) Recall by Cambridge University staff of departmental colloquia over the previous 2 years (Rubin & Baddeley, 1989). (B) Young children's conversations with parents about a pony ride arranged by the researchers over several days since the event (MacDonald & Hayne, 1996). (C) Correct recall of consonant trigrams over several seconds when rehearsal is prevented in Murdock's (1961) replication of Peterson and Peterson's classic (1959) study. (D) Accuracy of pigeons' reports of whether consecutively presented hues on the center key were the same or different (White, 1974) as a function of the delay between the first and second hue (also see Uruioli & DeMarse, 1997; White & McKenzie, 1982).

procedures, such as recall, recognition, and reproduction (cf. Luh, 1922). A wide range of retention intervals has been studied. Figure 1 shows examples of forgetting functions over years, days, and seconds.

Studies referred to below in the present paper have used the delayed matching-to-sample procedure to generate forgetting functions in nonhuman animals (Blough, 1959; Roberts & Grant, 1976; Wixted, 1989). These too provide instances of monotonic decrements in accuracy over the delay interval. Along with the many other studies reanalyzed by Rubin and Wenzel (1996), these studies suggest the possibility that the forgetting function has one particular form that is common to different ranges of retention-interval duration, different procedures, and different species. But is there a single best-fitting form of the forgetting function? And, if several forms fit, how do we decide between them? Should the form follow basic principles of the memory process? Answers to these questions are attempted below.

Is There a Best-Fitting Forgetting Function?

It is not the purpose of the present paper to compare the descriptive success of different functions, because

that exercise has been attempted by others (Laming & Scheiwiller, 1985; McCarthy & White, 1987; Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991). But it is useful to point to some of the difficulties associated with comparing functions on the basis of their goodness of fit.

A quantitative analysis of an extensive compilation of studies with at least five values of the retention interval was recently reported by Rubin and Wenzel (1996). Like Ebbinghaus (1885/1964), their emphasis was primarily empirical. They examined how well 210 data sets from a wide variety of prior studies were fitted by each of 105 two-parameter functions. Each function was fitted to the dependent measures reported in the original studies, and the original units of time for the retention interval were retained (ranging from seconds to years across different studies). They ranked the functions in terms of the number of times each function was one of the 10 best fits (in terms of r^2) to the 210 data sets. Functions that gave the best fits to the data included logarithmic, power, exponential, hyperbolic, and exponential and hyperbolic functions with time t scaled as \sqrt{t} .

The approach taken by Rubin and Wenzel (1996), to fit first and ask questions later, has merit in the context

of the search for order in forgetting functions. Rubin and Wenzel commented on a number of the procedural difficulties associated with such a task, such as the choice of dependent measures. Their approach has drawn criticism, however, such as the suggestion by Wickens (1998) that the fits may depend on the chosen measure. In practice, it does not seem to matter which dependent measure is used, with the exception of "percent savings" (Ebbinghaus, 1885/1964) and the "Ebbs" measure (Bahrick, 1965), where functions tend to be logarithmic. When percent savings or Ebbs measures are used, the function shows a large decrement to the first nonzero delay and then a slower rate of decrement (J. R. Anderson & Schooler, 1991). These measures, however, relativize performance so that the intercept is always 1.0 for percent savings or 0 for Ebbs. Functions that rely on variations in the intercept parameter are therefore constrained in their fits. Furthermore, the use of such measures is inconsistent with the conclusion that the intercept is a determinant value and one of the important properties of prototypical forgetting functions (see below).

The most commonly used measure, proportion correct, is limited by being prone to response bias and by being bounded at 1.0. By comparison, the ratio-based measures d' , $\log \alpha$, $\log d$, and the logit of proportion correct (logit p) do not have an upper bound and vary on an equal-interval scale. Additionally, the measures d' , $\log \alpha$, and $\log d$ estimate discriminability independently of response bias. Macmillan and Creelman (1991) provide a comprehensive account of the discriminability measure d' from signal detection theory and the discriminability measure $\log \alpha$ from Luce's (1963) choice theory. The discriminability measure $\log d$ from behavioral detection theory (Davison & Tustin, 1978; Nevin, 1981) is the same as $\log \alpha$ except that $\log d$ uses \log to the base 10. The transformation of percent correct, logit p (calculated as $\log(p/1-p)$) equals $\log d$ if there is no response bias. When the slopes of fitted functions are compared, it is essential to use equal-interval scale measures because differences in slope can result from the way a dependent variable can change over different ranges (Loftus, 1985; Wixted, 1990). More importantly, these equal-interval scale measures should be used in evaluating goodness of fit. Rubin and Wenzel (1996) did not have access to the necessary data for their analysis and so relied on whatever measures were described in the original studies.

In their Table 7, Rubin and Wenzel (1996) showed the results of fitting the set of best functions to data from studies using animals as subjects. This group of studies mainly used the delayed matching-to-sample procedure and included at least five retention intervals varying up to 40 sec. The mean r^2 values for hyperbolic, exponential, logarithmic, power, exponential \sqrt{t} , and hyperbolic \sqrt{t} functions were .85, .83, .87, .84, .88, and .88, respectively. The exponential was disadvantaged in this competition, unfortunately, because Rubin and Wenzel made the mistake of fitting the functions to percent correct measures on the assumption that they ranged from 0% to 100%, whereas the two-alternative forced-choice proce-

dures dictates a minimum of 50% (by chance). The exponential's eventual destiny is an asymptote at zero, an impossibility with chance at 50%. In contrast, all the other functions have a varying rate of decrement and are flexible enough to fit data points at long delays that are well above (or below) 50%.

A further problem is that procedural factors might contribute to the form of the forgetting function. For example, in Wickelgren's (1975) study of word recognition by sober or inebriated humans, there was a rapid drop in d' from about 2.8 to 1.2 over 2 through 50 sec in a continuous recognition memory task and a slower drop in d' from about 0.5 to 0.2 over 1 through 14 days. The procedure for Days 1–14 differed, however, from the procedure used for the first 2 h of the experiment in which delays ranged up to 50 sec.

In other examples, a combination of averaging and floor effects can produce hyperbolic-like functions. When pigeons were trained in delayed matching to sample at zero delays and tested at varying delays (Roberts, 1972) or trained with just one delay per condition (Harnett, McCarthy, & Davison, 1984), some birds responded at chance at all long delays. A power- or hyperbolic-like function then results from averaging monotonically decreasing functions with functions that have a high intercept but are at chance at all delays greater than zero. This averaging problem is related to the one explored by R. B. Anderson and Tweney (1997), who showed how power functions can result from the averaging of individual exponential functions. A thorough analysis of recall and recognition data by Wixted and Ebbesen (1997) showed, however, that such averaging artifacts are unlikely and that the form of the forgetting functions based on averaged data tends to reflect the form of functions for individual subjects.

As noted by both Rubin and Wenzel (1996) and Wickens (1998), certain functions can be eliminated from consideration owing to their inability to predict determinate values of performance at $t = 0$ or $t = \infty$. This is true of both logarithmic and power functions. The logarithmic function has often been used to fit a forgetting function to the data from the classic study by Ebbinghaus (1885/1964) and, as a result, tends to have been favored as a potential candidate for describing other data (Baddeley, 1997; Kintsch, 1970). In the study by Rubin and Wenzel, it was the most successful function, notwithstanding the difficulty of discriminating between functions in terms of the error variance in estimating their success. The power function was favored by J. R. Anderson and Schooler (1991) and by Wixted and Ebbesen (1991) but predicts infinitely accurate performance at $t = 0$. This property of the power function conflicts with the conclusion, outlined below, that performance at $t = 0$ may be high or low, depending on the attentional requirements of the task. The same difficulty applies to the logarithmic function, which is undefined at $t = 0$. A related problem with the power function is that, for a given value of a , $y = a$ at $t = 1$ for all values of b . That is, when b is varied the function pivots around the value $y = a$ at

$t = 1$, generating apparent intercept changes. The power function fitted by Wixted and Ebbesen (1991) assumed that $t = t + 1$, in order to avoid this problem. Laming and Scheiwiller (1985) compared several functions, including one based on a diffusion assumption, and concluded that, when a wide range of retention intervals is included, the functions tend to be logarithmic in form, although logarithmic functions are not mathematically correct because they tend to infinity as delays tend to zero, and, at finite values of delay, they can take negative values.

To summarize, a number of problems make it difficult to search for a best-fitting function. As noted by Rubin and Wenzel (1996), there tends to be a set of best-fitting functions, but it is not possible to decide between them on the basis of goodness-of-fit comparisons. In general, Rubin and Wenzel's conclusions about which might be the best-fitting functions seem convincing, but because the functions cannot easily be discriminated on empirical grounds, we are now at the point of asking theoretically oriented questions in order to justify the choice of function.

Theoretical Grounds for Forgetting Functions

The simple negative exponential $y = a \cdot \exp(-b \cdot t)$ is perhaps the most frequently used description of forgetting functions. In studies of animal working memory, fits of the negative exponential account for similar amounts of the variance in the data to many of the other best-fitting functions examined by Rubin and Wenzel (1996). Mathematically, the exponential is the only function that has a constant rate of decrement, and, intuitively, it lends itself to forgetting functions because it describes plausible time-related processes, such as decay in which the loss from a particular time is proportional to the level of performance at that time or interference in which there is a constant likelihood of encountering competing alternatives over time.

Simon (1966) argued that exponential forgetting was inconsistent with Jost's law, which says that the rate of forgetting slows down over time. There was little empirical evidence for Jost's law prior to 1954, however (Woodworth & Schlosberg, 1954, pp. 730–731). Some recent studies suggest that there might be less forgetting at longer retention intervals, and, indeed, a wide range of alternative mathematical functions have been applied to forgetting, such as the power exponential in which the rate of forgetting decreases as a power function of time (Wickelgren, 1972, 1974; Wickens, 1998). Squire (1989) addressed the issue of whether forgetting is continuous over very long times, as appears to be the case in studies of autobiographical memory (Rubin, 1982), or whether it quickly reaches a stable asymptote, as in memory for names and faces of classmates of up to 50 years ago (Bahrick, Bahrick, & Wittlinger, 1975). Squire's method was consistently applied to retention intervals ranging from 1 to 15 years. He concluded that the rate of forgetting was constant over such long times. Indeed, a simple exponential function fit his data well, but just as well as a power function—a function more consistent with Jost's law.

The functions fitted to the examples in Figure 1 illustrate the difficulty of deciding between functions solely on the basis of goodness of fit. The functions in Figure 1 are the version of the negative exponential with time scaled to \sqrt{t} (White & Harper, 1996; Wixted, 1990)—that is, $y = a \cdot \exp(-b \cdot \sqrt{t})$ —and the power function with an added constant, $y = a \cdot (t + 1)^{-b}$. The intercept, a , measures performance at time $t = 0$, in the absence of a memory requirement. The slope parameter, b , gives a measure of rate of forgetting. Figure 1 shows that the two functions provide similarly satisfactory, and practically identical, fits.

The power function was chosen for fitting the selection of data in Figure 1 because power functions are ubiquitous in psychophysics. If remembering is to be treated in the same terms as discrimination, power functions may be used to scale the effects of temporally distant stimuli (Fetterman, 1996). Indeed, power functions have been shown to provide very satisfactory fits to forgetting functions (J. R. Anderson & Schooler, 1991; Wixted & Ebbesen, 1991).

The exponential was chosen because it has the unique characteristic of the rate of decrement being constant. It might be argued that by raising time to the power of .5 in the exponential or by adding the constant 1 in the power function, a third parameter, c , is being added to each function. This third parameter, of course, affords an extra degree of freedom, but by setting it at a constant value for all fits, the fitted functions remain two-parameter functions. The three-parameter version of the exponential is the exponential-power function in which time is raised to a power c , interpreted as resistance to decay by Wickelgren (1972). That is, $y = a \cdot \exp(-b \cdot t^c)$. This latter function corresponds to the Weibull function championed by Wickens (1998). When the power is left to vary freely, it happens to take values that range between about .3 and .7 and average about .5 (Wickelgren, 1974). White and Harper (1996) chose this function to fit to a set of reanalyzed data because it scaled the time axis while leaving the properties of the exponential intact.

A possible justification for scaling the retention interval in terms of \sqrt{t} can be elaborated in terms of a diffusion metaphor in which the standard deviation, σ , of the discriminational processes (the discriminational dispersion in Thurstone's, 1927, terms) increases with increasing retention-interval duration (White & Wixted, 1999). The idea is not novel. Shepard (1958) suggested that generalization involved diffusion over time. A simple assumption that follows from the mathematics of diffusion (Berg, 1993) is that the standard deviation increases according to $\sigma = w \sqrt{t + 1}$. A diffusion process was assumed by Kinchla and Smyzer (1967), who concluded that decrements in performance over time were described by $y = a / \sqrt{1 + bt}$.

These ideas suggest that time is appropriately scaled to \sqrt{t} . They do not, however, provide a clear rationale for the mathematical form of the forgetting function. The exponential is chosen merely because of its constant-rate characteristic. Scaling the time dimension as \sqrt{t} allows the function to flatten out at longer delays in the same way

as the power function. If it can be demonstrated that rate of forgetting should be constant, as seems to be the case for long delays (Squire, 1989) and for short delays in animal studies (Rubin & Wenzel, 1996), Jost's law should be repealed. Both the simple exponential and the time-scaled exponential satisfy a constant rate-of-forgetting requirement.

Characteristics of Forgetting Functions

An entire forgetting function in which performance may vary considerably over a wide range of retention intervals can be summarized in terms of the parameters of a fitted function (White, 1985). By fitting negative exponential functions to a set of data from studies of animal memory, White (1985, 1991) concluded that forgetting functions have two characteristics that could be captured by the two parameters of the fitted function. The first characteristic is *initial discriminability*, or accuracy at $t = 0$, given by the value of the intercept. White (1985) interpreted the intercept as reflecting the effects of attentional demands or overall procedural requirements in the absence of a memorial requirement. In the animal studies, factors that render the short-term memory task more difficult result in a reduction in discriminability or accuracy at $t = 0$, without affecting the slope parameter of the fitted function (White, 1985). Such factors include decreasing the disparity between stimuli to be remembered, decreasing the number of repetitions of a stimulus on initial presentation, or decreasing the trial separation. At a zero retention interval, therefore, there is a determinate level of performance that varies systematically with factors that may affect attentional processes. That is, the y -intercept of the forgetting function represents one of the two characteristics of forgetting functions. It is not always infinity or undefined, as the two-parameter power or logarithmic functions suggest.

The second characteristic of forgetting functions is *rate of forgetting*, given by the slope parameter. When the negative exponential function was fitted to a set of data from animal working memory studies, rate of forgetting varied with factors related to retroactive interference and proactive interference, without influencing the value of the intercept parameter (Edhouse & White, 1988; Jones & White, 1994; White, 1985, 1991). White and Harper (1996) used the exponential to \sqrt{t} function to satisfactorily fit a large set of data pertaining to the effects of hippocampal lesions in monkeys. They demonstrated significant changes in the rate of forgetting parameter, without changes in the intercept. The original studies had concluded, however, that there was no change in the rate of forgetting (but without curve fitting). The earlier conclusions were based on the statistical evaluation of the interaction between lesion condition and retention interval. A different conclusion follows from rate of forgetting estimated directly as a parameter of a fitted function.

Bogartz (1990), Loftus (1985), and Loftus and Bamber (1990) have debated the issue of how to assess changes in forgetting rate independently of the level of original

learning. Wixted (1990) concluded that the best approach was to fit a mathematical function to the data relating performance to retention-interval duration and to use the slope parameter as a measure of rate of forgetting. This is consistent with the present approach. Slamecka and McElree (1983) had concluded that the rate of forgetting of lists of words was independent of the original level of learning. Their conclusion is easily confirmed by showing independent changes in the intercept and slope parameters of exponential functions fitted to their data (Loftus, 1985; Wixted, 1990). Similarly, Haist, Shimamura, and Squire (1992) showed that recognition and

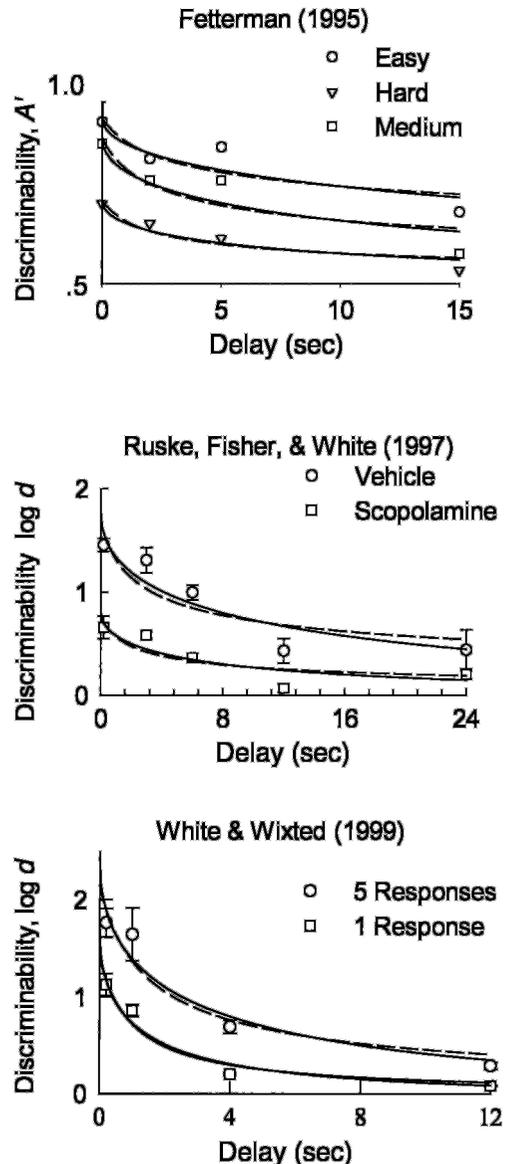


Figure 2. Examples of forgetting functions in which the level of initial discriminability (intercept of the fitted function) varies as a function of experimental condition, but with no change in the rate of forgetting. Solid lines are best-fitting exponential to \sqrt{t} functions, and dashed lines are best-fitting power functions.

recall accuracy as a function of delay was overall lower for amnesics than for matched controls, but they did not differ in rate of forgetting.

Figure 2 shows several examples in which the rate of forgetting parameter remains constant with changes in the intercept. The data in the top panel are from a study by Fetterman (1995) in which pigeons discriminated a set of three short durations from three long durations in a delayed matching-to-sample procedure. A nonparametric measure of discriminability (A') was overall high for the discrimination between the shortest and longest duration samples and was overall low for the discrimination between the middle-, short-, and long-duration samples. The forgetting functions were generally parallel and differed in intercept values but not in rate of forgetting. The middle panel of Figure 2 shows the results of administering scopolamine to pigeons working in a delayed matching-to-sample procedure (Ruske, Fisher, & White, 1997). The result was a systematic reduction in the intercept parameter of fitted exponential functions without a change in the rate of forgetting (also see Kirk, White, & McNaughton, 1988). Other drugs, such as chlorpromazine (Watson & Blampied, 1989) and phenobarbital (White, Harper, & Watson, 1994), increase rate of forgetting and reduce initial discriminability. There seem to be no drugs that affect rate of forgetting without also influencing initial discriminability (White, Ruske, & Colombo, 1996).

In the bottom panel of Figure 2, a sample ratio requirement of five responses generated higher initial discriminability than a requirement of one response, without changing the rate of forgetting (White & Wixted, 1999). The attentional component of forgetting functions, manifest in initial discriminability (the y -intercept of the function), is therefore adversely influenced by such factors as more difficult discriminations, scopolamine administration, and smaller sample ratio requirements.

Protection from such adverse effects results from administering glucose or procholinergic drugs, such as AF150 (Parkes & White, 2000; Ruske et al., 1997; Ruske & White, 1999). In a delayed matching-to-sample task with pigeons, Parkes and White (2000) showed the same effect of scopolamine on initial discriminability as did Kirk et al. (1988) in rats. They also showed the usual reduction in accuracy with a one-response sample ratio requirement relative to a five-response sample ratio requirement. Parkes and White showed that glucose administration countered the effects of both scopolamine administration and reducing the sample ratio requirement. The attenuation was manifest as an increase in the intercept (initial discriminability) of the fitted function without changing rate of forgetting. Their conclusion that glucose affected the attentional component of performance was consistent with the absence of an effect of glucose on the rate of forgetting when administered alone.

The clearest evidence for changes in rate of forgetting independently of the level of initial discriminability comes from studies of retroactive interference and proactive interference (Edhouse & White, 1988; Harper & White,

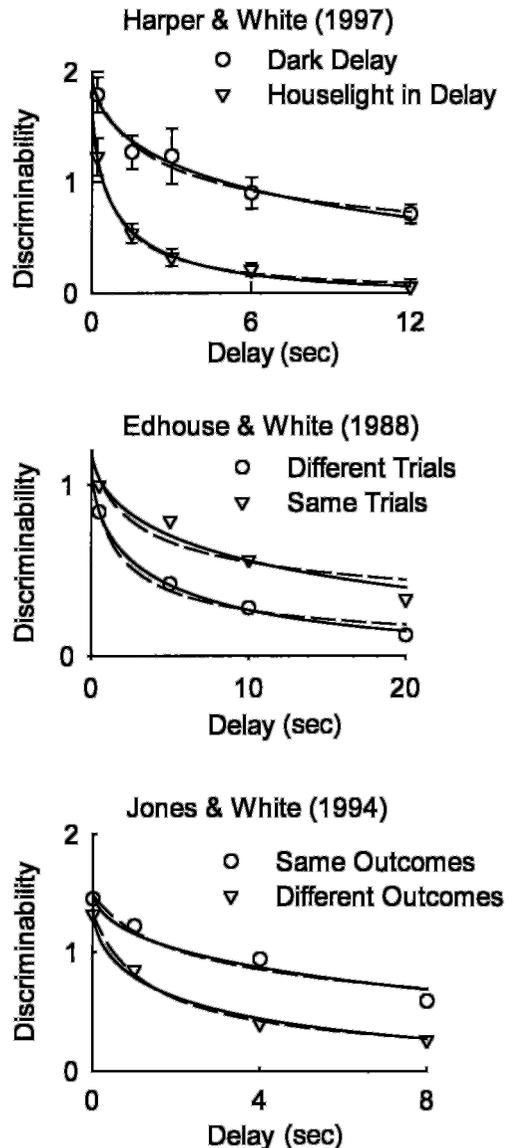


Figure 3. Examples of forgetting functions in which rate of forgetting varies with retroactive interference (top panel), proactive interference (middle panel), or differential outcomes (bottom panel), but without a significant change in initial discriminability. Solid lines are best-fitting exponential to \sqrt{t} functions, and dashed lines are best-fitting power functions.

1997). Figure 3 illustrates three cases. In the top panel, discriminability in a delayed matching-to-sample procedure decreased systematically with increasing duration of the retention interval and decreased at a much faster rate when houselight illumination was interpolated in the retention interval (Harper & White, 1997; Jans & Catania, 1980; White, 1985). In the middle panel of Figure 3, rate of forgetting on trials in which the sample differed from the sample on the preceding trial was about twice as fast as when the samples on consecutive trials were the same (Edhouse & White, 1988). This result illustrates the effect of intertrial agreement, or "local proactive in-

terference," which is independent of the trial-spacing effect (which influences initial discriminability). In the bottom panel, rate of forgetting on trials with signaled differential outcomes was less than that on trials with *same* outcomes (Jones & White, 1994).

The two functions fitted to the data in Figures 2 and 3 are the exponential to \sqrt{t} and the power functions, as in Figure 1. It is of interest to note, however, that the simple exponential provided the best fits to data in Figure 2 where intercepts differed across conditions, but the exponential to \sqrt{t} provided the best fits to the data in Figure 3 where rate of forgetting differed. There is no obvious reason for this difference, but it further illustrates the difficulty of deciding between the simple exponential and the exponential to \sqrt{t} functions.

Encoding Versus Retrieval

The advantages of describing forgetting functions in terms of the two parameters of fitted functions are both quantitative and theoretical. The two parameters serve as higher order measures of forgetting, and effects on initial discriminability, can be separated from effects on rate of forgetting. For example, some drugs influence initial discriminability whereas others also influence rate of forgetting (Parkes & White, 2000; White et al., 1996). In the context of delayed matching-to-sample procedures, the distinction is sometimes referred to as *delay-independent* versus *delay-dependent* effects.

The two parameters can also be identified in terms of the main processes of remembering: encoding and retrieval (Tulving, 1983). Factors such as stimulus duration, repetition, or salience influence encoding or attention to the sample in working memory procedures. Initial discriminability is also influenced by overall task requirements and by reinforcement factors (White & Wixted, 1999). Factors that influence retrieval, such as retroactive interference, proactive interference, and retrieval cue availability, have an effect on rate of forgetting. Thus, it seems reasonable to assume that initial discriminability may reflect the attentional or encoding component of performance in working memory tasks, whereas rate of forgetting is associated with the memorial component and is sensitive to conditions for retrieval (White & Harper, 1996).

A commonly held view is that encoding and retrieval are interdependent (Baddeley, 1997; Roediger & Gynn, 1996). This intuitively plausible view is strengthened by theoretical positions that rely on the interdependence of encoding and retrieval. Indeed, encoding can be expected to place constraints on successful retrieval. In Tulving and Thomson's (1973) principle of encoding specificity, retrieval depends on reinstating the operations that supported the successful encoding of the material to be remembered. According to global activation models, items are retrieved more easily when they are strongly associated with contextual cues present at the time of encoding (Clark & Gronlund, 1996). According to the levels-of-processing or transfer-appropriate processing approaches, retrieval is better with greater correspondence between task requirements at encoding and those at the time of

retrieval (Craig & Tulving, 1975; Morris, Bransford, & Franks, 1977).

Interdependence of encoding and retrieval is contradicted, however, by the finding that performance of a dual task at the time of encoding has a much greater effect on the amount recalled or recognized by human participants than when the dual task accompanies retrieval (Craig, Govoni, Naveh-Benjamin, & Anderson, 1996; Naveh-Benjamin, Craig, Gavrilescu, & Anderson, 2000). From this and other evidence, Craig et al. conclude that encoding and retrieval involve different processes. In terms of the present distinction between initial discriminability and rate of forgetting, the possibility of differences between encoding and retrieval suggests that the success of remembering at one particular time may be independent of the initial level of performance. In studies of delayed matching-to-sample in pigeons, rate of forgetting can vary independently of initial discriminability (White, 1985, 1991). Similarly, on the basis of the slopes of fitted functions, recall by human participants is independent of the original degree of learning (Slamecka & McElree, 1983). In other words, the level of performance or retrieval at a particular retention interval does not depend on the level of encoding. This conclusion may seem counterintuitive in that accuracy at a later retention interval should not be higher than that at zero delay. Evidence described below, however, indicates that, under some conditions, forgetting functions can be nonmonotonic. That is, accuracy at a later retention interval can be higher than that at zero delay (Sargisson & White, 2001).

Evidence from studies of working memory in pigeons suggests that performance at a given retention interval is specific to the conditions prevailing at the time of retrieval. White and Cooney (1996) trained pigeons in a delayed matching-to-sample procedure in which two delays (0.1 and 4 sec) were randomly (and unpredictably) mixed within sessions. Choices of red and green comparison stimuli at the short delay were reinforced with different probabilities, creating a strong bias to choose the comparison associated with the higher reinforcer probability. The bias generated at the short delay did not generalize to choices at the longer delay, which were nondifferentially reinforced. Similarly, in other conditions, differential bias at the long delay caused by varying the reinforcer differential at that delay did not generalize to the nondifferentially reinforced choice at the short delay. White and Cooney concluded that performance at one delay was independent of factors affecting remembering at another.

An analogous result was reported by Clayton and Dickinson (1999) for memory for food locations in scrub jays. The birds were trained to cache worms and nuts in different locations and to recover them in a later test. When tested 4 h later, the birds preferred worms to nuts. When tested 124 h later (or 100 h later in the replication by Clayton, Yu, & Dickinson, 2001), after the experimenter had replaced the worms with perished worms, the birds preferred nuts to worms. In other words, preference changed across the course of the retention interval in the

same way as in the study by White and Cooney (1996). Remembering is thus determined by the time of retrieval and is specific to the delay interval at which it is tested.

Remembering as Discrimination

The possibility that the success of retrieval at one time is independent of the level of performance at earlier times or even independent of the level of initial encoding is consistent with the view that remembering is specific to the time at which it is required. The discrimination made at one time may be independent (notwithstanding generalization) from the discrimination made at another, just as two discriminations about the spatial aspects of stimuli may be independent. Fetterman (1996) has summarized the advantages of treating remembering in the same terms as perceptual discriminations between proximal stimuli. In particular, this approach suggests that remembering is subject to psychophysical analysis and follows the same psychophysical principles as does other discriminative behavior (Fetterman, 1995; White, 1985, 1991; Wixted & Ebbesen, 1991). The approach does not necessarily rely on intervening or mediating variables supposed to bridge the temporal gap between stimulus and action (Watkins, 1990). Temporal distance is a dimension of the stimulus complex that influences behavior along with other physical dimensions of the event to be remembered and aspects of the stimulus context (cf. Fetterman, 1996).

The result reported by White and Cooney (1996) is inconsistent with approaches to nonhuman memory, such as trace decay or stored representations. These approaches rely on mediating processes that operate during the retention interval (Roitblat, 1982). For example, in White and Cooney's procedure, a prospective code set up at the time of the sample and involving choice of red at the end of the short delay would be inappropriate at the long delay. Yet the behavior showed temporary plasticity over the course of the delay. White and Cooney's result is reminiscent of the finding reported by Nevin (1970), who mixed two luminance stimuli in a signal detection task for pigeons. Nevin varied the reinforcer probability for correct reports in one luminance and held it constant in the other (and vice versa). Report probability in one luminance was independent of whether reinforcer probability varied in the other. Hence, there is a parallel between Nevin's result for proximal stimuli and White and Cooney's result for temporally distant events. Differences based on temporal dimensions can act in the same way as differences based on physical dimensions.

The perspective analogy elaborated by Staddon (1983) depends on an analogy of temporal distance to spatial distance. So too does the correspondence metaphor (Koriat & Goldsmith, 1996; White, 1996) and the notion that remembering follows Gibson's (1979) principles of direct perception (White, 1991). Thus, discriminations of temporally distant events may be difficult in the same way that discriminations of spatially distant events are. Temporal distance is therefore a property of the event.

TEMPORAL INDEPENDENCE: AN EXPERIMENT

Because forgetting functions typically decrease gradually as the retention interval lengthens, the inevitable conclusion is that the process of forgetting depends on time. If forgetting at a particular retention interval reflects a discrimination specific to a given delay, however, it should be possible to show that the level of performance at one delay is independent of whether performance at a shorter delay is higher or lower. Such temporal independence was demonstrated in the present experiment. In a delayed matching-to-sample procedure, reinforcers for correct choices were omitted at different delays in different conditions. The result of omitting reinforcers was that discriminability dropped at the delay where reinforcers were omitted. The question was whether discriminability would remain low at longer delays in the same retention interval following the reduction earlier in the interval or whether performance at later times is independent of performance earlier in the interval.

Method

Five adult homing pigeons, with several months of training in the baseline condition of the present procedure, were maintained throughout the experiment within 10 g either side of 85% of their free-feeding body weights, by supplementary feeding following experimental sessions. Water and grit were freely available in their living cages. Experimental sessions lasting 50 min were conducted for 7 days per week in chambers supplied by Med Associates, Inc. Each chamber was 29 cm high, 29 cm wide, and 24 cm deep. An opening in the center of one wall 4 cm above the grid floor allowed access to a hopper of grain, which could be raised for 2.5 sec. Three response keys above the hopper opening, 2.1 cm in diameter, 6 cm apart, and 21 cm above the floor, could be lit red or green. Experimental events were controlled and recorded by a Med-PC system on a personal computer.

Each daily session lasted for 80 trials. The trials were in a random order that allowed for equal frequencies of the different possible combinations of delay interval, sample stimuli, and left-right position of comparison stimuli. Trials were separated by 15-sec intervals, during which responses were ineffective and the chamber was dark. Each trial began with the presentation of red or green on the center response key. The fifth peck at this sample stimulus initiated a dark delay interval that lasted for 0.1, 2, 4, or 16 sec. The delays occurred equally often in a random order within each session. At the end of the delay, red and green comparison stimuli were presented on the left and right response keys (with left-right position changing randomly across trials). A single peck at the comparison stimulus that matched the red or green sample was reinforced with 2.5-sec access to grain. Incorrect responses produced a 2.5-sec blackout, followed by the intertrial interval.

In the baseline condition, every correct matching response at each of the four delays was reinforced. Including this condition, a total of six conditions was arranged. The third and fifth conditions were repetitions of the baseline condition. In Condition 2, the procedure was the same as for the baseline conditions with the single exception that correct choices at the 2-sec delay only were not reinforced but instead were followed by the 2.5-sec blackout and the intertrial interval. In Condition 4, correct choices at the 4-sec delay were not reinforced. In Condition 6, correct choices at the 0.1-sec delay were not reinforced. Each condition was conducted for a large number of sessions, in order to ensure stability of performance. The numbers of sessions completed in each condition were similar but not the

same for all birds, because a bird was run only if its weight was within the prescribed range. The average numbers of sessions completed in each of the three baseline conditions were 32, 30, and 55. The numbers of sessions completed in Conditions 2, 4, and 6 averaged 59, 38, and 55. The last 8 sessions of each condition were used for data analysis, so that there was a total of 80 trials for each sample at each of the four delays.

Accuracy of performance was assessed in terms of a measure of discriminability, $\log d$, which is given by the geometric mean of the ratios of correct (c) to error (e) responses following each sample. That is,

$$\log d = 0.5 \log [(c_r/e_r)/(c_g/e_g)].$$

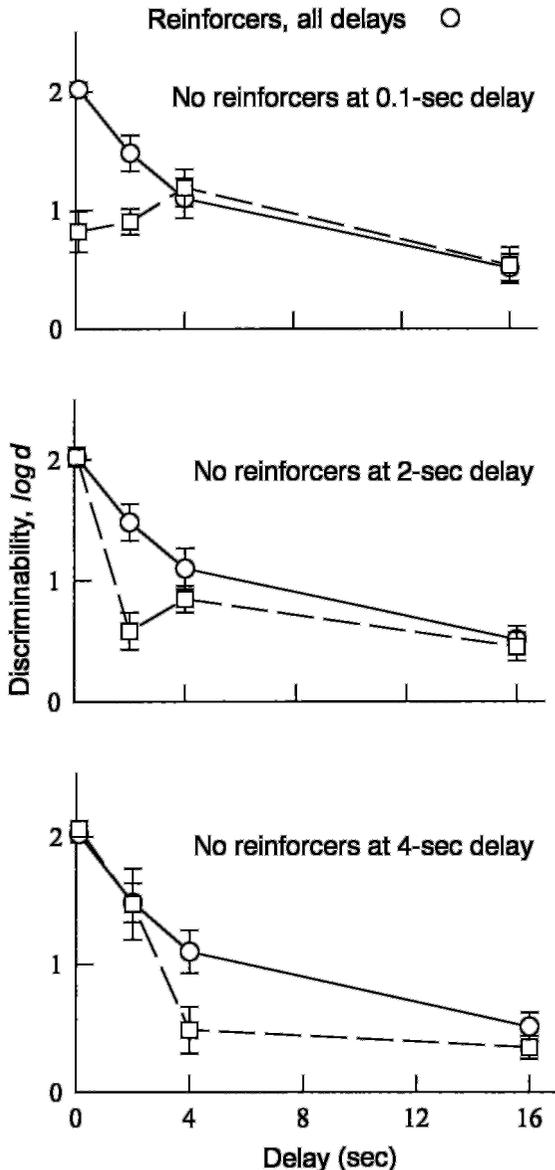


Figure 4. Mean discriminability (and standard errors) as a function of delay for baseline conditions of delayed matching-to-sample by pigeons (circles) and for conditions in which reinforcers were omitted for correct choices at one of the delays (squares).

This measure has similar properties to d' of signal detection theory (Macmillan & Creelman, 1991) and has been used in studies of animal memory (McCarthy & White, 1987; White, 1985). It has two advantages: (1) Unlike percent correct, it is not bounded at 1.0 and so is not susceptible to ceiling effects, and (2) it measures discriminability separately from response bias (Macmillan & Creelman, 1991). In order to take account of the few instances of zero frequencies where performance was perfect at the shortest delay, 0.5 was added to each cell in the signal-detection matrix of response totals, for all delays, as recommended by Hautus (1995).

Results

Figure 4 shows the mean $\log d$ measures of discriminability in the baseline conditions plotted as a function of delay-interval duration. Measures for the three baseline conditions were averaged, and the same function is plotted in each of the three panels of Figure 4 in order to provide a comparison for the effects of omitting reinforcers for correct responses at one of the delays. Because the important comparisons were between an individual reinforcer-omission condition and baseline performance, the tactic of generating a stable baseline meant that order effects did not contribute to the comparisons. The very small standard errors in Figure 4 indicate that the 5 birds were consistent in their performance, which remained stable across the 9 months of the experiment. The function for the baseline conditions shows the systematic negatively accelerated decrement in discriminability with increasing delay typical of forgetting functions generated by the delayed matching-to-sample procedure (White, Ruske, & Colombo, 1996).

Figure 4 shows the results of omitting reinforcers for correct responses at the 0.1-, 2-, and 4-sec delays. Again, the standard errors are small and indicate satisfactory between-bird consistency. Significant effects of delay, condition, and the delay \times condition interaction in repeated-measures analyses of variance confirmed the effects that are obvious in Figure 4. The F ratios for the delay \times condition interactions shown in Figure 4 for the 0.1-, 2-, and 4-sec delays were $F_s(3,12) = 18.79, 14.11, \text{ and } 5.69$, respectively (all p s $< .01$). In all three conditions in which reinforcers were omitted at a particular delay, discriminability at that delay was substantially lower than at the corresponding delay for the baseline condition. When there were no reinforcers for correct choices at the 2- and 4-sec delays, discriminability levels at the earlier delays were virtually identical to those for the baseline conditions.

The result of primary interest was that discriminability at retention intervals later than the delay at which reinforcers were omitted returned to the same level as for the baseline condition. This result is most marked for the conditions in which reinforcers were omitted at 0.1- and 2-sec delays. In both cases, performance increased, so that, at the 4-sec delay, it was at the same level as for the 4-sec delay in the baseline condition.

The result shown in Figure 4 demonstrates temporal independence in the specific respect that, at the 4- and 16-sec delays in the top two panels and at the 16-sec delay in the bottom panel, discriminability in the conditions with reinforcers omitted for correct responses at earlier delays was the same as at the same delays for the base-

line conditions. That is, discriminability at the later delays was unaffected by whether performance was reduced at the earlier delays by omitting reinforcers. The reason for the reduction in discriminability, or discrimination failure, at the delays where reinforcers were omitted is immaterial to temporal independence. Most simply, the absence of reinforcers at those delays resulted in near-random choices. The important result is the higher discriminability at longer delays. Because the delays at which remembering was tested were mixed randomly within sessions, a delay had an equal probability of terminating at, for example, 2 or 4 sec. Whether the discrimination was successful at 2 sec or had failed at 2 sec, owing to reinforcer omission had no effect on the discrimination at 4 sec.

Implications for Theory

In the present experiment, discriminability was specific to a specific delay. It seems that the task of remembering, at least for pigeons in a delayed-matching procedure, involved quite separate discriminations at the different retention intervals. Typically, the discriminations become more difficult as the delay progresses; however, the present result suggests that this may not always be the case and that the discrimination at one time is independent of that at another, notwithstanding the possible constraint placed on performance by the level of initial discriminability or by generalization across delays.

What are the implications of temporal independence for theories of forgetting? The notion that performance is directly determined by the strength of a memory trace or the relative strengths of competing traces has been popular in studies of animal memory (Grant, 1981; Roberts, 1972) and has many variants in the human literature. Temporal independence is inconsistent with the notion of trace decay, however, and joins other evidence that less directly questions notions based on trace decay (Edhouse & White, 1988). If low discriminability indicates that a trace has lost strength by a certain time in the retention interval, it is not plausible that it should gain strength later in the interval. Maintenance-rehearsal theory (Grant, 1981), drift theory (Roitblat, 1983), and the temporal discrimination hypothesis (D'Amato, 1973) also rely on temporally related processes and have difficulty in accounting for the temporal independence result (White & Cooney, 1996).

Temporal independence is more consistent with notions related to retrieval or discrimination difficulty (Tulving, 1983). Tulving and Pearlstone (1966) showed that cued recall resulted in higher accuracy than did free recall, thus demonstrating, in their terms, that words may have been available in memory but inaccessible. It might therefore be argued that information about the prior sample in the present study may have been available at any one time, but whether it was accessible depended on the conditions for discrimination at the time of retrieval. The distinction between availability and accessibility has been questioned by Watkins (1990), however, because there is no procedure to ascertain unavailability. That is, it is not possible to distinguish unavailability from inaccessibility.

Forgetting Functions as Generalization Gradients

In treating remembering as a process of discrimination, temporal distance is construed as a dimension of the sample or a feature of the stimulus complex (Fetterman, 1996; White, 1991). Supporting evidence for this view is provided by several studies. Fetterman (1995) arranged a delayed matching-to-sample procedure in which samples were short and long durations. Following 0-sec delay training, generalization tests included sample durations that ranged over several values between the short and long training durations and also several delays varying up to 18 sec. At each delay, the generalization gradients along the duration dimension were ogival. The influence of delay was shown by a general flattening of the gradients as the delay lengthened. Thus, remembering was governed by both sample duration and the temporal delay from the sample to choice. The greatest discrimination occurred for the original stimulus values used in training—that is, the short and long durations and the 0-sec delay.

In a delayed matching-to-sample task with durations as samples, Spetch (1987) trained 6 pigeons first with a 0-sec delay, then with a 10-sec delay, and, finally, with a 20-sec delay. Accuracy was tested at a range of delays following each retraining. Accuracy decreased with increasing delays following training with the 0-sec delay but remained constant across delays following training with the 10- and 20-sec delays. When Spetch and Rusak (1992) trained pigeons with long delays and then tested at a range of delays, the resulting forgetting function resembled a generalization gradient, whereby accuracy peaked at the most familiar delay interval.

Recently, Sargisson and White (2001) trained pigeons in delayed matching-to-sample tasks with delays of 0, 2, 4, or 6 sec from the outset of training. Following initial training in which responses on any of three white keys were immediately reinforced, the delayed-matching procedure was introduced with red and green sample stimuli and a fixed delay between sample and comparison presentations at the outset. For example, birds trained with the 6-sec delay never experienced shorter delays between sample and comparison during training. Forgetting functions were obtained following training with fixed delays in a single session in which half the trials included the training delay and were reinforced and the remaining trials were not reinforced and included probe delays varying from 0 to 10 sec.

Figure 5 shows that the longer the training delay, the flatter the function, the higher the overall accuracy, and the greater the tendency for the maximum accuracy to occur at the value of the training delay. The tendency for the maximum accuracy of the forgetting function to occur at the value of the training delay is reminiscent of the result of testing for generalization along the physical stimulus dimension of the training stimulus (Honig & Urciuoli, 1981). Hence, the decrement in accuracy with increasing delay in standard forgetting functions is at least in part the result of training with 0-sec delay. That is, the forgetting function is a generalization gradient along the

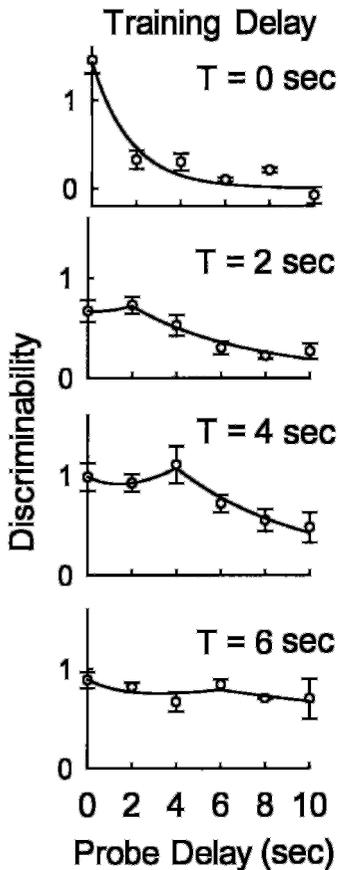


Figure 5. Forgetting functions obtained from probe test sessions in which delay varied, following training with a single delay. Continuous functions are predictions from a composite of a temporal distance component and a generalization component. From "Generalization of Delayed Matching-to-Sample Performance Following Training at Different Delays," by R. J. Sargisson and K. G. White, 2001, *Journal of the Experimental Analysis of Behavior*, 75, p. 12. Copyright 2001 by the Society for Experimental Analysis of Behavior. Adapted with permission.

temporal dimension of delay duration. In terms of temporal independence and the encoding/retrieval distinction, the important aspect of the result reported by Sargisson and White (2001) is that the forgetting function included levels of discriminability, typically at the training delay, that were higher than discriminability at 0-sec delay, and in the absence of any prior training with a 0-sec delay. That is, there was no evidence from the forgetting function or from prior training that performance at 0-sec delay reflected a higher level of accuracy at encoding than was seen at subsequent retrieval. In other words, a tentative but counterintuitive interpretation of the Sargisson and White result is that performance at retrieval can be superior to performance at encoding.

Constant-Rate Forgetting and Direct Remembering

The notion that remembering is delay-specific discriminative behavior was taken a step further by White

(1991), who proposed that remembering is direct, following Gibson's (1979) claim that perception is direct. In a theory of direct remembering, the remembered event is not assumed to be stored as a cognitive representation. Instead, it is likely that, at encoding, the individual system is set for later action through a neural mechanism, such as long-term potentiation, analogous to the way that reinforcers potentiate behavior. That is, the individual system is tuned to resonate to information available at the time of retrieval through the individual's prior learning experience and its evolutionary background. The information includes the retrieval context and signals provided by the context to initiate remembering, as well as the temporally distant event. Accordingly, because the conditions for retrieval are effective only at the time of remembering, remembering is specific to the time at which it occurs. Tulving (1997, p. 841) has recognized this possibility:

if the engram is a kind of entity that manifests itself only in activity, or retrieval, then we might conjecture that the physical changes resulting from experience do not exist as an engram in the absence of that activity. And we can also imagine that the engram, qua engram, is not detectable in its quiescent state, that is, in the absence of retrieval, with any physical technique.

White (1991) argued that the perception-memory continuity assumed by a theory of direct remembering requires temporal independence (also see Fetterman, 1996). Forgetting over a certain interval depends on the duration of that interval, but it is arbitrary as to where that interval begins. That is, the performance decrement from a certain time is independent of the level of performance at earlier times, as demonstrated above. The starting point of a function is arbitrary because, in principle, it is not possible to identify a point at which perception stops and memory begins. This was recognized by Gibson (1979, p. 253) in noting that "the stream of experience . . . is not a 'travelling razor's edge' dividing the past from the future." Identification of a zero point on the forgetting function is a convention related to our laboratory procedures, whereas everyday forgetting is part of the continuous stream of behavior. For monotonically decreasing functions, temporal independence is possible when the rate of forgetting is constant, but not otherwise. Mathematically, only the exponential has the constant-rate property that allows this.

The unique constant-rate property of the exponential function means that the decrement from a given time t_1 , to a second time t_2 , depends only on the distance, $t_1 - t_2$. That is, the proportional decrement from y_1 to y_2 is $y_2 / y_1 = \exp[b(t_2 - t_1)]$, where b is the slope constant. When the retention interval is scaled as the square root of time, and the exponential is even more successful in fitting a wide range of data (Rubin & Wenzel, 1996; White & Harper, 1996), the same conclusion applies. That is, $y_2 / y_1 = \exp[b(\sqrt{t_2} - \sqrt{t_1})]$. The important point is that the decrement depends not on the value t_1 of the retention interval from which the decrement occurs but

on the distance ($t_1 - t_2$), or on the distance ($\sqrt{t_1} - \sqrt{t_2}$). Similarly, the decrement from t_1 to t_2 does not depend on the extent of decrement from an earlier time to t_1 . With other functions, the absolute values of the retention interval determine the extent of decrement. As an obvious example, the rate of decrement for the hyperbola is rapid for short times soon after $t_1 = 0$ and slow at long times. Temporal independence, as demonstrated here, suggests that the proportional decrement from 4 to 16 sec, for example, depends on the interval of time from 4 to 16 sec, but not on the decrement that has occurred at times before 4 sec. Temporal independence is thus consistent with the constant-rate-of-forgetting property of the exponential.

As noted above, it is difficult practically to discriminate between logarithmic, power, exponential in \sqrt{t} , and hyperbolic in \sqrt{t} functions in terms of the variance they account for in the data (Rubin & Wenzel, 1996). That is, it is possible for the data to be equally well fit by several functions and, at the same time, for temporal independence to be established experimentally. The demonstration of temporal independence provides a way of deciding between the functions. Because temporal independence implies constant-rate forgetting, it provides experimentally based support for the exponential in t or \sqrt{t} as the preferred description for forgetting functions. These are the only functions associated with a constant rate of forgetting.

Consequences of Remembering

In an evolutionary approach to human memory, J. R. Anderson and Milson (1989) and J. R. Anderson and Schooler (1991) have argued that we are more likely to remember an item the more frequently it occurs. That is, remembering is adapted to relationships in the environment and the likelihood that it is needed. For example, R. B. Anderson, Tweney, Rivardo, and Duncan (1997) showed that the rate of decrement in recall of digits was a function of the probability that recall would be required at a given retention interval.

In the same way that signal presentation probability in signal detection procedures is linked to payoff probability (Alsop, 1998), need probability of remembered events is linked to payoff. Accordingly, remembering is sensitive to its differential consequences or payoffs (White & Cooney, 1996). Furthermore, the effects of differential payoffs in biasing delayed-matching choice between response alternatives interact with temporal distance of the sample in such a way that stronger biasing effects occur at longer delays (Jones & White, 1992).

White and Wixted (1999) proposed a model that predicts the interaction. On a trial-by-trial basis in the model, the individual choice between response alternatives is governed by the matching law. The choice is conditional on the value of the effective stimulus at the time the choice is made. The value of the effective stimulus varies along a "psychological dimension" and is associated with a prior history of payoffs or reinforcers for choices between response alternatives given the effective-stimulus value. To be consistent with direct remembering, the psychological dimension is a transformation of the physical

stimuli. As Luce (1994) has noted, however, the dimension may be interpreted in a variety of ways without affecting the main assumptions about the discriminational processes. The model predicts the relation between effective-stimulus value and reinforcer ratio by assuming that the sample stimuli have an effect that varies from trial to trial according to Thurstone's (1927) discriminational processes. It also assumes that the variance or discriminational dispersion of the processes increases with increasing retention-interval duration, perhaps according to a time-related process of diffusion (Shepard, 1958, 1988). White and Wixted did not, however, make any assumptions about the relation between standard deviation of the discriminational processes and time. If standard deviation, σ , is assumed to increase according to $\sigma = w\sqrt{t} + 1$, as predicted by a diffusion process (Berg, 1993), the forgetting function predicted by White and Wixted's model is very close to a negative exponential to \sqrt{t} , which fits the available data very well.

Exponential Forgetting

The treatment of remembering in the same terms as discrimination suggests that remembering is amenable to psychophysical analysis, in that the temporal distance of events can be scaled in terms of a mathematical function (Fetterman, 1996; White, 1985, 1991; Wixted & Ebbesen, 1991). If the functions have defined intercept and slope parameters, they can encompass the two characteristics of forgetting functions—initial discriminability and rate of forgetting. These characteristics are differentially affected by a range of experimental variables. The functions can be any that fit the data well (Rubin & Wenzel, 1996). However, the exponential is the only function that is consistent with temporal independence and constant-rate forgetting that follows from the assumption made by direct remembering that perception and memory are continuous.

The simple negative exponential is a very constrained function and does not cope well with data that appear to follow Jost's law. Although such data could result from procedural or measurement artifacts, as suggested above, they indicate the need for a search for alternative functions, such as the exponential to \sqrt{t} , which is one of the leading contenders in Rubin and Wenzel's (1996) hit parade of best-fitting functions. The treatment of remembering in the same terms as discrimination suggests yet another alternative. Shepard (1958, 1987) showed that the reduction in discrimination, d , with increasing psychological distance, D , between the stimulus values of a wide range of perceptual continua follows an exponential function. This *exponential law of generalization* is $d = \exp(-kD)$, where k is the rate constant, and is based on the assumption that "the psychophysical function that maps physical parameter space into a species' psychological space has been shaped over evolutionary history" (Shepard, 1987, p. 1319). If the retention interval is treated as a dimension of the sample or an aspect of the stimulus context that occasions later remembering, Shepard's exponential law of generalization describes how performance generalizes from the value of original training or

encoding (typically a 0-sec delay) to other nonzero delays (with temporal distance substituted for psychological distance). Generalization is therefore a component of the forgetting function. The other component follows from the direct-remembering assumption that the effect of the sample diminishes at a constant (exponential) rate with increasing temporal distance of the sample from the time of remembering. The empirical forgetting function expresses discriminability, $\log d$ or d' , as the average of the two components. That is,

$$\log d = 0.5[a.\exp(-b.t) + c.\exp(-k.|t - T|)],$$

where a and b are the intercept and slope parameters of the temporal distance component, c and k are the intercept and rate of decrement parameters of the generalization component, and T is the training delay from which generalization occurs. Figure 5 shows the composite equation fitted to data from the study by Sargisson and White (2001) in which the training delay in a delayed matching-to-sample procedure was fixed at different values for different groups of birds. Coincidentally, this composite of temporal distance and generalization components has the same form as the double exponential suggested by Wickelgren (1969). It can be shown to fit a wide range of data (Rubin, Hinton, & Wenzel, 1999). It fits very closely data that follow the exponential to \sqrt{t} and fits perfectly data that follow simple exponential functions when $T = 0$. The composite function retains the characteristics of independent variation in intercept and slope. The intercept of the resulting forgetting function is influenced by the intercept of the temporal distance component. It is also influenced by generalization from $t = 0$ of performance to delays associated with training. The slope of the resulting forgetting function is determined both by factors influencing the generalization of performance across delays and by such factors as retroactive interference and local proactive interference that influence the effect of the temporally distant sample at the time of retrieval (Edhouse & White, 1988).

Conclusion

In summary, the present approach treats remembering as a process of discrimination and treats the temporal distance of remembered events in the same terms as other physical dimensions. Forgetting functions typically, but not always, involve a monotonic decrement in performance with increasing temporal distance. They are described equally well by several mathematical functions, including power functions that seem consistent with a psychophysical power law and exponential functions with time scaled as \sqrt{t} . The issue of how the dimension of temporal distance should be scaled cannot easily be solved by curve fitting but requires a theoretical approach. A formal theory of remembering has yet to be developed that directly predicts the mathematical form of the forgetting function. In the meantime, two lines of converging evidence support the exponential to \sqrt{t} as the preferred descriptive function. First, by assuming diffusion in White and Wixted's (1999) model, it is appropriate to

scale the time dimension as \sqrt{t} . Second, the empirical evidence for temporal independence suggests that the effect of temporal distance be scaled in terms of an exponential function. When a generalization component is added to the temporal distance component, the resulting double exponential functions decrease in slope at longer delays and overlap the exponential to \sqrt{t} . The proposal that forgetting functions are a composite of two effects—exponential generalization according to Shepard's law and an exponential decrement owing to temporal distance—has promise in its ability to account for the main characteristics of forgetting functions. Research that attempts to identify the separate influence of the two components will facilitate the link between the study of remembering and psychophysical approaches.

REFERENCES

- ALSOP, B. (1998). Receiver operating characteristics from nonhuman animals: Some implications and directions for research with humans. *Psychonomic Bulletin & Review*, *5*, 239-252.
- ANDERSON, J. R., & MILSON, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*, 703-719.
- ANDERSON, J. R., & SCHOOLER, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*, 396-408.
- ANDERSON, R. B., & TWENEY, R. D. (1997). Artifactual power curves in forgetting. *Memory & Cognition*, *25*, 724-730.
- ANDERSON, R. B., TWENEY, R. D., RIVARDO, M., & DUNCAN, S. (1997). Need probability affects retention: A direct demonstration. *Memory & Cognition*, *25*, 867-872.
- BADDELEY, A. (1997). *Human memory*. Hove, U.K.: Psychology Press.
- BAHRICK, H. P. (1965). The Ebb of retention. *Psychological Review*, *72*, 60-73.
- BAHRICK, H. P., BAHRICK, P. O., & WITTLINGER, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, *104*, 54-75.
- BERG, H. C. (1993). *Random walks in biology*. Princeton, NJ: Princeton University Press.
- BLOUGH, D. S. (1959). Delayed matching in the pigeon. *Journal of the Experimental Analysis of Behavior*, *2*, 151-160.
- BOGARTZ, R. S. (1990). Evaluating forgetting curves psychologically. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 138-148.
- CLARK, S. E., & GRONLUND, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37-60.
- CLAYTON, N. S., & DICKINSON, A. (1999). Scrub jays (*Aphelocoma coerulescens*) remember the relative time of caching as well as the location and content of their caches. *Journal of Comparative Psychology*, *113*, 403-416.
- CLAYTON, N. S., YU, K. S., & DICKINSON, A. (2001). Scrub jays (*Aphelocoma coerulescens*) form integrated memories of the multiple features of caching episodes. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*, 17-29.
- CRAIK, F. I. M., GOVONI, R., NAVEH-BENJAMIN, M., & ANDERSON, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*, 159-180.
- CRAIK, F. I. M., & TULVING, E. L. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294.
- D'AMATO, M. R. (1973). Delayed matching and short-term memory in monkeys. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 7, pp. 227-269). New York: Academic Press.
- DARWIN, C. R. (1899). *The descent of man*. London: Murray. (Original work published 1871)

- DAVISON, M. C., & TUSTIN, R. D. (1978). The relation between the generalized matching law and signal-detection theory. *Journal of the Experimental Analysis of Behavior*, **29**, 331-336.
- EBBINGHAUS, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Original work published 1885)
- EDHOUSE, W. V., & WHITE, K. G. (1988). Sources of proactive interference in animal memory. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 56-71.
- FETTERMAN, J. G. (1995). The psychophysics of remembered duration. *Animal Learning & Behavior*, **23**, 49-62.
- FETTERMAN, J. G. (1996). Dimensions of stimulus complexity. *Journal of Experimental Psychology: Animal Behavior Processes*, **22**, 3-18.
- GIBSON, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- GRANT, D. S. (1981). Short-term memory in the pigeon. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 227-256). Hillsdale, NJ: Erlbaum.
- HAIST, F., SHIMAMURA, A. P., & SQUIRE, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 691-702.
- HARNETT, P., MCCARTHY, D. C., & DAVISON, M. C. (1984). Delayed signal detection, differential reinforcement, and short-term memory in the pigeon. *Journal of the Experimental Analysis of Behavior*, **42**, 87-111.
- HARPER, D. N., & WHITE, K. G. (1997). Retroactive interference and rate of forgetting in delayed matching-to-sample performance. *Animal Learning & Behavior*, **25**, 158-164.
- HAUTUS, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, **27**, 46-51.
- HONG, W. K., & URQUIOLI, P. J. (1981). The legacy of Guttman and Kalish (1956): 25 years of research on stimulus generalization. *Journal of the Experimental Analysis of Behavior*, **36**, 405-445.
- JANS, J. E., & CATANIA, A. C. (1980). Short-term remembering of discriminative stimuli in pigeons. *Journal of the Experimental Analysis of Behavior*, **34**, 177-183.
- JONES, B. M., & WHITE, K. G. (1992). Sample-stimulus discriminability and sensitivity to reinforcement in delayed matching to sample. *Journal of the Experimental Analysis of Behavior*, **58**, 159-172.
- JONES, B. M., & WHITE, K. G. (1994). An investigation of the differential-outcomes effect within sessions. *Journal of the Experimental Analysis of Behavior*, **61**, 389-406.
- KAMIL, A. C., & BALDA, R. P. (1990). Differential memory for cache sites in Clark's nutcrackers (*Nucifraga columbiana*). *Journal of Experimental Psychology: Animal Behavior Processes*, **16**, 162-168.
- KINCHLA, R. A., & SMYZER, F. (1967). A diffusion model of perceptual memory. *Perception & Psychophysics*, **2**, 219-229.
- KINTSCH, W. (1970). *Learning and memory*. New York: Wiley.
- KIRK, R. C., WHITE, K. G., & MCNAUGHTON, N. (1988). Low-dose scopolamine affects discriminability but not rate of forgetting in delayed matching to sample. *Psychopharmacology*, **96**, 541-546.
- KORIAT, A., & GOLDSMITH, M. (1996). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral & Brain Sciences*, **19**, 167-228.
- KRAEMER, P. J., & GOLDING, J. M. (1997). Adaptive forgetting in animals. *Psychonomic Bulletin & Review*, **4**, 480-491.
- LAMING, D., & SCHEIWILLER, P. (1985). Retention in perceptual memory: A review of models and data. *Perception & Psychophysics*, **37**, 189-197.
- LOFTUS, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 817-820.
- LOFTUS, G. R., & BAMBER, D. (1990). Learning-forgetting independence, unidimensional memory models, and feature models: Comment on Bogartz. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 916-926.
- LUCE, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103-189). New York: Wiley.
- LUCE, R. D. (1994). Thurstone and sensory scaling: Then and now. *Psychological Review*, **101**, 271-277.
- LUH, C. W. (1922). The conditions of retention. *Psychological Monographs*, **31** (Whole No. 142).
- MACDONALD, S., & HAYNE, H. (1996). Child-initiated conversations about the past and memory performance by preschoolers. *Cognitive Development*, **11**, 421-444.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MACPHAIL, E. M. (1996). Cognitive function in mammals: The evolutionary perspective. *Cognitive Brain Research*, **3**, 279-290.
- MCCARTHY, D., & WHITE, K. G. (1987). Behavioral models of delayed detection and their application to memory. In M. L. Commons, J. Mazur, J. A. Nevin, & H. C. Rachlin (Eds.), *Quantitative analysis of behavior: Vol. 5. The effect of delay and intervening events on reinforcement value* (pp. 29-54). Hillsdale, NJ: Erlbaum.
- MORRIS, C. D., BRANSFORD, J. D., & FRANKS, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, **16**, 519-533.
- MURDOCK, B. B. (1961). The retention of individual items. *Journal of Experimental Psychology*, **62**, 618-625.
- NAVEH-BENJAMIN, M., CRAIK, F. I. M., GAVRILESCU, D., & ANDERSON, N. D. (2000). Asymmetry between encoding and retrieval processes: Evidence from divided attention and a calibration analysis. *Memory & Cognition*, **28**, 965-976.
- NEVIN, J. A. (1970). On differential stimulation and differential reinforcement. In W. C. Stebbins (Ed.), *Animal psychophysics* (pp. 401-423). New York: Appleton-Century-Crofts.
- NEVIN, J. A. (1981). Psychophysics and reinforcement schedules: An integration. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analyses of behavior: Discriminative properties of reinforcement schedules* (Vol. 1, pp. 3-27). Cambridge, MA: Ballinger.
- OLSON, D. J. (1991). Species differences in spatial memory among Clark's nutcrackers, scrub jays, and pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **17**, 363-376.
- PARKES, M., & WHITE, K. G. (2000). Glucose attenuation of memory impairments. *Behavioral Neuroscience*, **114**, 307-319.
- PETERSON, L. R., & PETERSON, M. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, **58**, 193-198.
- ROBERTS, W. A. (1972). Short-term memory in the pigeon: Effects of repetition and spacing. *Journal of Experimental Psychology*, **94**, 74-83.
- ROBERTS, W. A., & GRANT, D. S. (1976). Studies of short-term memory in the pigeon using the delayed matching to sample procedure. In D. L. Medin, W. A. Roberts, & R. T. Davis (Eds.), *Processes of animal memory* (pp. 79-112). Hillsdale, NJ: Erlbaum.
- ROEDIGER, H. L., III, & GUYNN, M. J. (1996). Retrieval processes. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 197-236). San Diego: Academic Press.
- ROITBLAT, H. L. (1982). The meaning of representation in animal memory. *Behavioral & Brain Sciences*, **5**, 353-406.
- ROITBLAT, H. L. (1983). Pigeon working memory: Models for delayed matching-to-sample performance. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Vol. 4. Discrimination processes* (pp. 161-180). Cambridge, MA: Ballinger.
- RUBIN, D. C. (1982). On the retention function for autobiographical memory. *Journal of Verbal Learning & Verbal Behavior*, **21**, 21-38.
- RUBIN, D. C., & BADDELEY, A. D. (1989). Telescoping is not time compression: A model of dating autobiographical events. *Memory & Cognition*, **17**, 653-661.
- RUBIN, D. C., HINTON, S., & WENZEL, A. (1999). The precise time course of retention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1161-1176.
- RUBIN, D. C., & WENZEL, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, **103**, 734-760.
- RUSKE, A. C., FISHER, A., & WHITE, K. G. (1997). Attenuation of scopolamine-induced deficits in delayed-matching performance by a new muscarinic agonist. *Psychobiology*, **25**, 313-320.
- RUSKE, A. C., & WHITE, K. G. (1999). Attenuation of scopolamine-induced deficits in pigeons' attention to delayed matching-to-sample cues by D-cycloserine. *Psychobiology*, **27**, 507-512.
- SARGISSON, R. J., & WHITE, K. G. (2001). Generalization of delayed matching-to-sample performance following training at different delays. *Journal of the Experimental Analysis of Behavior*, **75**, 1-14.

- SHEPARD, R. N. (1958). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, **65**, 242-256.
- SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.
- SHEPARD, R. N. (1988). Time and distance in generalization and discrimination: Reply to Ennis (1988). *Journal of Experimental Psychology: General*, **117**, 415-416.
- SIMON, H. A. (1966). A note on Jost's law and exponential forgetting. *Psychometrika*, **31**, 505-506.
- SLAMECKA, N. J., & MCELREE, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 384-397.
- SPETCH, M. L. (1987). Systematic errors in pigeons' memory for event duration: Interaction between training and test delay. *Animal Learning & Behavior*, **15**, 1-5.
- SPETCH, M. L., & RUSAK, B. (1992). Temporal context effects in pigeons' memory for event duration. *Learning & Motivation*, **23**, 117-144.
- SQUIRE, L. R. (1989). On the course of forgetting in very long-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 241-245.
- STADDON, J. E. R. (1983). *Adaptive behavior and learning*. Cambridge: Cambridge University Press.
- THORNDIKE, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monograph Supplements*, **2**, 1-109.
- THURSTONE, L. L. (1927). A law of comparative judgment. *Psychological Review*, **34**, 273-286.
- TULVING, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- TULVING, E. (1997). Organization of memory: Quo vadis? In M. S. Gazzaniga (Ed.), *Conversations in the cognitive neurosciences* (pp. 839-847). Cambridge, MA: MIT Press.
- TULVING, E., & PEARLSTONE, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, **5**, 381-391.
- TULVING, E., & THOMSON, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80**, 352-373.
- URCUIOLI, P. J., & DEMARSE, T. B. (1997). Memory processes in delayed spatial discriminations: Response intentions or response mediation? *Journal of the Experimental Analysis of Behavior*, **67**, 323-336.
- WATKINS, M. J. (1990). Mediationism and the obfuscation of memory. *American Psychologist*, **45**, 328-335.
- WATSON, J. E., & BLAMPIED, N. M. (1989). Quantification of the effects of chlorpromazine on performance under delayed matching to sample in pigeons. *Journal of the Experimental Analysis of Behavior*, **51**, 317-328.
- WHITE, K. G. (1974). Temporal integration in the pigeon. *British Journal of Psychology*, **65**, 437-444.
- WHITE, K. G. (1985). Characteristics of forgetting functions in delayed matching to sample. *Journal of the Experimental Analysis of Behavior*, **44**, 15-34.
- WHITE, K. G. (1991). Psychophysics of direct remembering. In J. A. Commons, J. A. Nevin, & M. C. Davison (Eds.), *Models of behavior: Signal detection* (pp. 221-237). Hillsdale, NJ: Erlbaum.
- WHITE, K. G. (1996). Direct remembering and the correspondence metaphor. *Behavioral & Brain Sciences*, **19**, 208-209.
- WHITE, K. G., & COONEY, E. B. (1996). Consequences of remembering: Independence of performance at different retention intervals. *Journal of Experimental Psychology: Animal Behavior Processes*, **22**, 51-59.
- WHITE, K. G., & HARPER, D. N. (1996). Quantitative reanalysis of lesion effects on rate of forgetting in macaques. *Behavioural Brain Research*, **74**, 223-227.
- WHITE, K. G., HARPER, D. N., & WATSON, J. E. (1994). Effects of chronic phenobarbital administration on forgetting functions in pigeons. *Pharmacology, Biochemistry & Behavior*, **49**, 427-431.
- WHITE, K. G., JUHASZ, J. B., & WILSON, P. J. (1973). Is man no more than this? Evaluative bias in interspecies comparison. *Journal of the History of the Behavioral Sciences*, **9**, 203-212.
- WHITE, K. G., & MCKENZIE, J. (1982). Delayed stimulus control: Recall for single and relational stimuli. *Journal for the Experimental Analysis of Behavior*, **38**, 305-312.
- WHITE, K. G., RUSKE, A. C., & COLOMBO, M. (1996). Memory procedures, performance and processes in pigeons. *Cognitive Brain Research*, **3**, 309-317.
- WHITE, K. G., & WIXTED, J. T. (1999). Psychophysics of remembering. *Journal of the Experimental Analysis of Behavior*, **71**, 91-113.
- WICKELGREN, W. A. (1969). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, **6**, 13-61.
- WICKELGREN, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology*, **9**, 418-455.
- WICKELGREN, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition*, **2**, 775-780.
- WICKELGREN, W. A. (1975). Alcoholic intoxication and memory storage dynamics. *Memory & Cognition*, **3**, 385-389.
- WICKENS, T. D. (1998). On the form of the retention function: Comment on Rubin and Wenzel (1996): A quantitative description of retention. *Psychological Review*, **105**, 379-386.
- WIXTED, J. T. (1989). Nonhuman short-term memory: A quantitative reanalysis of selected findings. *Journal of the Experimental Analysis of Behavior*, **52**, 409-426.
- WIXTED, J. T. (1990). Analyzing the empirical course of forgetting. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 927-935.
- WIXTED, J. T., & EBBESEN, E. B. (1991). On the form of forgetting. *Psychological Science*, **6**, 409-415.
- WIXTED, J. T., & EBBESEN, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, **25**, 731-739.
- WOODWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental psychology: Revised edition*. New York: Holt, Rinehart & Winston.

(Manuscript received November 10, 2000;
revision accepted for publication May 2, 2001.)