Psychological Review
2006, Vol. 113, No. 1, 31–56


0033-295X/06/$12.00 DOI: 10.1037/0033-295X.113.1.31

# Memory Hazard Functions: A Vehicle for Theory Development and Test

Richard A. Chechile
Tufts University

A framework is developed to rigorously test an entire class of memory retention functions by examining hazard properties. Evidence is provided that the memory hazard function is not monotonically decreasing. Yet most of the proposals for retention functions, which have emerged from the psychological literature, imply that memory hazard is monotonically decreasing over the entire temporal range. Furthermore, the few remaining proposals, that do not have monotonically decreasing hazard, have difficulty fitting data over both short-term and long-term intervals. A new 2-trace hazard model is developed that successfully circumvents these difficulties. This new model is used to account for the size of memory span and the time course of proactive and retroactive interference effects. The model can fit the retention characteristics of H. M., the famous amnesic patient, as well as normal experimental participants. The model is also used to account for the time course of the misinformation effect.

*Keywords:* hazard functions, memory retention functions, memory storage dynamics, two-trace hazard model

Since the first experimental treatise on the topic of memory by Ebbinghaus (1885/1964), it has been clear that memory retention degrades in an orderly fashion subsequent to the learning of information. Yet one of the basic and still unresolved questions about memory pertains to the mathematical form of the retention function. This question in turn is connected to a number of other questions that also need to be addressed in order to ascertain the nature of the retention function. For example, how is memory to be measured, and how should a retention function be assessed? The resolution of questions such as these and the specification of the retention function are crucial milestones in the science of memory.

A number of proposals have been discussed in the psychological literature for the mathematical form of the memory retention function (e.g., Anderson & Schooler, 1991; Chechile, 1987; Ebbinghaus, 1885/1964; Rubin & Wenzel, 1996; Wickelgren, 1974; Wixted & Ebbesen, 1991). Even more functions, which are at least as good as ones in the psychological literature, can be generated. Clearly no consensus for the retention function has yet been achieved.

One reason for the lack of agreement has to do with a difference of opinion as to the approach for considering candidate functions. For some theorists, the retention function emerged from a more broad memory model (e.g., Atkinson & Shiffrin, 1968; Chechile, 1987; Wickelgren, 1974). Other researchers, however, deliberately eschewed a theoretical framework and empirically evaluated a wide range of functions by means of goodness-of-fit statistics. For example, Rubin and Wenzel (1996) examined a number of simple functional forms and fit these functions to an extensive set of memory data. Wixted and Ebbesen (1991) and Anderson and Schooler (1991) also used a similar approach to assess memory functions. However, the approach of assessing retention functions on the basis of goodness-of-fit has been strongly critiqued (see Roberts & Pashler, 2000; Wickens, 1998). Roberts and Pashler (2000) argued that goodness-of-fit is not persuasive in theory testing generally. Although many scientists would agree that fitting models to data is not enough to establish a successful theory, yet it is another matter to ignore the fit of the theory to data. A successful theory should not have a poor fit to existing data. After all, experimental data are a time-honored means for testing theories, and a good fit to data should not be ignored. Yet, theories that fit the data well in a restricted range might result in absurd predictions or a very poor fit when extrapolated beyond that range. Furthermore, many of the existing data sets are problematic, so fitting data requires careful consideration of the suitability of the data. Hence, additional principles need to be examined in the evaluation of theories beyond that of goodness of fit. This theoretical point has been stressed in the Wickens (1998) critique of the Rubin and Wenzel (1996) article. Wickens further argued for a principled examination as to why a particular memory retention function fails.

Although there currently is not a convergence among investigators on the mathematical form of the retention function, there is agreement that the retention function is a description of a stochastic process. Even Ebbinghaus, who conducted his research prior to the development of 20th century statistics, recognized that memory retention is not a deterministic process. According to Ebbinghaus (1885/1964), variability and uncertainty in predictions occurred because of "the constant flux and caprice of mental events" (p. 19). We must be content to characterize the change in the probability of remembering. Thus, the decline in memory performance with the retention interval indicates changes in the probability for remembering the previously encoded information.

In general, an ideal approach for understanding the dynamic changes of a stochastic process is to examine the corresponding hazard function. In this article, a wide range of candidate functions are critically examined. Both functions generated from memory

I thank Jean-Claude Falmagne, Mario Fific, Janet Metcalfe, David Riefer, James Townsend, and John Wixted for providing valuable comments and suggestions about the work reported in this article.

Correspondence concerning this article should be addressed to Richard A. Chechile, Psychology Department, Tufts University, Medford, MA 02155. E-mail: richard.chechile@tufts.edu

theories as well as simple descriptive functions produced in the absence of a theory are considered. A number of criteria are used, but the analysis is largely based on hazard functions. New developments in the area of hazard function analysis (see Chechile, 2003) provide a means to obtain further insights about the mathematical form of the retention function. In fact, a hazard analysis results in a powerful set of constraints on memory theory—constraints that are problematic for most of the proposals for the memory retention function. However, it is crucial to first discuss a number of subtle features about hazard functions and the general linkage of memory retention functions to subprobability distributions. These points are covered in the brief general tutorial provided in the next section.

## Probability, Subprobability, and Hazard Functions

### *Continuous Stochastic Process*

Three theoretical constructs are central for understanding continuous stochastic processes. Depending on the investigator's primary purpose, each conceptual tool can be ideal. These constructs are probability density, $f(x)$, the cumulative probability, $F(x)$, and the hazard function, $\lambda(x)$. Probability density functions (PDF), such as the familiar bell-shaped normal distribution, are ideal for representing the shape and location of the distribution. However, probability is not directly shown in a PDF display. Probability corresponds to the area between points on a PDF display. The cumulative probability that an event occurred prior to the current value of the random variable, x, is $F(x)$, and this probability is the left-tail area of the PDF distribution. Thus, the cumulative probability and the probability density are interrelated by the calculus operations of integration and differentiation; see Figure 1 for an illustration of this point. The cumulative probability is the integrated probability density function from the lower limit of the random variable to the current value of x, and the probability density at x is the derivative of the cumulative function, that is, $f(x) = dF(x)/dx$.

If the distribution corresponds to the lifetimes of various memory traces, then the left-tail area of the PDF display represents the proportion of forgotten items by time $t$. The probability of trace storage is denoted as $\theta_S(t)$. Moreover, the cumulative probability, $F(t)$, equals $1 - \theta_S(t)$. In terms of this representation, it is clear that there must be a linkage between a model of how storage changes with time and the underlying stochastic model for memory lifetimes. If a memory researcher posits a mathematical form for memory retention as a function of time, then the researcher is implicitly proposing a stochastic distribution for the lifetime of memory traces.

The hazard for a continuous stochastic process is defined as the likelihood that a critical event occurs exactly at x and did not occur prior to that point. For a general variable x, the hazard is the probability density at x divided by $1 - F(x)$. It is the dividing by $1 - F(x)$ that reconditions the probability density to represent the fact that the critical event has not yet occurred. In the application of hazard functions to memory retention, consideration is limited to the case where the variable x is equal to time. The hazard function, $\lambda(t)$, is thus defined as

$$\lambda(t) = \frac{f(t)}{1 - F(t)} . \quad (1)$$

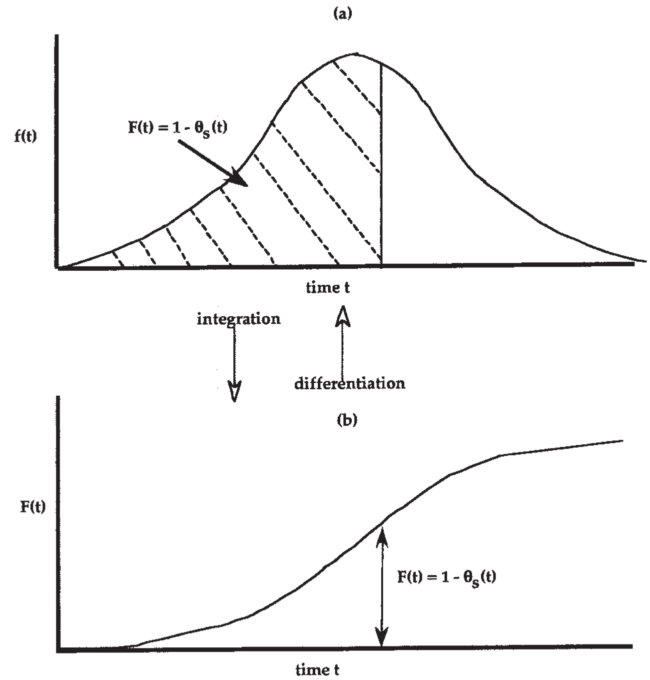


*Figure 1.* A hypothetical probability density function $f(t)$ for (a) memory lifetimes and (b) the resulting cumulative distribution, $F(t)$, which corresponds to the proportion of items lost from storage by time $t$. Memory storage at time $t$ is denoted as $\theta_S(t)$.

Hazard functions are ideally suited for studying the dynamic changes in stochastic processes. Because the critical event is quite general, hazard functions have been useful in a number of scientific disciplines. For example, hazard functions have been employed in the health sciences (Klein & Moeschberger, 1997) as well as in actuarial and product reliability studies (Bain, 1978). For these applications, the critical event is typically either patient death or product failure. Hazard functions have also been valuable in psychological research. Townsend and Ashby (1978) helped promote the use of hazard functions in theoretical analyses of information processing capacity. Here the critical event corresponds to the completion of a task by the participant. This definition of the critical event leads to hazard functions for various stochastic models for understanding response time (e.g., Colonius, 1988; Luce, 1986; Townsend & Ashby, 1983). In other cases, the critical event is taken to be the initiation of an eye movement (e.g., Peterson, Kramer, Wang, Irwin, & McCarley, 2001; Yang & McConkie, 2001). In yet other psychological applications, the critical event corresponds to the loss of a memory trace (e.g., Chechile, 1987; Wickens, 1998, 1999). It is this definition of the critical event that is used in this paper to explore the problem of the memory retention function. A recent general compendium of mathematical results about hazard functions can be found in Chechile (2003). Some of the results in Chechile (2003) were previously known, but many new theorems for understanding hazard functions are also developed in that paper.

Given a specified hazard function, it is possible to derive the density and cumulative distributions. To connect hazard with these other basic constructs, it is convenient to provide a representation of the left-tail area of the hazard function, that is, the integrated hazard. The integrated hazard from the lower limit of the variable to the current value of x is denoted as H(x). It is well known that the cumulative probability is in general

$$F(x) = 1 - \exp(-H[x]). \quad (2)$$

See Chechile (2003, p. 486) for a proof of Equation 2. Because $f(x)$ is the derivative of $F(x)$, it also follows that the probability density is given as

$$f(x) = \lambda(x)\exp(-H[x]). \quad (3)$$

Chechile (2003) showed how the theorist can directly model the hazard function of a psychological process and then use Equations 2–3 to specify the resulting $F(x)$ and $f(x)$ functions. In fact, several new stochastic models were developed in this fashion in Chechile (2003). Hazard functions are thus an alternative window for examining stochastic processes.

The interval [0, U] will be referred to as the *domain of support* for the variable; hence, U is the upper limit for the random variable. For a probability distribution it must be the case that $F(U) = 1.0$ because the integration of all the probability densities over the full domain of support is 1.0. From Equation 2 it follows that H(U) must diverge to infinity in order for the distribution to be a probability distribution. Thus, the integrated hazard (the area underneath the hazard function curve) must grow without limit for a probability distribution. This point is illustrated in Figure 2 for a unique distribution from the hazard function perspective—the exponential distribution. The probability density function for the exponential is $f(t) = a\exp(-at)$, and is shown in Figure 2A. The exponential distribution is the only probability density function that has constant hazard, and that constant equals the value of the parameter shown in the density display; see Figure 2B. Again, in this paper the variable is time, that is, $x = t$. In Figure 2C, the area beneath the hazard function, that is, the integrated hazard, $H(t)$, is plotted as a function of time. Note that $H(t) = at$ and grows without limit. For all continuous probability distributions, $H(t)$ diverges to positive infinity as $t$ approaches U, and this corresponds to $F(U) = 1$.

The case where $F(U) = 1$ corresponds to the statement that the critical event will occur eventually in the domain of support. This type of hazard model is reasonable for cases where the critical event corresponds to the completion of a speeded task because a response is expected on each test trial. However, there are other applications where the critical event might not always occur. In such cases, the value for $F(U)$ will be less than 1.0. In fact, $F(U)$ is less than 1.0 whenever the integrated hazard does not diverge to infinity, see Equation 2. In mathematics, such stochastic models are called *subprobability distributions* (Bertoin & LeJan, 1992: Graczyk, 1994). Subprobability theory is very similar to probability theory except that the theorems from probability theory that depend crucially on the sum of the probabilities being equal to 1.0 are not valid for subprobability theory. However, there are many theorems from probability theory that do not depend on stipulating that the probabilities add to 1.0. Those theorems are also valid for subprobability theory. Thus, probability theory can be considered as the special case of subprobability theory where the sum of the probabilities is 1.0.

The distinction between probability and subprobability is particularly germane for the study of memory loss where the critical event is the loss of a memory trace. If a probability model is used for the trace lifetime distribution, then permanent memory is impossible because all memory traces would be assumed to fail. But suppose some memories are preserved over the entire life of the individual. The death of the individual is outside of the domain of interest for memory researchers who are focused on the fate of previously learned items. Consequently, the upper limit, U, for the time variable is not infinity but rather is the life span of the individual. If some traces are preserved over that life span, then the critical event did not always occur. Because $F(U)$ is the probability of the critical event occurring over the life span of the individual, then $F(U)$ is less than 1.0 whenever there are some permanent memories. Hence, if there is an allowance for the possibility of permanent memory, then a subprobability model for the trace lifetime distribution must be used rather than a probability distribution.

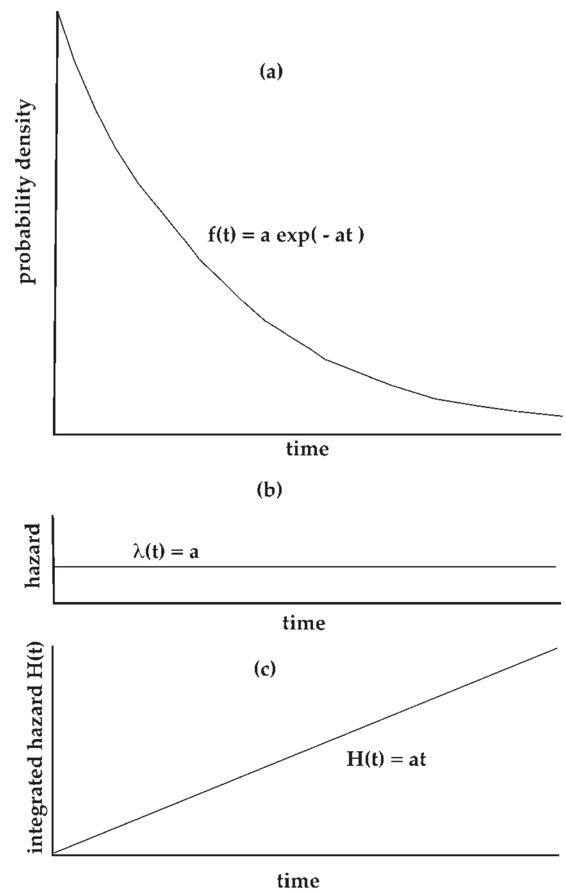


*Figure 2.* The probability density function for (a) an exponential distribution; (b) the corresponding hazard function, $\lambda(t)$; and (c) the resulting integrated hazard, $H(t)$.

However, some might argue that a probability distribution, such as the exponential distribution shown in Figure 2, has an infinite domain of support for the random variable; hence, over the finite time of a person's life there would be some permanent memory. For example, the trace survival value, $1 - F(t)$, would be nonzero after a long period such as 50 years. However, if memory encoding were sufficiently well established so as to enable 99.9% of the memory traces to survive the first 15 s, then the exponential lifetime probability model would result in a probability of $10^{-45677}$ for memory survival after 50 years. Although this value is technically nonzero, it effectively results in no realistic chance for any of the traces to survive a 50-year interval. If one does not want to presume that all memories fail over the life of the individual, then an exponential lifetime model for the traces cannot be used.

However, suppose that the probability density of the exponential distribution is multiplied by a positive constant ($b$) that is less than 1.0, that is, $f(t) = b\, a \exp(-at)$. For this distribution $F(\text{U}) = b < 1.0$. This modification represents a change from a probability model to a subprobability model. The shape of the exponential subprobability distribution is the same as the exponential, that is, the subprobability densities are $b$ times the corresponding exponential densities. For the proportion of the memory traces that fail, the expected value of the lifetime distribution is the same as for the exponential lifetime distribution. Of course, the unconditional expected value is different because $1 - b$ proportion of the memory traces survive over the entire lifetime of the individual. Note that the exponential subprobability model for memory lifetimes is a more general representation. A special case of this model includes the exponential lifetime model, that is, the special case of $b = 1$. Also, note that the modification of the exponential distribution to create the exponential subprobability changes the hazard function. The exponential subprobability model has the following hazard function:

$$\lambda(t) = \frac{a}{1 + \frac{1-b}{b}\exp(at)}. \tag{4}$$

The denominator of Equation 4 grows with increasing time, so the hazard is no longer a constant, but rather decreases monotonically with time. This hazard function also provides for the possibility that some memories survive over the entire life of the person.

### *Discrete Stochastic Processes*

If the stochastic process is discrete rather than continuous, then there is an analysis similar to the above discussion. For example, if memory retention were a function of the number of events after encoding ($n$) rather than time ($t$), then the trace lifetime distribution could be represented by a discrete probability distribution. For a discrete probability distribution, there are probabilities for each discrete event. The definition of hazard is different from Equation 1 for the case of a discrete probability distribution since there cannot be probability densities (see Chechile, 2003). Given a discrete distribution, where the probability of the critical event occurring for event $j$ is $p_j$, the hazard is defined as

$$\lambda_n = \begin{cases} p_0, & \text{for } n = 0 \\ \dfrac{p_n}{1 - \sum_{j=0}^{j=n-1} p_i}, & \text{for } n > 0 \end{cases}. \tag{5}$$

The discrete hazard function can also be the starting point for the stochastic modeling because the theorist can derive the probability distribution from the hazard function. Chechile (2003) showed that discrete probabilities are expressed as a function of the hazard via the following equation:

$$p_n = \lambda_n \prod_{i=0}^{n-1} (1 - \lambda_i). \tag{6}$$

There is also a special discrete probability distribution corresponding to the case of constant hazard, the geometric distribution; that is, $p_n = a(1 - a)^{n-1}$ where $\lambda_n = a$ for all $n$. If the lifetime distribution is a discrete probability distribution, then the sum of all the probabilities is 1.0. However, if the sum of all the probabilities is less than 1.0, then the distribution is a discrete subprobability distribution. Again in order to represent the case of permanent memory for some of the memory targets, a subprobability distribution must be used. For example, if a geometric distribution (with constant hazard) has a parameter value that results in 99.9% of the memory targets surviving the first 30 interpolated events, then the probability of memory after 50 years (assuming one interpolated event per second) would be less than $10^{-228413}$. Clearly this probability model does not realistically allow for permanent memory. However, the geometric distribution can be modified to be a subprobability distribution by multiplying the discrete probabilities by a positive constant $b$ that is less than 1.0, that is, $p_n = b\, a(1 - a)^{n-1}$. As with the exponential subprobability distribution, the modification of the geometric distribution changes the hazard function, which now is

$$\lambda_n = \frac{a}{1 + \frac{1-b}{b(1-a)^{n-1}}}. \tag{7}$$

As $n$ increases in value, the second term in the denominator of Equation 7 increases, so the modification of the geometric distribution results in changing the hazard function from a constant to a function that decreases with an increasing number of interpolated items. As with the modification of the exponential distribution, a decreasing hazard, subprobability model makes long-term memory possible.

## Candidate Functions and Model Constraints

A number of candidate functions for representing the change in retention are provided in Table 1. For each case, the dependent measure (i.e., the left-hand side of a retention equation) is $\theta_S$, which is defined as the probability of trace storage. The right-hand side of the retention equation is shown in Table 1 as a function of either time ($t$) or the number intervening items ($n$) and free fitting parameters, that is, the $a$, $b$, $c$, and $d$ parameters. Table 1 is designed to be inclusive of proposals that have emerged from the psychological literature, as well as modifications to address problems with particular functional forms. The Ebbinghaus function originated from the first experimental treatise about memory, that is, Ebbinghaus (1885/1964). The hyperbolic, exponential, and logarithm functions are taken from Rubin and Wenzel (1996); although the hyperbolic and exponential functions shown in Table 1 are modified from the ones in Rubin and Wenzel (1996) in order to constrain the functions to the value of 1.0 when $t = 0$ because errors at $t = 0$ should be called perceptual–encoding errors rather

Table 1
*Candidate Retention Functions and Their Corresponding Hazard Property*

| Name | Function | Hazard property |
|---|---|---|
| Ebbinghaus[d] | $\frac{b}{b + [\log(t+1)]^c}$ | Peaked function |
| Hyperbolic[e] | $\frac{1}{at+1}$ | Monotonic decreasing |
| PM–hyperbolic[c] | $1 - b + \frac{b}{at+1}$ | Monotonic decreasing |
| Exponential[e] | $\exp(-at)$ | Constant |
| PM–exponential[c] | $1 - b + b\exp(-at)$ | Monotonic decreasing |
| Logarithm[e] | $b - a\log(t)$ | Ill defined at $t = 0$ |
| Modified logarithm[c] | $1 - a\log(t+1)$ | Monotonic decreasing |
| Power[g] | $b/t^c$ | Ill defined at $t = 0$ |
| Modified PM–power[c] | $1 - b + \frac{b}{(t+1)^c}$ | Monotonic decreasing |
| AS–power[a] | $\frac{b}{b+t^c}$ | Peaked if $c > 1$; decreasing if $0 < c \leq 1$ |
| PM–AS–power[c] | $1 - b + \frac{b^2}{b+t^c}$ | Peaked if $c > 1$; decreasing if $0 < c \leq 1$ |
| STFT[f] | $\frac{\exp(-a\,t)}{(1+b\,t)^c}$ | Monotonic decreasing |
| PM–STFT[c] | $1 - d + \frac{d\exp(-at)}{(1+bt)^c}$ | Monotonic decreasing |
| Multiple-store model[c] | $1 - b(1 - a^n)$ | Monotonic decreasing |
| Trace susceptibility theory[b] | $1 - b + b\exp(-at^2)$ | Peaked function |

*Note.* PM = permanent memory; AS = Anderson–Schooler; STFT = single-trace fragility theory.
[a] Anderson and Schooler (1991). [b] Chechile (1987). [c] Chechile, present article. [d] Ebbinghaus (1885/1964). [e] Rubin and Wenzel (1996). [f] Wickelgren (1974). [g] Wixted and Ebbesen (1991).

than memory errors. (In order to keep the focus on changes in memory performance, it is expected that experimenters have already removed the effect of any initially misperceived targets.) The power function was taken from the Wixted and Ebbesen (1991) article, and the AS–power model was from the Anderson and Schooler (1991) article. The single-trace fragility theory (STFT) was developed by Wickelgren (1974); although the function in Table 1 is reexpressed to predict the probability of storage as opposed to $d'$, that is, Wickelgren intended memory at $t = 0$ to have strength equal to $\rho$, but by setting $\rho$ to 1.0 the STFT function is reexpressed to predict a probability value on the (0, 1) interval. The function for the multiple-store model (MSM) is derived in Part A of the Appendix and is based on a simple Markov model form of the Atkinson and Shiffrin (1968) model when there is no further rehearsal of the memory target after the initial encoding. The trace susceptibility theory (TST) function is taken from Chechile (1987). The other six functions (i.e., permanent memory [PM]–hyperbolic, PM–exponential, modified logarithm, modified PM–power, PM–AS–power, and PM–STFT) were developed here to correct for problematic features in other candidate functions. The PM–hyperbolic, PM–exponential, PM–AS–power, and PM–STFT models are based on subprobability distributions so as to account for long-term retention. In each case, the new function is more general because it allows for the possibility of some permanent memory. For example, if $b = 1$, then the PM–exponential model is equivalent to the exponential model. However, if $b$ is less than 1.0, then $1 - b$ percent of the memory items are permanently stored. The modified logarithm and the modified PM–power functions were developed to rectify problems with the respective logarithm and power functions. In both the logarithm and power models, the mathematical form of the retention function, as expressed in the original papers, is ill defined at $t = 0$. Those models predict an infinite probability of survival at $t = 0$; clearly, probability cannot exceed 1.0. Those models also have ill behaved hazard at $t = 0$. Hence, in this article, the logarithm and power functions are not considered further because each function has been replaced with a more powerful and properly defined function. For the other functions, the analysis of the general property of the hazard function is provided in Part A of the Appendix.

The retention functions listed in Table 1 differ in a number of ways. For example, the models differ in regard to the number of free parameters. For the hyperbolic, exponential, and modified-logarithm functions, there is a single parameter to fit the retention data. The STFT function has three free parameters, and the PM–STFT function has four free parameters. The other retention functions have two free parameters.

The functions shown in Table 1 also differ in regard to their theoretical grounding. Only the last four models were developed by means of principled arguments. For example, Wickelgren (1974) developed the STFT function by solving several coupled differential equations. Although the use of principles to establish a functional relationship is an intellectually sound approach, it has been difficult to reach a consensus in the area of memory research as to what the principles are. In contradistinction to the principled models, the first 11 functions are ad hoc proposals. Although many scientists eschew ad hoc models in general, an initially ad hoc model might be precisely correct or at least a valuable approximation. For example, the Bohr (1913) quantum model for the hydrogen atom can be described as an ad hoc proposal, but it nonetheless has proved to be a valuable approximation and an important contribution. Also, it might be possible that all of the principled theories in a particular research area are erroneous. Consequently, each of the proposed functions will be critically examined, regardless of the degree of the theoretical justification for the retention function.

Another difference among the functions is the nature of the independent variable of the retention interval, that is, either $t$ or $n$. All the models other than the MSM are expressed in terms of time, $t$, rather than the number of intervening events, $n$. The mechanism of storage loss in the MSM formulation is item "knock out" from a short-term memory as a result of the attention to intervening items. However, the passage of time is usually highly correlated with the number of intervening events. Thus, the linear transformation $t = n\tau$ should accurately represent the retention function in terms of time rather than events. This point is discussed later in the paper.

One clear difference among the functions listed in Table 1 is the ability of the function to represent the case in which participants can permanently remember some items over their entire life. As discussed previously, the exponential distribution with its constant hazard cannot effectively model this case. However, the PM–exponential model can deal with the permanent memory of some traces. The inability of a function to account for the possibility of permanent retention is a serious failing. For example, Bahrick (1984) studied the memory retention of information taught in high school over a 50-year period. For most participants, Bahrick (1984) reported above-chance levels of retention on many performance measures with "no significant rehearsal effect" (p.1), that is, the participants retained some knowledge of their training after a lengthy interval with little or no rehearsal. Of course there are massive individual differences in the rate of long-term retention. Although some participants demonstrate substantial memory loss, others have an extraordinary memory. Wilding and Valentine (1997) documented many cases of individuals with superior memory retention. Luria (1968) provided a detailed exploration of one person who remembered so much that it was a problem at times. Consequently, any retention function in Table 1 that cannot account for the possibility of long-term retention should be abandoned in favor of another more general model as discussed above in terms of the exponential model and the PM–exponential model. Hence, six of the functions lists in Table 1 are disregarded either because of the failure to be defined at $t = 0$ or because of the failure to account for the possibility of long-term memory. The surviving nine functions that satisfy these two constraints are displayed in Table 2.

A clear qualitative difference among the functions is the general form of the corresponding hazard function. Three functions (i.e., Ebbinghaus, PM–AS–power, and TST) have a peaked-shaped hazard function. For these functions, the hazard at $t = 0$ is low and increases with time, but after reaching a peak the hazard then decreases for longer times. However, the other six functions displayed in Table 2 have a hazard profile that strictly decreases monotonically for increasing time. For these functions, the hazard is at a maximum when $t = 0$, and it decreases over the entire time domain. Note that for longer retention intervals, all the functions displayed in Table 2 have the same general behavior of decreasing hazard with increasing time; however, in the shorter temporal domain the functions are very different. Consequently, considerable information about the mathematical form of the memory retention function can be obtained by exploring the memory hazard function in the early time period.

## A Method to Ascertain the General Shape of the Memory Hazard Function

Although hazard functions are an ideal way to explore the dynamic properties of a stochastic process, there are some serious difficulties with developing an empirical hazard function. The numerator of hazard is a probability density; consequently, the empirical determination of the hazard at time $t$ requires a density

Table 2
*Candidate Retention Functions with Long-Term Capability and a Rational $t = 0$ Value*

| Name | Function | Hazard property |
|---|---|---|
| Ebbinghaus | $\frac{b}{b + [\log(t+1)]^c}$ | Peaked function |
| PM–hyperbolic | $1 - b + \frac{b}{at + 1}$ | Monotonic decreasing |
| PM–exponential | $1 - b + b\exp(-at)$ | Monotonic decreasing |
| Modified logarithm | $1 - a\log(t + 1)$ | Monotonic decreasing |
| Modified PM–power | $1 - b + \frac{b}{(t+1)^c}$ | Monotonic decreasing |
| PM–AS–power | $1 - b + \frac{b^2}{b + t^c}$ | Peaked if $c > 1$; decreasing if $0 < c \leq 1$ |
| PM–STFT | $1 - d + \frac{d\exp(-at)}{(1 + bt)^c}$ | Monotonic decreasing |
| Multiple-store model | $1 - b(1 - a^n)$ | Monotonic decreasing |
| Trace susceptibility theory | $1 - b + b\exp(-at^2)$ | Peaked function |

*Note.* PM = permanent memory; AS = Anderson–Schooler; STFT = single-trace fragility theory.

estimate at time $t$. Density estimation is difficult because it requires retention data for two times that are very close in value. For example, to estimate empirically the hazard at 5 s requires independent retention estimates for both 5 s and another time close in value, such as 6 s. A substantial number of observations are needed for each of those two time intervals in order to obtain an accurate value of the difference in the probability of memory retention. Probability densities are not directly observable but can be approximated by the difference in the retention between the two intervals divided by the length of the interval (1 s in the case of the example). However, a hazard function is a representation of hazard as a function of time. Consequently, to see if the hazard in the 5- to 6-s range is increasing or decreasing, it is necessary to also find the hazard at 6 s, which in turn requires the estimation of memory retention at 7 s. Of course the change in hazard would need to be assessed at many other points to approximate the general shape of the hazard function. Clearly the empirical assessment of the hazard function is a formidable experimental challenge. Additional problems with statistical estimates of hazard functions are discussed in Chechile (2003). In light of the problems with empirical hazard function assessment, some investigators have worked with the integrated hazard function (e.g., Townsend & Nozawa, 1995). Other investigators have focused on a proportional hazard analysis that is designed to detect the ordering of two hazard functions (e.g., Wenger & Gibson, 2004). Yet a great deal can be learned about hazard functions from mathematical analyses, and these results do not depend on empirical observations that are subject to statistical error.

Many definitive results can be obtained about hazard functions without the necessity of conducting brute-force empirical measurement of hazard. For example, it has been proven that all probability and subprobability distributions that have monotonically decreasing hazard (as well as the exponential distribution with constant hazard) must have a monotonically decreasing probability density (see the proof to Corollary 1 in Chechile [2003] for the rationale for this conclusion). Consequently, the monotonically decreasing density illustrated in Figure 2A is typical of all probability and subprobability distributions that have monotonically decreasing hazard. Many other results can be established about hazard functions via mathematical proofs (see Chechile, 2003). In this paper, a new theorem is established that provides a means to experimentally test if the memory hazard function is monotonically decreasing. Instead of estimating the hazard at various times, one only need find the probability of memory storage after various time delays.

In order to set the stage for the theorem, consider the variable $u(t)$, which is defined to be equal to $F(t)/t$. Recall that the cumulative probability for a trace lifetime distribution is $F(t)$ and is equal to $1 - \theta_S(t)$ where $\theta_S(t)$ is the probability of memory storage at time $t$. The theorem deals with the peak of the $u(t)$ function for the case of probability and subprobability distributions that have monotonically decreasing hazard.

*Theorem 1*: For any continuous probability or subprobability distribution for $t \geq 0$ that has monotonically decreasing hazard, the function $u(t) = F(t)/t$ is at a maximum for $t = 0$. (See Part B of the Appendix for the proof.)

The above theorem provides a means to experimentally test an entire class of retention functions. If the peak of the function $u(t) = F(t)/t$ is not at $t = 0$, then there is clear evidence against the entire class of retention functions that result in monotonically decreasing hazard. However, if $u(t)$ keeps increasing as the retention interval gets shorter, then the class of retention functions that predict peaked hazard should be rejected. Furthermore, the $u(t)$ function can be readily explored without the abovementioned difficulties with empirical hazard function measurement. However, there are a few additional design issues that need to be addressed before the peak of the $u(t)$ function can be assessed.

Several design issues have to do with the sampling of retention intervals and the type of experimental task. Suppose that the peak of the $u(t)$ function is at 5 s, but the investigator only obtained estimates of $u(t)$ for longer times. Such a study would miss the $u(t)$ peak at 5 s. Conversely, suppose the "true" retention function does have monotonically decreasing hazard, then it is important to observe the behavior of $u(t)$ near the value of 0 s. This reasoning leads to the conclusion that it is necessary to use a procedure that is suitable for examining short retention intervals. It is also important to prevent the participant from rehearsing the target item during the retention interval. If a participant is allowed to rehearse a to-be-remembered (TBR) item during the retention interval, then the length of the effective retention interval can be quite different from $t$. Fortunately, a short-term memory task such as the Brown–Peterson procedure (see Peterson & Peterson, 1959) is well suited to achieve these design requirements. The Brown–Peterson procedure involves a memory test for subspan target items. For a typical trial, a nonsense item like *BXT* is briefly presented, and is followed by material designed to keep the participant busy during the retention interval. For example, the participant might be required to "shadow" or repeat an auditory string of digits that is presented in the retention interval. In the last stage of a Brown–Peterson trial, the participant is tested for their knowledge of the TBR item.

Another design issue pertains to the measurement of memory. Ebbinghaus (1885/1964) used a saving measure to assess learning. Saving is defined as $1 - (T_r/T_i)$ where $T_i$ is the time to learn a list of items minus the time for recitation and $T_r$ is the time to later relearn the list minus the time for recitation. Although a saving measure is likely to be sensitive to target storage, it is also a problematic measure on at least two accounts. Ebbinghaus's own performance on list learning demonstrated a sizable effect of time of the day on his time to learn information. Time-based performance will be affected by shifts in the participant's phasic alertness during the day (see Van Zomeren & Brouwer, 1994). Furthermore, a list-learning paradigm is not well suited to test short time delays.

Another possible candidate for a performance measure is $d'$ from signal detection theory (SDT). However, $d'$ is not a probability measure, and it is probability that is needed to assess the memory hazard function. Moreover, Chechile (1978) delineated a number of problems caused by the application of SDT to memory research and developed a case for the use of multinomial processing tree (MPT) models instead. Clearly the topic of the relative merits of SDT and MPT models is worthy of further discussion, but that discussion is beyond the scope of the present article. The position argued here is that MPT models have been highly successful in measuring storage, and MPT models result in the type of probability measure that can link memory storage with hazard.

Batchelder and Riefer (1999) have provided an extensive review of MPT measurement models in psychology. The essential feature of these models is the explicit description of the categories of multinomial data in terms of latent psychological parameters. Chechile (1998) and Hu and Phillips (1999) have provided general methods for statistically estimating the latent psychological parameters of interest. A number of MPT models have been used to obtain separate measures for storage and retrieval processes (e.g., Batchelder & Riefer, 1980; Chechile, 2004; Chechile & Meyer, 1976; Chechile & Soraci, 1999; Rouder & Batchelder, 1998). In each case, the MPT model is closely linked to a rather specific experimental task that results in categorical observations. For example, Chechile and Meyer (1976) and Chechile (2004) developed MPT models that are designed for a task that is easily executed in conjunction with a short-term memory task like the Brown–Peterson procedure. The experimental task for these MPT models are subsequently referred to as the *Chechile–Meyer testing procedure*.

The Chechile–Meyer testing procedure is based on the random intermixing of recall and *yes–no* recognition test trials. The random intermixing is a design feature to ensure that participants do not process and retain information differently whenever they are tested with a recall probe as opposed to a recognition probe. The Chechile–Meyer testing procedure also requires the participants to provide a 3-point confidence rating for their recognition response, where 3 denotes a *certain response*, 2 denotes an *educated guess*, and 1 denotes a *blind guess*.

The model for the recall trials is provided in Figure 3. If the TBR item is stored insufficiently (i.e., either partial storage or no storage at all), then there will be incorrect recall due to a storage deficiency. The probability for this case is $1 - \theta_S$. Yet, if there is sufficient storage (with probability $\theta_S$) then correct recall still requires the successful retrieval of the TBR item at the time of test. The rate of successful retrieval of sufficiently stored targets is denoted as $\theta_r$. Note that this task is designed for TBR items that have a trivially small probability of correct recall without storage, that is, the TBR items are selected from a very large pool of potential items. Hence, correct recall requires both sufficient storage and successful retrieval, that is, the correct recall rate is equal to the product $\theta_S\theta_r$. All of the models developed in Chechile and Meyer (1976) and developed in Chechile (2004) have used the above MPT representation for the recall trials. However, there are differences among the models in regard to the MPT representation for the old and new recognition test trials.

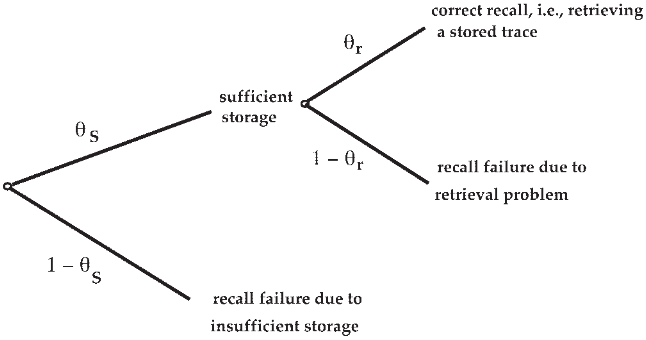


*Figure 3.* The storage and retrieval processes that underlie the recall task.

In this paper, Model 6P from Chechile (2004) is used for the recognition test trials. The 6P model for the old and new recognition test trials is shown in Figure 4. This model is different in two ways from the earlier Chechile and Meyer (1976) models. First, it ignores the difference between the two lowest levels of confidence ratings, that is, the 1 and 2 ratings. Second, it represents new recognition test trials (i.e., foil recognition trials) differently than the earlier MPT models.

Because of the random intermixing of test probes, the same storage probability is also applicable on old recognition trials. When there is sufficient storage, it is assumed that the participant can gain access to the stored encoding on old recognition tests and give the *yes* 3 response. Yet, when there is insufficient storage, the participant can also give the *yes* 3 response because of guessing and being overly confident. The other response possibilities in old recognition are delineated in the tree. These responses depend on the old recognition guessing parameter ($\theta_g$) and the parameter ($\theta_1$) for the use of the high-confidence rating without full target knowledge of the target. A similar process tree is provided in Figure 4 for the foil recognition trials. Note that a new parameter, $\theta_k$, is used for foil recognition trials. This parameter represents the proportion of test trials where the participant can recollect enough of the TBR item to reject the foil with high confidence. Note that there is a different guessing parameter, $\theta_{g'}$, used for foil recognition trials, but the same $\theta_1$ parameter is used as in the old recognition trials.

Chechile (2004) provided an extensive discussion of parameter estimation for the 6P model as well as a discussion of the evidence for model validation. Only a few remarks about model validation are provided here. For example, Chechile (2004) showed that when participants were given two additional seconds to recall the TBR item at the time of test, there was a significant improvement in the correct recall rate. It is important to note that only the retrieval parameter changed to account for this difference in the recall rate. Increasing the time permitted for a recall response cannot meaningfully result in an improvement in storage, but it did provide more time for the participant to retrieve the TBR item from memory. Hence, the results from that study strongly support the success of the model to factor the correct recall rate into separate memory components for storage and retrieval processes. Moreover, the retrieval measure did vary precisely in the fashion expected given the basic conceptualization of retrieval. Chechile (2004) provided additional validation evidence by identifying other independent variables that only influenced storage. Thus, there is support for the validity of the 6P model. In this paper, the test of the $u(t)$ function requires a measure of trace storage. The $\theta_S$ measure from the 6P model is used as the storage measure; hence, $u(t)$ will be taken to be $(1 - \theta_S)/t$.

## Experimental Tests of the Memory Hazard Function

Given the discussion in the previous section, there have been only a few experiments that have collected the requisite data for determining the peak of the $u(t)$ function in the short temporal domain. One such study is Experiment 2 from Chechile (1987). Although the original published analysis did not report the $u(t)$ function and used an older storage-retrieval measurement model, the data from that study can, nonetheless, be reexamined to assess the location of the maximum of the $u(t)$ function. Six retention intervals were studied in an experiment that used the Brown–

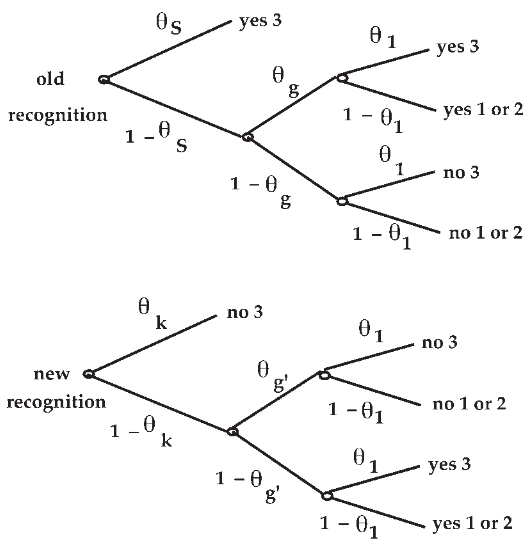


*Figure 4.* The multinomial processing tree model for the old and new recognition trials.

Peterson experimental paradigm. These intervals were filled with interpolated digits that the participants had to repeat (shadow) verbally. The digits were presented at the rate of one digit per .33 s. The retention intervals were filled with 1, 4, 12, 36, 95, or 225 interpolated digits. The memory targets were nonsense triads of letters. Instructions emphasized the avoidance of rehearsal and stressed the accuracy in shadowing the digits presented in the retention interval. Hence, rehearsal during the interpolated interval was unlikely. Furthermore, the Chechile–Meyer testing procedure was used, so the 6P model can be used to obtain a value for $u(t)$ for each retention interval. Additional details about the experiment are described in Chechile (1987).

The assessment of the peak of the $u(t)$ function can be done either on a group-averaged basis or on an individual basis. This point warrants further clarification. Later in this article, it is argued that tests of the retention function require the examination of individuals rather than retention functions based on group-averaged data. However, in this section we are interested in determining whether the memory hazard function is monotonically decreasing or is peaked. Theorem 1 establishes a means for assessing if memory has monotonically decreasing hazard. Chechile (2003) proved for probability models that a mixture of monotonically decreasing hazard functions is also a monotonically decreasing hazard function. In Part B of the Appendix, a more general proof is provided that a mixture of monotonically decreasing hazard functions is monotonically decreasing. The proof is more general because it applies to both probability and subprobability hazard models. The theorem is also more general because it applies for a mixture of components with constant hazard.

The averaging across participants is a mixture operation. If the individual memory hazard functions are monotonically decreasing, then the group average memory hazard function must also be a monotonically decreasing function and the peak of the $u(t)$ function will be at $t = 0$. Conversely, if the $u(t)$ function has a peak at a nonzero retention interval, then the memory hazard function cannot be monotonically decreasing. Enhanced statistical precision is achieved by performing the analysis on a group-averaged basis.

The $u(t)$ values for the group-averaged data from Chechile (1987, Exp. 2) are displayed in Figure 5. There is a clear peak shown at the 4-s retention interval, that is, the 12 interpolated digit condition. A 95% Bayesian probability interval for $u(t)$ at the .33-s retention interval is (.016, .074). A similarly computed 95% probability interval for $u(t)$ at the 4-s retention interval is (.087, .112). Given that these 95% probability intervals are nonoverlapping indicates that it is highly probable that $u(t)$ is larger in the 4-s condition relative to the .33-s condition, that is, the probability of an increase in $u(t)$ between the conditions exceeds .999. Hence, a sophisticated and powerful statistical assessment analysis of the group-averaged data results in evidence that is contrary to the prediction of memory models with monotonically decreasing hazard. It is also possible to test on an individual-participant basis that the $u(t)$ function is greater at the 4-s condition than at the .33-s condition, although this test has reduced statistical power. To do this test, the 6P model was used separately for each of the 30 participants in the study. The value for $u(t)$ at 4 s was compared with $u(t)$ at .33 s. A sign test is statistically significant and indicates that $u(t)$ is larger in the 4-s condition compared with the .33-s condition, $p < .021$. Thus, even this low power statistical test indicates that $u(t)$ has a peak for a nonzero time. Given this result and Theorem 1, it follows that the memory hazard function cannot be monotonically decreasing.

Experiment 1 from Chechile (1987) is another experiment that meets the aforementioned design constraints for assessing the peak of the $u$(t) function. This study also used the Brown-Peterson experimental paradigm along with the Chechile–Meyer testing procedure. In this experiment, the participants' task during the retention interval consisted of verbally shadowing a string of digits that filled the retention interval. The shadowing rate was either fast or slow. In the fast condition the interpolated digits were presented at the rate of one digit per .33 s, whereas in the slow condition, the digits were presented at the rate of one digit per .66 s. Furthermore, the participants were provided with a means to express any rehearsals that occurred while they were shadowing digits during the retention interval, that is, a button was provided that signaled to the experimenter the occurrence of a rehearsal. The data from any rehearsal trials was omitted. There were four retention intervals of 1.33 s, 2.67 s, 5.33 s, and 10.67 s. The group-averaged $u(t)$ values at each retention interval are displayed in Figure 6. The (.155, .175) interval and the (.107, .137) interval are the 95% Bayesian probability intervals for $u(t)$ at the 5.33-s retention condition for the respective fast and slow conditions. In both conditions, the entire 95% probability interval is larger than the mean $u(t)$ at the 1.33-s condition. Thus, the $u(t)$ function for this experiment, which used a very strict means to eliminate rehearsal of the TBR target, also demonstrated a nonzero peak.

## The $u(t)$ Peak and the Rehearsal Hypothesis

Given the finding of a peak for the $u(t)$ function, it is important to explore if this peak is due to any uncontrolled experimental factors. For example, one might hypothesize that the peak of the $u(t)$ function is a consequence of surreptitious rehearsal while the participants were shadowing the interpolated digits. Rehearsal here is understood to be the concurrent attention to both the interpolated

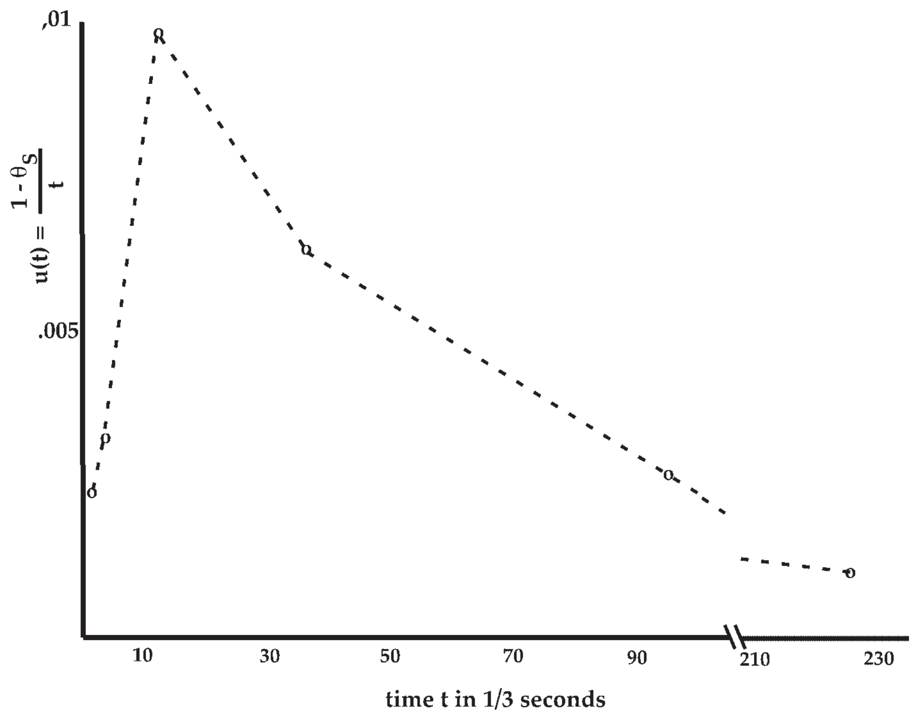


*Figure 5.* The group average $u(t) = (1 - \theta_S)/t$ as a function of time ($t$) for the data from Experiment 2 from Chechile (1987).

digit-shadowing task and the continuing conscious focus on the TBR target. If the participants were always able to concurrently rehearse while performing the interpolated task, then they would not have demonstrated any memory failures for the short retention intervals, which is contrary to the findings of the Chechile (1987) study. However, if one supposes that there were a mixture of rehearsal trials and nonrehearsal trials, then it might be theoretically possible to explain the delayed peak of the $u(t)$ function as a consequence of rehearsal during the short retention intervals. According to this rehearsal account, the $u(t)$ value for the short retention intervals is a mixture of some rehearsal trials where $u(t) = 0$ and of some nonrehearsal trials where $u(t)$ is very large—so large as to be consistent with the hypothesis of a monotonically decreasing $u(t)$ function. If $p_R$ denotes the proportion of rehearsal trials and $1 - p_R$ denotes the proportion of nonrehearsal trials, then this rehearsal account would predict that the observed $u(t)$ equals $(1 - p_R)\, u(t)_{NR}$ where $u(t)_{NR}$ has a large enough value to be consistent with a true $u(t)$ peak at $t = 0$.

In order to assess the likelihood of the rehearsal account for the peak of the $u(t)$ function, it is important to estimate the proportion of rehearsal trials required in order to support the hypothesis that the $u(t)$ function is monotonically decreasing. If only a small percentage of rehearsal trials can result in a delayed $u(t)$ peak, then the rehearsal hypothesis would be a viable explanation for the effect. However, if there must be a high rehearsal rate in order to be consistent with a monotonically decreasing $u(t)$ function, then the rehearsal hypothesis is less likely. The value for $u(t)$ that would be consistent with a monotonically decreasing hazard function is denoted as $u(t)_{NR}$ and it can be estimated by extrapolating from the post-peak slope of the $u(t)$ function. For example, the $u(t)$ peak for the Chechile (1987, Exp. 2) study occurred for the 4-s retention interval, that is, the 12 interpolated digit condition shown in Figure 5. The slope for $u(t)$ in the 4-s to 12-s range is $-.0043$, that is, $u(t)$ decreased at this rate as time increased from 4 s to 12 s. If that rate is used to extrapolate into the range of the two short retention intervals of .33 s and 1.33 s, then there would be estimates for $u(t)_{NR}$ that can be used to assess the rehearsal mixture model. The resulting extrapolation estimates for $u(t)_{NR}$ are .1138 and .1094 for the respective .33-s and 1.33-s conditions. The corresponding actual values for $u(t)$ are .0244 and .0316 for the respective .33-s and 1.33-s conditions. Consequently, the rehearsal mixture hypothesis would need to postulate that the rehearsal rate was 79% and 71% for the respective .33-s and 1.33s retention intervals in order to account for the peak-shaped $u(t)$ function. Although it is theoretically possible that the participants were surreptitiously rehearsing at a high rate for these two short retention intervals, it does not seem likely because the rate of shadowing for the auditory stream of digits was quite rapid and the instructions clearly directed the participants to shadow accurately the digits and not to rehearse the TBR item.

In the above analysis of the presumed rehearsal rate to account for the peaked $u(t)$ function as a consequence of a mixture of rehearsal and nonrehearsal trials, the post-peak slope of the $u(t)$ function was used to extrapolate $u(t)_{NR}$ values for the prepeak retention intervals. However, suppose instead of extrapolating one simply stipulates that the prepeak $u(t)_{NR}$ values are at least as large as the peak value. With this weaker assumption an even more conservative estimate of the presumed rehearsal rate can be calculated. This conservative method results in the presumed re-

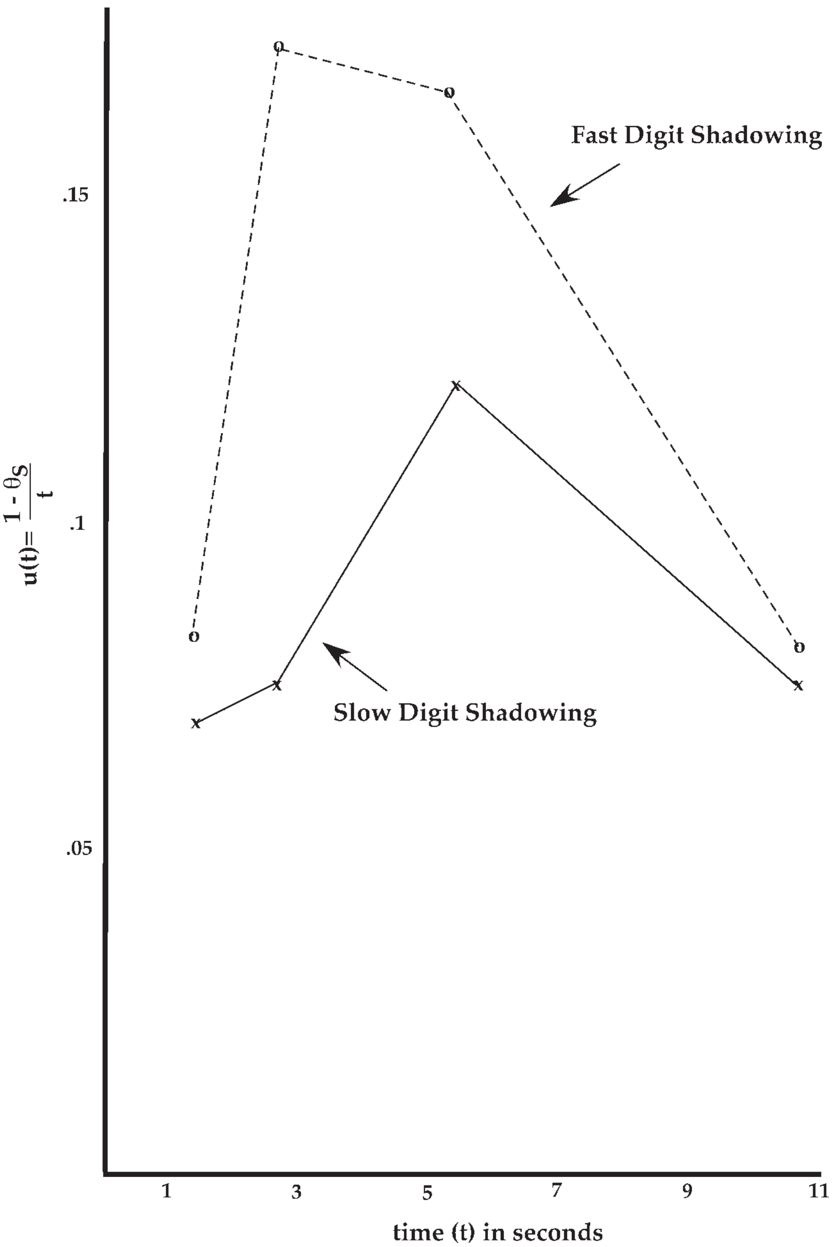


*Figure 6.* The group average $u(t)$ as a function of time ($t$) for fast and slow digit shadowing for the data from Experiment 1 from Chechile (1987).

hearsal rates of 75% and 68% for the respective .33-s and 1.33-s retention intervals. Furthermore, when this same conservative analysis is done for the data from Chechile (1987, Exp. 1; shown in Figure 6), the presumed rehearsal rates for the 1.33-s retention interval for the respective fast and slow conditions are 52% and 44%. Yet, in Chechile (1987, Exp. 1), the participants could indicate when they did rehearse, and those admitted rehearsal trials were already removed. The participants were not reluctant to exercise this option of indicating a rehearsal because approximately 5.5% of the trials were culled because of rehearsal. Consequently, the rehearsal mixture hypothesis has a serious problem because the presumed rehearsal rate required to account for the peaked $u(t)$ function is unreasonably large in the context of the experimental controls implemented in the Chechile (1987, Exp. 1) study.

## Implication of the $u(t)$ Peak Studies

Given the strong evidence that the $u(t)$ function has a peak for some nonzero time and the difficulty to account for this finding in terms of the rehearsal hypothesis, it is clear that six of the models shown in Table 2 are problematic because they are based on a monotonically decreasing hazard model. More specifically, the PM–hyperbolic, PM–exponential, modified logarithm, modified

PM–power, PM–STFT, and MSM models are not consistent with the experimental findings because these retention functions do not have the correct memory hazard profile. Furthermore, any memory theory that does not result in a peaked-shaped hazard function will similarly have a problem because a nonzero peak, for $u(t)$ implies that the memory hazard function is not monotonically decreasing for all retention intervals.

One might argue that some of the functions that result in monotonically decreasing hazard are not “wrong” because the functions were not intended for short retention intervals. Yet it is rare to find stipulated restrictions on the time scale for proposed functions, and in some cases the investigator was definitely interested in describing short as well as the long retention intervals (e.g., Wickelgren, 1974). Moreover, it is not clear why one should limit a retention function to exclude short retention intervals. The challenge here is to find a general retention function that works over both short and long time frames. Such a retention model should predict a peaked-shaped hazard function. Of the functions originally listed in Table 1, only three meet the constraints of allowing for long-term memory and having a peaked-shaped hazard function, that is, the Ebbinghaus function, the modification of the Anderson and Schooler function with $c > 1$, and the TST function.

## Goodness-of-Fit Constraints on Possible Candidate Functions

A further narrowing of the candidate functions requires additional assessment criteria. One key criterion is the relative goodness-of-fit of the three models. To conduct this assessment, the Ebbinghaus function, the PM–AS–power function, and the TST function were each fit to the data from Chechile (1987, Exp. 2). That study was selected because there are ample data for each participant to obtain fairly accurate storage estimates at each of six retention intervals. It is essential that the retention function be examined on an individual-participant basis. This point has been stressed by a number of other investigators (e.g., Anderson & Tweney, 1997; Estes, 1956; Myung, Kim, & Pitt, 2000; Sidman, 1952; Underwood, 1949). From the perspective of peak-shaped hazard functions, it is easy to see why an average across participants can distort the retention function. If the location and amplitude of the hazard peaks varies across individuals, then the hazard function for the group-average data might not reflect any of the individuals.

One potential problem with goodness-of-fit is the difficulty of comparing fits that vary in the number of free parameters. However, the Ebbinghaus, the PM–AS–power, and the TST function all have two parameters. Consequently, the relative fit of these models does not confound the number of model parameters with the quality of the fit. The measure used to assess the models is the $r^2$ value between the model-predicted $\theta_S$ values and the $\theta_S$ values obtained from the 6P measurement model. In general, the $r^2$ measure can be a misleading model comparison statistic when one is working with nonlinear models that differ in the number of adjustment parameters. However, the $r^2$ measure used here is between the predicted and observed $\theta_S$ values, and on that basis the relationship should be linear. Although more sophisticated model comparison statistics exist (see Myung, Forster, & Browne, 2000), the $r^2$ statistic is a reasonable model comparison measure for the present application.

For each of the 30 participants, the PM–AS–power model was not a good fit, that is, for each participant either the Ebbinghaus function or the TST model was superior. For 22 participants, the best fit was for the TST function, and for the other 8 participants, the Ebbinghaus function was the best fit. A sign test indicates that the TST function is better than the Ebbinghaus function, $p < .019$. The mean $r^2$ values for the TST, Ebbinghaus, and PM–AS–power models are .938, .905, and .846, respectively. Thus, in terms of the goodness-of-fit criterion, the TST function is clearly superior to the other two surviving candidate retention functions.

An additional criterion for assessing the candidate functions should be discussed. This criterion deals with the principle that a change in measurement units (or scale) should not affect the functional form.[1] For example, the TST function can be expressed in terms of various time scales and the mathematical form of the function is not changed. As an illustration, consider the TST model where $a = .25$ and time is expressed in units of seconds, that is, $1 - b + b\exp(-.25t^2)$. If time is reexpressed in units of .1 s, then the predictions of the model are the same by simply setting $a = .0025$. Thus, the TST formulation is insensitive to the units of measurement. However, the Ebbinghaus function is altered whenever there is a change in the temporal measurement units. Consequently, there is an additional rationale for rejecting the Ebbinghaus function. This problem also occurs for the modified logarithm, the modified PM–power, and the PM–AS–power functions. However, in all of these cases the scale-sensitive function can be revised to be scale invariant by replacing the term $(t)$ with $(dt)$, where $d$ is an additional parameter.

Although the TST model holds up well and is the sole survivor from the original set of 15 functions, it can be shown that this function has other problems. These difficulties will be explored in the next section.

## Problems With the TST Model as a General Retention Function Model

The Chechile (1987) TST is based upon a Weibull subprobability model for the distribution of trace lifetimes. The Weibull probability distribution has been extensively used in product reliability modeling. In its general form, the distribution has a shape parameter ($c$) and a scale factor ($a$). The cumulative distribution is given as

$$F(t) = 1 - \exp(-at^c)$$

Because this distribution is a probability model, the cumulative distribution approaches 1.0 as the time becomes large. It is well known (see Chechile, 2003) that this distribution has monotonically increasing hazard when the shape parameter $c$ is greater than 1.0 and has monotonically decreasing hazard for $c < 1$. When $c = 1$, the Weibull distribution is equal to the exponential distribution and has constant hazard. In fact, one reason why this distribution is important in industrial product reliability is the flexibility of the distribution to model a wide range of hazard environments.

[1] I thank Jean-Claude Falmagne for pointing out the problem of scale sensitivity.

The TST function is based on a Weibull subprobability model as opposed to a Weibull probability model. As with the previously discussed difference between the exponential probability model and the exponential subprobability model, the subprobability densities are *b* times the usual Weibull densities, where *b* is a parameter that is less than 1.0. With the TST function, the shape parameter *c* is set equal to 2.0. The resulting cumulative distribution, which is equivalent to $1 - \theta_S(t)$, is given as

$$F(t) = 1 - \theta_S(t) = b[1 - \exp(-at^2)].$$

In general, a Weibull subprobability model has peaked hazard when the shape parameter is greater than 1.0 and has monotonically decreasing hazard when the shape parameter is less than or equal to 1.0. Since the shape parameter is fixed to a value greater than 1.0, the TST formulation results in predicting a peaked hazard profile.

Although there is an excellent fit of this model to short-term memory data, the value of 2.0 for the shape parameter does not result in good fits for other data sets where memory is examined over a longer time frame. In general, there are only a few suitable data sets for the assessment of memory retention that enable the examination on an individual-participant basis and over a wide temporal range. In most cases, the memory investigator has regrettably grouped over participants and over items of differing levels of complexity. Such mixtures make it impossible to assess nonlinear retention functions because the model for the mixture might not represent any of the participants. Fortunately, however, there are several excellent data sets that can be used to assess the TST function over a longer temporal range. These experiments also provide ample evidence against the application of the TST formulation for longer retention intervals.

One such experiment is the Strong (1913) study. That study involved the extensive testing of 2 individuals. There were also data for several other participants, but the records were incomplete for those participants. Hence, only the individuals with the complete records were reanalyzed. For a typical test trial in the Strong experiment, the participants were shown a list of 20 words, and later the participants were provided with a test list of 40 words. The participants picked the words that they believed were on the original study list, and they also indicated their degree of confidence in the recognition decision. Strong examined 13 retention intervals over a temporal range of 15 s to 7 days. Hence, the Strong data set involved retention intervals that varied over five orders of temporal magnitude.

It is possible to construct an MPT model for the data collected in the Strong (1913) study. The MPT model is portrayed in Figure 7 in terms of a storage parameter and a response bias parameter, respectively $\theta_S$ and $\theta_x$. According to the MPT model, if there was sufficient storage of a TBR item, then the participant would have picked that item from the test list with high confidence. Without sufficient storage, the participant could still have responded correctly and used (with a probability of $\theta_x$) the high-confidence rating. An estimate of $\theta_x$ can be taken from the proportion of nonstudied words selected and rated with high confidence. The resulting estimates for $\theta_S$ are shown in Table 3.

Transformations were done to create a linear structure in order to examine whether a Weibull subprobability model is appropriate for the Strong (1913) data. If the *x* variable is taken to be log (*t*) and the *y* variable is defined as $\log[\log(b/(b - 1 + \theta_S))]$, then a

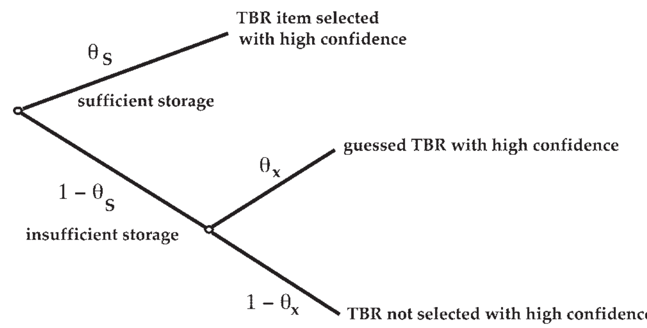


*Figure 7.* The multinomial processing tree model for the Strong (1913) study.

Weibull subprobability lifetime model should result in a linear plot. That is, if

$$\theta_S(t) = 1 - b + b\exp(-at^c),$$

then the above *x* and *y* transformations should produce a straight line. Moreover, the slope of the line is an estimate of the Weibull subprobability shape parameter, *c*. These linear plots are displayed in Figure 8 for Participants A and B from the Strong study. The linear fit is excellent for both participants, but the slope estimates for the *c* parameter are .29 and .23 for Participants A and B, respectively. These values are clearly inconsistent with the TST formulation in which the shape parameter *c* is set to 2.0.

Another data set to examine is the Ebbinghaus (1885/1964) retention experiment. Ebbinghaus reported the average saving measure for relearning a list of items after delays that ranged from 19 min to 31 days. Thus, the temporal delays ranged over more than four orders of magnitude. Although a memory saving measure is likely to be sensitive to both whole item storage as well as partial storage, it will, nonetheless, be taken as a measure of sufficient storage for the purpose of estimating the Weibull subprobability shape parameter. The single subject tested was Ebbinghaus himself.

The same transforms as described above for the Strong (1913) study are implemented for the Ebbinghaus retention data. The resulting graph is displayed in Figure 9. The transformed data are well described by a straight line. For these data the *b* parameter is estimated as .85, and the shape parameter *c* is .164. As with the analysis of the Strong (1913) data, it is clear that the predicted value for the shape parameter according to the TST formulation is not supported.

The Strong (1913) and Ebbinghaus (1885/1964) data sets provide powerful evidence that the TST retention function fails because of the value of the Weibull subprobability shape parameter. The TST formulation sets the shape parameter to a constant value of 2.0. In the short temporal domain, the TST retention function with this value for the shape parameter is an excellent description. Moreover, when the shape parameter is greater than 1.0, then a Weibull subprobability model for memory lifetimes has the property of having the desired peaked-shaped hazard profile. A single memory trace cannot simultaneously have the Weibull subprobability shape parameter less than 1.0 and greater than 1.0. If one uses a single-trace system with a Weibull subprobability shape parameter that is less than 1.0, then one cannot account for the

Table 3
*Storage Estimates Obtained for Participants A and B from the Strong (1913) Study*

| Delay (min) | Storage estimate for A | Storage estimate for B |
|---|---|---|
| .25 | .920 | .914 |
| 5 | .727 | .806 |
| 15 | .758 | .743 |
| 30 | .663 | .696 |
| 60 | .677 | .642 |
| 120 | .612 | .581 |
| 240 | .606 | .515 |
| 480 | .515 | .445 |
| 720 | .463 | .402 |
| 1,440 | .219 | .329 |
| 2,880 | .210 | .257 |
| 5,760 | .186 | .191 |
| 10,080 | .134 | .144 |

short-term retention data and the peaked shaped memory hazard function, but one would be able to accurately describe memory retention over a long time frame. Clearly there is a need to consider a multiple-trace model.

There is another reason to believe that there is more than a single post-perceptual memory trace as posited in the TST. Usually time and the number of intervening items are perfectly correlated in retention studies because the interpolated items occur at a stable rate. However, time and items can be decoupled if the interpolated information is presented at different rates. Chechile (1987, Exp. 1) examined the role of items and time in the context of the Chechile–Meyer task. The participants in that study shadowed interpolated digits during the retention interval of a Brown–Peterson experiment. The digits were presented either at a fast rate of one digit per .33 s or at a slow rate of one digit per .66 s. The other experimental factor was the temporal duration of the retention interval, that is, 1.33 s, 2.67 s, 5.33 s, or 10.67 s. See Table 4 for the storage values for each combination of retention interval duration and number of interpolated digits. The Model 6P analysis results in a Bayesian posterior probability distribution for the storage parameter for each condition. From these distributions, it is possible to compute directly the probability for various hypotheses about condition differences (see Chechile, 1998, 2004). Apart from a ceiling effect for the two 1.33-s conditions, there is clear effect for items given fixed time. For example, for fixed time equal to 2.67 s, it is highly probable that there is an effect for items (i.e., the probability of a difference is .9964). Similarly, for the two conditions where $t \geq 5.33$ s, it is highly probable that there is an effect for items (i.e., the probability of a difference is .9998). However, when $n$ is fixed at 16 interpolated items, time did not result in a decrease of performance; if anything the storage is higher for the longer duration condition. Hence, for $t \geq 5.33$ s, there is only support for the number of items as the factor influencing memory retention. However, for $t < 5.33$ s there is support for the role of both time and the number of interpolated items. This point is illustrated by the difference between the two conditions where $n = 8$ (i.e., the .538

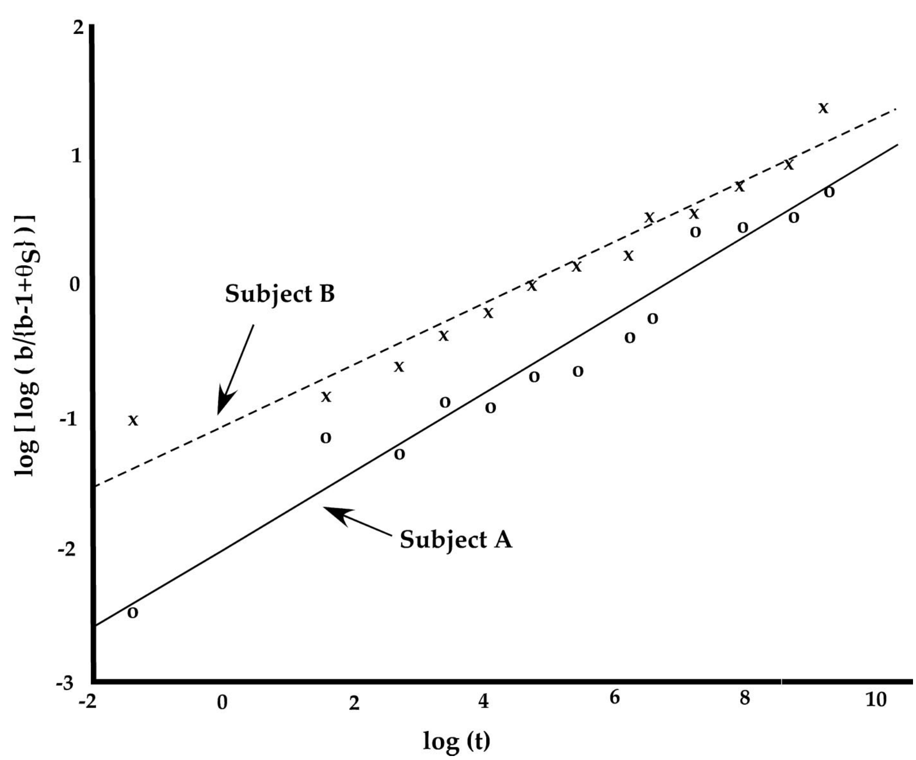


*Figure 8.* A plot of transformed storage estimates from the Strong (1913) study for Participants A and B versus the logarithm of the time of the retention interval. The ordinate is $\log[\log\{b/(b - 1 + \theta_S)\}]$. Plots should be linear if the retention functions are consistent with a Weibull subprobability model for memory trace lifetimes. The slope of the plot is an estimate of the shape parameter for the Weibull subprobability.

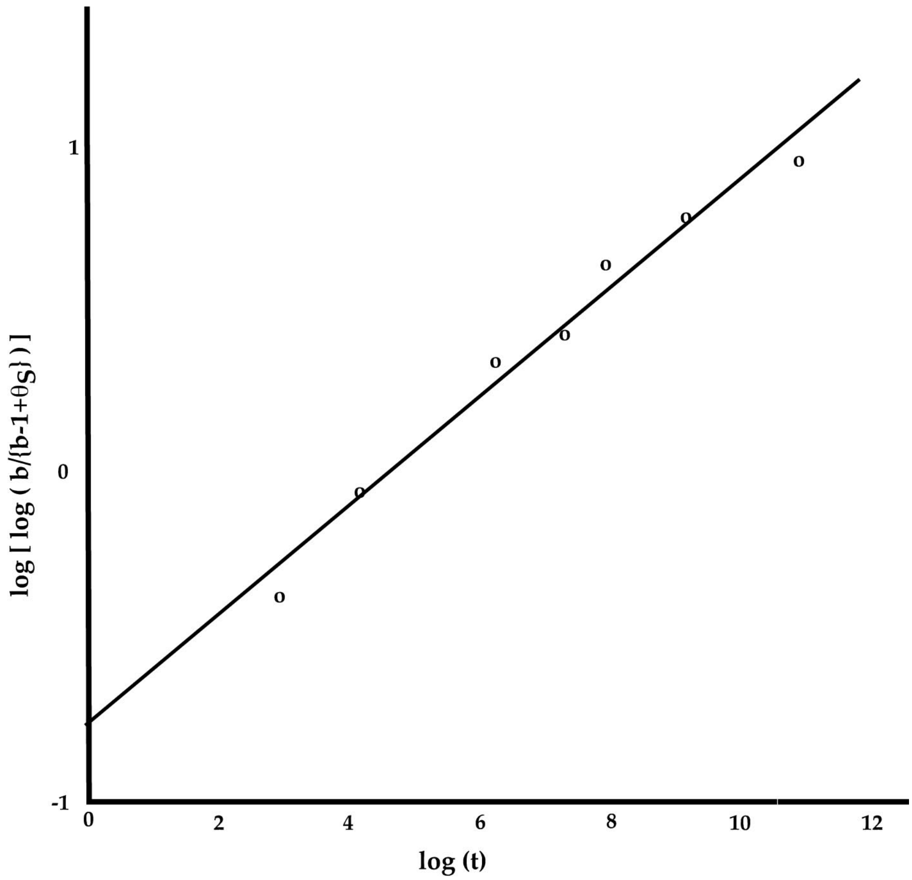


*Figure 9.* A plot of the transformed data from the Ebbinghaus (1885/1964) study of memory retention. The $x$-axis is log ($t$) and the $y$-axis is $\log[\log\{b/(b - 1 + \theta_S)\}]$ where $\theta_S$ is taken to be the average savings value. The graph should be linear if retention is characterized by a Weibull subprobability model for memory lifetimes. The slope of the graph is an estimate of the shape parameter for the Weibull subprobability distribution.

value for storage for the fast rate and the .345 mean for storage for the slow rate). When the time of the retention interval is short (i.e., $t = 2.67$ s and $n = 8$) there is more storage in comparison to the case with a longer retention interval (i.e., $t = 5.33$ s and $n = 8$). These results are problematic for a single memory trace model. If there were a single trace, how can time be an important variable in only the short time frame whereas the number of items is important in both the short and long time frames?

Table 4
*Storage $\theta_S$ as a Function of Items n and Time t (From Chechile, 1987, Experiment 1)*

| Time ($t$) in seconds | Interpolated digits ($n$) | Storage estimate |
|---|---|---|
| 1.33 | 4 | .890 |
| 1.33 | 2 | .908 |
| 2.67 | 8 | .538 |
| 2.67 | 4 | .801 |
| 5.33 | 16 | .114 |
| 5.33 | 8 | .345 |
| 10.67 | 32 | .136 |
| 10.67 | 16 | .206 |

What is needed is a retention function that has the property of a peaked hazard function as well as having a shape parameter consistent with both the studies of long-term and short-term retention. Such a model is described in the next section.

## The Two-Trace Hazard Model

Consider a system of two independent traces where both traces can be destroyed. This two-trace system is markedly different from the classical multiple-store model as exemplified by the Atkinson and Shiffrin (1968) model. The long-term store in the MSM has zero risk of a storage loss, that is, the long-term store is a permanent memory store. For the two-trace hazard model that is advanced here, both traces are susceptible to storage loss.

Trace 1 is based on a Weibull probability distribution with a shape parameter of 2.0 for the trace lifetime model. Hence, for this representation of the TBR item, the cumulative distribution is

$$F_1(t) = 1 - \exp(-dt^2). \quad (8)$$

This first trace has a monotonically increasing hazard rate (see Chechile, 2003). In fact, the hazard is linearly increasing with $t$, that is, the hazard is $2dt$. Because of the increasing hazard rate, it is unlikely that this representation of the TBR item will be available for longer retention intervals. Trace 1 cannot represent long-term memory retention.

Trace 2 is based on a Weibull subprobability distribution with a shape parameter $c$ that is less than 1.0. For this trace, the cumulative distribution is

$$F_2(t) = b[1 - \exp(-at^c)]. \quad (9)$$

Because the Weibull subprobability shape parameter is less than 1.0, this trace has monotonically decreasing hazard. If a memory survives long after the original encoding, then it will be because this trace is still available.

The storage of a TBR item is lost when both traces are lost, that is,

$$F(t) = 1 - \theta_S(t) = F_1(t)F_2(t), \quad (10)$$

where $F_1(t)$ and $F_2(t)$ are specified in Equations 8 and 9, respectively. In Part C of the Appendix, it is proved that the above system of two independent traces has a peaked hazard function. This result is particularly noteworthy given that neither trace has a peaked hazard function. However, because the memory of a target is only lost when both traces are lost, the system of traces for the TBR item has peaked-shaped hazard. It is also possible that some targets survive over the entire lifetime of the individual.

Some of the features of the two-trace hazard model are illustrated in Figure 10. Given the parameter values of $a = .25$, $b = .92$, $c = .5$, and $d = .002$, Figure 10A is a display of the corresponding hazard for (a) Trace 1 alone, (b) Trace 2 alone, and (c) the system of two traces. Note that the hazard system has a peak, despite the fact that neither of the component traces has peaked hazard. In Figure 10B, the corresponding lifetime density functions are illustrated.

It is important to clarify that neither of the traces in this two-trace hazard model deal with the phenomenon of perceptual persistence. There is a clear consensus that there is a very brief sensory persistence of visual and auditory information. For visual targets, the sensory persistence is called the *iconic image* (Averbach & Coriell, 1961; Spencer & Shuntich, 1970; Sperling, 1960). Yet this perceptual persistence can be eliminated if the target is masked. In fact, masking studies demonstrated that the duration of the iconic image is about 200 ms (Spencer & Shuntich, 1970). It is also clear that there is a persistence of auditory targets, that is, the *echoic trace* (Efron, 1970; Massaro, 1970). Although there is some debate as to the duration of the echoic trace, masking studies have indicated that the duration is about 250 ms (Massaro, 1975). Because most memory retention studies provide a masking stimulus immediately after the presentation of the TBR item, it is reasonable to ignore the very brief perceptual persistence of the target. Both of the traces in the above two-trace hazard model are regarded as independent post-perceptual representations of the TBR item.

Before exploring the properties of the two-trace hazard model, another simplification should be discussed. As noted above, for most experiments the time of retention interval is nearly perfectly correlated with the number of events that occur in the interval. Mathematically, however, time is a continuous variable and the number of events is a discrete variable. Generally it is more convenient to work with a continuous measure rather than with a discrete measure. Furthermore, one cannot be sure as to the value of the number of interpolated events, especially for studies with very long retention intervals. Yet information about time is readily available. Consequently, time is used as the critical variable rather than the number of interpolated events.

Of course when the retention interval is filled with events that occur at a more rapid rate in comparison to other retention intervals, then it would be expected to result in a higher rate of storage loss. Yet this difference in the rate of interpolated events can be accounted for by the model parameters, that is, the $a$ and $d$ parameters should increase along with the rate of interpolated events. Also, if the interpolated events are similar but different than the TBR item, then storage interference is expected. The magnitude of the interference can vary as a function of the density of similar intervening items. In this case, the $a$, $b$, and $d$ parameters are expected to capture the differences in the type of interfering events. Thus, the general functional form of the two-trace hazard model can handle the different levels of interference by means of the values of the aforementioned parameters. Consequently, time is used as the single predictor of storage lost, but in reality the vigor of the storage interference is captured by the fitting parameters.

## Fits of Two-Trace Hazard Model

How well does the two-trace hazard model fit experimental data for both the long and short time frames? First, how does the two-trace hazard model handle memory studies with long retention intervals? For example, the shortest retention interval for the classic Ebbinghaus (1885/1964) study was 19 min. It is highly probable that Trace 1 would have failed long before 19 min. Consequently, for the Ebbinghaus data the two-trace hazard model becomes effectively a one-trace model, that is, Trace 2. Note that it was already demonstrated that a Weibull subprobability model with a shape parameter less than 1.0 is an excellent description for the Ebbinghaus data (see Figure 9).

Another example of a study with long retention intervals is the Strong (1913) experiment. This study involved the learning of lists; therefore, the length of the retention interval is not precisely known. What is known is the time between the end of list presentation and the beginning of testing. The shortest duration for this delay was 15 s for the Strong experiment. However, the lag between item presentation and testing was considerably longer than 15 s. Thus, for the Strong (1913) experiment, it is likely that the two-trace hazard model would be effectively a one-trace model, that is, Trace 2. It was previously demonstrated that a Weibull subprobability model with a shape parameter less than 1.0 was an excellent description for the Strong data (see Figure 8). Consequently, the two-trace hazard model can account for the changes in retention found in long-term studies.

To examine how well the two-trace hazard model fits short-term memory data, let us examine the fit for the Chechile (1987, Exp 2) data. Recall that this data set was highlighted earlier when the relative fit of the TST, Ebbinghaus, and the AS–PM–power functions was examined. This data set has six retention intervals and has sufficient sample size to enable reliable storage estimates for each retention interval on an individual-participant basis.

A search algorithm was used to locate a parameter fit of the two-trace hazard model for each participant. For 29 of the 30

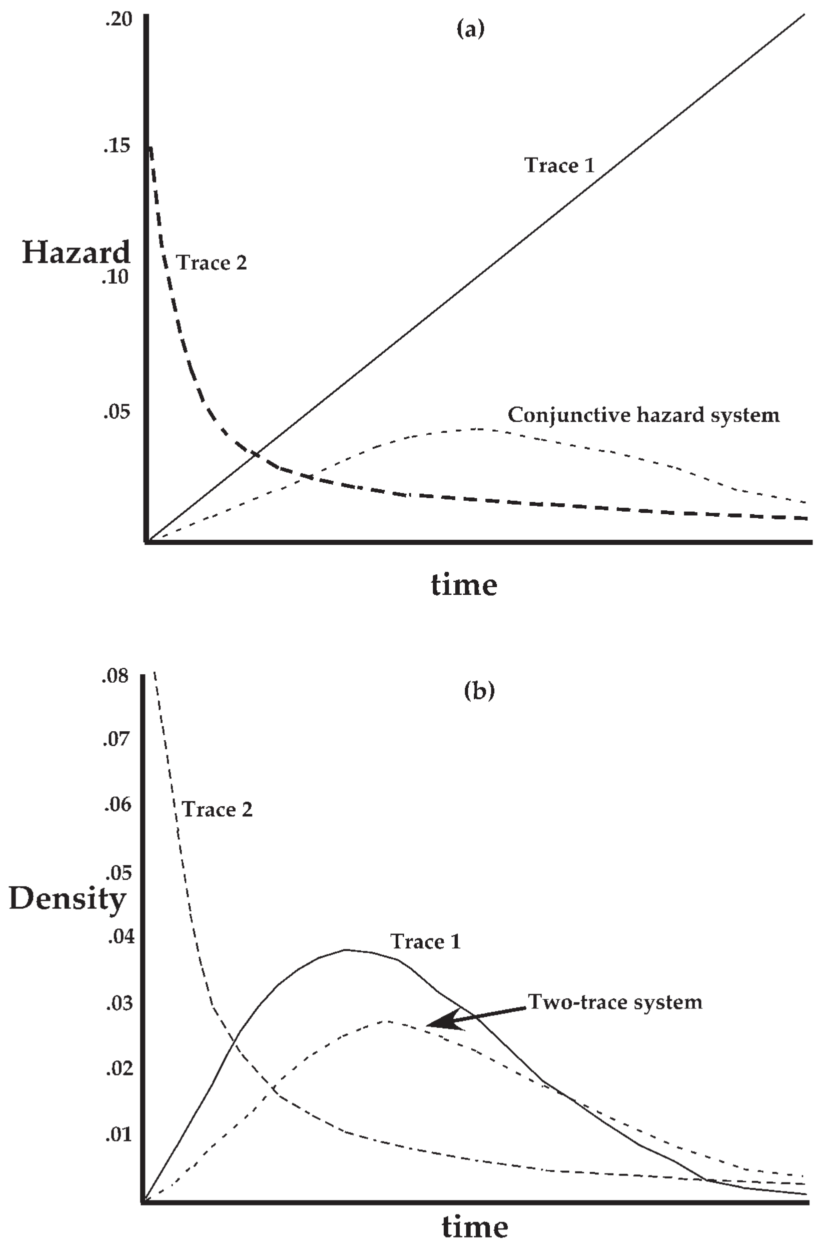


*Figure 10.* (a) Plots of the hazard functions for Trace 1, Trace 2, and the two-trace system where the parameter values of the model are $a = .25$, $b = .92$, $c = .5$, and $d = .002$, and (b) illustrates the corresponding lifetime distributions for Trace 1, Trace 2, and the system of two traces.

participants, the $r^2$ between the fitted and estimated $\theta_S$ values were higher than any of the other functions, that is, the TST, Ebbinghaus, and PM–AS–power functions. The data for 1 participant were still better fit with the TST function; however, no effort was made to expand the range of the parameter search algorithm for the two-trace hazard model so as to improve the fit. Also the $r^2$ value for that participant was .92, which is still a reasonable fit. The respective median and mean $r^2$ values for the full set of 30 participants are .996 and .973. Recall that the TST function had respective median and mean $r^2$ values of .980 and .93. Clearly, the two-trace hazard model does a good job of fitting short-term retention changes, although the model does have more adjustment parameters.

## Within-Participant Parameter Variability

It is important to see how the two-trace hazard model can describe the effect of trial-by-trial variation in the model fitting parameters. The two-trace hazard model is a stochastic model with a number of fitting parameters. Though care has been taken to fit the model to individuals rather than to group averages, there can still be variability across trials. Some items might vary in encoding quality. Also, the interpolated interference might vary in vigor across trials. How does trial-by-trial parameter variability influence the model?

First consider the rate coefficient *d* for Trace 1. Suppose there is variation in this parameter across trials for a given participant. The hazard function for Trace 1 is a linear function of time with the slope being 2*d*. The mean of a linear function is also a linear function. It thus follows that *d* is one half of the mean slope of the trial-by-trial Trace 1 hazard functions. Hence, the best fitting *d* parameter can be understood in terms of this statistical property of the set of trial-by-trial *d* values.

Next consider the *b* parameter in terms of the variability issue. Since the *b* parameter is the proportion of the traces that are eventually lost, this parameter is already regarded as a statistical property of a collection of trials. Slight random variation in this parameter can be ignored and modeled with a fixed value. However, a shift in the *b* parameter can occur whenever there is a systematic change in either the encoding quality or the vigor of the postencoding interference. Consequently, the *b* parameter for a given condition is regarded as a measure of the overall encoding quality and the postencoding storage interference.

Instead of the above interpretation of the *b* parameter, one might posit that the initial encoding of memory targets results in an extreme mixture of items, that is, there are two extreme states of susceptibility to storage loss. In one state all the items are lost, whereas in the other state none of the items are lost. The proportion of items in the susceptibility state is b, and $1 - b$ is the proportion of the items in the permanent state. This extreme-state mixture interpretation of the *b* parameter can result in the same equation for the retention function as used in the two-trace hazard model. But the extreme-state mixture hypothesis is testable. Although it is possible to cause *b* to change by varying the strength of the initial encoding, it is also possible that a homogeneous set of encoded items can be subjected to different levels of storage interference. From the two-trace hazard model perspective, it is expected that more memory traces will be lost if item encoding is followed with an environment that is designed to produce more storage interference in comparison to the standard set of interpolated events. Müller and Pilzecker (1900) did find greater forgetting when highly interfering material was presented early in the retention interval as opposed to later in the interval. Wixted (2004) discussed additional experimental findings that demonstrated greater forgetting when interference occurs early in the retention interval. Findings of this type pose a problem for the extreme-state mixture hypothesis. Moreover, the extreme-state mixture interpretation is a form of the MSM, which will be critiqued later in this paper. Rather than the extreme-state mixture account, the *b* parameter is treated as a value applied to all target items, and the value of *b* reflects the overall encoding quality and the postencoding storage interference.

Finally, let us consider the variability in the *a* and *c* parameters. These parameters affect the rate of memory loss. The variability of these parameters is illustrated in Figure 11. Clearly any trial-by-trial variation in these parameters will contribute to “error.” Yet, the variability of these parameters is a function of time. For short time values it is difficult to discriminate the retention functions in Figure 11. Also, for very long times, the retention functions approach the asymptote of 1-b and are less affected by the variation in the *a* and *c* parameters. In the temporal range illustrated in Figure 11, parameter variability influences the rate of change of the retention function. Yet, the best fitting values for the *a* and *c* parameters still represent the central profile for the rate of memory

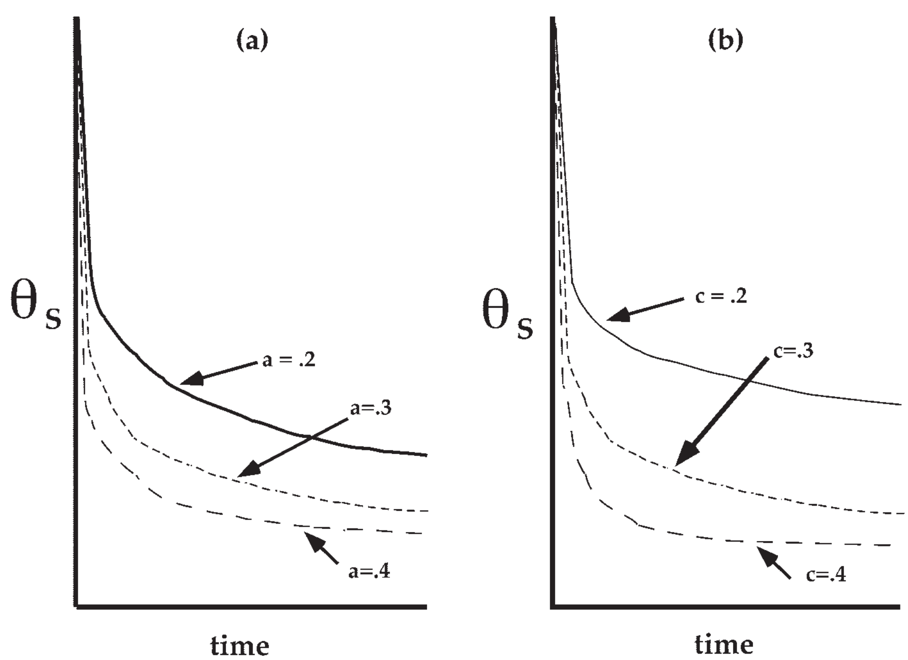


*Figure 11.* (a) Plots of the storage retention for various values for the *a* parameter when $b = .9$, $d = .002$, and $c = .3$, and (b) illustrates storage retention plots for various values for the *c* parameter when $b = .9$, $d = .002$, and $a = .3$.

loss for the aggregated data for a given participant. Hence, all the parameters in the two-trace hazard model are regarded as overall measures that apply to all the trials for a given participant in a given condition.

## Differences Between the Two-Trace Model and the MSM

The two-trace hazard model might appear to be similar to the classic multiple memory-store model as advanced by Atkinson and Shiffrin (1968). Both models can account for some permanent memory and account for the rapid loss of other items. However, the models are quite different. Although the MSM has two post perceptual systems, it does not represent all items as having two representations as does the two-trace hazard model. In the MSM, only after an item is "transferred" from the short-term store to the long-term store would there be a long-term representation. Items that were never transferred to the long-term store would only have been represented in the short-term system; all of these items would be lost. Furthermore, once an item is deposited into the long-term store, then no further storage degradation would occur for the TBR item. On reflecting about the presumed permanence of the long-term store of the Atkinson and Shiffrin model, Shiffrin (1999) pointed out that the search of associative memory (SAM) model (Raaijmakers & Shiffrin, 1980, 1981) represents a more advanced account of long-term forgetting. Shiffrin (1999) also stated, "In the SAM model, then, forgetting occurs when retrieval does not succeed" (p. 27). Hence, the idea that all items that are initially transferred to the long-term store are available to be remembered remained a central premise of the MSM. Yet it has already been shown that long-term storage changes have been found (see Figures 8 and 9). With the two-trace hazard model, no items are assumed to be permanently stored initially; rather "permanent memory" emerges as a statistical property of some traces to survive an entire lifetime memory hazard.

Moreover, data from another experiment can clearly demonstrate storage changes for items that should have been transferred to the long-term store (if the MSM were correct). Chechile and Ehrensbeck (1983) used a paired-associate task that provided a means to examine whether there are changes in storage for items that should be in the presumed long-term store. The participants were required to shadow digits for 20 s between the study and test phase of each trial of the paired-associate task. Given this long shadowing period before the test phase of the trial, the information can only be remembered from the long-term store according to the MSM. Training continued in this fashion until the participants were correct for 87% of the test items. The data collected in this experiment can be analyzed by means of the Chechile (2004) 6P model that is discussed above. Chechile and Ehrensbeck (1983) used for analysis an older MPT model that has been eclipsed by the 6P model. The resulting mean storage probability ($\theta_S$) was found to be .870 for the last training trial. According to the MSM, this storage probability would have to represent the average proportion of the list that has been transferred to the long-term store. Chechile and Ehrensbeck (1983) also required the participants to learn an entirely new paired-associate list. Training on the second list continued for 14.5 minutes. Finally the first list was tested again so as to obtain an estimate of $\theta_S$ after the 14.5-min retention interval. The 6P analysis of the final test results in a mean storage probability of .645. This value represents a statistically reliable decrease in the probability of storage. This decline in storage underscores a key problem with the MSM, that is, there are storage losses from the presumed permanent memory store.[2]

The two-trace hazard model and the MSM also differ in regard to the shape of the memory hazard function. As shown in Part A of the Appendix, the MSM predicts a monotonically decreasing hazard function; whereas, the two-trace hazard model results in a peaked-shaped hazard function (see Part C of the Appendix and Figure 10 a). Since it is now established that the memory hazard function is peaked shaped, it is clear that the MSM has a major problem in predicting the correct hazard function.

One of the major arguments for the MSM is the idea that short-term memory is the pathway for new learning. The famous case of H. M. has been offered as support for the idea that new learning requires a transfer from the short-term to the long-term store. H. M. is the unfortunate patient who had bilateral hippocampus surgery (see Hilts, 1995, for a detailed account). Although H. M. demonstrated a good memory of his life prior to his operation, new learning was apparently nonexistent, supporting the idea that short-term memory or working memory is an information-processing bottleneck for new learning. However, the data from one of the most extensive behavioral studies on H. M. can be reanalyzed and modeled with an MPT model to demonstrate that the original view about H. M. is inaccurate. Freed, Corkin, and Cohen (1987) tested H. M. over five experiments by using 1,200 complex, novel photographs taken from foreign language magazines. Each photograph on the list was studied for 10 s, and the list was presented once again. Freed et al. (1987) later tested H. M. on a two-alternative force-choice recognition test after retention intervals of either 10 min, 1 day, 3 days, or 7 days. A simple MPT model for this task is provided in Figure 12. If there was any stored information about the TBR item, then the correct photograph should have been selected on a two-alternative force-choice recognition test. If there was absolutely no information stored about the target, H. M. would be expected to be at a 50% chance level. The probability of no storage is denoted as $\theta_n$, and the probability of some storage (either partial or sufficient storage)

---

[2] On a less formal level, examples can be offered for cases in which there is a loss of information that should meet all the criteria for being in the long-term store of the MSM. For persons who are more than 20 years past their freshman year in high school, can they recognize their course schedule? Was English taught at 10:00? What was the 1:00 class? Many are not able to recognize their schedule despite the fact it was once well known for a long time. In light of this "thought experiment," the results of the Chechile and Ehrensbeck (1983) study are not surprising.

Another personal example can be provided that illustrates a loss of information that should meet all definitions of being in the long-term store of the MSM. In a recent move to a new building, I had to pack my files. In the process, I came across a manuscript that was misfiled, and I opened to the middle of the paper. The paper described an extensive review of transfer of learning experiments in the field of developmental psychology, and there were proposals for several new studies to address problems raised in the review. I did not recognize this paper and was curious as to why it was in my possession. I was not familiar with the content, the references, and the style of writing. On turning the paper back to the first page, I was surprised to discover that I had written the paper. After the fact, I surmised that the paper was written for a graduate seminar on developmental learning taken 30 years earlier. Even upon rereading the paper, I could not recognize any aspects of it.

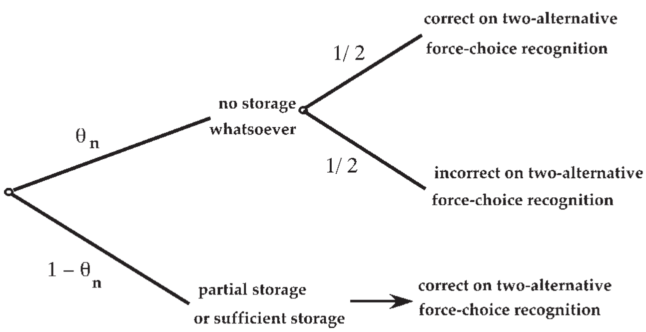


*Figure 12.* Multinomial processing tree model for the two-alternative, forced-choice recognition test.

is $1 - \theta_n$. The probability of no storage is estimated as twice the rate of incorrect forced-choice recognition. The estimated value for some storage (i.e., $1 - \theta_n$) is .72, .54, .47, and .55 for the respective retention intervals of 10 min, 1 day, 3 days, and 7 days. Clearly, H. M. demonstrated some new learning but also showed evidence of long-term memory loss. Although this pattern is not consistent with the MSM, the two-trace hazard model can describe the pattern accurately. It is reasonable to assume that Trace 1 was no longer available after a 10-min delay because Trace 1 has linearly increasing hazard. Hence, only Trace 2 needs to be considered. The use of values $b = .63$, $a = .446$, and $c = .14$ results in predicted values for partial or complete storage of .71, .55, 52, and .50 for the respective retention intervals of 10 min, 1 day, 3 days, and 7 days. The two-trace hazard model parameter values are reasonable in comparison to those of other participants who have been examined previously. In fact, Freed et al. (1987) designed the initial encoding conditions for H.M. to give him a reasonable chance to learn the photographs. Control participants in their experiment only had 1 s to encode the pictures compared with the time of 20 s per picture given to H. M. If H. M. was given enough time to encode the stimuli, he demonstrated some learning. More recently, O'Kane, Kensinger, and Corkin (2004) provided evidence that H. M. could produce accurate descriptions about people who became famous after the onset of his amnesia. Clearly, H. M. has difficulty learning new information, but he can learn new information and can be characterized by the two-trace hazard model.

Finally, it is important to discuss the two-trace hazard model in light of a form of an MSM referred to as *working memory*. Working memory is a framework developed by Baddeley and Hitch (1974). Working memory is a form of short-term memory where there is an interacting set of modules, that is, a visuospatial sketch pad, a phonological loop, and a central executive. As discussed previously, Trace 1 in the two-trace hazard model is not expected to be available for very long because it has monotonically increasing hazard. One might think of Trace 1 as a short-term or working memory. However, the working memory framework is a more detailed representation about how information is processed during encoding and executing a set of simple tasks. In the two-trace hazard theory, the focus is on the fate of information after encoding and not on the representation of the processing of information. Also Trace 1 is treated as a post encoding representation of the TBR item.

## The Two-Trace Hazard Model and Storage Interference

The question arises about how the two-trace hazard model accounts for the effects of interference on the probability of storage. Although both traces are expressed as functions of time, in reality Trace 1 is regarded as a function of time, and Trace 2 is considered to be a function of intervening events that are correlated with time. This effect can be described as a retroactive interference effect on information storage. Furthermore, this retroactive interference can vary in its vigor. For example, more weakly encoded items are expected to result in a greater rate of memory loss. Changes in the rate of loss can be captured by the $b$, $a$, and $d$ parameters in the two-trace hazard model.

Proactive interference (PI) also affects information storage (Chechile, 1987; Chechile & Butler, 1975). How can the two-trace hazard model account for PI? It is important to note that there is a fundamental asymmetry between retroactive and PI. Retroactive interference can occur without any PI (see Baddeley & Scott, 1971), but PI does not occur in the absence of retroactive interference. Without the passage of time filled with interpolated items, the participants do not forget the most recent item even if there are similar previous memory targets. Thus, it is reasonable to account for PI by a stronger retroactive interference. In a typical short-term memory PI experiment, the participant must learn a target such as *grape cherry watermelon*. The memory targets for the subsequent trials also come from the same category. This repeated use of a common category results in storage PI (see Chechile & Butler, 1975). It is reasonable to regard storage PI as a product of a weaker encoding of the distinctive features of the target because the items are encoded with a greater stress on the general category. In terms of the current article, the two-trace hazard model can handle differing rates of retroactive interference by the values of the fitting parameters; thus the model can account for storage PI.

## Rationale for the Memory Hazard Properties

One might ask why the two component traces have their particular hazard function. Why does Trace 1 have a linearly increasing hazard function? Why does Trace 2 have a monotonically decreasing hazard function?

The hazard of Trace 1 can be modeled as a function of the strength of the trace. At $t = 0$, let the strength be denoted as $s_0$. Moreover, it is reasonable to assume that as the strength is reduced, then the hazard increases. Let us stipulate that the hazard is related to the strength via the equation,

$$\lambda(t) = \ln\left[\frac{s_0}{s(t)}\right].$$

Note when $t = 0$ then $s(t) = s_0$, and the hazard is zero. Also note that as the strength of the trace decreases, then hazard increases. Furthermore, if one models the strength for $t > 0$ as

$$s(t) = s_0 \exp(-Dt),$$

then it follows that the hazard function is linearly increasing function with time. In other words, the hazard function for Trace 1 can be regarded as a consequence of the time-based decay of trace strength.

For a sufficiently long time, all the memory retention theories (even the ones that have peaked-shaped hazard functions) predict a decreasing hazard rate. For example, all nine functions in Table 2, as well as the two-trace hazard function, predict that the hazard is decreasing for sufficiently long time. Trace 2 from the two-trace hazard model has monotonically decreasing hazard for all $t$ values. Why should hazard decrease with increasing time?

One possible interpretation of the declining hazard is the consolidation hypothesis, which was first advanced by Müller and Pilzecker (1900) and later refined by Hebb (1949). According to the consolidation hypothesis, new memories are initially fragile and require some time to harden into a consolidated trace. From this perspective, older memories are less susceptible to loss because the trace itself has become strengthened. Supporters for this hypothesis have noted that the general clinical finding in retrograde amnesia cases is that the more recent memories are lost while older memories are preserved (Ribot, 1881). Some investigators have proposed that newly acquired memories need to be converted from a neural transmission format into a maintenance format that represents structural and/or biochemical changes (Bliss & Collingridge, 1993). In the extreme form of the consolidation hypothesis, the hardened trace is permanently stored. However, that position is another form of the MSM that was critiqued earlier. Furthermore, it is not necessary to claim that older traces are impervious to storage loss; instead one might simply claim that older traces are less susceptible to storage loss. That idea was a key feature of the Chechile (1987) TST. More recently Wixted (2004) has linked Ribot's law of retrograde amnesia with Jost's (1897) law of forgetting and discussed these laws in terms of consolidation theory. Jost's law dealt with the case in which there are two memory traces of equal strength but unequal age, that is, Jost's law is the hypothesis that there is more decay of strength for the younger trace. If the change in strength is taken to mean the proportion of items lost at time $t$ of the pool of available items, then Jost's law is effectively an assertion that the memory hazard decreases with time. Thus, for both the TST and the consolidation theory, the memory trace becomes hardened with age.

Although a consolidation hypothesis can explain the decline of Trace 2 hazard with time, there are reasons for doubting the trace consolidation interpretation. For example, Misanin, Miller, and Lewis (1968) showed that electroconvulsive shock (ECS) does not affect an old memory that is learned 24 hr earlier. However, they also showed that an old (and presumably consolidated memory) was disrupted if the animal received an additional training trial just prior to ECS administration. This finding presents a major challenge to the consolidation hypothesis and has led some investigators to assume that a relearning trial converts a consolidated trace back to a fragile trace (Nader, Schafe, & Le Doux, 2000). Yet that proposal would imply that additional training keeps a memory in a susceptible state, which is a paradoxical position to hold. It is more reasonable to assume that additional training strengthens the memory trace. To account for the above findings one can instead focus on the differential influence of the events that occur in the retention interval. Without further learning trials, it is likely that Trace 2 becomes weaker with age. If the weakening of the trace were the only factor, then a decreasing hazard rate would not be expected. However, Trace 2 can be interfered by intervening items, and the strength of that retroactive interference is a function of time, that is, intervening items can have a differential impact depending on the amount of elapsed time. Later in the retention interval, an interpolated item is more likely to affect another recently presented interpolated item rather than affect the target. Thus, the potency of interpolated events to affect the TBR item should decline with time. If the reduced potency of the intervening events changes more rapidly than the reduction of strength of Trace 2, then an overall decreasing hazard rate will result.

Another factor that influences the fate of the memory trace is the similarity of the interpolated item to the memory target. Suppose an interpolated event is very close to the TBR item. If it is identical to the TBR item, then the event should be a new learning trial, that is, Trace 1 for the TBR item is reestablished and the strength of Trace 2 is augmented. However, suppose the item is very close to the TBR item but is different. It is possible that the new item can replace the old item. This type of memory distortion has been demonstrated in misinformation effect experiments (Belli, 1989; Loftus, Miller, & Burns, 1978). In a typical misinformation effect experiment, the participant might see a photograph of an item from a category, for example, a hammer from the tool category. A week later the participant would be told that the tool was a screwdriver. Even later the participant is asked to recall the picture of the tool. A number of the participants falsely report the tool to be a screwdriver. However, from the two-trace hazard perspective this distortion effect might be expected only if there was a delay after the initial encoding of the TBR item. In fact, Belli, Windschitl, McCarthy, and Winfrey (1992) found that the misinformation effect only occurred for long retention intervals. If the delay after the initial encoding is short, the trace is sufficiently strong to withstand the attempt at distortion. However, if the distinctive features of the target have been weakened or lost, then the misleading information can be effective to alter the nature of the encoded item.

## Discussion and Conclusion

In this paper, the careful assessment of the memory retention function has led to a new model of memory and a new retention function. The resulting function is

$$\theta_S(t) = 1 - b[1 - \exp(-dt^2)][1 - \exp(-at^c)].$$

The model does well in fitting appropriate data sets that represent both short-term and long-term retention changes. However, the argument for this new model does not rely solely on the good fit of the model to data.

Theorems 1 and 2 are powerful new results that provide for the testing of an entire class of memory retention functions. The function $u(t) = (1 - \theta_S)/t$ must have a peak at $t = 0$ if the memory hazard function is monotonically decreasing (Theorem 1). Furthermore, a mixture of monotonically decreasing hazard functions is also monotonically decreasing (Theorem 2). In the past, the problem of mixtures has bedeviled research examining the mathematical form of the memory retention function. Mixtures of items that vary in difficulty and mixtures of participants who vary in their memory ability confound the testing of a nonlinear retention function. Nonetheless, if the function has monotonically decreasing hazard, then any mixture is also monotonically decreasing. Hence, if the memory theory implies a hazard function that is monotonically decreasing, then the $u(t)$ function must have a peak at $t = 0$, regardless of whether the assessment is conducted on an individual basis or on a group-average basis. Furthermore, the assessment of hazard is best done by an indirect method such as the assessment of the peak of the $u(t)$ function. Hazard functions are notoriously difficult to estimate. In fact, it was because of the difficulties in direct empirical hazard function assessment that has led to the development of a set of rigorous mathematical tools for hazard function analysis that do not rely on empirical hazard measurement (see Chechile, 2003). Consequently, the method of examining the properties of the $u(t)$ function circumvents a number of difficulties. The clear experimental finding is that the $u(t)$ function is peaked at a time that is greater than zero. It is unlikely that the peak of the $u(t)$ function is an artifact of surreptitious rehearsal. It thus follows that any memory theory is problematic if the theory predicts strictly monotonically decreasing hazard for all time scales. In fact, most of the existing retention functions predict monotonically decreasing hazard. Consequently, the finding that $u(t)$ is not peaked at $t = 0$ has an enormous impact on a wide class of theories.

It is clear that not all memories are lost over the lifetime of the individual. In fact, some participants have remarkable memory over their entire lives (Luria, 1968; Wilding & Valentine, 1997). Permanent memory is inconsistent with memory hazard functions that have monotonically increasing hazard. It thus follows that the memory hazard function can only be a peaked-shaped function, that is, increasing to a maximum and decreasing for long time values. Few retention functions have a peaked-shaped hazard function. Remarkably the original Ebbinghaus retention function does produce a peaked-shaped hazard function. However, the Ebbinghaus function did not fit the Chechile (1987, Exp. 2) data as well as either the TST function or the two-trace hazard model.

The TST function is a formulation that is based on a Weibull subprobability model for a single memory trace. The Weibull probability distribution has a very special place in statistical theory because it is so flexible for representing different types of hazard environments. Depending on the value of a shape parameter, the Weibull probability distribution can describe systems that have monotonically decreasing hazard, constant hazard, or monotonically increasing hazard. This flexibility is a key reason why the distribution has been used extensively in reliability engineering. However, in all of these cases long-term retention would not be possible because the Weibull probability distribution implies that all items are lost. Yet the TST function is based a Weibull subprobability distribution; hence some memories can be permanent. The proportion of lost items is $b$. Depending on the shape parameter, a Weibull subprobability distribution can describe either a peaked-shaped hazard function or a monotonically decreasing hazard function. The TST model can account for the peaked-shape by using a shape parameter that is greater than 1.0; however, experiments that examine retention past the short-term time frame result in values for the shape parameter that are less than 1.0, for example, the Strong (1913) study. If the TST shape parameter is adjusted to account for these long-term experiments, then the property of a peaked-shaped hazard function is lost. Despite the power and flexibility of the Weibull hazard model, the constraints of peaked-shape hazard along with the findings from short-term and long-term memory studies are more than a single-trace system can handle. However, the two-trace hazard model circumvents these problems.

The question arises as to the purpose and character of the two traces in the two-trace hazard model. Why two traces? Redundancy in complex systems can reduce the risk of failure. Such systems are referred to as *conjunctive hazard systems*; the system only fails if both components fail. The study of retention functions has led to the model in which one of the systems is rather temporary, that is, Trace 1. Nonetheless, the existence of this other representation of the TBR item reduces the risk of memory loss, particularly in the time period immediately following the encoding of the TBR item. This reduced risk can lead to predictions about the size of immediate memory, that is, memory span. For example, by using the Chechile (1987, Exp. 2) data set, it is possible to demonstrate that a reasonable memory span requires more than Trace 2. In Chechile (1987, Exp. 2), the TBR item was a subspan triad of letters that was immediately followed by digits that the participant had to shadow. Recall that the two-trace hazard model was fit to each of the 30 participants in that experiment, and the two-trace model fit the data better than the other functions that also have a peaked-shaped hazard. For each participant, the values for the model parameters have been previously determined, and these values can be used to predict the memory span. The predicted memory span is taken to be $n_p + 3$, where $n_p$ is an integer and is the largest $n$ value that corresponds to $\theta_S$ being greater than .975. The value of three in the predicted memory span equation is required because the TBR item is recalled without error on an immediate memory test, and the TBR item has three letters. For 28 of the 30 participants, the theoretically predicted memory span is $6 \pm 2$, which is reasonably close to reported memory span values (Miller, 1956); the other 2 participants had spans greater than 8. However, if only Trace 2 is used and the predicted memory span is recomputed, then the predicted span would be 3 for 28 of the 30 participants and would be 4 for the other 2 participants. Clearly, Trace 1 reduces the risk of an immediate memory loss and provides for the possibility of longer immediate memory, that is, it increases the memory span. Consequently, Trace 1 has important adaptive significance despite the relatively short duration of the trace. Without an immediate memory span of some reasonable length, it is difficult to imagine effective information processing. The development of a temporary backup trace thus represents an important evolutionary advance in the ability to process information.

It is important to stress that the inference of a two-trace system for post-perceptual memory has been made from the analysis of human data. The generality of the two-trace hazard model is still an open question. It is possible that some animals might lack a secondary memory representation. If a species only has a single-trace representation, then a peaked memory hazard function and large immediate memory span would not be expected.

Clearly questions about the mathematical form for the rate of remembering information has led to a new rich framework for conceptualizing the memory system. The finding of a peaked-shaped memory hazard function is a powerful constraint on memory theory. The use of multinomial modeling methods for short-term and long-term retention studies has further led toward a compelling new framework for understanding memory—the two-trace hazard model. There are a number of differences between the two-trace hazard model and the MSM. Unlike the MSM, the two-trace hazard model can deal with the dynamic changes in memory storage that occur both over a short temporal scale as well as over a longer term. The model can handle enormous individual differences because it can fit normal participants as well as patients such as H. M. who have massive memory deficits. The two-trace hazard model cannot only account for the memory retention function but can also provide an explanation of other memory properties such as memory span size and the time course of proactive and retroactive interference.

## References


Anderson, J. R., & Schooler, L. J. (1991). Reflections on 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.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation, Vol. 2* (pp. 89–195). New York: Academic Press.

Averbach, I., & Coriell, A. S. (1961). Short-term memory in vision. *Bell Systems Technical Journal, 40,* 309–328.

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

Baddeley, A., & Scott, D. (1971). Short-term forgetting in the absence of proactive interference. *Quarterly Journal of Experimental Psychology, 23,* 275–283.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113,* 1–29.

Bain, L. J. (1978). *Statistical analysis of reliability and life-time models.* New York: Marcel Dekker.

Batchelder, W. H., & Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review, 87,* 375–397.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6,* 57–86.

Belli, R. F. (1989). Influence of misleading post event information: Misinformation, interference, and acceptance. *Journal of Experimental Psychology: General, 118,* 72–85.

Belli, R. F., Windschitl, P. D., McCarthy, T. T., & Winfrey, S. E. (1992). Detecting memory impairment with a modified test procedure: Manipulating retention interval with centrally presented event items. *Journal of Experimmental Psychology: Learning, Memory, & Cognition, 18,* 356–367.

Bertoin, J., & LeJan, Y. (1992). Representations of measures by balayage from a regular point. *Annals of Probability, 20,* 538–548.

Bliss, T. V., & Collingridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature, 361,* 310–319.

Bohr, N. (1913, December). *On the spectrum of hydrogen.* Address to the Physical Society of Copenhagen. Copenhagen, Denmark.

Chechile, R. A. (1978). Is $d_S'$ a suitable measure of recognition memory "strength?" *Bulletin of the Psychonomic Society, 12,* 152–154.

Chechile, R. A. (1987). Trace susceptibility theory. *Journal of Experimental Psychology: General, 116,* 203–222.

Chechile, R. A. (1998). A new method for estimating model parameters for multinomial data. *Journal of Mathematical Psychology, 42,* 432–471.

Chechile, R. A. (2003). Mathematical tools for hazard function analysis. *Journal of Mathematical Psychology, 47,* 478–494.

Chechile, R. A. (2004). New multinomial models for the Chechile–Meyer task. *Journal of Mathematical Psychology, 48,* 364–384.

Chechile, R., & Butler, K. (1975). Storage and retrieval changes that occur in the development and release of PI. *Journal of Memory and Language, 14,* 430–437.

Chechile, R. A., & Ehrensbeck, K. (1983). Long-term storage losses: A dilemma for multistore models. *Journal of General Psychology, 109,* 15–30.

Chechile, R. A., & Meyer, D. L. (1976). A Bayesian procedure for separately estimating storage and retrieval components of forgetting. *Journal of Mathematical Psychology, 13,* 269–295.

Chechile, R. A., & Soraci, S. A. (1999). Evidence for a multiple-process account of the generation effect. *Memory, 7,* 483–508.

Colonius, H. (1988). Modeling the redundant signals effect by specifying the hazard function. *Perception and Psychophysics, 47,* 604–606.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology.* New York: Dover. (Original work published 1885).

Efron, R. (1970). Effect of stimulus duration on perceptual onset and offset latencies. *Perception & Psychophysics, 8,* 231–234.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53,* 134–140.

Freed, D. M., Corkin, S., & Cohen, N. J. (1987). Forgetting in H. M.: A second look. *Neuropsychologia, 25,* 561–571.

Graczyk, P. (1994). Cramér theorem on symmetric spaces of noncompact type. *Journal of Theoretical Probability, 7,* 609–613.

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

Hilts, P. F. (1995). *Memory's ghost: The strange tale of Mr. M.* New York: Simon & Schuster.

Hu, X., & Phillips, G. A. (1999). GPT. EXE: A powerful tool for the visualization and analysis of general processing tree models. *Behavior, Research, Methods, Instruments & Computers, 31,* 220–234.

Jost, A. (1987). Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen [The strength of associations in their dependence on the distribution of repetitions]. *Zeitschrift fur Psychologie und Physiologie der Sinnesorgane, 14,* 436–472.

Klein, J. P., & Moeschberger, M. L. (1997). *Survival analysis: Techniques for censored data.* New York: Springer.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 4,* 19–31.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Luria, A. R. (1968). *The mind of a mnemonist.* New York: Basic Books.

Massaro, D. W. (1970). Preperceptual auditory images. *Journal of Experimental Psychology, 85,* 411–417.

Massaro, D. W. (1975). *Experimental psychology and information processing.* Chicago: Rand McNally.

Miller, G. A. (1956). The magical number seven, plus or minus two. *Psychological Review, 63,* 81–97.

Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science, 160,* 203–204.

Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis [Experimental contributions to the science of memory]. *Zeitschrift für Psychologie Ergänzungsband, 1,* 1–300.

Myung, I. J., Forster, M. R., & Browne, M. W. (2000). Guest Eds.' introduction: Special issue on model selection. *Journal of Mathematical Psychology, 44,* 1–2.

Myung, I. J., Kim, C., & Pitt, M. A. (2000). Towards an explanation of the power-law artifact: Insights from response surface analysis. *Memory & Cognition, 28,* 832–840.

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature, 206,* 722–726.

O'Kane, G., Kensinger, E. A., & Corkin, S. (2004). Evidence for semantic learning in profound amnesia: An investigation with patient H. M. *Hippocampus, 14,* 417–425.

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley. (2001). Visual search has memory. *Psychological Science, 12,* 287–292.

Peterson, L., & Peterson, J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198.

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

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

Ribot, T. (1881). *Les maladies de la mémoire* [Diseases of memory]. Paris: Germer Bailliere.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit: A comment on theory testing. *Psychological Review, 107,* 358–367.

Rouder, J. N., & Batchelder, W. H. (1998). Multinomial models for measuring storage and retrieval processes in paired associate learning. In Dowling, C. E., Roberts, F. S., & Theuns, P. (Eds.), *Recent progress in mathematical psychology: Psychophysics, knowledge representation, cognition, and measurement* (pp. 195–225). Mahwah, NJ: Erlbaum.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103,* 734–760.

Shiffrin, R. M. (1999). 30 years of memory. In Izawa, C. (Ed.). *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson–Shiffrin model* (pp. 17–33). Mahwah,NJ: Erlbaum.

Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin, 49,* 263–269.

Spencer, C. J., & Shuntich, R. (1970). Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology, 85,* 198–203.

Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs,* 74 (Whole No. 498).

Strong, E. K. (1913). The effect of time-interval upon recognition memory. *Psychological Review, 20,* 339–372.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In Castellan, N. J., & Restle, F. (Eds.). *Cognitive theory Vol. 3* (pp. 199–239). Hillsdale, NJ: Erlbaum.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* Cambridge, U.K: Cambridge University Press.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39,* 321–359.

Underwood, B. (1949). *Experimental psychology.* New York: Appleton-Century Crofts.

Van Zomeren, A. H., & Brouwer, W. H. (1994). *Clinical neuropsychology of attention.* New York: Oxford University Press.

Wenger, M. J., & Gibson, B. S. (2004). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 708–719.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition, 2,* 775–780.

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.

Wickens, T. D. (1999). Measuring the time course of retention. In Izawa, C. (Ed.). *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson–Shiffrin model* (pp. 245–266). Mahwah, NJ: Erlbaum.

Wilding, J., & Valentine, E. (1997). *Superior memory.* East Sussex, U.K: Psychology Press.

Wixted, J. T. (2004). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychological Review, 111,* 864–879.

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

Yang, S. N., & McConkie, G. W. (2001). Eye movements during reading: A theory of saccade initiation times. *Vision Research, 41,* 3567–3585.

# Appendix

## Hazard Properties of Memory Retention Functions and Proofs for Theorems

### Part A: Hazard Properties of the Candidate Retention Functions

Ebbinghaus: The hazard definition (Equation 1), $\lambda(t)$, for the Ebbinghaus function is given as

$$\lambda(t) = \frac{c[\log(t+1)]^{c-1}}{(t+1)(b+[\log(t+1)]^{c})} .$$

Note that $\lambda(0) = 0$, and $\lambda(t) > 0$ for $t > 0$. Yet for large $t$

$$\lambda(t) \rightarrow \frac{c}{(t+1)\log(t+1)} ,$$

which is monotonically decreasing with increasing $t$. Thus, there must be a hazard peak for $t > 0$.

#### *Hyperbolic*

For this model,

$$\lambda(t) = \frac{a}{at+1} .$$

For $t = 0$, $\lambda(0) = a$, and for $t > 0$, $\lambda(t)$ is monotonically decreasing.

#### *PM–Hyperbolic*

The hazard function for this model is

$$\lambda(t) = \frac{ab}{(at+1)^2(1-b)+b(at+1)} .$$

Because the denominator of $\lambda(t)$ increases with $t$, $\lambda(t)$ is monotonically decreasing.

#### *Modified Logarithm*

For probability distributions, Chechile (2003) showed that the general shape of the hazard function could be inferred from the behavior of the function $g(t)$, which is defined as

$$g(t) := \frac{-f'(t)}{f(t)} ,$$

where $f'(t)$ is the derivative of $f(t)$. The modified logarithm retention function is based on a probability distribution model for memory lifetimes. For this distribution, $g(t) = (t+1)^{-1}$ and $g'(t) = -(t+1)^{-2}$. Note that $g'(t) < 0$ for all $t$. From Theorem 2 of Chechile (2003), if $g'(t) < 0$ for all $t$, then the hazard is monotonically decreasing.

### Modified PM–Power

For this function,

$$\lambda(t) = \frac{bc}{(1-b)(t+1)^{c+1} + b(t+1)}.$$

Because the denominator of $\lambda(t)$ increases with $t$, $\lambda(t)$ is monotonically decreasing.

### AS–Power

For the AS–power model, the condition that $c > 0$ is required to avoid an absurd $F(t)$ function at $t = 0$. In general for this model,

$$\lambda(t) = \frac{c}{t\left(\frac{b}{t^c} + 1\right)}.$$

Let us first consider the case where $0 < c \leq 1$. For this case, the denominator of $\lambda(t)$ grows with $t$; hence, $\lambda(t)$ is monotonically decreasing. Next consider the case for $c > 1$. Note in this case that $\lambda(t) \to 0$ as $t \to 0$; yet for large $t$, $\lambda(t) \to c/(t)$, which is decreasing with $t$. However, for all $t >$ 0, $\lambda(t)$ is positive; hence, $\lambda(t)$ has a peak for $t > 0$.

### PM–AS–Power

For this model it is required that $c$ be positive in order to avoid an absurd $F(t)$ function at $t = 0$, and

$$\lambda(t) = \frac{cb^2}{t\left[\frac{(1-b)(b+t^c)^2}{t^c} + b^2\left(\frac{b}{t^c} + 1\right)\right]}.$$

For $0 < c \leq 1$, the denominator of $\lambda(t)$ grows with $t$; thus, $\lambda(t)$ is monotonically decreasing. For $c > 1$, $\lambda(t) \to 0$ as $t \to 0$. For large $t$,

$$\lambda(t) \to \frac{cb^2}{(1-b)t^{c+1}},$$

and is monotonically decreasing with $t$. Yet, for all $t > 0$, $\lambda(t)$ is positive; hence, $\lambda(t)$ has a peak for $t > 0$.

### STFT

For this model,

$$\lambda(t) = a + \frac{cb}{1+bt}.$$

As $t$ increases, the hazard decreases for all $t$.

### PM–STFT

The hazard function for this model is

$$\lambda(t) = \frac{da + \frac{cbd}{1+bt}}{d + e^{at}(1-d)(1+bt)^c}.$$

As $t$ increases the numerator decreases and the denominator increases; thus, the hazard is monotonically decreasing.

### MSM

First, let us develop the discrete retention function for the multistore Markov model. After the study period, let $1 - b$ denote the proportion of memory traces that are stored in the permanent, long-term store, and let $b$ represent the proportion of traces that are only stored in the short-term store. If postencoding rehearsal is prevented, then it will be assumed that no further transfer occurs from the short-term store to the long-term store. Furthermore, it will be assumed that each interpolated event has a chance (with probability $1 - a$) of causing the TBR to be lost from the short-term store. The interpolated events do not affect a target encoding that has already been stored in the long-term store during the study period. Hence, the probability of a storage loss, $Q_n$, after $n$ interpolated events is

$$Q_n = b(1-a) + b(1-a)a + b(1-a)a^2 + \cdots + b(1-a)a^{n-1}.$$

Upon simplification,

$$Q_n = b(1-a^n).$$

Hence,

$$\theta_S(n) = 1 - b(1-a^n),$$

which is the form of the retention function for this model as shown in Tables 1 and 2.

Note that the probability of a memory loss on the $n$th interpolated event is $q_n = b(1-a)a^{n-1}$. From Equation 5 it follows that

$$\lambda_n = \frac{b(1-a)a^{n-1}}{1 - b(1-a^{n-1})}.$$

$$= \frac{1-a}{1 + \frac{1-b}{ba^{n-1}}},$$

which is in the same form as Equation 7, that is, the MSM is based on a geometric subprobability distribution for trace lifetimes. Further, note that as $n$ increases the hazard decreases.

### TST

The TST retention function is based on a Weibull (shape factor equal to 2) subprobability distribution of memory lifetimes. Hence, it follows

$$f(t) = 2abte^{-at^2}.$$

$$\lambda(t) = \frac{2abt\exp(-at^2)}{1 - b + b\exp(-at^2)}.$$

Note that $\lambda(0) = 0$. Recall the definition of $g(t) = -f'(t)/f(t)$; it follows that

$$g(t) = 2at - \frac{1}{t}.$$

From Theorem 1 from Chechile (2003), the hazard is peaked at $t = \tau$ if $\lambda(\tau) = g(\tau)$. For the TST function, this condition occurs when

$$1 - b + b\exp(-a\tau^2) = 2a(1-b)\tau^2.$$

Approximating the $\exp(-a\,\tau^2)$ term results in the following approximate solution for the time of the peak in the hazard function:

$$\tau \approx \frac{1}{2a(1-b)}.$$

Thus, the hazard function is peaked.

Finally, the hazard properties of the power and logarithm functions shown in Table 1 are not meaningful because those two retention functions are not based on a sensible stochastic lifetime model. Consequently, no hazard analyses are provided for those retention functions.

(*Appendix continues*)

## Part B: Proofs for Theorems 1 and 2

Below is a restatement of Theorem 1 along with its proof.

*Theorem 1.* For any continuous probability or subprobability distribution for t≥0 that has monotonically decreasing hazard, the function $u(t)=F(t)/t$ is at a maximum for $t = 0$.

*Proof.* Let us first find $u(t)=F(t)/t$ for the special case of $t = 0$. The limit of $u(t)$ as $t$ approaches 0 can be obtained by means of L'Hospital's rule, that is,

$$\lim_{t\to 0} u(t) = \lim_{t\to 0} f(t) = f(0).$$

L'Hospital's rule can be used because the derivative of $F(t)$—or the probability density—will exist for all time values for continuous probability and subprobability distributions. Now let us find $u(t)$ for $t > 0$. By using the mean value theorem of integral calculus, the $F(t)$ term in the numerator of $u(t)$ can be rewritten as $f(t_m)t$, where $0 < t_m \leq t$. Thus, it follows that $u(t) = f(t_m)$. As mentioned above and from the proof of Corollary 1 in Chechile (2003), monotonically decreasing hazard implies $f(t_1) > f(t_2)$ if $t_1 < t_2$. In a similar fashion, it follows that $u(t + \Delta t) = f(t_{m^*})$ where $t_{m^*} \leq t + \Delta t$. Because the rate of growth of $F(t)$ is decreasing with $t$ for a monotonically decreasing hazard function, it follows that the mean value point $t_{m^*}$ is greater than $t_m$. Thus $u(t + \Delta t) < u(t)$ and the largest value of $u(t)$ is at $t = 0$.

The next theorem deals with a mixture of monotonically decreasing hazard functions. Chechile (2003) established that a mixture of $k$ continuous distributions has monotonically decreasing hazard if all of the component distributions have monotonically decreasing hazard. However, that theorem (Theorem 6 from Chechile, 2003) and its proof are limited to the case in which each component is based on a probability distribution. Yet for the application of memory hazard functions, many of the distributions of interest are based on subprobability distributions rather than probability distributions. Consequently, there is a need to determine whether that earlier theorem can be generalized to the case in which the components are based either on probability or subprobability models. The next theorem establishes the generalization of the earlier Theorem 6 from Chechile (2003). Not only does the theorem handle the case of a mixture with monotonically decreasing hazard, but it also applies for the case of a mixture of components that have constant hazard.

*Theorem 2.* The hazard function is monotonically decreasing for a mixture of continuous density functions that individually have strictly nonincreasing hazard.

*Proof.* Let us first consider a mixture of two continuous distributions that have strictly nonincreasing hazard. Furthermore, let us define $\lambda_2(t)$ as the component with the greater hazard, i.e., $\lambda_2(t)>\lambda_1(t)$. Let us also define $w$ as the mixture weight for the second component. Thus,

$$f(t) = (1 - w)f_1(t) + wf_2(t),$$

where $f_i(t)$, $i = 1,2$, are the two component density functions for the mixture. Furthermore, the survivor function is

$$1 - F(t) = 1 - (1 - w)F_1(t) - wF_2(t).$$

Next consider the hazard for the mixture at the special case of $t = 0$. For $t = 0$, $\lambda_1(0) = f_1(0)$; $\lambda_2(0) = f_2(0)$, and $1 - F(0) = 1$. Thus, it follows that

$$\lambda(0) = (1 - w)\lambda_1(0) + w\lambda_2(0).$$

Clearly, $\lambda(0)$ is between the two individual hazard components, that is, $\lambda_1(0)<\lambda(0)<\lambda_2(0)$, and the location between the two components depends on the mixture weight. For example, if $w = .75$, then the location of $\lambda(0)$ is 75% of the way through the interval between the individual hazard components, and the ratio of the weights is 3 to 1. As the ratio of weight is reduced, the location of the hazard is closer to the lower endpoint. Next consider the mixture hazard for some small value for time. In this case the hazard can be expressed as

$$\lambda(t) = \frac{(1 - w)f_1(t) + wf_2(t)}{(1 - w)[1 - F_1(t)] + w[1 - F_2(t)]}.$$

Because $f_j(t) = \lambda_j(t)\,[1 - F_j(t)]$ for $j = 1, 2$, it follows that

$$\lambda(t) = \frac{(1 - w)[1 - F_1(t)]\lambda_1(t) + w[1 - F_2(t)]\lambda_2(t)}{(1 - w)[1 - F_1(t)] + w[1 - F_2(t)]}.$$

$$\lambda(t) = \frac{W_1\lambda_1(t) + W_2\lambda_2(t)}{W_1 + W_2},$$

where $W_1 = (1 - w)[1 - F_1(t)]$ and $W_2 = w[1 - F_2(t)]$. For $t > 0$, the hazard for the mixture is again between the two component hazards, that is, $\lambda_1(t) < \lambda(t) < \lambda_2(t)$. Because $\lambda_2 > \lambda_1$, it also follows that $F_2(t) > F_1(t)$ and $1 - F_2(t) < 1 - F_1(t)$. Note that the ratio of the weights at this small positive time is

$$\frac{W_2(t)}{W_1(t)} = \frac{w}{1 - w}\,\frac{1 - F_2(t)}{1 - F_1(t)} < \frac{w}{1 - w}.$$

Consequently, in comparison to the $t = 0$ case, the hazard is now shifted closer to the smaller hazard component. Because both hazard components are strictly nonincreasing, each of the endpoint components' hazard values are nonincreasing with increasing time. Thus, it follows that $\lambda(t) < \lambda(0)$. This argument can be continued for each increasing value of time. Hence, the theorem is established for the case of a mixture of two components. The general proof follows by means of mathematical induction. The mixture of $k$ components can be treated as a mixture of two components; the first distribution is a mixture of $k - 1$ components and the second distribution is the $k$th component.

## Part C: Proof That the Two-Trace Hazard Model Has Peaked Hazard

For the two-trace hazard model described by Equations 8–10, the general form for $\lambda(t)$ is given as

$$\lambda(t) = \frac{F_1'(t)F_2(t) + F_1(t)F_2'(t)}{1 - F_1(t)F_2(t)}, \qquad \text{(A1)}$$

where $F_1(t)$ and $F_2(t)$ are defined by Equations 8 and 9, respectively. Differentiating these functions yields

$$F_1'(t) = 2dt\exp(-dt^2),$$

$$F_2'(t) = \frac{cba\exp(-at^c)}{t^{1-c}}.$$

It is clear from Equations 8 and 9 that $F_1(0) = 0$, $F_2(0) = 0$; thus it follows from Equation A1 that $\lambda(0) = 0$. Now for large $t$, $F_1(t) \to 1$, $F_1'(t) \to 0$, and

$$\lambda(t) \to \frac{F_2'(t)}{1 - F_2(t)} = \lambda_2(t),$$

where $\lambda_2(t)$ is the hazard function of the second component and is a monotonically decreasing function. Thus, the hazard begins at zero for $t = 0$ but is a positive decreasing function for large $t$. Hence, it follows that the hazard function is peaked shaped.

Received January 6, 2005
Revision received June 10, 2005
Accepted June 21, 2005 ■