AppliedMath

MDPI

Article

Using Hazard and Surrogate Functions for Understanding Memory and Forgetting

Richard A. Chechile

Psychology and Cognitive and Brain Science, Tufts University, Medford, MA 02155, USA

Abstract: The retention of human memory is a process that can be understood from a hazard-function perspective. Hazard is the conditional probability of a state change at time t given that the state change did not yet occur. After reviewing the underlying mathematical results of hazard functions in general, there is an analysis of the hazard properties associated with nine theories of memory that emerged from psychological science. Five theories predict strictly monotonically decreasing hazard whereas the other four theories predict a peaked-shaped hazard function that rises initially to a peak and then decreases for longer time periods. Thus, the behavior of hazard shortly after the initial encoding is the critical difference among the theories. Several theorems provide a basis to explore hazard for the initial time period after encoding in terms of a more practical surrogate function that is linked to the behavior of the hazard function. Evidence for a peak-shaped hazard function is provided and a case is made for one particular psychological theory of memory that posits that memory encoding produces two redundant representations that have different hazard properties. One memory representation has increasing hazard while the other representation has decreasing hazard.

Keywords: mathematics of human forgetting; memory modeling; hazard functions; modeling forgetting; mathematical psychology; memory hazard functions

MSC: 62N05; 62P15; 91E40

check for updates

Citation: Chechile, R.A. Using Hazard and Surrogate Functions for Understanding Memory and Forgetting. *AppliedMath* **2022**, *2*, 518–546. https://doi.org/10.3390/appliedmath2040031

Academic Editor: Takashi Yamauchi

Received: 25 July 2022
Accepted: 22 September 2022
Published: 4 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The hazard function first emerged from the context of actuarial studies that dealt with product failure or death [1]. Given a system that can undergo a state change, such as going from being alive to being deceased, the hazard $h(t)$ at time t is defined as the instantaneous current risk of the state change, conditioned by the fact that the transition has not yet occurred. Thus,

$$h(t) = \frac{f(t)}{S(t)}, \tag{1}$$

where $f(t)$ is the probability density function of the distribution of lifetimes and where $S(t)$ is the proportion of all the cases at time t where the system remains in its initial state. $S(t)$ is thus called the *survivor function*, and it is the right tail of the lifetime distribution. The cumulative probability for the cases where the state transition has occurred is $F(t) = \int_0^t f(x)dx$, which is the left tail of the lifetime distribution. Furthermore, since the density function is equal to $-\frac{dS}{dx}$, it follows from (1) that $h(x) = \frac{-dS}{S\,dx}$. Hence,

$$-\int_0^t h(x)\,dx = \int_0^t \frac{dS}{S}. \tag{2}$$

Also, because the initial value for the survivor function is $S(0) = 1$, it follows from (2) that

$$S(t) = e^{-\int_0^t h(x)\,dx}. \tag{3}$$

From (1) we know $f(t) = h(t)\,S(t)$; thus the probability density can be re-expressed using Equation (3) as the following function of hazard.

$$f(t) = h(t)e^{-\int_0^t h(x)\,dx}. \tag{4}$$

Hazard can also be defined for a discrete random variable; however, in this paper hazard is used for only the continuous case.

For a comprehensive exploration of the mathematical and statistical properties of hazard functions see [2–5]. In mathematical psychology, a number of researchers have developed and applied hazard functions [6–10]. The topics of human memory retention and forgetting are the focus of attention in this paper, and the analysis of hazard functions provides a powerful framework for achieving important insights about these processes. However, first to aid the reader, it is useful to briefly describe the organization of the paper.

Because some readers are unfamiliar with the mathematical facts about hazard and hazard-surrogate functions, there is a review provided in Section 2 that discusses how these functions have the potential for uncovering essential properties about human memory. This section provides the proofs for five essential theorems and two corollaries. Although these results have been previously established, they are reworked here in a common notation because some of the results are relatively obscure and counterintuitive. Nine candidate models for a memory retention function, which were previously developed from the psychological literature, are discussed in Section 3 from a hazard function perspective, while hazard functions for the various models differ and thereby provide a means for discriminating the theories, the hazard function per se is statistically difficult to empirically measure directly. However, in Section 4, another hazard-surrogate function is discussed, along with two theorems. This function provides for a statistically sensitive method for critically testing the hazard function for human memory. This section also contains the findings from three experiments that use the surrogate function to uncover the key property of peak-shaped hazard. Finally in Section 5, there is a theoretical discussion about the memory structure that is consistent with the mathematical and empirical findings.

2. Fundamentals for the Hazard Analysis of Memory and Forgetting

2.1. Linking Forgetting to the Product Failure Context

To see why hazard functions are connected to memory research, let us consider the basic question of forgetting as studied in psychological experiments. Carefully controlled experiments that examine forgetting first need to train the research participants to learn some novel facts or associations. If at a later time the participants are correct when tested about that information, then we know that the memory, which was originally established at time t_0, has survived until time t. However, for the duration to be a valid quantity, it is important (1) that the to-be-remembered targets are novel and not part of the person's pre-experimental knowledge and (2) that the post-encoding processes did not involve rehearsal of the target information. If there were a surreptitious rehearsal, say one minute after encoding, then a measured retention interval of say two minutes would be mistaken. Thus, to avoid post-encoding rehearsal, it is important that the research participants are kept busy doing other activities over the retention interval that is unrelated to the target information, although for very long retention intervals it is not possible to control all the post-encoding experiences of the participant (Hazard models for manufactured products are different when the product is repaired or adjusted subsequent to the initial production. Such repairs prolong the survival of the product and is analogous to the post-encoding rehearsal of a memory target item.). Psychologists have invented a number of procedures that enable a study of memory retention as a function of time. Experimental tasks such

as the serial-learning procedure [11], paired-associate training [12], the Brown-Peterson task [13,14], and the continuous testing method [15] are standard memory procedures that meet the above mentioned design requirements.

Given the experimental procedures developed by psychologists to study memory, it is clear that memory retention is a similar process to that of the survival of a manufactured product. In both cases there is something new created that is initially in a state of functionality, but later only a proportion $S(t)$ of the original items survived to time t. For both cases, models for the survival will be assumed to be a function of time; yet, it is very likely in both cases that it is not time per se that is responsible for the failures. A manufactured product usually fails because of the wear that occurs from usage, so without usage the product's lifetime is discontinuously increased. Additionally, the survival of a manufactured product varies with the roughness of usage. Thus, when we model product reliability as a function of time, it is only an idealization for the events that occur in time. Given more time, there is more product usage and wear. By using time as a proxy for the amount of wear, the system is characterized without delineating exactly the sequence of events that occurred with product use. Similarly memory loss over the retention interval could either be due to other interfering events that occurred subsequent to studying the target or due to the amount of time elapsed. Some investigators [16,17] varied the amount of information per unit time during the retention interval to see the relative contribution of interfering events and time itself. Although there is a small unique role for time, most of the variability in memory loss is due to the number of intervening events that occurred. Nonetheless in many studies with very long retention intervals, we do not know how many events occurred, so time is used as an approximation for the amount of interfering events.

2.2. Initial Conditions and Rationality Constraints for Memory Reliability Modeling

To assure the initial condition that $S(0) = 1$, the presentation of the memory targets during the learning phase of a memory experiment needs to be sufficiently long to assure that the items are adequately encoded. This condition requires experience with the stimuli class. In the real world, people are bombarded with information. For example, in a one-second time slice of a naturalistic real-world scene, there are more than 14,000 bits of information that are available for processing, but on average people can only acquire about 19 bits per second [18]. However, if only three letters (without consecutive letter repetitions), such as LDQ, is presented as a target item, then this unit has $\frac{\log(26 \cdot 25^2)}{\log 2} \approx 14$ bits of information, which is within the memory processing capacity limit, so it should be encoded completely within one second. In general researchers need to be careful to have knowledge about how to present the memory targets to assure near perfect initial encoding for the type of stimuli used [19,20].

There is also a complicating measurement issue for memory retention studies because there is not a perfect dependent variable that is a pure measure of the probability of target storage. The dependent variables obtained in behavioral studies involve a host of entangled underlying psychological processes such as storage, retrieval, and guessing [18]. For example, a correct response might be a lucky guess, but an incorrect response might be only a temporary retrieval problem of recovering a successfully stored memory. Although no process-pure measure is known, it is nonetheless possible to model how the underlying processes tap a set of separate dependent variables. With a cognitive psychometric model, a validated measurement of the desired target storage probability at time t can be extracted. There are a number of these model-based measurement methods developed in psychological science [21–25].

By following the above mentioned experimental and psychometric methods, it is possible for memory researchers to obtain data on the survival of memory as a function of time. However, a rational mathematical model must also satisfy a few constraints. First, any proposed theory must satisfy the initial condition that $S(0) = 1$. Second, since the probability density function is equal to $-\frac{dS(t)}{dt}$, it follows that any proposed mathematical function for $S(t)$ must satisfy the constraint that $-\frac{dS(t)}{dt} \geq 0, \ \forall \ t \geq 0$.

2.3. Probability and Subprobability Hazard Models

It is important to distinguish between probability and subprobability models for hazard. For both model types, the support for the random variable is on a continuous interval, which is denoted generally as $[L, U]$. The bounds of L and U on the support interval can be either limits such as, respectively, $-\infty$ and $+\infty$ or they can finite real values. For a probability model of product survival as function of time (t), $L = 0$ and the boundary conditions are: (1) $S(0) = 1$ and $F(0) = 0$, which corresponds to all units functioning at $t = 0$, and (2) $S(U) = 0$ and $F(U) = 1$, which corresponds to all units failing by time $t = U$. Yet there are applications where the upper limit of the support is finite and some units are still functioning at the upper limit of the support. For human memory there are some memories that are still stored over the entire lifetime of the individual. For example, Bahrick reported cases of memories that were stored over a 50-year duration and where the participants reported not having rehearsed or used the information during the retention interval [26]. A subprobability model can capture cases where some units do not fail. *Subprobability* is the term that a number of mathematicians have used for a probabilistic measure where the integration of the density function over the support of the random variable is less than 1.0 (i.e., $F(U) < 1$) [27–29]. For a subprobability model of product survival, the boundary conditions are: (1) $S(0) = 1$ and $F(0) = 0$, and (2) $S(U) = 1 - \delta$ and $F(U) = \delta$ where $1 > \delta > 0$. The equation $S(t) = 1 - F(t)$ holds for both probability and subprobability models.

For any probability model it is possible to produce a subprobability counterpart model by multiplying the regular normalized density function $f_*(t)$ by a positive constant $\delta < 1$, $f(t) = \delta f_*(t)$. It thus follows that $F(t) = \delta F_*(t)$ and $S(t) = 1 - \delta F_*(t)$. Although the multiplication by a positive δ results in linear changes in the density function as well as in the cumulative and survivor functions, it has important consequences for the hazard and the integrated hazard. To see this point, consider the simplest hazard probability model of an exponential distribution where $f_*(x) = k\,e^{-kx}$ for x on the support of the nonnegative reals and for $k > 0$. The value for the cumulative distribution is $F_*(t) = \int_0^t k\,e^{-kx}\,dx = 1 - e^{-kt}$. Thus, the hazard is $h_*(t) = \frac{k\,e^{-kt}}{e^{-kt}} = k$. This property of constant hazard is a well known and unique feature of the exponential distribution [30]. Although the hazard for the exponential probability model is a constant, the hazard for the counterpart exponential subprobability model is not a constant. For the exponential subprobability model, the hazard is

$$\begin{aligned} h(t) &= \frac{\delta\,k\,e^{-kt}}{1 - \delta\int_0^t k\,e^{-kx}\,dx}. \\ &= \frac{k}{\frac{1-\delta}{\delta}\,e^{kt} + 1}. \end{aligned}$$

Thus, the hazard for the subprobability exponential model is monotonically decreasing, whereas the exponential model has constant hazard.

Subprobability models are not uncommon because stochastic processes are often right-tail censored for practical limitations, such as employing a time limit for a measured response. Right-tail censoring has also led some theorists to use the concept of *reverse hazard*, which is defined as $r(x) = \frac{f(x)}{F(x)}$ since reverse hazard is not a function of δ when $F(U) < 1$ [31–34]. Consequently reverse hazard is especially useful when the quantity δ is unknown [35]. However, in this paper reverse hazard is not needed for learning some critical details about human memory retention.

2.4. Weibull Distribution and Hazard

The Weibull distribution is an important survival model because it can capture monotonically decreasing hazard, constant hazard, or monotonically increasing hazard depend-

ing on the value of a shape parameter. This model will also play an important role for describing the survivor function for memory.

The Weibull cumulative function is $F(t) = 1 - \exp(-kt^c)$ on the support of the non-negative reals and where k is a positive scale factor and c is a positive shape parameter [30]. The resulting hazard function is $h(t) = ckt^{c-1}$. Note that at the value of $c = 1$, the Weibull distribution becomes the exponential distribution and has constant hazard. For $0 < c < 1$, the Weibull has monotonically decreasing hazard, and for $c > 1$, the hazard is monotonically increasing.

As with the exponential distribution, there is also a Weibull subprobability model where the cumulative probability model is $F(t) = \delta[1 - \exp(-kt^c)]$ for $0 < \delta < 1$. The hazard for the Weibull subprobability model is

$$h(t) \quad = \quad \frac{c\,k\,t^{c-1}}{\frac{1-\delta}{\delta}\,e^{kt} + 1}\,. \tag{5}$$

For the Weibull subprobability model, the hazard is monotonically decreasing for $0 < c \leq 1$, but for $c > 1$, the hazard has a peak shape. That is, $h(0) = 0$, but for some small t values the hazard is increasing until reaching a maximum at t_p. For $t > t_p$, the Weibull subprobability model has monotonically decreasing hazard. For example, the case where $\delta = 0.5$, $c = 2$, and $k = 1$ results in a hazard peak at approximately $t = 1.278$.

2.5. A Surrogate Function and Proofs of Monotonicity

For the analysis of the monotonicity and other properties of the hazard function, it is sometimes helpful to use other functions. For cases that have a known density function $f(x)$ for $x \in [L, U]$, it has been shown that there is a convenient surrogate function for understanding monotonicity properties [6].This function is

$$g(x) \quad = \quad -\frac{f'(x)}{f(x)}\,. \tag{6}$$

The following two general theorems about hazard functions illustrate the importance of the $g(x)$ function. These theorems are variations of several other previously published proofs that deal with a class of functions where the hazard either has a monotone rate or is a constant [9].

Theorem 1. *For a continuous differentiable density function $f(x)$ on support $[L, U]$, the hazard is: monotonically increasing if and only if $h(x) > g(x)\ \forall\ x$, monotonically decreasing if and only if $h(x) < g(x)\ \forall\ x$, and a constant if and only if $h(x) = g(x)\ \forall\ x$.*

Proof. It follows from the definition of hazard that

$$h'(x) \quad = \quad \frac{[f(x)]^2 + f'(x)\,S(x)}{[S(x)]^2}\,, \tag{7}$$

$$= \quad [h(x)]^2 + \frac{f'(x)f(x)}{S(x)f(x)}\,, \tag{8}$$

$$= \quad h(x)[h(x) - g(x)]\,, \tag{9}$$

$$\left(\frac{S(x)}{f(x)}\right) h'(x) \quad = \quad h(x) - g(x)\,. \tag{10}$$

Since the term $\left(\frac{S(x)}{f(x)}\right)$ is non-negative for all $x \in [L, U]$, it follows that the sign of $h'(x)$ is directly a function of the quantity $h(x) - g(x)$. Thus, if $h(x) > g(x)\ \forall\ x \implies h'(x) > 0$, and conversely if $h'(x) > 0\ \forall\ x \implies h(x) > g(x)$. Similarly if h(x)<g(x) $\forall\ x \implies h'(x) < 0$, and conversely if $h'(x) < 0\ \forall\ x \implies h(x) < g(x)$. Lastly if $h(x) = g(x)\ \forall\ x \implies h'(x) = 0$, and conversely if $h'(x) = 0\ \forall\ x \implies h(x) = g(x)$. □

Theorem 1 is quite general because it only requires a continuous differentiable density function. It applies for both probability models and subprobability models. For example, using the theorem for the case of the exponential distribution $f(t) = ke^{-kt}$ for non-negative t, we find that both $h(t)$ and $g(t)$ are equal to k, so the hazard is a constant. However, for the subprobability version of the exponential when $f(t) = \delta ke^{-kt}$, we now find that $h(t) - g(t) = k[\frac{1}{\frac{1-\delta}{\delta}e^{kt}+1} - 1] < 0$, which results in correctly concluding that the hazard is monotonically decreasing. Theorem 1 is limited as a tool when there is not a closed form solution for the cumulative distribution that enables a closed-form formula for $h(x)$. However, the next theorem is useful even when the cumulative distribution function does not have a closed-form solution.

Theorem 2. *For a continuous differentiable density function $f(x)$ for $x \in [L, U]$ with $f(U) = 0$, the hazard rate is: either monotonically increasing if and only if $g'(x) > 0 \; \forall \; x$, monotonically decreasing if and only if $g'(x) < 0 \; \forall \; x$, or a constant hazard rate if and only if $g'(x) = 0 \; \forall \; x$.*

Proof. It is convenient to find the expected value for $g(y)$ for $y \in [x, U]$. The conditional expected value is

$$E(g(y) \,|\, y \geq x) = \frac{\int_x^U g(y)\, f(y) dy}{\int_x^U f(y) dy}, \tag{11}$$

$$= \frac{-\int_x^U f'(y)\, dy}{S(x)} = \frac{0 + f(x)}{S(x)} = h(x)\,. \tag{12}$$

Note that $E(g(y) \,|\, y \geq x) = h(x)$ even for subprobability models because if a normalized density function were multiplied by δ where $0 < \delta \leq 1$, then the resulting $g(x)$ function is invariant to this change to the density function. Thus, it follows from (10) that

$$\left(\frac{S(x)}{f(x)}\right) h'(x) = E(g(y) \,|\, y \geq x) - g(x)\,. \tag{13}$$

The sign of $h'(x)$ is thus directly related to the quantity $E(g(y) \,|\, y \geq x) - g(x)$, provided that $f(U) = 0$. If $g'(x) > 0 \; \forall \; x$, then $E(g(y) \,|\, y \geq x)$ and thus $h(x)$ is growing with x since the smallest values of $g(x)$ are being trimmed as x increases, which implies that $h'(x) > 0$. Conversely if $h'(x) > 0 \; \forall \; x$ implies that $\frac{dE(g(y) \,|\, y \geq x)}{dx} > 0$, so $g'(x) > 0$. Similarly if $g'(x) < 0 \; \forall \; x$, then $h(x) = E(g(y) \,|\, y \geq x) < g(x)$, so $h'(x) < 0$. Again the converse is also true. Lastly if $g'(x) = 0 \; \forall \; x \implies g(x) = k$, and because $h(x) = E(k) = k \implies h'(x) = 0$. Again the converse also holds. □

Several useful corollaries follow from the above theorems.

Corollary 1. *The density function is monotonically decreasing for distributions that have monotonically decreasing hazard, and the lower limit for the support L must be finite where $f(x)$ has its maximum value.*

Proof. In general if the hazard rate is monotonically decreasing $\forall \; x \in [L, U]$, then from Theorem 1 it follows that $g(x) > h(x)$. Thus, $-\frac{f'(x)}{f(x)} > \frac{f(x)}{S(X)}$, so $-f'(x) > \frac{(f(x))^2}{S(x)} > 0$. So, $f'(x) < 0 \; \forall \; x$. Hence, the density function must be monotonically decreasing. Moreover, if the lower limit of support were $-\infty$, then the cumulative probability would grow without limit. Therefore L must be finite with $f(L)$ as the maximum density value. □

Although a monotonically decreasing hazard function has monotonically decreasing probability density, the converse statement does not follow because there can be monotonically decreasing density functions that have monotonically increasing hazard. Consider the case where $f(x) = \frac{3}{2} - x$ on the support interval of $[0, 1]$. For this density function

$h(x) = (\frac{3}{2} - x)(1 - \frac{3x}{2} + \frac{x^2}{2})^{-1}$ and $g(x) = (\frac{3}{2} - x)^{-1}$, so it follows that $\frac{h(x)}{g(h)} = \frac{(\frac{3}{2}-x)^2}{1-\frac{3}{2}x+\frac{x^2}{2}} > 1$ for $x \in [0, 1]$. Thus, from Theorem 1, $h(x)$ is monotonically increasing because $h(x) > g(x)$ for all x in the support.

The next corollary deals with the cases when hazard monotonicity does not hold.

Corollary 2. *For a continuous differentiable density function $f(x)$ for $x \in [L, U]$ that does not have either monotonically increasing hazard, monotonically decreasing hazard, or constant hazard, there is at least one point x_0 that is either a local hazard peak, valley or an inflection point. It is a hazard peak if $g(x_0) = h(x_0)$ with $g'(x_0)g(x_0) > 0$, it is a hazard valley if $g(x_0) = h(x_0)$ with $g'(x_0)g(x_0) < 0$, and it is an inflection point if $g(x_0) = h(x_0)$ with $g'(x_0)g(x_0) = 0$.*

Proof. If $f(x)$ does not have a monotonic hazard rate, then there must be at least one x_0 that corresponds to $h'(x_o) = 0$. From Theorem 1, it is clear that a hazard peak, valley or an inflection point occurs if $g(x_0) = h(x_0)$. Additionally, from Equation (10), it follows that

$$g(x) = h(x) - \frac{h'(x)}{h(x)}, \tag{14}$$

$$g'(x) = h'(x) - \frac{h''(x)h(x) - [h'(x)]^2}{[h(x)]^2}. \tag{15}$$

From Theorem 1 and Equation (15), a hazard function has a peak, valley, or inflection point when $g(x_0) = h(x_0)$ and when $h'(x_0) = 0$, which corresponds to when $g'(x_0) = -\frac{h''(x_0)}{h(x_o)} = -\frac{h''(x_0)}{g(x_0)}$. Thus, $h''(x_0) = -g'(x_0)g(x_0)$. So, at x_0 the hazard has either a peak if $g'(x_0)g(x_0) > 0$, a valley if $g'(x_0)g(x_0) < 0$, or an inflection point if $g'(x_0)g(x_0) = 0$. □

As an illustration of the utility of the $g(x)$ surrogate function, consider the case of the Gaussian distribution, which has a density function of $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]$, but it lacks a closed form-function for $S(x)$ and $h(x)$. The support for the distribution is the real numbers, and clearly $f(U = \infty) = 0$. For this distribution, $g(x) = \frac{x-\mu}{\sigma^2}$, so $g'(x) = \frac{1}{\sigma^2}$, which is positive for all x. Thus, from Theorem 2 it is clear that the Gaussian distribution has monotonically increasing hazard.

2.6. Using Hazard to Create Novel Probability Models

Because the hazard function captures the change in the tendency for a state change, it is possible to design new probability models that have a particular desired hazard rate. As an illustration of this operation, suppose that the stochastic theorist wants a continuous density function that has a hazard function of $h(x) = \frac{1}{6}(x^3 - 6x^2 + 9x + 6)$ over the support interval of $[0, \infty)$. This hazard function rises to a peak at $x = 1$, decreases to a valley at $x = 3$, and increases for $x > 3$. From Equation (3) the survivor function for this desired distribution is

$$S(x) = \exp\left(-\frac{x^4}{24} + \frac{x^3}{3} - \frac{3x^2}{4} - x\right).$$

From Equation (4) the corresponding density for the distribution is

$$f(x) = \frac{1}{6}(x^3 - 6x^2 + 9x + 6)\exp\left(-\frac{x^4}{24} + \frac{x^3}{3} - \frac{3x^2}{4} - x\right).$$

2.7. Hazard for Mixture Models

Up to this point, the reliability models assumed that all units are identically and independently distributed. For mixture models the assumption of identically distributed is relaxed. Mixture models can exhibit hazard properties that are not shared by any of the component distributions. Suppose for example that there is a mixture with two Gaussian

components (i.e., θ proportion of the units have a Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$, whereas the rest of the units have a Gaussian distribution with mean $\mu = \Delta > 0$ and standard deviation $\sigma = 1$). As shown above, a Gaussian random variable has monotonically increasing hazard regardless of its mean and standard deviation. However, a mixture of say two components can exhibit a hazard rate that is either monotonically increasing or it can have a local peak and a local valley [9]. Which one of these two different hazard profiles occurs depends on the mixing rate (i.e., the value for θ) and depends on the separation between the means (Δ). For example, when $\theta = 0.5$ and $\Delta < 2.70$ the mixture has monotonically increasing hazard, but if $\theta = 0.5$ and $\Delta \geq 2.70$, then the hazard has a profile that initially rises to a local peak, decreases to a local valley, and increases again [36]. However, when the mixing rates are $\theta = 0.05$ and $1 - \theta = 0.95$, the critical separation between the two Gaussian components has to be $\Delta \geq 3.61$ before the phase change in hazard profile occurs. These general features about a hazard function for a mixture has been demonstrated for other distributions where the components of the mixture have monotonically increasing hazard [36]. Furthermore, this phase change has also been shown for sufficiently separate components that consists of more than two components. For example, a mixture of three components has been shown to exhibit a profile that increases to a local peak, decreases to a local valley, increases to a second local peak, decreases to a second local valley, and increases for larger values of x [9]. Consequently the finding of a hazard function that has multiple peaks and valleys is a potential signal that the distribution consists of a mixture, and it can suggest how many components there are in the mixture. However, as shown in the previous subsection about creating novel probability models, there can also be a homogeneous random process that has the same hazard profile.

Although a hazard rate for a mixture of components that individually have an increasing hazard rate is complicated, the mixture of component distributions that individually have non-increasing hazard has a monotonic rate [9]. A formal proof of this assertion is given below.

Theorem 3. *The hazard is monotonically decreasing for a mixture of components that individually have non-increasing hazard with* $f_i(U) = 0$.

Proof. First consider the case where there are two components in the mixture. Moreover, if one or both of the components has constant hazard, then it is an exponential distribution $f_i(x) = k_i \exp[-k_i\,(x - L)]$, so $f_i(U = \infty) = 0$. Since $f_i(U) = 0$ it follows from Theorem 2 that $g_i'(x) = (\frac{1}{f_i(x)})^2\, H_i(x) \leq 0$ for $i = 1, 2$, where

$$H_i(x) \;\; = \;\; [f_i'(x)]^2 - f_i(x) f_i''(x) \leq 0\,. \tag{16}$$

Moreover, the equations for $f(x)$, $f'(x)$, and $f''(x)$ for the mixture are:

$$\begin{aligned} f(x) &= \theta f_1(x) + (1-\theta) f_2(x)\,, \\ f'(x) &= \theta f_1'(x) + (1-\theta) f_2'(x)\,, \\ f''(x) &= \theta f_1''(x) + (1-\theta) f_2''(x)\,. \end{aligned}$$

It thus follows that $H(x)$ for the mixture is

$$H(x) \;\; = \;\; \theta^2 H_1(x) + (1-\theta)^2 H_2(x) + \theta(1-\theta) f_1(x) f_2(x)\, G(x)\,, \tag{17}$$

$$G(x) \;\; := \;\; \frac{2 f_1'(x) f_2'(x)}{f_1(x) f_2(x)} - \frac{f_1''(x)}{f_1(x)} - \frac{f_2''(x)}{f_2(x)}. \tag{18}$$

Because $H_i(x) \leq 0$, it follows that for $i = 1, 2$ that $f_i(x)f_i''(x) \geq [f_i'(x)]^2$. Hence, $\frac{f_i''(x)}{f_i(x)} \geq \left[\frac{f_i'(x)}{f_i(x)}\right]^2$. Thus, $\frac{f_i''(x)}{f_i(x)} = \left[\frac{f_i'(x)}{f_i(x)}\right]^2 + a_i$ where $a_i \geq 0$. Substituting this last equation into Equation (18) yields

$$G(x) = \frac{2f_1'(x)f_2'(x)}{f_1(x)f_2(x)} - \left[\frac{f_1'(x)}{f_1(x)}\right]^2 - a_1 - \left[\frac{f_2'(x)}{f_2(x)}\right]^2 - a_2\,, \tag{19}$$

$$= -\left(\frac{f_1'(x)}{f_1(x)} - \frac{f_2'(x)}{f_2(x)}\right)^2 - a_1 - a_2\,. \tag{20}$$

Even if both components of the mixture has constant hazard (i.e., the distributions are exponential distributions of the from $f_i(x) = k_i \exp[-k_i(x - L)]$), the first term in Equation (20) is negative, so $G(x) < 0$ and thus $H(x)$ from Equation (17) is also negative. Hence, the theorem holds for any two-component mixtures where the components have non-increasing hazard because $g'(x) < 0$. The general proof is established by mathematical induction. That is, assume the theorem holds for a mixture of $k - 1$ components. The case with k components is then considered as a mixture of two components where the first component is a mixture of $k - 1$ components and the second component is the kth component. The same argument as above results in concluding that the k-component mixture has monotonically decreasing hazard. □

Theorem 3 will play an important role for subsequent applications of hazard for understanding models for memory retention.

Finally, one possible interpretation for a subprobability model is as a mixture of two distributions. For $1 - \delta$ proportion of the units, the survivor function is 1. Thus, the probability density for the lifetime failure distribution for this proportion is 0, so the value of $F(U)$ for these failure-resistant units is 0.

2.8. Disjunctive Hazard Systems

Virtually all units consist of essential constituent parts that must all work properly for the unit to be functional. A failure of any of the subsystems results in an overall system failure. Units of this type can be called a *disjunctive hazard* system or more informally a *weakest-link* system. For this type of product, it does matter what the average lifetime of the constituent subsystems is because the system fails when the weakest link fails. Statisticians and probability theorists recognized that there were three limiting or asymptotic probability distributions for problems of this type [37–39]. That is, the distribution of the minimum value for a system of n components has an asymptotic limit for large n that must be one of three distributions, which are denoted as $L^{(1)}$, $L^{(2)}$, and $L^{(3)}$. Hence, all probability distributions belong to one of three domains of attraction for the limiting case for the weakest link. For example, if the subsystems have a Gaussian distribution, then the weakest link has a $L^{(1)}$ limiting form; whereas if the subsystems have either an exponential or uniform distribution, then the weakest link has a $L^{(2)}$ limiting form [40]. However, for lifetime models where t_i is the survival time for the ith component, only the $L^{(1)}$ and $L^{(2)}$ limiting distributions are suitable, and these distributions are, respectively, the Gumbel and Weibull distributions [41]. Thus, it is not accidental that the Gumbel and Weibull models are frequently used to capture the survivor function for manufactured products. It was shown previously in this paper that the Weibull distribution hazard for $t \geq 0$ is $h(t) = ckt^{c-1}$ for $c > 0$, so this function can be either monotonically decreasing, constant hazard, or monotonically increasing hazard, depending on the value of the shape parameter c. A standard Gumbel distribution is on the support of the real line $(-\infty, \infty)$. A truncated Gumbel distribution on the support of $t \geq 0$, with a scale factor of b, has the density of $f(t) = \frac{\exp[-\frac{t}{b} - \exp(-\frac{t}{b})]}{b[1-\exp(-1)]}$. The resulting $g(t)$ function is $\frac{1}{b}[1 - \exp(-\frac{t}{b})]$. So, $g'(t) = \frac{\exp(-\frac{t}{b})}{b} > 0 \; \forall \; t \in [0, \infty)$, and it thus follows from Theorem 2 that the truncated Gumbel distribution has monotonically increasing hazard. Note that a subprobability version of

the truncated Gumbel distribution has the same value for $g'(t) > 0 \; \forall \; t \; \in \; [0, \infty)$, so it too has monotonically increasing hazard. These facts about the Gumbel distribution will be discussed later in the application of hazard functions to the question of memory retention.

If $S_i(x)$ is the survivor function for the ith component of a disjunctive hazard system, then a key property for disjunctive hazard systems follows when the $S_i(x)$ are independent.

Theorem 4. *The hazard for a disjunctive hazard system of n independent component is* $h_{dis} = \sum_{i=1}^{n} h_i(x)$*, where* $h_i(x)$ *is the hazard for the ith component.*

Proof. The survivor function for a system of n independent components is $S_{dis} = \prod_{i=1}^{n} S_i(x)$. Thus, it follows that

$$\begin{aligned} f_{dis}(x) &= \sum_{i=1}^{n} -S_i'(x) \prod_{j \neq i} S_j(x)\,, \\ &= \sum_{i=1}^{n} f_i(x) \prod_{j \neq i} S_j(x)\,, \\ h_{dis}(x) &= \frac{\sum_{i=1}^{n} f_i(x) \prod_{j \neq i} S_j(x)}{\prod_{i=1}^{n} S_i(x)}\,, \\ &= \sum_{i=1}^{n} h_i(x)\,. \end{aligned}$$

□

If each unit of the system has the same unit hazard function (i.e., $h_i(x) = h(x)$), then the hazard for a disjunctive system where all n are required is $h_{dis}(x) = n\,h(x)$. Clearly complex systems with many essential constituent components have a short lifetime. For example, if we define the half-life duration $\tau_{0.5}$ as the time a system survivor function $S_{dis}(\tau_{0.5}) = \frac{1}{2}$, the value of $\tau_{0.5}$ will be smaller as the number of essential components increases. For example, if $S_i(t) = e^{-kt}$ for each component, then $S_{dis} = e^{-nk\tau_{0.5}} = \frac{1}{2}$, so $\tau_{0.5} = \frac{\log 2}{kn}$.

2.9. Conjunctive Hazard Systems

A key idea underlying the engineering of a highly reliable system is having independent redundant systems. For example, an air-traffic control system has one or more backup systems. If the main system fails, then the backup system is able to maintain the functionality of air-traffic guidance. For systems of this type, the failure of the system requires the failure of all the independent subsystems. Let us call such arrangements a *conjunctive hazard system*. The cumulative distribution for a conjunction arrangement is $F_{con}(x) = \prod_{i=1}^{n} F_i(x)$.

Theorem 5. *The hazard of a conjunctive system, which only fails when all the of the n independent component systems fail, is* $h_{con}(x) = \frac{\sum_{i=1}^{n} f_i(x) \prod_{j \neq i} F_j(x)}{1 - \prod_{i=1}^{n} F_i(x)}$.

Proof. The probability density for the conjunction system is $f_{con}(x) = \sum_{i=1}^{n} f_i(x) \prod_{j \neq i} F_j(x)$. It thus follows that $h_{con}(x) = \frac{\sum_{i=1}^{n} f_i(x) \prod_{j \neq i} F_j(x)}{1 - \prod_{i=1}^{n} F_i(x)}$. □

To see how a conjunctive hazard system improves the half-life duration, let us consider the case where each subsystem has a survivor function of $S_i(t) = e^{-kt}$. The resulting survivor function for the conjunctive system at the system half-life is $S_{con}(\tau_{0.5}) = 1 - (1 - e^{-k\tau_{0.5}})^n = \frac{1}{2}$. So, it follows that $\tau_{0.5} = \frac{1}{k} \log[\frac{1}{1-(\frac{1}{2})^{\frac{1}{n}}}]$. To see the benefit for the case of a single redundant backup (i.e., $n = 2$), let us compute the quantity $\frac{\tau_{0.5}(n=2)}{\tau_{0.5}(n=1)}$. This ratio for

the exponential individual survivor function is approximately 1.772. Thus, the backup results in a 77.2-percent increase in the system half-life duration.

It is also instructive to see how the hazard of a two-component conjunctive system compares to the hazard rate of the individual components. From Theorem 5, if both components are identical, then

$$\begin{aligned} h_{con} &= \frac{2f(x)F(x)}{1-[F(x)]^2}, \\ &= \frac{2f(x)F(x)}{[1-F(x)][1+F(x)]}, \\ &= \frac{2F(x)}{1+F(x)}h(x). \end{aligned}$$

Note for all values for the cumulative distribution where $0 < F(x) < 1$, $h_{con}(x) < h(x)$, but the magnitude of this inequality is dependent on the value for $F(x)$. When the overall risk is low, say when $F(x) = 0.05$, then $h_{con}(x)$ is approximately $0.0952h(x)$. However, when the overall risk is high, say when $F(x) = 0.8$, then $h_{con}(x)$ is approximately $0.8889h(x)$. Thus, a conjunctive system is particularly advantageous for reducing system failure when the overall risk is low.

3. Hazard Functions for the Candidate Models for Memory Retention

Since the original experimental study of human memory by Ebbinghaus [11] in 1885, there have been more than 200 published papers in the psychological literature that have studied memory retention as a function of the time after encoding. See [42,43] for reviews of this literature and for a discussion about some models for a memory retention function. Nine candidate models for the survivor function $S(t)$ are provided in Table 1 along with the corresponding hazard function. Each of these models meet the aforementioned constraints of $S(0) = 1$ and $S(U) = 1 - \delta$ for $0 \leq \delta \leq 1$ because we want to study the survival of memories that were initially encoded correctly at time $t = 0$ and because we want to incorporate the feature that some proportion of the memories might survive over the entire lifetime of the individual. Consequently some of the models listed in Table 1 are modified from the form that was originally proposed so as to satisfy the above two boundary-conditions. In this section each of the candidate models is discussed briefly.

Table 1. Nine candidate models for the memory survivor function $S(t)$ are shown along with the corresponding hazard function. The parameters a, b, c and d are positive. The cumulative functions for the two respective traces of the two-trace hazard model are $F_1(t) = 1 - \exp(-d\,t^2)$ and $F_2(t) = b(1 - \exp(-at^c)$.

Model Name	Survivor Function	Hazard Function
modified hyperbolic	$1 - b + \frac{b}{at+1}$	$\frac{ab}{(at+1)^2[1-b+\frac{b}{at+1}]}$
modified exponential	$1 - b + be^{-at}$	$\frac{a}{\frac{1-b}{b}e^{at}+1}$
modified logarithm	$1 - b\frac{\log(t+1)}{\log(U+1)}$	$\frac{b}{(t+1)\log\frac{(U+1)}{(t+1)^b}}$
modified power	$1 - b + \frac{b}{(t+1)^c}$	$\frac{cb}{(1-b)(t+1)^{c+1}+b(t+1)}$
modified single-trace fragility	$1 - d + \frac{de^{-at}}{(1+bt)^c}$	$\frac{da+\frac{cbd}{1+bt}}{d+e^{at}(1-d)(1+bt)^c}$
modified Anderson-Schooler	$1 - d + \frac{db}{b+t^c}$	$\frac{dbc}{t^{1-c}[(1-d)(b+t^c)^2+db(b+t^c)]}$
Ebbinghaus	$\frac{b}{b+[\log(t+1)]^c}$	$\frac{c[\log(t+1)]^{c-1}}{(t+1)(b+[\log(t+1)]^c)}$
trace susceptibility theory	$1 - b + be^{-at^2}$	$\frac{2abte^{-at^2}}{1-b+be^{-at^2}}$
two-trace hazard theory	$1 - F_1(t)F_2(t)$	$\frac{2dte^{-dt^2}F_2(t)+bact^ce^{-at^c}F_1(t)}{1-F_1(t)F_2(t)}$

3.1. Modified-Hyperbolic Model

Several researchers in animal learning proposed the retention function of $S(t) = \frac{1}{at+c}$ [44,45]. Note that for $c > 1$ this function does not satisfy the constraint that $S(0) = 1$. This feature is likely to be intentional because animals rarely perform the task perfectly initially. These nonverbal organisms need to be trained to do the task. Yet after extensive training the animals will often not be 100 percent consistent in executing the task properly. To modify the original equation to meet the $S(0) = 1$ constraint, the value of c can be set to 1. Chechile [43] further modified the hyperbolic model to include for the possibility of long-term memory retention for a subset of the memories. Hence, in the modified model the survivor function is $1 - b + \frac{b}{at+1}$. If $b = 1$, then there are no lifetime memories, but for $0 < b < 1$, there are $1 - b$ long-term memories.

The hazard function for the modified-hyperbolic model has a constant numerator along with a denominator that increases with time. Consequently, the hazard is monotonically decreasing with time. Because the hazard is monotonically decreasing, it also follows from Corollary 1 that the density $f(t)$ is also monotonically decreasing. Moreover $\lim_{t \to U} f(t) \to 0$.

The fact that memory hazard for long times is decreasing is unlike the typical hazard for manufactured products. This property about memory has been articulated by Jost as far back as 1897 [46]. Jost's law is the claim that if there are two memories of equal strength, then it is more likely the more recent memory will be lost.

3.2. Modified Exponential Model and the Multiple-Store Model

The modified exponential model was discussed previously in Section 2.3 as an example of a subprobability model. Recall that the exponential model has constant hazard, so it is inconsistent with Jost's law. However, the subprobability version of the exponential model had monotonically decreasing hazard and thus is consistent with Jost's law. However, there is another more important reason for justifying the modified exponential model as a viable candidate for the retention function because it represents a form of a highly regarded multiple-store model of memory [47,48].

For the multiple-store memory framework, the initial encoding results in either (1) the information being transferred from a temporary storage buffer to a permanent store, which is postulated to last over the lifespan of the person, or (2) the information remains in a short-term memory. An item in the short-term store can survive for awhile until some new information knocks it out. This theory is a mixture model where $1 - b$ proportion of the memories are immune to memory loss, and b proportion of the studied items are 100-percent susceptible to loss eventually [47]. Thus, *the multiple-store model does not have two representations that are both susceptible to loss of information [47]*. Memory failures for information in the long-term store is assumed to be a temporary retrieval problem; the target is still assumed to be available for recall at a later time. Thus, storage loss occurs when information is not transferred to the long-term store, and it is knocked out of the short-term store by other incoming information. A Markov-chain model can be used to characterize the storage loss from the short-term store. Suppose b proportion of the targets are exclusively in the short-term store, so $1 - b$ proportion of the targets are in the long-term store. Furthermore, suppose each post-encoding competing event has a probability of $1 - q$ for displacing the target in the short-term store. The resulting cumulative discrete probability for memory loss as a function of n post-encoding events is

$$\begin{aligned} F(n) &= b(1-q)[1 + q + q^2 + \cdots + q^{n-1}], \\ &= b(1-q)\frac{(1-q^n)}{1-q} = b(1-q^n). \end{aligned}$$

Since in general we do not know the value of n, it is reasonable to assume that $n = mt$ where $m > 0$. This assumption enables a re-expression of the cumulative probability as a function of time. It is also convenient to express the term q^{mt} as e^{-at}, where $a = m \log \frac{1}{q} > 0$.

Thus, the survivor function is $S(t) = 1 - b + be^{-at}$. As shown in Section 2.3, the resulting hazard function is monotonically decreasing. It also follows from Corollary 1 that the density function is monotonically decreasing with $\lim_{t\to U} f(t) \to 0$.

3.3. Modified Logarithm Model

In a classic book on experimental psychology [49], the memory retention function was hypothesized to be $b - m\log(t)$, but this proposal does not satisfy two memory boundary conditions. The modified logarithm version shown in Table 1 corrects the original proposal so as to enable $S(0) = 1$ and to allow for some long-term memory. The modified survivor function is $1 - \frac{b\log(t+1)}{\log(U+1)}$ for $t \in [0, U]$. The corresponding density function is $f(t) = \frac{b}{(t+1)\log(U+1)}$ and $\lim_{t\to U} f(t) = \lim_{t\to U} \frac{b}{(t+1)\log(U+1)} \to 0$. Furthermore the corresponding $g(t) := \frac{-f'(t)}{f(t)}$ function is simply $\frac{1}{(t+1)}$. Because $g'(t) = \frac{-1}{(t+1)^2} < 0 \; \forall \; t \in [0, U]$, it follows from Theorem 2 that the hazard is monotonically decreasing.

3.4. Modified Power Model

Several investigators (e.g., [50,51]) argued for a power function survivor model of the form $\frac{b}{t^c}$. Similar to the other modified models, the original function is modified to satisfy the two memory boundary conditions. The resulting hazard function shown in Table 1 is monotonically decreasing because the numerator is a constant and the denominator increases with t. It also follows from Corollary 1 that the density function is monotonically decreasing with $\lim_{t\to U} f(t) = \lim_{t\to U} \frac{bc}{(t+1)^{c+1}} \to 0$.

3.5. Modified Single-Trace Fragility Theory

Wickelgren [52] developed a single-trace model where the strength decreased as a function of time. If the maximum strength at $t = 0$ is rescaled to be 1 rather than a parameter ρ, then the single-trace fragility theory would predict a survivor function of $\frac{\exp(-at)}{(1+bt)^c}$. For this form of the single-trace fragility model, the hazard function would be $a + \frac{cb}{1+bt}$, which is monotonically decreasing. However, this form of the survivor function does not allow for any lifetime memories, but with the modification shown in Table 1, $1 - d$ proportion of the traces survive indefinitely. Nonetheless the hazard function shown in the table is still monotonically decreasing in time because the numerator is decreasing with time while the denominator is increasing with increasing time. Because of Corollary 1, it is known that the density function is monotonically decreasing. Moreover $\lim_{t\to U} f(t) \to 0$.

3.6. Modified Anderson-Schooler Model

Anderson and Schooler [53] proposed another version of a power model for a memory retention function. These authors suggest the function $S(t) = \frac{b}{b+t^c}$, which has the hazard function of $h(t) = \frac{c}{t+bt^{1-c}}$. The hazard is monotonically decreasing for $0 < c \leq 1$ because the denominator is increasing with time and the numerator is a constant. For $c > 1$, $h(0) = 0$, but for larger values of t, the hazard is decreasing because the denominator is approximately t. Consequently, for $c > 1$ this function has a peak-shaped hazard profile. However, the Anderson-Schooler version of the retention function does not enable the possibility of some lifetime memories. Hence, the modified survivor function shown in Table 1 provides for long-term memories. Despite this modification to the Anderson-Schooler function, the hazard has the same properties as did the original model. That is, the hazard is monotonically decreasing if $0 < c \leq 1$, but it has a peak shape for $c > 1$.

3.7. Ebbinghaus Model

This model is the original memory retention function from the first experimental psychologist to examine memory [11], although the logarithm term has been adjusted to $\log(t + 1)$ rather than $\log t$ to avoid the problem of the term being undefined at $t = 0$. The resulting hazard function shown in Table 1 is only defined for $c > 1$ because if c were

less than 1, then the hazard would diverge at $t = 0$. However, with $c > 1$, the hazard function is zero at $t = 0$. Moreover, for large t, the hazard is monotonically decreasing. Thus, the hazard function has a peak shape for $c > 1$.

3.8. Trace Susceptibility Theory

Unlike the previously discussed models, the trace susceptibility theory was designed with hazard in mind from the outset [17]. This theory is based on the premise that the memory consists of a network of features that becomes insufficiently stored when the weakest link is destroyed. Given the fact that the Weibull distribution is one of the two asymptotic limiting forms for a disjunction collection of features, the trace susceptibility theory is based on an assumed Weibull model, but to enable lifetime memories, it is a Weibull subprobability model. The shape parameter of the Weibull subprobability model is stipulated to be 2. Recall that a Weibull probability model with a shape parameter greater than 1 has increasing hazard, but a Weibull subprobability model with a shape parameter greater than 1 has peak-shaped hazard as discussed in Section 2.4.

3.9. Two-Trace Hazard Model

The previously discussed hazard models for memory were based on a single memory encoding or trace. In contrast to these single-representation models, the two-trace hazard model is a conjunctive hazard system [43]. Thus, a key feature of this model is the redundant encoding of the information. Survival of either of two traces is sufficient for the memory of an event to be retained. Trace 1 is hypothesized to be a Weibull probability model with a shape parameter of 2. The survivor function for Trace 1 is $S_1(t) = e^{-dt^2}$ and the hazard function for this trace is $h_1(t) = 2dt$, which is clearly an increasing hazard function. Thus, Trace 1 is not likely to survive over a long time period. Trace 2 is based on a Weibull subprobability model with a shape parameter $0 < c \leq 1$. The survivor function for Trace 2 is $S_2(t) = 1 - b + be^{-at^c}$. The corresponding hazard function for Trace 2 is

$$h_2(t) = \frac{ac}{t^{1-c}\left(e^{at}\frac{1-b}{b} + 1\right)}.$$

The hazard for Trace 2 is monotonically decreasing because the numerator is a constant and the denominator is an increasing function of time. Consequently only Trace 2 is consistent with Jost's law. So what is the hazard function for the system of two traces? Given $F_1(t) = 1 - S_1(t)$ and $F_2(t) = 1 - S_2(t)$, the resulting hazard function is

$$h(t) = \frac{2dte^{-dt^2}F_2(t) + bact^c e^{-at^c}F_1(t)}{1 - F_1(t)F_2(t)},$$

Because $F_1(0) = 0$ and $F_2(0) = 0$, it follows that $h(0) = 0$. However, for larger values for t, the conjunctive hazard converges to $h_2(t)$, which is a monotonically decreasing function of t. Thus, the hazard profile for the system of two traces is peak shaped (i.e., rising from zero initially before eventually decreasing for increasing time).

3.10. Scale Invariance and the Number of Fitting Parameters

Although all the survivor functions listed in Table 1 are a function of t, not all of the functions are scale invariant. The modified exponential model is an example of a model that is time-scale invariant. Suppose time is measured in units of seconds and with the parameter $a = 0.3$, the same survivor function can also represent the case where time is measured in units of minutes and where $a = 18$ because $a = \frac{0.3}{sec} \times \frac{60\ sec}{min} = \frac{18}{min}$. The value of the a parameter is scale sensitive, but the survivor function is free of measurement units. As in the physical sciences, it is important for the parameters to have measurement-scale dependence rather than the physical law having scale dependence. For example, in the classical formula for gravitation, $F = G\frac{m_1 m_2}{r^2}$, the force F, the masses m_1 and m_2, and the

distance r can be expressed in a variety of unit systems without changing the law. The only thing that does change is the value and the units of the G parameter. Five models in Table 1 are time-scale invariant in a similar fashion, but the other four models in the table are dependent on the units of time measurement employed. Consequently, these four models for memory hazard ought to be modified to be time-scale invariant.

The four models that require recasting for time-scale invariance are: the modified logarithm, the modified power, the modified Anderson-Schooler, and the Ebbinghaus function. The adjusted survivor functions for these models are shown in Table 2 along with the other functions.

Table 2. The nine candidate models for the memory survivor function $S(t)$ that are time-scale invariant. The parameters a, b, c and d are positive. The cumulative functions for the two respective traces of the two-trace hazard model are $F_1(t) = 1 - \exp(-d\,t^2)$ and $F_2(t) = b(1 - \exp(-at^c)$.

Model Name	Survivor Function	# Parameters
modified hyperbolic	$1 - b + \frac{b}{at+1}$	2
modified exponential	$1 - b + be^{-at}$	2
modified logarithm	$1 - b\frac{\log(at+1)}{\log(aU+1)}$	3
modified power	$1 - b + \frac{b}{(at+1)^c}$	3
modified single-trace fragility	$1 - d + \frac{de^{-at}}{(1+bt)^c}$	4
modified Anderson-Schooler	$1 - d + \frac{db}{b+(at)^c}$	4
Ebbinghaus	$\frac{b}{b+[\log(at+1)]^c}$	3
trace susceptibility theory	$1 - b + be^{-at^2}$	2
two-trace hazard theory	$1 - F_1(t)F_2(t)$	4

The adjustment to the four models to meet the criterion of being time-scale invariant alters the specific hazard function, but not so much to change either their monotonicity or peakedness. The hazard for the time-scale invariant form of the modified logarithm model is

$$h(t) \quad = \quad \frac{ab}{(at+1)\log\frac{aU+1}{(at+1)^b}}\,.$$

The hazard for the time-scale invariant version of the modified power model is

$$h(t) \quad = \quad \frac{abc}{(1-b)(at+1)^{c+1} + b(at+1)}\,.$$

The hazard for the time-scale invariant form of the modified Anderson-Schooler model is

$$h(t) \quad = \quad \frac{dbca^c}{t^{1-c}((1-d)[b+(at)^c]^2 + db[b+(at)^c])}\,.$$

Finally the time-scale invariant version of the Ebbinghaus function is

$$h(t) \quad = \quad \frac{ac[\log(at+1)]^{c-1}}{(at+1)(b+[\log(at+1)]^c)}\,.$$

4. Some Empirical Evidence About Memory Hazard

The nine proposals for the memory hazard were informed by results from experiments of different types that span different time frames [42]. Nonetheless, there is one agreed upon pattern – after some initial duration, the hazard is considered to be decreasing. The first five functions shown in Table 1 hypothesize a monotonically decreasing hazard function for all time, but the last four functions predict a peak-shaped hazard function. Consequently, a critical question to assess is: does hazard immediately after encoding increase or does

hazard decrease monotonically? To answer this question, experiments need to focus on the time period shortly after memory encoding because that is the temporal region where the models differ.

Yet it is technically challenging to obtain an empirical hazard value. If an experiment measures the memory after five different times, say $t_1, \cdots, t_5$, then it is possible to find the proportion of times where the target information is correct for each time (i.e, values for $S_i(t_i)$ for $i = 1, \cdots, 5$ can be measured). However, to obtain a valid $S(t)$ measurement, corrections are needed to take into account guessing processes and to account for the possibility that the performance might be also a function of retrieval processes. Thus, to obtain credible values for $S(t_i)$ requires a validated psychometric model so as to extract a measure of the probability of target storage [18]. However, even with these correction procedures, an empirical measurement of hazard has further challenges to obtain an estimate of $f(t)$, which is required to compute hazard. To estimate $f(t)$ a second $S(t_i + \Delta t)$ measurement is required. Empirical hazard estimates have potentially high statistical error because there is likely a large error in estimating $S(t_i) - S(t_i + \Delta t_i)$ that is compounded by dividing by $S(t_i) \times \Delta t$. Consequently, valid and statistically reliable empirical hazard estimates are very difficult to obtain in practice. However, it is possible to use another surrogate function to glean some useful information about the hazard function without having to actually estimate the hazard function directly.

4.1. Using a Surrogate Function for Assessing the Hazard Function

The function $u(t) := \frac{F(t)}{t}$ can be used to ascertain if the early part of the hazard function is increasing or decreasing [43,54]. Experimentally $u(t)$ can be readily estimated because t is known as is $F(t) = 1 - S(t)$. The rationale for how $u(t)$ can be used as a surrogate function for ascertaining hazard properties is established with the following two theorems.

Theorem 6. *If a continuous probability or subprobability on support $t \geq 0$ has monotonically decreasing hazard, then $u(t) = \frac{F(t)}{t}$ is monotonically decreasing with a maximum value of $u(0) = f(0)$.*

Proof. Given a monotonically decreasing hazard function on the support $t \geq 0$, it is known from Corollary 1 that the density $f(t)$ is monotonically decreasing for all t with the maximum at $t = 0$. From L'Hospital's rule it follows that $\lim_{t\to 0} u(t) = \lim_{t\to 0} \frac{F(t)}{t} = f(0)$. Furthermore,

$$u'(t) = \frac{F'(t)t - F(t)}{t^2} \tag{21}$$

$$= \frac{1}{t}[f(t) - u(t)], \tag{22}$$

For any $t > 0$, $u(t) = \frac{\int_0^t f(r)\,dr}{t}$. However, from the mean-value theorem of integral calculus, the integral $\int_0^t f(r)\,dr = tf(t_m)$, where $0 < t_m < t$, so $u(t) = \frac{t f(t_m)}{t} = f(t_m)$. Thus, from (22), it follows that the sign of $u'(t)$ is directly a function of the term $f(t) - f(t_m)$. Because $f(t)$ is monotonically decreasing and because $t_m < t$, it follows that $u'(t) < 0$. Hence for monotonically decreasing hazard functions on support of $t \geq 0$, $u(t)$ must be monotonically decreasing for all t with a maximum at $u(0)$. □

Theorem 7. *If the function $u(t) = \frac{F(t)}{t}$ is monotonically increasing for $t \in [0, t_p]$, then the hazard $h(t)$ is increasing in that interval.*

Proof. For all $t \in [0, t_p]$, $u'(t) = \frac{1}{t}[f(t) - u(t)]$ with $u(t) = \frac{t f(t_m)}{t} = f(t_m)$ where $0 < t_m < t_p$. Because $u(t)$ is increasing over the interval $[0, t_p]$, it follows that $f(t) - f(t_m) > 0$

over this interval. Thus, $f'(t) > 0$ over this interval, and it follows that $g(t) = \frac{-f'(t)}{f(t)} < 0$. However, from Equation (10), it follows that $h(t)$ is increasing because $h(t) > g(t)$ in this interval. □

Theorems 6 and 7 indicate that there is a way to critically falsify any theory for memory hazard that predicts strictly monotonically decreasing hazard. Namely if evidence can be found that $u(t)$ is initially increasing, then that finding would be falsifying evidence for theories that predict monotonically decreasing hazard. Yet to find a statistically reliable difference in $u(t)$ estimates requires a reasonably large sample size, which leads to the practical reality that experimentation on $u(t)$ requires pooling data over different research participants and test trials. However, from Theorem 3, it is known that a mixture of non-increasing hazard functions results in a monotonically decreasing hazard. Consequently if all participants in a group have monotonically decreasing hazard, then the hazard function for the pooled data is also monotonically decreasing. Thus, an experiment that has a large sample size for each retention interval and has research protocols that enable model-based correction estimates for the probability for memory storage has a chance to see if $u(t)$ is increasing in the early part of the retention interval. Of course if the experiment does not examine time values in the critical period where a hazard function peak exits, then the study might miss detecting the hazard peak.

Experiment 3 from [25] was designed in a fashion to enable estimates of $u(t)$ for various short retention intervals, although the measurement of $u(t)$ was not the purpose of the study.(The study addressed a different set of questions associated with implicit and explicit memory. The experiment was designed to see if implicit memory had a different retention function than explicit memory. Implicit memory is exemplified when a participant cannot recall or confidently recognize an item but nonetheless can correctly select the item from a lineup list.) The study examined the memory of novel nonsense letters taken from a stimulus set that was scaled to be of low meaningfulness [55]. Following the presentation of a target letter triad, there was a series of digits that participants had to verbally repeat as soon as they heard the digits. The target letters and the digits were presented over headphones. The testing involved a decision about four visually displayed letter triads. For half of the test probes none of the stimuli matched the target item for that trial, and the other half had one of the stimuli match the target. The participant had to respond *yes* or *no* if any of the triads on the four-item list of triads was the target for that trial. The participants also had to indicate if they were of high or low confidence in their decision.(Given that the original paper was focused on measuring explicit and implicit memory, there was also a series of forced-choice tests after the yes/no decision; however, those forced-choice data are not pertinent for the $u(t)$ analysis.) For each retention interval there were 696 tests where the target is on the list and another 696 tests where the target was not on the list.

The analysis in the current paper is based on model 7B from [23] for estimating the storage probability for each retention interval. The data needed for using model 7B is based on a yes-no recognition memory test along with high-versus-low confidence judgments. Table 3 defines the response cells frequencies n_i associated with the experimental task. The frequencies for old recognition tests $n_1, \cdots, n_4$ form a four-cell multinomial that has in the population the proportions $\phi_1, \cdots, \phi_4$ where $\sum_1^4 \phi_i = 1$. Similarly for the new recognition test there are population proportions $\phi_5, \cdots, \phi_8$ where $\sum_5^8 \phi_i = 1$. The proportions for the cells of the two multinomials are a function of memory and task factors. Psychological processes such as sufficient storage, guessing when target storage is insufficient, and rating processes influence the proportions in the multinomial outcome cells. Model 7B is a psychometric model that describes how those psychological processes map onto the multinomial outcomes. The parameters of the 7B model are denoted by subscripted θ symbols for processes such as for target storage (θ_S), guessing *yes* on old recognition (θ_g), guessing *no* on new recognition ($\theta_{g'}$). There are also parameters for using the *sure* versus *unsure* rating when guessing. The model structure consists of two probability trees for the

old and new recognition testing procedures. Psychometric models of this type have come to be called multinomial processing tree (MPT) models.

Table 3. The frequencies labels are shown for the various response categories for the experimental task that is associated with model 7B from [23]. Participants responded either *yes* or *no* to the test probe and rated their decision either as *sure* or *unsure*. The test probes were either *old* or *new* corresponding to the probe, respectively, being a memory target or a novel item. For each outcome in the Table there is also a population proportion ϕ_i, $i = 1, \cdots, 8$.

Test Type	No Sure	No Unsure	Yes Unsure	Yes Sure
old	n_1	n_2	n_3	n_4
new	n_5	n_6	n_7	n_8

The latent θ parameters can be statistically estimated by means of a procedure called population parameter mapping (PPM) [56–58]. The parameters of model 7B can also be estimated with either a frequentist maximum likelihood method or with a conventional Bayesian method, but the PPM procedure provides additional information about model coherence and is computationally faster [23]. PPM is a specialized Bayesian Monte Carlo sampling procedure developed to estimate the latent parameters of a scientific model for multinomial data. This procedure was invented as a Bayesian Monte Carlo sampling method that avoided the lower efficiency of Markov chain Monte Carlo (MCMC) algorithms. Unlike MCMC procedures, the PPM procedure is not an approximate method, and it does not require a burn-in period [58]. Moreover, the sampled vectors are independent, so the issue of correlated samples, which can occur with MCMC algorithms, is not a problem. With multinomial data there is an exact Monte Carlo sampling method for sampling vectors of points (ϕ_i) from the posterior distribution for sets of multinomial data [56–58]. Each vector from ϕ spaced is mapped to a corresponding θ-space vector for the scientific model parameters. The collection of coherently mapped points from ϕ space to θ space are used for point and interval estimation of the latent model parameters. Chechile [23] proved that the 7B model is identifiable and demonstrated that the PPM estimation method out performed (in terms of accuracy) the frequentist maximum likelihood estimation, especially for sample sizes less than 1000. For sample sizes greater than or equal to 1000, the maximum likelihood estimate and the PPM estimate are virtually the same. See the Appendix A for additional details about the PPM method for estimating the storage parameter for model 7B as well as for seeing the software for implementing the estimation.

The frequencies for the various temporal delays for experiment 3 from [25] are provided in Table 4. The time values shown in the table are one second longer than the digit-repeating interval in order to give the participants a chance to read the four-item visual list to which they had to respond either *yes* or *no*. The participants were instructed to respond *yes* if any of the four items on the test probe was the target stimulus for that trial. Perfect performance on this task would correspond to 696 trials of correct performance with high confidence on the target-present tests and 696 trials of correct performance with high confidence on the target-absent tests. These values were used to create an initial set of $\theta_S(0)$ values as a baseline for finding $u(t)$.

Table 4. The table provides the response frequencies from experiment 3 in [25] for each temporal delay.

Time	$(\mathbf{n_1}, \cdots, \mathbf{n_4})$	$(\mathbf{n_5}, \cdots, \mathbf{n_8})$
1.33	(13, 4, 9, 670)	(658, 28, 4, 6)
2.33	(33, 21, 39, 603)	(610, 69, 6, 11)
5	(81, 79, 165, 371)	(378, 223, 76, 19)
13	(42, 194, 275, 185)	(163, 341, 177, 15)
31	(36, 294, 265, 101)	(123, 369, 188, 16)

The key information for the $u(t)$ analysis for the data from experiment 3 in [25] is provided in Table 5. For each t value the posterior median for $\theta_S(t)$ is provided. The corresponding $u(t)$ is the median of the set of mapped values for $u(t) = \frac{1}{t}[\theta_S(0) - \theta_S(t)]$. The 95-percent interval for each condition is based on the quantiles for this set of $u(t)$ values. See the Appendix A for the details about the storage estimation software.

Table 5. Estimates for $u(t)$ for data from experiment 3 in [25].

Time in s.	Storage Estimate	u(t)	95 % u(t) Interval
1.33	0.938	0.041	[0.002, 0.088]
2.33	0.764	0.098	[0.061, 0.178]
5	0.470	0.104	[0.091, 0.116]
13	0.230	0.059	[0.054, 0.062]
31	0.111	0.028	[0.027, 0.030]

There is a clear maximum for $u(t)$ for $t > 1.33$. There is a posterior probability of 0.9788 that $u(t = 2.33) > u$(t=1.33), and there is a probability of 0.9891 that $u(t = 5) > u(t = 1.33)$. For times greater than $t = 5$, there is a decline in $u(t)$ because (1) the probability that $u(t = 5) > u(t = 13)$ is 0.9975 and (2) because the probability that $u(t = 13) > u(t = 31)$ is 0.9994.

Thus, via Theorem 6 the data from this experiment are inconsistent with any theory that predicts strictly monotonically decreasing hazard. Additionally, via Theorem 7, it is clear that the hazard is increasing over a brief initial interval of approximately five seconds after the initial encoding. Thus, this experiment has falsifying evidence against the first five models listed in Table 2. However, more strongly, the experiment is inconsistent with any model (either an existing theory or a yet to be generated theory) that predicts strictly monotonically decreasing hazard. Yet, given the import of this strong conclusion, it is prudent to explore if this general conclusion can be replicated and supported by other experiments.

Experiment 2 from [17] is a study that provides an assessment of $u(t)$ over a time frame from 1.33 to 76 s. In this experiment the targets are again letter triads of low meaningfulness. Additionally, the retention interval is filled with digits that had to be repeated (i.e., letter shadowing). The target stimulus, the test probe, and the interpolated digits are all auditory, but the gender for the articulation of the target item and test probe is different from the gender used for the interpolated digits. The probe stimulus for each recognition test was either the original target or a novel triad of nonsense letters. The data for each condition is provided in Table 6.

Table 6. The table provides the response frequencies from experiment 2 in [17] for each temporal delay.

Time	$(\mathbf{n_1}, \cdots, \mathbf{n_4})$	$(\mathbf{n_5}, \cdots, \mathbf{n_8})$
1.33	(1, 0, 1, 898)	(880, 6, 9, 5)
2.33	(8, 3, 13, 876)	(858, 26, 15, 1)
5	(49, 80, 142, 629)	(646, 179, 53, 22)
13	(74, 226, 283, 317)	(434, 329, 99, 38)
32.67	(63, 281, 314, 242)	(296, 412, 143, 49)
76	(61, 322, 309, 208)	(244, 462, 166, 28)

Similar to the analysis in Table 5, the key results for experiment 2 from [17] are provided. The time values listed in the table are one second longer than the time for digit shadowing in order for the participant to hear the probe stimulus. Perfect performance in this experiment would consist of 900 trials with high confidence correct old recognition, and 900 trials with high confidence correct rejection of the foil probes. Those values were used for generating the vector of baseline $\theta_S(0)$ values need for the $u(t)$ calculations.

For each temporal delay the values for posterior median for $\theta_S(t)$, $u(t)$ and the 95-percent interval for $u(t)$ are provided. There is a clear maximum for $u(t)$ at $t = 5$. The posterior probability that $u(t = 5) > u(t = 2.33)$ is greater than 0.9999, and the probability that $u(t = 5) > u(t = 13)$ is 0.9971. Furthermore there is a probability of 0.9994 that $u(t = 13) > u(t = 32.67)$, and a probability of 0.9997 that $u(t = 32.67) > u(t = 76)$. Thus, this experiment provides further support for the conclusion that memory hazard initially increases before the hazard declines with time.

Experiment 1 from [17] is a third study for which $u(t)$ can be estimated for several times in the critical time period after memory encoding. The experimental protocols for this experiment are the same as for the study upon which Table 7 is based, except for the sample size. In this study, there were in each condition 240 old recognition tests and 240 new recognition tests.(There is another set of four retention intervals that were examined with a slower rate for the interpolated digit-shadowing task. These four conditions are omitted because that rate of presenting digits in the retention interval was too slow to prevent participants from rehearsing the target items [17].) For this study there is not a five-second condition, but instead there are two intervals bracketing the five-second retention interval (i.e., intervals of $t = 3.67$ and $t = 6.33$). The frequency data are in Table 8 and the $u(t)$ results are in Table 9. Interestingly $u(t)$ is not statistically different between the two retention intervals that bracket five seconds (i.e., the $t = 3.67$ and $t = 6.33$ conditions). Both of these conditions did have a reliably larger $u(t)$ than the $u(t)$ for the $t = 2.33$ condition (with probability greater than 0.985 for both cases), and both of these conditions had a $u(t)$ larger than $u(t)$ for the $t = 11.67$ condition (with a probability greater than 0.986 for both cases).

Table 7. Estimates for $u(t)$ for data from experiment 2 in [17].

Time in s.	Storage Estimate	u(t)	95 % u(t) Interval
1.33	0.994	0.001	[0.000, 0.007]
2.33	0.969	0.011	[0.000, 0.017]
5	0.630	0.073	[0.062, 0.084]
13	0.230	0.059	[0.054, 0.064]
32.67	0.148	0.026	[0.024, 0.028]
76	0.172	0.011	[0.010, 0.011]

Table 8. The table provides the response frequencies from experiment 1 in [17] for each temporal delay.

Time	$(\mathbf{n_1}, \cdots, \mathbf{n_4})$	$(\mathbf{n_5}, \cdots, \mathbf{n_8})$
2.33	(5, 5, 6, 224)	(224, 13, 1, 0)
3.67	(10, 29, 48, 153)	(187, 52, 1, 0)
6.33	(26, 55, 95, 64)	(114, 103, 20, 3)
11.67	(9, 83, 106, 42)	(45, 133, 59, 3)

Table 9. Estimates for $u(t)$ for data from experiment 1 in [17].

Time in s.	Storage Estimate	u(t)	95 % u(t) Interval
2.33	0.908	0.031	[0.000, 0.071]
3.67	0.551	0.116	[0.075, 0.233]
6.33	0.194	0.124	[0.102, 0.144]
11.67	0.147	0.071	[0.060, 0.077]

The above three experiments indicate that the human memory hazard function is not strictly monotonically decreasing as predicted by the first five models listed in Table 2. These studies when analyzed in light of Theorems 6 and 7 establish a framework for using an experimentally practical surrogate function $u(t) = \frac{F(t)}{t}$ to ascertain that the human memory system has a peaked-shaped hazard function. This approach has the advantage of learning

about hazard empirically without actually estimating hazard per se. The method thus avoids the technical and statistical difficulties of estimating a density-based metric. Yet the research still has the limitation of only examining human participants. It might be possible that some animal species might exhibit a different memory hazard function. However, it is difficult to measure the $u(t)$ surrogate function in a careful fashion with animals because they need to be trained to do a complex recognition model task. Human participants are capable of following complex instructions that allow for model-based corrections to obtain a measure of memory storage after various retention intervals. Nonetheless it is important to understand the properties of the human memory system even if humans differ from other animals. In the next subsection a case is made for a particular function that has peak-shaped hazard for human memory.

4.2. Evidence for the Two-Trace Hazard Model

Given the results from the previous subsection, attention is focused on the four models that predict a peak-shaped hazard function (i.e., the modified Anderson-Schooler model, the Ebbinghaus function, the trace susceptibility theory, and the two-trace hazard model). For each theory there is a survivor function that predicts target storage as a function of time, and each theory has fitting parameters that can be adjusted for each individual. Although the pooled data were used to assess the $u(t)$ function, the averaging of nonlinear functions with different fitting parameters can be misleading as shown by Estes [59]. Thus, for testing the fit quality of the various models, it is important to do the testing separately for each person. Moreover, to test the fit quality on an individual basis requires a sizable sample size for each retention interval to reduce the uncertainty in the estimated storage probability. It is also important to assess the storage probability of memory targets over both the short-term and the long-term temporal regions.

In the short-term time frame, experiment 2 from [17] is particularly important because each of 30 participants was evaluated over six retention intervals with 90 replication tests per interval; thus the data in this study enables a reasonably precise estimate of target storage for each condition and for each participant. For this analysis the two-trace hazard model, with its four parameters, was omitted because it could always be as effective as the trace susceptibility theory. Both the two-trace hazard model and the trace susceptibility theory employ a Weibull model. Consequently, if the b parameter of both models are equivalent and if the d parameter of the two-trace hazard model is equated to the a parameter of the trace susceptibility theory, then there are always values for the extra a and c parameters of the two-trace hazard model that effectively match the prediction equation of the trace susceptibility theory.

The measure of fit quality for each model was the correlation between the estimated storage probability θ_S and the predicted value from the theory. The r^2 means for the trace susceptibility theory, the Ebbinghaus function, and the modified Anderson-Schooler model are, respectively, 0.938, 0.905, and 0.846 [43]. Importantly, the modified Anderson-Schooler model was not the best fit for any of the 30 participants. The trace susceptibility theory was the best fit for 22 participants and the Ebbinghaus function was the best fit for the other 8 participants. Given the fact that the Ebbinghaus function had a better fit than the trace susceptibility theory for 8 participants, these people were also examined in terms of the two-trace hazard model. The two-trace hazard model had a better fit for each of those 8 participants than did the Ebbinghaus function. Thus, in the short-term time frame, the trace susceptibility and the two-trace hazard model have a better fit to the storage probability than either the modified Anderson-Schooler or the Ebbinghaus model.

There have not been very many studies that report data for individuals in the long-term temporal domain. Nearly all the data sets are based on averaging data over participants. Yet Estes [59] showed that learning and retention curves based on grouped data are problematic. Consequently, Chechile [43] was only able to discuss two studies that had individual-participant data that enabled estimates of the storage probability over a long-term temporal region [11,60]. Chechile [43], who developed both the trace susceptibility theory and the

two-trace hazard model, reported that the trace susceptibility theory failed to fit this long-term data, but the two-trace hazard model did fit each participant well. However, the two studies reported by Chechile only had results from three people, who were either the author or a relative of the author [11,60]. However, another study by Sloboda [61] examined nine participants over ten retention intervals in a fashion that enables the estimation of the model 7B storage probability value for each person for each interval. The ten intervals examined were: 30 s, 5 min, 15 min, 30 min, 1 h, 3 h, 9 h, 24 h, 72 h, and 144 h. The target stimuli in the Sloboda experiment were meaningful words selected from a word frequency norm [62]. There were 120 tests for each of the ten retention intervals (i.e., 60 target-present trials and 60 target-absent trials). Sloboda assumed that trace 1 of the two-trace hazard model was no longer available for retention intervals of 5 min or longer. If we denote the storage probabilities for traces 1 and 2 as, respectively, θ_{S_1} and θ_{S_2}, then $\theta_{S_1} = e^{-dt^2}$ and $\theta_{S_2} = 1 - b(1 - e^{-at^c})$. The mean value for d was 12.96 min^{-2} for the previously reported 30 participants in the short-term time domain; consequently by 5 min the value for θ_{S_1} would be less than 10^{-140}. Thus, Sloboda only fitted $\theta_{S_2} = 1 - b(1 - e^{-at^c})$ to the long-term storage probability values. The mean and standard deviation for the fits of the model for the a, b, and c parameters, respectively, are: $1.9(1.1)\,\text{min}^{-2}$, $0.991(0.009)$, and $0.238(0.043)$. The resulting fit is excellent for each of the nine participants; the mean correlation between the model predicted values and storage probabilities is 0.947.

Thus, the currently available data are sufficient to indicate that the modified Anderson-Schooler model, the Ebbinghaus model, and the trace-susceptibility theory have problems fitting all the participants; whereas the two-trace hazard model has an excellent fit to each participant. Consequently, until a better theoretical function is found, the two-trace hazard model is the current best option for a model of memory retention as a function of time. In this model, the storage of the target is $\theta_S = 1 - (1 - \theta_{S_1})(1 - \theta_{S_2})$ with $\theta_{S_1} = e^{-dt^2}$ and $\theta_{S_2} = 1 - b(1 - e^{-at^c})$.

5. Discussion

The basis of any human memory is the development of an internal representation of experienced events. However, after the initial encoding there are dynamic changes to the internal representation that alter the likelihood for remembering the original event. The hazard function is a natural mathematical tool for studying the dynamic changes of a stochastic system. Thus, much can be learned about the properties of the mental representation and the process of forgetting by an examination of the memory hazard function. This system-level approach for studying the memory retention process avoids neuroscience measurements of the memory process, but nonetheless the system-level approach results in powerful constraints on memory theory and the structure of the memory representation.

In this paper, the hazard properties of nine models for memory retention were studied; each of these theories emerged from previous research in psychological science. For all of these theories the hazard decreases for long delays, but four of the models predicted hazard is initially increasing in the early stages of memory retention before a peak is reached. By examining a surrogate function, which is easier to experimentally measure than the hazard function, a case is made that hazard has a peak shape. Furthermore, goodness-of-fit analyses of selective studies found that there was evidence against three of the four theories that correctly predicted peak-shape hazard. The sole surviving model is the two-trace hazard model. A central assumption for this theory is that the initial encoding establishes two independent representations or traces. Trace 1 is assumed to have storage probability at time t that is $\theta_{S_1} = e^{-dt^2}$, which is a Weibull distribution with a shape parameter of 2. The hazard for trace 1 is an increasing function of time, so this trace is not likely to survive very long. Trace 2 is assumed to have a storage probability at time t that is $\theta_{S_2} = 1 - b(1 - e^{-at^c})$. Trace 2 is a Weibull subprobability distribution with a shape parameter of $0 < c \leq 1$. Trace 2 has a monotonically decreasing hazard function. It is possible that some proportion of the trace 2 representations survive over the lifetime of the individual (i.e., $1 - b$ proportion of the events can survive where $0 \leq b \leq 1$).

The redundant memory traces, where the availability of either trace is sufficient for target memory retention, results in a system that has peak-shaped hazard. It is also a system that has an increased memory span for retaining a set of m concurrent items without memory loss. Given this theory there are a number of interesting questions to discuss.

To begin, it is reasonable to ask: why are there two independent encodings, and why do these traces have survivor functions that are related to a Weibull distribution? The answer to the second of these questions has to do with special properties of the Weibull distribution as a survivor model. Products, which are complex systems with many components that must all function properly, will fail whenever one component of the system fails. It is reasonable to consider a memory encoding as a complex combination of sensory features activated by the target stimuli along with an elaborated set of participant-generated cognitive processes that interpret the current stimulus and link it to temporal and spatial markers as well as link it to previous associations [18]. The encoding is not simply a mapping of an external stimulus to the activation of sensory features, but rather it involves a wide range of participant-generated elaborations that situates the current event into a larger context. This system of activated features and associations is bound together. Although some features of the encoding are trivial aspects of the event, which can be lost without losing the memory of the event, there is a set of essential components that must be retained to have a complete memory of the event. With the loss of any one of the critical components of this bound network results in a change from the initial state of complete memory storage. For a complex system such as a memory representation, the failure of the system is not captured by the mean of the distribution of the component lifetimes; rather it is the distribution for the weakest component. Since time is a nonnegative value, there are two mathematical distributions for the asymptotic limiting form for a minimum lifetime (i.e., the Gumbel and Weibull distributions [41]). However, the Gumbel distribution and a Gumbel subprobability distribution have monotonically increasing hazard for all t, so it is not a suitable model for memory failure because all the theories of memory hazard predict that the hazard is decreasing for long temporal delays. However, the Weibull distribution can have monotonically decreasing hazard, constant hazard, or monotonically increasing hazard, depending on the value of a shape parameter. Thus, the model for trace 1 is a Weibull probability model with a shape parameter set equal to 2, so this trace has increasing hazard. The model for trace 2 is a Weibull subprobability model that has a shape parameter of $0 < c < 1$, so it has monotonically decreasing hazard. Note that a single Weibull distribution does not have a peak-shaped hazard function, but a conjunctive system of two independent Weibull-related traces with different hazard monotonicity has a peaked-shape hazard function. Hence, one reason for having encoding redundancy is it enables a combination of Weibull systems to have a peak-shaped hazard function.

There is a neuropsychological model of memory encoding that posits multiple representations with different temporal scales (i.e., the Teyler-DiScenna index theory [63]). This neural model describes memory encoding as an interaction between two different brain regions– the hippocampus (a deeper brain structure in both hemispheres) and the neocortex. Previous research has linked the hippocampus with learning (e.g., [64]); whereas other researchers have shown an important role of the neocortex in learning (e.g., [65]). In the indexing theory, the sensory detection of the stimulus activates neocortical cells, but these activated neocortical cells map to the two hippocampi, which in turn can reactivate at a later time the neocortical cells associated with the stimulus. It is outside the scope of this paper to elaborate in detail the indexing theory, but the theory is mentioned here to support the plausibility of two representations of the target event being established.

The two-trace hazard model also has some similarity to an information-processing model of learning and memory – namely the multiple-store model that was discussed in Section 3.2 [47,48]. As the name implies, the multiple-store or the working-memory model has two stores, but the long-term store is hypothesized as a permanent storage system, so not all targets are stored in this system. The essential differences between the two-trace hazard model and the multiple-store model are: (1) all targets in the two-trace hazard

model, which have been encoded so that $S(0) = 1$, have two representations, (2) each representation in the two-trace model is susceptible to memory loss rather than only the items in the short-term store, and (3) the overall hazard for the two-trace model has a peak shape, whereas the multiple-store model has monotonically decreasing hazard.

It is also reasonable to ask: why is it that memory hazard is decreasing after the initial short-term increase? Manufactured products tend to wear in time after an initial high-risk, so the hazard for these products is typically a U-shaped function rather than a peak-shaped function. Memory retention has clearly different hazard properties from manufactured products. Two hypotheses have been proposed to explain this difference. The first hypothesis proposes that the memories that survive early threats from competing new information become hardened or immune to the effects of interference. This hypothesis is typically called *memory consolidation theory* [66–68]. Consolidation theory assumes that new encodings are fragile like a fresh painting. However, with the passage of time, the encoding is hardened into a permanent representation like the drying of wet paint. While consolidation theory can account for the reduced tendency for forgetting for older memories, this theory has been critiqued by several investigators [18,69,70]. Among the reasons for doubting the consolidation hypothesis is the forgetting that has been observed when the memory targets are given more training [69]. This effect has caused consolidation theorists to argue that additional learning softens a previously hardened trace [68]. However, critics find this position theoretically problematic. Another reason for doubting the consolidation hypothesis is its failure to account for the phenomenon of storage proactive interference. Proactive interference is the higher rate for forgetting of the recent memories that share some similarity to prior events [17,71]. The hardening of a memory trace with time alone is insufficient to account for proactive interference.

The second hypothesis to account for the reduced risk of older memories stresses the fact that time itself is part of the encoding of information [18]. For example, if the target is the nonsense triad such as BXT, then this stimulus with its ordinal letter positions are linked to a larger spatial and temporal context. Subsequent events are also encoded in terms of their stimulus features along with its temporal/spatial cues. New events compete with the prior events for storage, and the vigor of the interference is related to the degree of similarity between the events. Time is a feature of this assumed similarity-based competition, so events after a much greater time are more likely to compete with other events in that time frame and not with the older memories. Hence the hazard for an older item is reduced from that of the more recent item. Thus, this alternative account for the reduced risk with time is based on the importance of time as an important aspect of memory encoding. Yet the study of time as a separate aspect of memory encoding is still at an early stage of development, so further research is needed to delineate its role in memory organization and recollection [72].

Finally it is important to acknowledge the limitations to the conclusions of this paper. Memory is a basic property of living systems. It is important for all organisms to model the environment to increase the likelihood for survival. However, the hazard properties of memory studied in this paper has been limited to a single species – humans. It may well be the case that other species, which have a divergent evolutionary history, have different memory mechanisms. Consequently the two-trace hazard model might not be valid for other species. Given the importance of having a good model of the environment, it is reasonable to expect that evolutionary mutations that improve memory are likely to emerge. For example, having two representations of events increases the chance that the information is preserved longer to enable more complex cognitive operations. Yet not all evolutionary paths might have this information-processing advantage. The study of animal cognitive processes is methodologically challenging, but a memory hazard function is quite general as are the theorems highlighted in this paper.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in Section 4.1 are provided in Tables 4, 6, and 8, and these data were taken from previously published open sources that are cited.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MCMC	Markov chain Monte Carlo
MPT	multinomial processing tree
PPM	population parameter mapping

Appendix A. PPM Details for Model 7B and the Software for θ_S Estimation

Model 7B is a MPT representation that accounts for how the proportions associated with the response outcomes shown in Table 3 ϕ_i, $i = 1, \cdots, 8$, are linked to underlying latent parameters that are involved in the experimental task [23]. Parameters ϕ_i, $i = 1, \cdots, 4$ correspond to the parameters of the Dirichlet distribution for the old recognition test multinomial, and ϕ_i, $i = 5, \cdots, 8$ correspond to the parameters of the Dirichlet distribution for the new recognition test multinomial. The statistical estimation of the latent parameters disentangles the important cognitive processes that are involved. Since this paper is focused on the underlying storage parameter, a simplified version of model 7B is described here where only ϕ_3, ϕ_4, ϕ_7, and ϕ_8 proportions are used to estimate θ_S. The model 7B representation for these proportions is shown below.

$$\begin{aligned}
\phi_3 &= (1-\theta_S)\theta_g(1-\theta_2)\,,\\
\phi_4 &= \theta_S + (1-\theta_S)\theta_g\theta_2\,,\\
\phi_7 &= (1-\theta_k)(1-\theta_{g'})(1-\theta_2)\,,\\
\phi_8 &= (1-\theta_k)(1-\theta_{g'})\theta_2\,,
\end{aligned}$$

where θ_k, θ_g,$\theta_{g'}$, and θ_2, respectively, are: the probability of confident foil rejection based on some memory target knowledge, the probability of guessing *yes* on an old recognition test given insufficient storage, the probability of guessing *no* on a new recognition test given the lack of target knowledge, and the probability of giving a *sure* confidence rating following a *yes* despite having only insufficient target storage. The PPM mapping equation for θ_S is $\theta_S = \phi_4 - \frac{\phi_3\phi_8}{\phi_7}$. It is well known how to randomly sample a vector of values for the proportions of a Dirichlet distribution (e.g., see Theorem 3.5 on page 167 in [58]). Each of thousands of vectors for the Dirichlet parameters is obtained by the following set of operations:

$$\begin{aligned}
\phi_4 &= b_4\,,\\
\phi_3 &= (1-b_4)b_3\,,\\
\phi_8 &= b_8\,,\\
\phi_7 &= (1-b_8)b_7\,,
\end{aligned}$$

where

$$\begin{aligned} b_4 &\sim beta(\alpha_4, \alpha_{To} - \alpha_4)\,, \\ b_3 &\sim beta(\alpha_3, \alpha_{To} - \alpha_3 - \alpha_4)\,, \\ b_8 &\sim beta(\alpha_8, \alpha_{Tn} - \alpha_8)\,, \\ b_7 &\sim beta(\alpha_7, \alpha_{Tn} - \alpha_7 - \alpha_8)\,, \\ \alpha_i &= n_i' + n_i\,, \\ \alpha_{To} &= \sum_{i=1}^{4} \alpha_i\,, \\ \alpha_{Tn} &= \sum_{i=5}^{8} \alpha_i\,, \end{aligned}$$

where $\sim$ denotes *'is distributed as'* and where n_i' are the prior values of the Dirichlet distribution. The prior used for this paper is the non-informative flat prior, which corresponds to the case where $n_i' = 1$ for $i = 1, \cdots, 8$. For each mapping the values for b_3, b_4, b_7, and b_8 are based on a set of random scores sampled from the specified beta distributions shown above. Random scores from a beta distribution can be done using the built-in base R command **rbeta()**.

It is possible that the mapping equation $\theta_S = \phi_4 - \frac{\phi_3\phi_8}{\phi_7}$ can result in an incoherent result when $\frac{\phi_3\phi_8}{\phi_7} > \phi_4$, but those ϕ-space points are rejected and are not mapped. Thus, the PPM method can estimate the probability of coherence for the model itself by the number of coherent mapping divided by the total number of mapping attempts. For this paper the collection of posterior values sampled for any experimental condition was implemented by a R program. The R function called **storage_7B()** was created for this task. Some comments are included in the source code shown below.

```
storage_7B<-function(prior=rep(1,8),freq,samples=50000){
#Function is to find thetaS for model 7B given recognition
#memory data. The~data is stored in a 8 cell
#vector for frequencies for old and new recognition.
#The first four cells correspond to the respective
# frequencies for no_sure, no_unsure, yes_unsure, yes_sure,
#for the old recognition tests, and~next four cells are
#corresponding frequencies for the new recognition tests.
#The only coherence condition is that thetaS is nonnegative.
#The prior vector is the default of c(1,1,1,1,1,1,1,1) that is
#joint flat Dirichlet prior. If~the user has an informative
#prior, then the prior vector needs to be inputted.
#The default value for samples is 50000, but~the user
#can enter their preferred value for the number of
#Monte Carlo samples.
if (length(freq)!=8)
{stop("data vector must have length 8 for the data frequencies.")}
thetaS=rep(0,samples)
postphin=prior+freq
alpha4=postphin[4]
alpha3=postphin[3]
alphaTo=postphin[1]+postphin[2]+postphin[3]+postphin[4]
alphaTn=postphin[5]+postphin[6]+postphin[7]+postphin[8]
alpha8=postphin[8]
alpha7=postphin[7]
b4=rbeta(samples,alpha4,alphaTo-alpha4)
```

```
b3=rbeta(samples,alpha3,alphaTo-alpha3-alpha4)
b8=rbeta(samples,alpha8,alphaTn-alpha8)
b7=rbeta(samples,alpha7,alphaTn-alpha7-alpha8)
phi4=b4
phi3=(1-b4)*b3
phi8=b8
phi7=(1-b8)*b7
c=0
for (i in 1:samples){
 tstry=phi4[i]-((phi3[i]*phi8[i])/phi7[i])
 if (tstry>=0){
 c=c+1
 thetaS[c]=tstry} else {c=c}
  }
thetaS=thetaS[1:c]
probcoh=c/samples
outlist<-list(c=c,probcoh=probcoh,thetaS=thetaS)
#outlist is the list of output values and vectors
}
```

After loading the above R code, one can obtain a vector of random values from the posterior distribution for θ_S by using the function. As an example let us consider the case where the frequencies as defined in Table 3 are, respectively, $(36, 294, 265, 101)$ and $(123, 369, 188, 16)$, and the user accepts the default flat prior and the default number for *samples* of $50,000$. In this case the user can implement the estimation at the R prompt by the following command:

A<- storage_7B(freq=c(36,294,265,101,123,369,188,16))

The output from this command is stored in a new R object called **A**, which resides in the R workspace. The user can see the number of coherent mapping from the command **A$c**. The proportion of coherent samples is obtained via the command **A$probcoh**. The vector of θ_S values can be further examined by the standard base-R commands. For example, the posterior median of the distribution can be found from the command **median(A$thetaS)**. Additionally, the command **quantile(A$thetaS,prob=0.975)** results in the value for the 97.5-percentile. If the user wants to have the results from two different conditions resident in the R workspace at the same time then a different name should be employed when calling the function again with different data (e.g, **A1<- storage_7B()**).

References

1. Steffensen, J.F. *Some Recent Researches in the Theory of Statistics and Actuarial Sciences*; Cambridge University Press: New York, NY, USA, 1930.
2. Bain, L.J. *Statistical Analysis of Reliability and Life-Time Models*; Marcel Dekker: New York, NY, USA, 1978.
3. Barlow, R.E.; Proschan, F. *Mathematical theory of reliability*; Wiley: New York, NY, USA, 1965.
4. Gross, A.J.; Clark, V.A. *Survival Distributions: Reliability Applications in the Biomedical Science*; Wiley: New York, NY, USA, 1975.
5. Mann, N.R.; Schafer, R.E.; Singpurwalla, N.D. *Methods for Statistical Analysis of Reliability and Life Data*; Wiley: New York, NY, USA, 1974.
6. Thomas, E.A.C. Sufficient Conditions for Monotone Hazard Rate and Application to Latency-Probability Curves. *J. Math. Psychol.* **1971**, *8*, 303–332. [CrossRef]
7. Townsend, J.T.; Ashby, F.G. Methods of Modeling Capacity in Simple Processing Systems. In *Cognitive Theory*; Castellan, N.J., Restle, F., Eds.; Erlbaum: Hillsdale, NJ, USA, 1978; Volume 3, pp. 199–239.
8. Luce, R.D. *Response Times: Their Role in Inferring Elementary Mental Organization*; Oxford University Press: New York, NY, USA, 1986.
9. Chechile, R.A. Mathematical Tools for Hazard Function Analysis. *J. Math. Psychol.* **2003**, *47*, 478–494. [CrossRef]
10. Lappin, J.S.; Morse, D.L.; Seiffert, A.E. The Channel Capacity of Visual Awareness Divided Among Multiple Moving Objects. *Atten. Percept. Psychophys* **2016**, *78*, 2469–2493. [CrossRef] [PubMed]
11. Ebbinghaus, H. *Memory: A Contribution to Experimental Psychology*; Teachers College Press: New York, NY, USA, 1885.

12. Münsterberg, H.; Campbell, W.W. Studies from the Harvard Psychological Laboratory II. *Psychol. Rev.* **1894**, *1*, 441–495. [CrossRef]
13. Brown, A.S. Some Tests of the Decay Theory of Immediate Memory. *Q. J. Exp. Psychol.* **1958**, *10*, 127–181. [CrossRef]
14. Peterson, L.R.; Peterson, M.J. Short-Term Retention of Individual Verbal Items. *J. Exp. Psychol.* **1959**, *58*, 193–198. [CrossRef]
15. Shepard, R.N.; Teghtsoonian, M. Retention of Information Under Conditions Approaching a Steady State. *J. Exp. Psychol.* **1961**, *62*, 302–309. [CrossRef]
16. Levy, C.; Jowaisas, D. Short-Term Memory: Storage Interference or Storage Decay? *J. Exp. Psychol.* **1971**, *88*, 189–195. [CrossRef]
17. Chechile, R.A. Trace Susceptibility Theory. *J. Exp. Psychol. Gen.* **1987**, *116*, 203–222. [CrossRef]
18. Chechile, R.A. *Analyzing Memory: The Formation, Retention, and Measurement of Memory*; MIT Press: Cambridge, MA, USA, 2018.
19. Crannell, C.W.; Parrish, J.M. A Comparison of Immediate Memory Span for Digits, Letters, and Words. *J. Psychol. Interdiscip. Appl.* **1957**, *44*, 319–327. [CrossRef]
20. Warren, E.L. Memory Capacity and Storage. Ph.D. Dissertation, Tufts University, Medford, MA, USA, 2015.
21. Chechile, R.A.; Meyer, D.L. A Bayesian Procedure for Separately Estimating Storage and Retrieval Components of Forgetting. *J. Math. Psychol.* **1976**, *13*, 269–295. [CrossRef]
22. Chechile, R.A.; Soraci, S.A. Evidence for a Multiple-Process Account of the Generation Effect. *Memory* **1999**, *7*, 483–508. [CrossRef]
23. Chechile, R.A. New Multinomial Models for the Chechile-Meyer Task. *J. Math. Psychol.* **2004**, *48*, 364–384. [CrossRef]
24. Bröder, A.; Schötz, J. Recognition ROCs are Curvilinear – or are They? On Premature Arguments Against the Two-High-Threshold Model of Recognition. *J. Exp. Psychol. LMC* **2009**, *35*, 587–606. [CrossRef] [PubMed]
25. Chechile, R.A.; Sloboda, L.N.; Chamberland, J.R. Obtaining Separate Measures for Implicit and Explicit Memory. *J. Math. Psychol.* **2012**, *56*, 35–53. [CrossRef]
26. Bahrick, H.P. Semantic Memory Content in Permastore: Fifty Years of Memory for Spanish Learned in School. *J. Exp. Psychol. Gen.* **1984**, *113*, 1–29. [CrossRef]
27. Rao, C.R.; Shanbhag, D.N. An Elementary Proof for an Extended Version of the Choquet-Deny Theorem. *J. Multiv. Anal.* **1991**, *38*, 141–148. [CrossRef]
28. Bertoin, J.; LeJan, Y. Representation of Measures by Balayage from a Regular Point. *Ann. Prob.* **1992**, *20*, 538–548. [CrossRef]
29. Graczyk, P. Cramér Theorem on Symmetric Spaces of Noncompact Type. *J. Theor. Prob.* **1994**, *7*, 609–613. [CrossRef]
30. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; Wiley: New York, NY, USA, 1994; Volume 1.
31. Block, H.W.; Savits, T.H.; Singh, H. The Reverse Hazard Rate Function. *Prob. Eng. Inf. Sci.* **1998**, *12*, 69–90. [CrossRef]
32. Chandra, N.K.; Roy, D. Some Results on Reverse Hazard Rate. *Prob. Eng. Inf. Sci.* **2001**, *15*, 95–102. [CrossRef]
33. Finkelstein, M.S. On the Reverse Hazard Rate. *Rel. Eng. Sys. Saf.* **2002**, *78*, 71–75. [CrossRef]
34. Chechile, R.A. Properties of Reverse Hazard Functions. *J. Math. Psychol.* **2011**, *2011*, 203–222. [CrossRef]
35. Woodroofe, M. Estimating a Distribution Function with Truncated Data. *Ann. Stat.* **1985**, *13*, 163–177. [CrossRef]
36. Chechile, R.A. A Novel Method for Assessing Rival Models of Recognition Memory. *J. Math. Psychol.* **2013**, *57*, 196–214. [CrossRef]
37. Fréchet, M. Sur la loi de Probabilité de l'écart Maximum. *Ann. Soc. Pol. Math. Cracovie* **1927**, *6*, 93–116.
38. Fisher, R.A.; Tippett, L.H.C. Limiting Forms of the Frequency Distributions of the Largest or Smallest Member of the Sample. *Math. Proc. Camb. Philos. Soc.* **1928**, *24*, 180–190. [CrossRef]
39. Gnedenko, B.V. Sur la Distribution Limite du terme Maximum d'une Série Alétoire. *Ann. Math.* **1943**, *44*, 423–453. [CrossRef]
40. Mann, N.P.; Singpurwalla, N.D. Extreme-Value Distributions. In *Encyclopedia of Statistical Sciences*; Kotz, S., Johnson, N.L., Eds.; Wiley: New York, NY, USA, 1982; Volume 2, pp. 606–613.
41. Galambos, J. *The Asymptotic Theory of Extreme Order Statistics*; Wiley: New York, NY, USA, 1978.
42. Rubin, D.C.; Wenzel, A.E. One Hundred Years of Forgetting: A Quantitative Description of Retention. *Psychol. Rev.* **1996**, *103*, 734–760. [CrossRef]
43. Chechile, R.A. Memory Hazard Functions: A Vehicle for Theory Development and Test. *Psychol. Rev.* **2006**, *113*, 31–56. [CrossRef]
44. Staddon, J.E.R. *Adaptive Behavior and Learning*; Cambridge University Press: Cambridge, UK, 1983.
45. Harnett, P.; McCarthy, D.; Davison, M. Delayed Signal Detection, Differential Reinforcement, and Short-Term Memory in the Pigeon. *J. Exp. Anal. Behav.* **1984**, *42*, 87–111. [CrossRef] [PubMed]
46. Jost, A. Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen. *Zeit. Psychol.* **1897**, *14*, 436–472.
47. Atkinson, R.C.; Shiffrin, R.M. Human Memory: A Proposed System and its Control Processes. In *The Psychology of Learning and Motivation: Advances in Research and Theory*; Spence, K.W., Spence, J.T., Eds.; Academic Press: New York, NY, USA, 1968; Volume 2, pp. 89–195.
48. Baddeley, A.D.; Hitch, G. Working Memory. In *The Psychology of Learning and Motivation*; Bower, G.A., Ed.; Erlbaum: Hillsdale, NJ, USA, 1974; Volume 8, pp. 176–189.
49. Woodworth, R.S. *Experimental Psychology*; Henry Holt: New York, NY, USA, 1938.
50. Rubin, D.C. On the Retention Function for Autobiographical Memory. *J. Verb. Learn. Verb. Behav.* **1982**, *21*, 21–38. [CrossRef]
51. Wixted, J.T.; Ebbesen, E.B. On the Form of Forgetting. *Psychol. Sci.* **1991**, *2*, 409–415. [CrossRef]
52. Wickelgren, W.A. Single-Trace Fragility Theory of Memory Dynamics. *Mem. Cogn.* **1974**, *2*, 775–780. [CrossRef]
53. Anderson, J.R.; Schooler, L.J. Reflections on the Environment in Memory. *Psychol. Sci.* **1991**, *2*, 396–408. [CrossRef]
54. Chechile, R.A.; Sloboda, L.N. Reformulating Markovian Processes for Learning and Memory from a Hazard Function Framework. *J. Math. Psychol.* **2014**, *59*, 65–81. [CrossRef]
55. Witmer, L.R. The Association Value of Three-Place Consonant Syllables. *J. Genet. Psychol.* **1935**, *47*, 337–359. [CrossRef]

56. Chechile, R.A. A New Method for Estimating Model Parameters for Multinomial Data. *J. Math. Psychol.* **1998**, *42*, 432–471. [CrossRef]
57. Chechile, R.A. A Novel Bayesian Parameter Mapping Method for Estimating the Parameters of an Underlying Scientific Model. *Commun. Stat. Theory Methods* **2010**, *39*, 1190–1201. [CrossRef]
58. Chechile, R.A. *Bayesian Statistics for Experimental Scientists: A General Introduction Using Distribution-Free Methods*; MIT Press: Cambridge, MA, USA, 2020.
59. Estes, W.K. The Problem of Inference from Curves Based on Group Data. *Psychol. Bul.* **1956**, *53*, 134–140. [CrossRef] [PubMed]
60. Strong, E.K. The Effect of Time-Interval upon Recognition Memory. *Psychol. Rev.* **1913**, *20*, 339–372. [CrossRef]
61. Sloboda, L.N. The Quantitative Measurement of Explicit and Implicit Memory and its Applications an Aging Population. Ph.D. Dissertation, Tufts University, Medford, MA, USA, 2012.
62. Kucera, H.; Francis, W.N. *Computational Analysis of Present-Day American English*; Brown University Press: Providence, RI, USA, 1967.
63. Teyler, T.J.; DiScenna, P. The Hippocampal Memory Indexing Theory. *Behav. Neurosci.* **1986**, *100*, 147–154. [CrossRef]
64. O'Keefe, J.; Nadel, L. *The Hippocampus as a Spatial Map*; Oxford University Press: Oxford, UK, 1978.
65. Olton, D.S.; Becker, J.T.; Handelmann, G.E. Hippocampal Function: Working Memory or Cognitive Mapping? *Physiol. Psychol.* **1980**, *8*, 239–246. [CrossRef]
66. Müller, G.J.; Pilzecker, A. Experimentelle Beiträge zur Lehre von Gedächintz. *Zeit. Psychol.* **1900**, *1*, 1–300.
67. Fanselow, M.S. Factors Governing One-Trial Contextual Conditioning. *Anim. Learn. Behav.* **1980**, *18*, 264–270. [CrossRef]
68. Nader, K.; Schafe, G.E.; Le Doux, J.E. Fear Memories Require Protein Synthesis in the Amygdala for Reconsolidation after Retrieval. *Nature* **2000**, *406*, 722–726. [CrossRef]
69. Misanin, J.R.; Miller, R.R.; Lewis, D.J. Retrograde Amnesia Produced by Electroconvulsive Shock after Reactivation of a Consolidated Memory Trace. *Science* **1968**, *160*, 554–555. [CrossRef]
70. Power, A.E.; Berlau, D.J.; McGaugh, J.L. Anisomycin Infused into the Hippocampus Fails to Block "Reconsolidating" but Impairs Extinction: The Role of Re-Exposure Duration. *Learn. Mem.* **2006**, *13*, 27–34. [CrossRef]
71. Chechile, R.; Butler, K. Storage and Retrieval Changes that Occur in the Development and Release of PI. *J. Verb. Learn. Behav.* **1975**, *14*, 430–437. [CrossRef]
72. Chechile, R.A.; Pintea, G.I. Measuring Components of the Memory of Order. *J. Math. Psychol.* **2021**, *100*, 102476. [CrossRef]