Psychological Review
2004, Vol. 111, No. 4, 864–879

0033-295X/04/$12.00 DOI: 10.1037/0033-295X.111.4.864

On Common Ground: Jost's (1897) Law of Forgetting and Ribot's (1881) Law of Retrograde Amnesia

John T. Wixted
University of California, San Diego

T. Ribot's (1881) law of retrograde amnesia states that brain damage impairs recently formed memories to a greater extent than older memories, which is generally taken to imply that memories need time to consolidate. A. Jost's (1897) law of forgetting states that if 2 memories are of the same strength but different ages, the older will decay more slowly than the younger. The main theoretical implication of this venerable law has never been worked out, but it may be the same as that implied by Ribot's law. A consolidation interpretation of Jost's law implies an interference theory of forgetting that is altogether different from the cue-overload view that has dominated thinking in the field of psychology for decades.

Two seemingly unrelated psychological laws, both enacted in the late 1800s, seem to have stood the test of time. One law was advanced by Ribot in 1881 in a book entitled *Les Maladies de la Memoire* (The Diseases of Memory). On the basis of clinical observations, Ribot (1881, 1882) observed that brain injury affects premorbid memories in the reverse order of their formation such that newly formed memories (i.e., those formed just prior to brain injury) are impaired to a greater extent than older memories. Today, Ribot's law is referred to as temporally graded retrograde amnesia, and much evidence involving both human and animal subjects supports its validity (Squire, Clark, & Knowlton, 2001). The main theoretical implication of Ribot's law is that memories need time to consolidate. As they do, they become less vulnerable to the forces of traumatic brain injury. Although the field of experimental psychology has yet to fully embrace this way of thinking, consolidation theory is regarded as a standard model in the field of neuroscience (e.g., McGaugh, 2000; Squire, 1992).

Ribot's law may be intimately related to one of two ideas advanced by Jost in 1897. Jost's laws hold that that if two associations (i.e., two memory traces) are of equal strength but different ages, the older one will (a) benefit more from an additional learning trial and (b) decay less rapidly in a given period of time than the younger one (Youtz, 1941). The second of these two laws (Jost's law of forgetting) is the main focus of this article. Jost's second law typically manifests itself in forgetting functions that represent levels of retention aggregated over items, as illustrated in Figure 1. In this hypothetical example, List 1 was learned to an initial level of 10 items recalled. After 20 units of time, during which some of the List 1 items were forgotten, a second list (List 2) was learned to a level of 8 items recalled. As shown in the figure, the younger traces are lost at a relatively fast rate such that, at some point, the number of items recalled from the young list equals the number of items recalled from the old list. Jost's law of forgetting states that from that moment on, the younger function will decline more rapidly than the older one (although, in truth, the same was also true before their levels coincided).

Correspondence concerning this article should be addressed to John T. Wixted, Department of Psychology, 0109, University of California, San Diego, La Jolla, CA 92093-0109. E-mail: jwixted@ucsd.edu

Both of Jost's laws have long been held in high regard by the field. Slamecka (1985), for example, referred to them as Jost's "two famous laws" (p. 815), and Estes (1979) remarked,

> It is rather remarkable that by the end of the first decade of experimental psychology laws of the growth and decay of traces or associations had been formulated in a way that has remained virtually unchanged through the remainder of the century and has entered into nearly every theory of learning and memory. Credit for the penetrating insight responsible for these formulations evidently belongs to Adolf Jost, a student of Georg Mueller; hence "Jost's laws" have earned a place among the most enduring and ubiquitous principles of memory. (pp. 643–644)

In spite of the fact that Jost's law of forgetting is held in the highest regard and has stood unchallenged for more than a century, its theoretical significance remains virtually unexplored. That the field of experimental psychology would be comfortable with this state of affairs is remarkable given that the theoretical implications of other well-established laws, such as Weber's law, have been repeatedly and thoroughly considered. By contrast, not one detailed effort to grapple with the theoretical implications of Jost's law can be identified in the 20th-century literature (or, so far, in the 21st-century literature).

What explains this unusual absence of curiosity? One possibility is that natural, but incorrect, intuitions about Jost's law create the impression that it amounts to little more than another way of stating the obvious. For example, it might be imagined that Jost's law follows merely from the fact that curvilinear forgetting functions necessarily fall through a greater range when memory traces are young compared with when they are old. In his influential text on experimental psychology, Woodworth (1938) may have fallen prey to this intuition when he said,

> This law of Jost's can even be deduced from the general shape of a retention curve. As a lesson becomes old it reaches a flatter part of the curve and its further decline will be slow. Therefore, a young lesson momentarily at the same retention level as an old one is on a steeper part of the curve and doomed to decline more rapidly. (p. 59; Woodworth & Schlosberg, 1954, p. 730)

It is not entirely clear what Woodworth (1938) meant by the "general shape of a retention curve," but the most straightforward

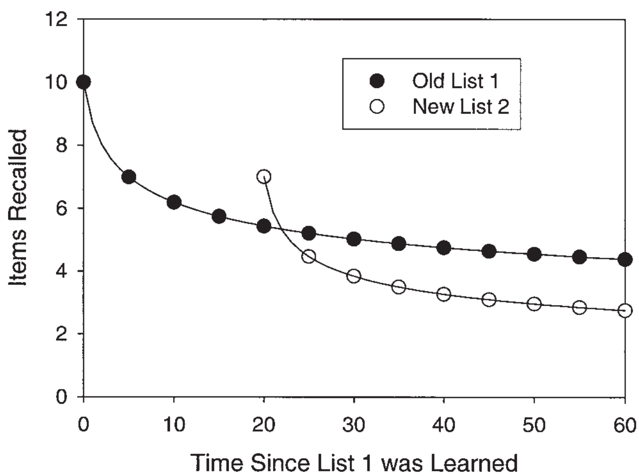

Figure 1. An illustration of Jost's (1897) second law.

interpretation of this statement is that Jost's law follows from the mere fact that forgetting functions are curvilinear. Youtz (1941) appeared to have the same idea in mind when she observed that Jost's law follows from the fact that empirical forgetting functions are "negatively accelerated" (p. 33), which is how curvilinear functions are often described. If Jost's law merely implies that forgetting functions are curvilinear, then it is hard to imagine that it has much theoretical significance.

Although Woodworth's (1938) intuition about Jost's law is quite natural, it is also quite wrong. Curvilinear forgetting functions do, of course, fall through a greater range when memory traces are young compared with when they are old, but this is not why Jost's law is true. Consider, for example, the hypothetical forgetting function shown in Figure 2. This function declines from 16 items recalled to 8 items recalled over the 1st day (an absolute loss of 8 items), from 8 items recalled to 4 items recalled over the 2nd day (a loss of 4 items), and from 4 items recalled to 2 items recalled over the 3rd day (a loss of 2 items). Thus, although the function's absolute drop decreases with age, which is the property that Woodworth (1938) emphasized, its proportional drop remains constant regardless of the age of the memory trace. In this example, the function will always fall from whatever its level happens to be at time t to 50% of that level at time $t + 1$ day (which is to say that its *half-life* is equal to 1 day). The constant proportional loss illustrated in Figure 2 is the defining characteristic of the exponential, and that property figures prominently in the discussion of Jost's law to follow. Other downward curvilinear functions exhibit a faster or slower proportional decline as a function of time.

In a short but insightful note about Jost's law, Herbert Simon (1966) observed that if the old and new forgetting functions both have the same half-life (i.e., the same proportional decline per unit time), then Jost's law is *incompatible* with exponential forgetting. As shown in Figure 2, the exponential is a perfectly adequate curvilinear function. Thus, if Jost's law is incompatible with exponential forgetting, as Simon (1966) showed, it cannot also be true that Jost's law follows merely from the fact that forgetting functions are curvilinear, as Woodworth (1938) implied. The reason why Jost's law is incompatible with exponential forgetting is that old and new exponential functions with the same half-lives would necessarily decline by the exact same amount if their strengths ever did happen to coincide. If, for example, both functions had a half-life of 1 day, then they would both decline by 50% from their common level in 1 day's time, which means that the new function would forever be coincident with the old one (an outcome that does not correspond to Jost's law).

What Properties of Empirical Forgetting Functions Yield Jost's Law?

These considerations reveal that the mere curvilinearity of forgetting functions is not the property that underlies Jost's law. What properties of empirical forgetting functions do underlie Jost's law? As Simon (1966) observed, one possibility is that the higher degree of learning associated with the older forgetting function yields a slower rate of forgetting (i.e., a longer half-life), in which case Jost's law would be compatible with exponential forgetting after all. That is, even though the proportional rate of decay remains constant as a function of time for both the old and the new exponential forgetting functions, the older function will decay less than the new one (in both proportional and absolute terms) from their point of common strength. A second possibility is that the degree of learning has no effect on the rate of forgetting, but as a trace ages it loses strength at an ever-slower proportional rate. If so, then the older trace would have a slower rate of forgetting precisely because it is older than the younger trace, and forgetting would not be exponential in form.

Figure 3 illustrates these two possibilities. The upper panel shows two exponential functions of the form $R(t) = N_0e^{-kt}$, where $R(t)$ represents the number of items remembered at time t, N_0 represents the degree of learning (i.e., the number of items encoded at $t = 0$), and k represents the time constant (the value of which determines the function's half-life). As indicated above, if memories decay exponentially, the passage of time does not affect the proportional rate of loss. Even so, Jost's law holds in this example because a higher degree of learning is associated with a slower proportional rate of loss, and the older trace was initially learned to a higher degree. The younger forgetting function in this example has a half-life of approximately 5 units of time, whereas the older

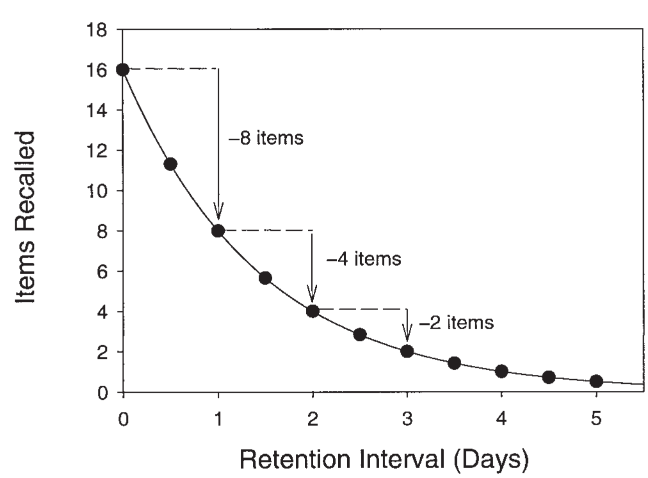

Figure 2. An illustration of the decreasing absolute rate of loss associated with curvilinear forgetting functions.

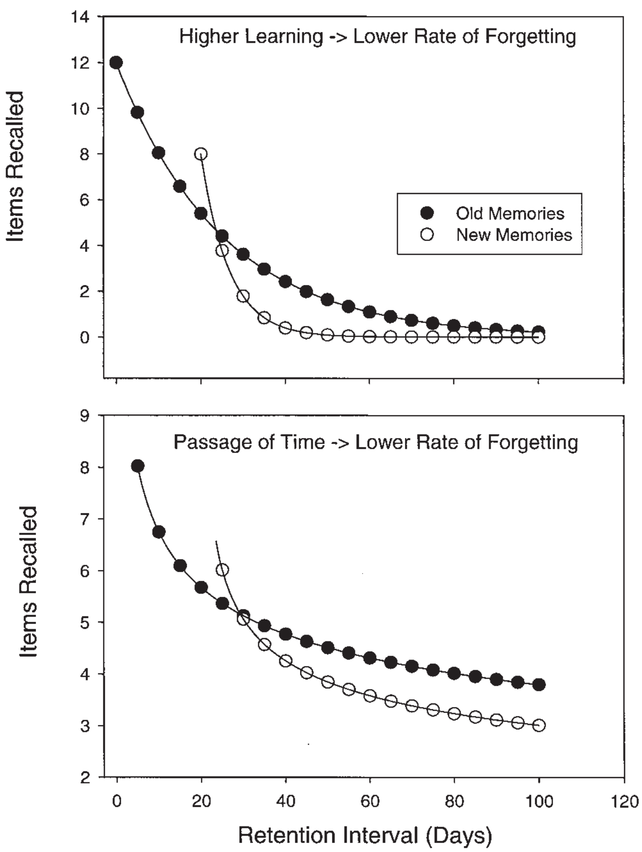

Figure 3. Top: An illustration of Jost's law arising because a higher degree of learning results in a lower rate of forgetting (but the passage of time does not). Bottom: An illustration of Jost's law arising because the passage of time results in a lower rate of forgetting (but a higher degree of learning does not).

forgetting function has a half-life of approximately 20 units of time. Thus, when the functions intersect, the old function will require much more time to fall to 50% of that common level of performance than the younger one (in accordance with Jost's law).

The lower panel of Figure 3 illustrates Jost's law arising because the proportional rate of forgetting decreases with age. Both functions drop by 25% in the first 15 units of time, and both require more than twice that amount of time to drop another 25% (i.e., the rate of forgetting slows with the passage of time). Because the proportional rate of loss is not constant over time, these functions are not exponential in form. For this illustrative example, power functions of the form $R(t) = N_1 t^{-k}$ were used, though many other functions could have been used instead (such as the logarithm). The parameter N_1 represents the number of items remembered after 1 unit of time, and the exponent k governs the initial rate of decay, with a larger value of k yielding a faster proportional decline. Unlike the situation depicted in the upper panel, the degree of learning in this example is not assumed to affect the rate of forgetting (i.e., both the older and younger functions have the same exponent, k). Instead, Jost's law holds because the older trace, by virtue of its advanced age, has a slower proportional rate of forgetting by the time the new trace is formed.

So which is it? Does a higher degree of learning reduce the rate of forgetting, in which case Jost's law could hold even if forgetting were exponential in form, or are forgetting functions characterized by an ever-decreasing rate of forgetting? Figure 4 presents representative forgetting function data from Slamecka and McElree (1983) that help to answer that question. As shown in the figure, the higher degree of learning condition exhibits a 36% loss per day over the 1st day (from 11.75 to 7.50 items recalled) and an 11.5% loss per day over the next 5 days (from 7.5 to 4.0 items recalled). Thus, these data do not exhibit the constant proportional rate of loss that characterizes the exponential. Instead, the proportional rate of forgetting decreases dramatically with the passage of time. The same phenomenon is evident in the lower degree of learning condition, which exhibits a 51% loss per day over the 1st day (from 9.25 to 4.50 items recalled) and a much slower 13.8% loss per day, on average, over the next 4 days (from 4.5 to 2.0 items recalled).

The ever-slowing rate of decay evident in the forgetting functions reported by Slamecka and McElree (1983) has been repeatedly confirmed by investigations into the mathematical form of empirical forgetting functions. Wickelgren (1974, 1977), Anderson and Schooler (1991), and Wixted and Ebbesen (1991, 1997) all suggested that empirical forgetting functions are most accurately described by a power function, though Wixted and Ebbesen (1991) also showed that the logarithm was a very close rival. Rubin and Wenzel (1996) argued that four different functions were equally viable (the logarithm, the power function, a hyperbolic-power function, and an exponential-power function). Wickens (1998) and White (2001) have also advocated the exponential-power function. Although no single function stands out as being definitely superior to the others, all of these investigations agree that (a) the simple exponential (which requires a constant proportional rate of decay) is not a viable candidate and (b) every viable candidate has the

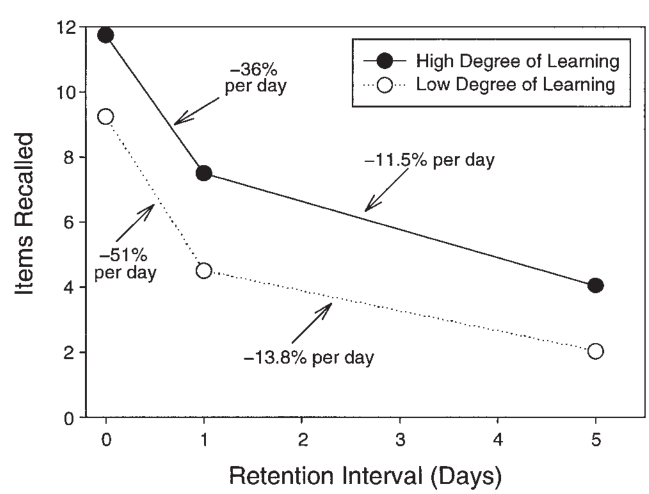

Figure 4. High and low degree of learning data as a function of retention interval. The data illustrate the ever-decreasing proportional rate of loss, as well as the lower proportional rate of loss associated with a higher degree of learning. From "Normal Forgetting of Verbal Lists as a Function of Their Degree of Learning," by N. J. Slamecka and B. McElree, 1983, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* p. 392. Copyright 1983 by the American Psychological Association.

property that yields Jost's law of forgetting, namely, the proportional rate of decay decreases with the passage of time (as illustrated in the lower panel of Figure 3).

Thus, it seems virtually certain that one of the properties of empirical forgetting functions that Simon (1966) identified as possibly contributing to Jost's law of forgetting (viz., an ever-slowing rate of forgetting) does, in fact, hold. What about the other property? Does a higher degree of learning induce a slower rate of forgetting? If so, that would also play a role because the older function referred to in Jost's law was initially learned to a higher degree than the new one (cf. Figure 1). The empirical data shown in Figure 4 suggest that the effect of the degree of learning may, indeed, be relevant. Over the 1st day, the lower degree of learning function drops by 51%, whereas the higher degree of learning function drops by only 36%. After Day 1, the difference in the proportional rate of decay as a function of degree of learning is rather small (11.5% per day in the high degree of learning condition and 13.8% per day in the low degree of learning condition), but the difference is still evident, and all of the forgetting functions reported by Slamecka and McElree (1983) show similar proportional trends. On the other hand, other data in the literature (e.g., Wixted & Ebbesen, 1991) exhibit a barely detectable effect of the degree of learning on the rate of forgetting. After reviewing evidence like this, Anderson (2000) concluded that a higher degree of learning does not reliably result in a slower rate of forgetting, but my own reading of the same evidence suggests that it does, though the effect is modest (cf. Underwood & Keppel, 1963).

From data like these, we can conclude that Jost's law may arise for both of the reasons considered by Simon (1966), but the effect of the passage of time on the rate of forgetting is dramatic and has been repeatedly replicated, whereas the effect of degree of learning on the rate of forgetting is smaller and less clearly established in the literature. Thus, to unravel the theoretical significance of Jost's law, the main task is to explain why the rate of forgetting so reliably decreases with the passage of time. That issue is the focus of the rest of this article, and the possible explanations for that phenomenon that have been advanced in the literature are considered next.

Why Does the Rate of Decay Decrease With the Passage of Time?

The fact that empirical forgetting functions exhibit a decreasing rate of decay does not necessarily mean that the underlying memory traces decay in a similar manner. In fact, of the four explanations that have been proposed for the ever-slowing rate of forgetting associated with empirical forgetting functions, three are consistent with the idea that the underlying memory traces decay exponentially.

One possibility is that empirical forgetting functions exhibit a decreasing rate of decay with the passage of time (and fail to show a large effect of the degree of learning) only because of the nature of the measurement scale. At the level of the underlying memory traces, the exact opposite might be true. This could occur, for example, if the measurement scale (e.g., percent correct) were more sensitive to change when memory strength was high compared with when it was low. As strength weakens with time, the measurement scale would become progressively less sensitive, thereby creating the false impression that the proportional rate of change decreases with time. Similarly, a higher degree of learning might induce a lower rate of forgetting at the level of the underlying memory trace, but this would be obscured by the fact that the measurement scale is more sensitive to change when strength is high. The idea that, in spite of appearances to the contrary, memory traces decay exponentially and a higher degree of learning results in a lower rate of forgetting was advanced by Loftus (1985b). According to this idea, the first of the two possibilities mentioned by Simon (1966) applies at the level of underlying memory traces (upper panel of Figure 3), even though empirical forgetting functions correspond to the second of those two possibilities (lower panel of Figure 3) because of a nonlinear measurement scale. An important implication of this account is that the rate of decay associated with individual memory traces does not change over time (i.e., they decay exponentially), even though the rate of decay associated with the empirical forgetting function does.

A second explanation for the ever-decreasing rate of forgetting associated with empirical forgetting functions is that the individual memory traces that comprise an aggregate forgetting function each decay exponentially but with vastly different half-lives (Sikström, 1999, 2002). If so, then the ever-slowing proportional rate of decay evident in empirical forgetting functions would simply reflect the survival of the fittest memory traces. According to this view, Jost's law arises because the old forgetting function is composed mainly of slowly decaying traces (because the rapidly decaying traces have already faded away), whereas the young forgetting function is composed of a mixture of rapidly and slowly decaying traces (so it decays relatively rapidly, on average). This intuitively plausible explanation for Jost's law was, in fact, taken for granted by Slamecka (1985) when he said "one can view an older list as having endured a more extended item-attrition experience than a younger list, with the consequence that its remaining items are hardier and more resistant to loss than are those of the younger list" (p. 815). Again, the important implication of this view for the present analysis is that the rate of decay associated with individual memory traces does not change over time, even though the rate of decay associated with the aggregate empirical forgetting function does.

A variant of this idea yields a third (albeit very similar) explanation for Jost's law. The core assumption of this account is that forgetting functions descend to an asymptote greater than zero. Jost's law is explained by assuming that a higher degree of learning results in a higher asymptotic level of memory performance. An old forgetting function, having decreased considerably with the passage of time, would be relatively close to its high asymptote (and, so, would essentially decay no further). By contrast, a new forgetting function that is momentarily at the same level of strength as the old forgetting function would, with the further passage of time, descend towards its own (lower) asymptote. Hence, Jost's law.

The idea that forgetting functions descend to an asymptote greater than zero was recently proposed by Rubin, Hinton, and Wenzel (1999). They showed that forgetting from long-term memory (LTM) can be accurately described by an exponential function that includes a third parameter that represents a nonzero asymptote. In fact, as described in more detail later, the three-parameter exponential (unlike its two-parameter counterpart) describes empirical forgetting functions every bit as well as the power law does.

At the level of underlying memory traces, the simplest model for an exponential decay to nonzero asymptote would hold that some traces decay exponentially, perhaps all with the same half-life, whereas other traces are permanent and do not decay at all. Unlike the continuous distribution of half-lives discussed above, the half-life distribution in this case would be bimodal, with one mode being a finite half-life and the other “mode” being an infinitely long half-life. As before, the relevant implication of this account is that one need not assume that the rate of forgetting associated with individual traces changes in any way with the passage of time. The permanent traces are forever permanent, and the decaying traces always decay exponentially with an unchanging half-life.

A fourth and quite distinct possibility is that the memory traces themselves decay to a lesser degree the longer they manage to survive. According to this view, the properties of the empirical forgetting function actually do correspond to the properties of the underlying memory traces. If so, young traces all fade relatively rapidly, but they fade less rapidly as they age (in which case forgetting would not be exponential in form, even at the level of the underlying trace). One reason why memories might decay less rapidly with age is because they become less vulnerable to the forces of retroactive interference (RI) as a result of consolidation. The first empirical report suggesting that traces do, in fact, become less vulnerable to RI over time was published by Adolf Jost’s mentor, Georg Müller, in 1900 (Müller & Pilzecker, 1900). However, shortly thereafter, the field of experimental psychology adopted the view that forgetting occurs mainly when new associations are added to a retrieval cue—that is, that forgetting is caused by cue overload (Watkins & Watkins, 1975)—not because fragile traces are compromised before they have had a chance to consolidate. In fact, for the most part, interference theorists explicitly rejected the idea that hypothetical consolidation process has anything at all to contribute to the understanding of forgetting. But the notion of consolidation may provide the key to understanding Jost’s law, in which case the dominant cue-overload view of forgetting may need to be reevaluated.

These competing explanations for the ever-slowing rate of decay associated with empirical forgetting functions (i.e., these competing explanations of Jost’s law) have never been differentially evaluated. In fact, all previous theoretical treatments of Jost’s law have amounted to little more than a sentence or two (e.g., Sikström, 2002; Slamecka, 1985; Wixted, 2004; Woodworth, 1938; Woodworth & Schlosberg, 1954). This is a curious state of affairs because my own reading of the literature is that a century of work in the fields of psychology and neuroscience has produced a body of knowledge that is sufficiently complete to finally distinguish between them. In what follows, each of these four possible explanations for the ever-slowing rate of forgetting is considered in more detail and in light of the relevant empirical evidence.

Measurement Scale Artifacts

The possibility that Jost’s law arises because of a measurement scale artifact is considered only briefly because it has already been considered in great detail in the literature (Bogartz, 1990; Loftus, 1985a, 1985b; Loftus & Bamber, 1990; Slamecka, 1985; Wixted, 1990). As indicated above, Loftus (1985b) argued that the strength of memories may not decay at an ever-decreasing rate, even though empirical forgetting functions exhibit that property. Instead, the passage of time may have no effect on the proportional rate of decay (i.e., memories might decay exponentially). According to this view, Jost’s law holds for the reasons depicted in the upper panel of Figure 3, even though it appears to hold for the reasons depicted in the lower panel. The illusion arises because the measurement scale is more sensitive to change when strength is high compared with when it is low (much as a fuel gauge might indicate that the level of gasoline is changing a rate that does not correspond to the true rate of change).

The possibility that Jost’s law arises only because a higher degree of learning results in a slower rate of forgetting at the level of the underlying memory trace and not because of an ever-slowing rate of decay at that level is challenged by several observations. First, almost all forgetting functions exhibit a pronounced decreasing rate of decay with the passage of time no matter what the measurement scale happens to be (percent correct, savings, priming, hits minus false alarms, etc.). It is, of course, possible that all of these scales create the same kind of measurement distortion, but such a coincidence would be surprising. Second, some measurement scales are theoretically linear with respect to the underlying trace. Given the assumptions of signal detection theory, for example, d' provides a linear measurement scale (Loftus, 1985b). A signal-detection account of recognition memory is not universally accepted, but it has been the dominant view for decades. And recognition forgetting functions using d' as a dependent measure exhibit the same decreasing rate of forgetting that is observed when any other dependent measure is used (Wixted & Ebbesen, 1997). This holds true even when one allows for the possibility of an unequal-variance detection model. Thus, given the assumptions of signal detection theory, at least, the ever-decreasing rate of decay evident in empirical forgetting functions does not appear to be a measurement scale artifact. Instead, memory traces, at least in the aggregate (which is what forgetting functions represent), appear to exhibit a decreasing proportional rate of decay with the passage of time.

Item Variability

Even if empirical forgetting functions faithfully reflect the aggregate strength of underlying traces (i.e., even if measurement scale issues do not complicate the analysis), they may not faithfully reflect the properties of individual memory traces. The second explanation for Jost’s law of forgetting holds that traces decay exponentially, all toward an asymptote of zero, but they do so with widely varying half-lives. According to this view, the ever-slowing proportional rate of decay evident in most aggregate forgetting functions merely reflects survival of the fittest memory traces. After a time, all of the surviving memory traces have a low likelihood of failure, not because they once were fragile and now are sturdy, but because they were sturdier than the other (now failed) traces all along.

The idea that items differ in durability may be driven by the incontrovertible intuition that some items on a list are much more memorable than others. Although this would seem to indicate that items differ greatly in their underlying temporal properties, it might instead mean that they differ greatly in the degree to which they are initially learned. As an analogy, amnesic patients are often said to forget very quickly because they cannot remember a meeting that took place 30 min ago (Wixted, in press). But these patients actually have a normal rate of forgetting (McKee & Squire, 1992). Their impairment lies mainly in an inability to encode new information in the first place. Thus, even though intuition suggests that unimpaired subjects have a vastly lower rate of forgetting than patients with amnesia, the difference actually lies in the degree of initial registration. The same might apply to items that differ greatly in memorability. The mere fact that some items are much more likely to be remembered after a delay than others does not in and of itself demonstrate that items differ greatly in the rate at which they are forgotten.

If items do differ substantially in their decay rates in addition to differing in their degree of initial registration, then one might expect to see dramatically different decay rates for classes of items that differ substantially in how meaningful they are (or that differ along some other important dimension). Underwood and Postman (1960) performed a relevant experiment by comparing the rates of forgetting over a period of 1 week for a list of three-letter common words versus a list of meaningless three-letter trigrams. To their surprise, they found that the rate of forgetting for the meaningful words was the same as that for the meaningless trigrams. The lists differed mainly in how easily they were initially learned, with words taking significantly less time to encode than nonwords. Keppel (1968) reviewed a broader array of evidence suggesting that, with degree of learning equated, variables such as word frequency, meaningfulness, and association value have no effect on the rate of forgetting. Slamecka and McElree (1983) also conducted an item analysis and found that the most easily encoded words had about the same rate of decay as their less easily encoded counterparts, even when the degree of learning was not equated. Thus, direct evidence that items, once encoded, have substantially different rates of decay is hard to come by. By contrast, direct evidence that they differ in the degree of initial learning is widespread.

Findings like these raise the intriguing possibility that items do not differ greatly in their rates of decay even though they differ greatly in the degree to which they are initially learned (which accounts for differences in apparent durability). Still, one might wonder about how much variability in the rates of decay across items one needs to assume in order to accommodate some of the basic facts about the time course of forgetting.

An analysis of empirical forgetting functions often yields a pattern that is hard to explain by assuming that individual items decay exponentially with varying half-lives. That pattern consists of a forgetting function that is fit extremely well by a power function with a small exponent (such as -0.13) while, at the same time, being very poorly described by the exponential (which provides a fit that barely exceeds the fit provided by a straight line). This pattern is not always observed, but it is observed often enough that it needs to be explained in order for the item variability account to remain viable. It is, for example, the pattern that corresponds to Ebbinghaus's (1885/1913) famous savings function (Anderson, 2000, p. 227) and to the word recall and face recognition forgetting functions reported by Wixted and Ebbesen (1991).

To investigate this matter, a variety of Monte Carlo simulations were performed in which exponential half-lives were assumed to be distributed according to particular distributions. In each simulation, a large number of exponential functions with half-lives drawn from a particular distribution were averaged together (just as averaging is involved in the analysis of empirical data), and the resulting aggregate function was fit with a power function, an exponential function, and a straight line. Table 1 presents repre-

Table 1
Aggregate Forgetting Function Characteristics for Various Settings of the Scale and Shape Parameters of the Weibull Distribution of Exponential Half-Lives

Scale	Shape	$\theta_{.95}/\theta_{.05}$	Line	Exp	Power	*a*	*b*
1	0.4	2.6×10^{4}	72.5	96.4	99.0	0.45	0.68
1	0.3	7.7×10^{5}	78.9	93.5	99.2	0.41	0.49
1	0.2	6.8×10^{8}	83.8	90.8	99.5	0.39	0.30
1	0.1	4.6×10^{17}	86.6	88.5	99.9	0.37	0.12
10	0.4	2.6×10^{4}	85.8	94.2	98.4	0.75	0.36
10	0.3	7.7×10^{5}	86.0	92.3	99.1	0.65	0.27
10	0.2	6.8×10^{8}	86.6	91.0	99.4	0.55	0.19
10	0.1	4.6×10^{17}	86.4	88.7	99.9	0.45	0.10
100	0.4	2.6×10^{4}	91.3	94.4	98.0	0.90	0.17
100	0.3	7.7×10^{5}	90.2	92.9	98.7	0.81	0.15
100	0.2	6.8×10^{8}	89.0	91.6	99.3	0.68	0.13
100	0.1	4.6×10^{17}	87.5	88.9	99.8	0.53	0.08
1,000	0.4	2.6×10^{4}	93.4	95.0	97.5	0.97	0.07
1,000	0.3	7.7×10^{5}	91.8	93.5	98.6	0.91	0.08
1,000	0.2	6.8×10^{8}	89.9	92.5	99.3	0.79	0.08
1,000	0.1	4.6×10^{17}	88.2	90.7	99.7	0.61	0.06
10,000	0.4	2.6×10^{4}	94.5	95.0	97.9	0.99	0.03
10,000	0.3	7.7×10^{5}	93.0	93.7	98.6	0.95	0.04
10,000	0.2	6.8×10^{8}	90.5	91.5	99.4	0.86	0.05
10,000	0.1	4.6×10^{17}	87.6	88.6	99.9	0.68	0.05

Note. The table shows the ratio of exponential half-lives at the 95th and 5th percentiles of the distribution ($\theta_{.95}/\theta_{.05}$), the percentage of variance accounted for by a straight line, the exponential (Exp), and the power function when fit to the resulting aggregate forgetting function, as well as the values of the two parameters (*a* and *b*) of the best-fitting power function of the form at^{-b}.

sentative results of one such analysis. In this analysis, the half-lives of exponential forgetting functions were assumed to be distributed according to a two-parameter Weibull distribution. After setting the shape and scale parameters of the Weibull distribution, 2,000 exponential forgetting functions with half-lives drawn from that distribution were generated and then averaged together to create the aggregate function. As shown in Table 1, settings for the Weibull distribution of exponential half-lives can be found that yield an aggregate forgetting function that is extremely well fit by the power function with a small exponent and that is so poorly fit by the exponential that it barely outperforms a straight line. However, the required variability is rather large. To provide an idea of how much variability in exponential half lives is required for this pattern to emerge, the table also shows the ratio of the half-life value at the 95th percentile of the cumulative Weibull distribution ($\theta_{.95}$) to the half-life value at the 5th percentile of the distribution ($\theta_{.05}$). These values can be obtained by solving the following cumulative Weibull distribution function for θ (the half-life), first with $p = .95$ and then with $p = .05$: $p = 1 - \exp[(-\theta/\text{scale})^{\text{shape}}]$.

As an example, in the first line of Table 1, the scale parameter of the Weibull was set to 1 and the shape parameter to 0.4. Thus, the half-life at the 95th percentile of the cumulative Weibull is found by solving the following equation for $\theta_{.95}$: $.95 = 1 - \exp[(-\theta_{.95}/1)^{0.4}]$, and the half-life at the 5th percentile of the cumulative Weibull is found by solving the following equation for $\theta_{.05}$: $.05 = 1 - \exp[(-\theta_{.05}/1)^{.4}]$.

In this case, $\theta_{.95}$ is found to be approximately 15.5 and $\theta_{.05}$ is found to be approximately 0.0006 (the time scale is arbitrary). Thus, the ratio of these two half-lives ($\theta_{.95}/\theta_{.05}$) is approximately 26,068. Generally speaking, for the two-parameter Weibull distribution, $\theta_{.95}/\theta_{.05} = [\log(.05)/\log(.95)]^{1/\text{shape}}$ (i.e., relative variability in half-lives is a function of the shape parameter only).

In spite of the rather extreme degree of half-life variability when the scale parameter is set to 1 and the shape parameter to 0.4, the pattern of interest (i.e., the power function fitting the aggregate function extremely well, with the exponential barely outperforming the straight line) is not in evidence. That is, in the first line of Table 1 (and in every case in which the shape parameter is set to 0.4), the power function outperforms the exponential, but it exhibits systematic deviations from the aggregate forgetting function, and the exponential still fits the data reasonably well (and quite a bit better than the straight line). However, the pattern of interest does begin to emerge when the variability in exponential half-lives becomes even greater as the shape parameter becomes smaller. When the half-life variability is extreme, the pattern is clearly evident. Thus, averaged exponentials can yield a power function like those often seen in real data (e.g., Ebbinghaus, 1885/1913), but extreme variability is needed. Similar conclusions are reached when the underlying distribution of exponential half-lives is assumed to follow a log-normal or a pareto distribution. Sikström (1999, 2002) investigated a similar issue in detail as well, and he also found that a great deal of variability in exponential half-lives of item features was needed to produce a credible power function in the aggregate.

The important point to make here is that the large variability in exponential forgetting rates that appears to be needed to produce a power function with a small exponent must be considered against empirical evidence suggesting that items with very different properties tend to decay at the same rate (Keppel, 1968). Perhaps some distribution of exponential half lives with only a small amount of variability will eventually be found that yields a credible power function with a small exponent. Or perhaps some property of items other than meaningfulness (or frequency) will be found which, when manipulated, yields very large differences in forgetting rates. In the absence of those developments, however, the idea that a typical empirical forgetting function reflects averaged exponentials seems hard to sustain.

Do Forgetting Functions Decay to Nonzero Asymptotes?

A third possible reason why the proportional rate of forgetting decreases with the passage of time is that forgetting functions decay (exponentially perhaps) to a nonzero asymptote (Rubin et al., 1999). According to this view, Jost's law of forgetting follows from the fact that the older forgetting function, having been learned to a higher degree, descends to a relatively high asymptote. Thus, when the strengths of the old and new forgetting functions coincide, the old one has little room to decay any further.

The simplest explanation for a function that decays exponentially to a nonzero asymptote is that some items decay with a constant half-life whereas other items do not decay at all. Like the Weibull distribution of half-lives considered earlier, the half-life distribution in this case would still be extremely variable, but the distribution of half-lives would be quite simple (namely, a bimodal distribution, with one mode being a finite half-life and the other being an infinitely long half-life). The idea that forgetting functions descend to an asymptote greater than zero has recently been advanced in the literature (Rubin et al., 1999), so it is worth considering the relevant evidence in some detail.

As indicated earlier, empirical forgetting functions are often accurately described by the power function and poorly described by the exponential. However, if one allows for the possibility of a nonzero asymptote by adding a third parameter, then the exponential rivals the power law in terms of its ability to fit the data. Data reported by Wixted and Ebbesen (1991), which are reproduced in Figure 5, illustrate this point. In this experiment, subjects studied short lists of words that were followed by a filled retention interval ranging from 2.5 to 40.0 s. Degree of learning was varied by manipulating rate of word presentation. The upper panel of Figure 5 shows that both the high degree of learning data and the low degree of learning data are extremely well fit by the power function. Note that these data also reinforce a point made earlier about the effect of the degree of learning on the rate of forgetting. Specifically, with regard to empirical forgetting functions, a higher degree of learning does not have a large effect on the rate of forgetting (though it does have some effect). The power-function exponent in the high degree of learning condition is slightly lower than the corresponding value in the low degree of learning condition, but, in this case, the difference is not significant.

The lower panel of Figure 5 shows how well the three-parameter exponential-decay-to-nonzero-asymptote describes the same recall data. It is obvious that the addition of a nonzero asymptote allows the exponential to fit these data very well (every bit as well as the power function), and this is not merely due to the fact that it has more free parameters. When the power function was fit to these data, two parameters were estimated for each condition (for a total

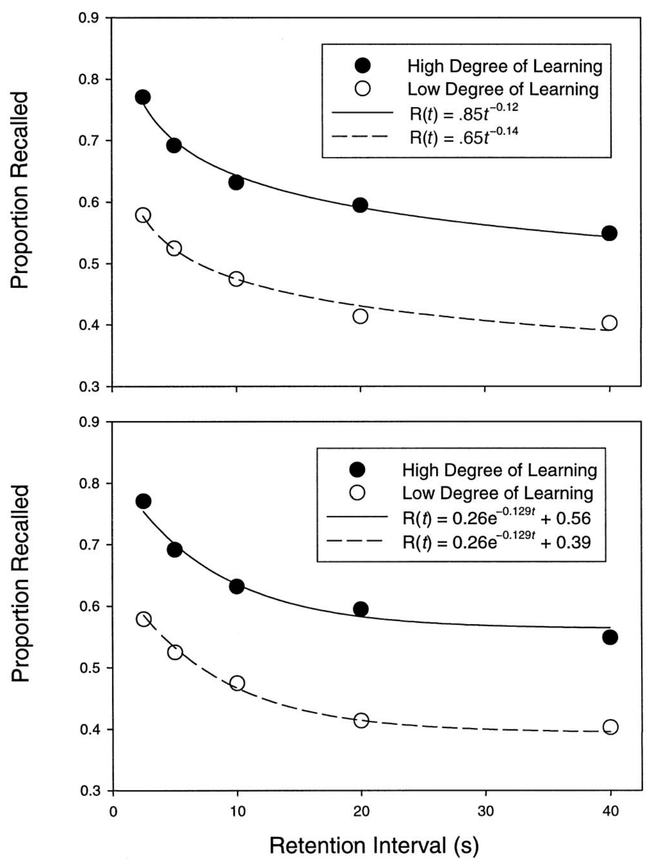

Figure 5. Top: Least-squares fits of the power function to high and low degree of learning data reported by Wixted and Ebbesen (1991). Bottom: Least-squares fit of the exponential-decay-to-nonzero-asymptote to the forgetting data reported by Wixted and Ebbesen (1991). From "On the Form of Forgetting," by J. T. Wixted and E. Ebbesen, 1991, *Psychological Science, 2,* p. 411. Copyright 1991 by Blackwell. Adapted with permission.

of four free parameters across the two fits). Only four free parameters were used across the two fits for the exponential functions as well. Although the exponential has three free parameters, two were constrained to have the same value across conditions for these fits. Only the asymptote was free to differ for the low and high degree of learning conditions. The fits are undeniably excellent, but they suggest that the upper function declines to an astonishingly high asymptote of 56% correct and the lower function to a less astonishing, but still remarkably high, asymptote of 39% correct. Perhaps some other fitted function would project to low, but still nonzero, asymptotes that would seem less implausible, but I have yet to find such a function.

Results like these may explain why the default assumption in the field is usually that forgetting functions decline toward an asymptote of zero. In fact, this seems to be the view of almost everyone who has ever investigated the mathematical form of forgetting (e.g., Anderson & Schooler, 1991; Ebbinghaus, 1885/1913; Rubin & Wenzel, 1996; White, 2002; Wickelgren 1974, 1977; Wickens, 1998; Wixted & Ebbesen, 1991, 1997), all of whom considered only functions that eventually decline to zero. On the other hand, Rubin et al. (1999) and, before him, Bahrick (1984), seriously considered the possibility that forgetting functions descend toward an asymptote greater than zero. To those who find the idea of nonzero asymptotes appealing, the burden of proof might seem to rest with those who believe otherwise. It would be nice if some empirical evidence could be brought to bear on the issue, and there are at least two sets of data in the literature that can be exploited for this purpose.

Rubin et al. (1999) conducted a one-of-a-kind study that involved 300 subjects and 10 different retention intervals. Subjects were exposed to a continuous paired-associates task, and recall for a previously presented pair was tested at lags ranging from 0 intervening pairs to 99 intervening pairs. The result was the most complete and most precise forgetting function in the literature. Rubin et al. (1999) fit several two-parameter functions to the data and found that although the power function accounted more variance than its competitors (97%), it exhibited obvious systematic deviations and could be rejected on those grounds. A five-parameter sum of exponentials was suggested as an alternative because it fit the data extremely well and without any apparent systematic deviations. The five-parameter exponential tentatively proposed by these authors was as follows: $R(t) = p_1 e^{-p_2 t} + p_3 e^{-p_4 t} + p_5$, where p_1 through p_5 represent free parameters.

As Rubin et al. (1999) observed, there are several reasons for believing that the first two-parameter exponential function (involving parameters p_1 and p_2) reflects a rapid decay to zero from short-term memory (STM), whereas the second, three-parameter exponential function (involving parameters p_3, p_4, and p_5) reflects forgetting to a nonzero asymptote from LTM. Because of its large time constant (p_2 was approximately equal to 0.87), the first exponential function mainly affects the fit over the first two retention intervals only, and those two retention intervals are so short (zero or one intervening item) that few would doubt that retrieval from STM was involved. In addition, the reaction times associated with those two retention intervals were qualitatively faster than the reaction times for the remaining eight retention intervals, and it is known from prior research that retrieval from STM is faster than retrieval from LTM. As Rubin et al. (1999) put it, "[t]he reaction time data also support the claim that Lag 0, and possibly in the recall conditions Lag 1, depend heavily on working memory. . .the reaction times from these lags are much shorter than those from all the later lags, which are similar to each other" (pp. 1172–1173). As shown in the upper panel of Figure 6, when the first two points are excluded, the remaining eight points are extremely well fit by the three-parameter exponential that, hypothetically, captures decay from LTM. This is the same exponential function that accurately describes the LTM data from Wixted and Ebbesen (1991), as shown in Figure 5.

The power law of forgetting is typically applied to data that reflect forgetting from LTM, so one might wonder about its ability to describe these 8 data points as well. A three-parameter version of the power law that was originally proposed by Wickelgren (1974), one that can be fairly compared to the three-parameter exponential, is $R(t) = N_0(\alpha t + 1)^{-k}$, where N_0 is the degree of learning parameter at $t = 0$, k is the proportional rate of forgetting parameter, and α is a scaling constant that is needed because time

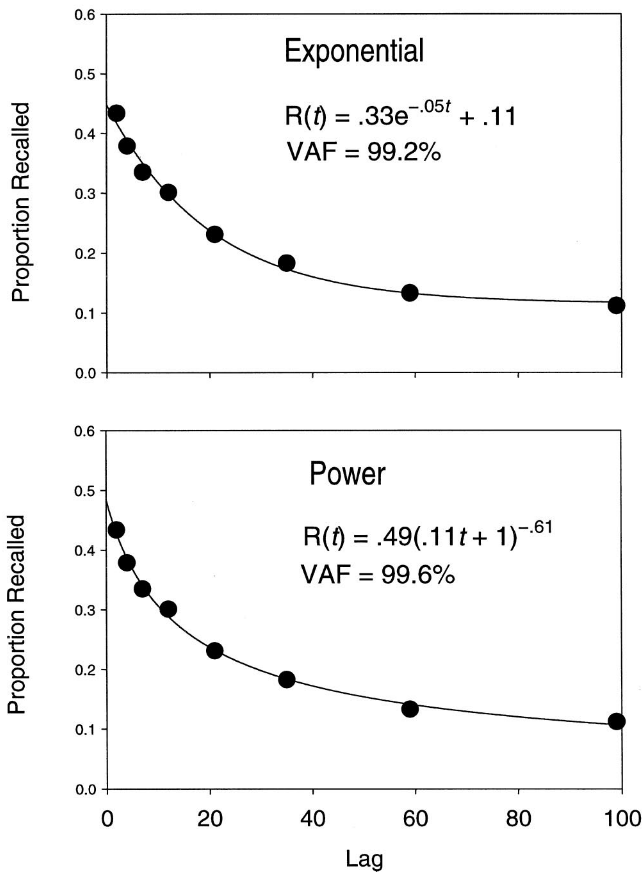

Figure 6. Least-squares fits of the three-parameter exponential (top) and three-parameter power function (bottom) to forgetting data reported by Rubin et al. (1999). Lag is measured in the number of intervening items. VAF = variance accounted for.

is measured in arbitrary units.[1] This function declines to an asymptote of zero, and the lower panel of Figure 6 shows the least squares fit of this function to the 8 relevant data points from Rubin et al. (1999). Obviously, the three-parameter power function fits the data as well as the three-parameter exponential, so a choice between them based on goodness of fit is not possible.

Figure 6 reveals that we face here the same dilemma we faced earlier with respect to the data reported by Wixted and Ebbesen (1991). Specifically, two functions that fit the data equally well are in fundamental conflict on one theoretically fascinating point. The exponential implies that this forgetting function declines to a nonzero asymptote, whereas the power function implies that it will eventually decline to zero. And in this case, the exponential's estimated asymptote of 11% correct may not seem unreasonably large (though it does to me). To determine which equation provides the better estimate of the projected level of performance, one would like to see one more long retention interval added to the design. Although we obviously cannot add anything to the study now, there are enough data points that we can retrospectively perform this exact kind of experiment.

The predictive capabilities of the three-parameter exponential and the three-parameter power function were tested by fitting both to the first 5 of the 8 LTM data points to see how well they predicted where the next 3 data points would fall. The equations were then fit to the first 6 data points to see how well they predicted where the next 2 data points would fall, and then to the first 7 to see how well they predicted where the final data point would fall.

As described in detail by Myung and Pitt (2002), a test of the predictive abilities of competing functions is much more compelling than the traditional goodness-of-fit test based on the percentage of data variance accounted for. The reason is that even functions with the same number of free parameters differ in their ability to adjust themselves to accommodate error variance (i.e., functions differ in flexibility). A function that accounts for a high percentage of variance because of its high flexibility will pay a price in its ability to predict the future course of the data. As Myung and Pitt (2002; Pitt, Kim, & Myung, 2003) showed, this problem can be illustrated by fitting a function with n free parameters to n data points. The fitted function will account for 100% of the variance, but because some of the variability in the data is attributable to error variance, even the correct function will have accommodated itself to that error in order to account for all the data variance. As a result, it will make incorrect predictions.

As the number of data points to be fit exceeds the number of free parameters, an incorrect but flexible function will continue to capitalize on error variance (and will continue to make incorrect predictions about the future path of the data). However, the correct function will settle down and begin to make more accurate predictions. As shown below, fitting the three-parameter power function and the three-parameter exponential to at least 5 data points appears to be sufficient for the former to make accurate predictions.

Figure 7 shows the results of this fitting exercise for the three-parameter exponential. The top panel shows the least squares fit of

[1] It is worth noting that the three-parameter power function is often indistinguishable from the two-parameter power function that is more commonly fit to forgetting data. The functions differ only when t is small because as αt becomes large relative to 1 as t increases, the parenthetical sum, $\alpha t + 1$, becomes approximately equal to αt, so the equation reduces to $R(t) \approx N_0(\alpha t)^{-k}$, which can be rewritten in the more familiar form: $R(t) \approx N_1 t^{-k}$, where $N_1 = N_0 \alpha^{-k}$. Depending on the specific values of α and k, the two- and three-parameter versions of the power function may converge rapidly or slowly as t increases. For the data reported by Wixted and Ebbesen (1991), the parameter values are such that the functions converge quite quickly, so the two-parameter version fits the data about as well as the three-parameter version. For the data reported by Rubin et al. (1999), the three-parameter version of the power function is needed to fit the data because the parameters are such that $\alpha t \approx \alpha t + 1$ only when t becomes fairly large. This does not seem problematic here because the competing function (namely, the exponential) also has thee parameters. The three-parameter power function has desirable properties that address a common criticism of the two-parameter version of this equation (e.g., Rubin et al., 1999; Wickens, 1998). Whereas the two-parameter version explodes to infinity as t approaches zero (which is problematic), the three-parameter version contains a true "degree of learning" parameter (N_0) that reflects the estimated level of performance at $t = 0$ based on retrieval from LTM. The three-parameter version is the conceptually complete version of the power function, but the two-parameter version is often used in practice simply because it can be without any appreciable loss of descriptive accuracy.

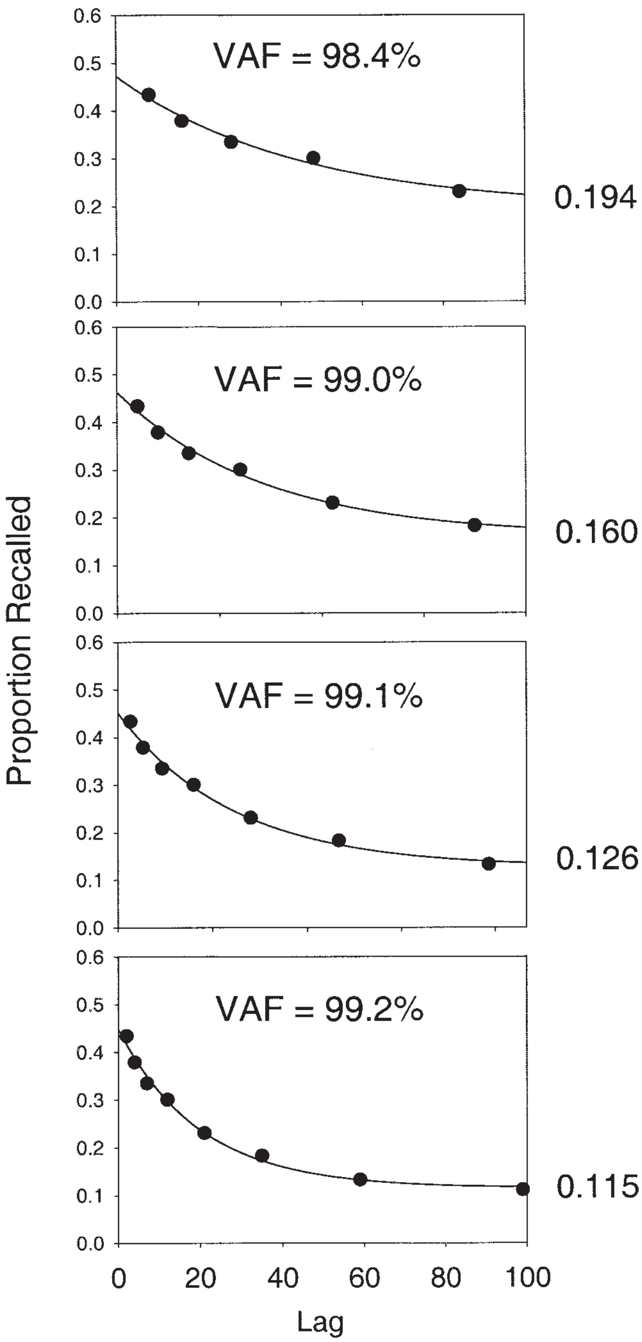

Figure 7. Successive fits of the three-parameter exponential to different subsets of the 8 data points reported by Rubin et al. (1999, Table A1, Column "All 3") that reflect forgetting from long-term memory. The top panel shows a fit through the first 5 points, and the next three panels show how the fits change each time an additional data point is added. Lag is measured in the number of intervening items. VAF = variance accounted for.

this function through the first 5 data points. The next three panels show the least squares fit of this function to 6, 7, and 8 data points, respectively. Two important points need to be emphasized about these fits. First, the three-parameter exponential provides an excellent fit (accounting for nearly 100% of the variance) without systematic deviations in every case. Thus, one would never consider rejecting this function on those grounds (though, in light of recent work by Myung & Pitt, 2002, one might not be terribly impressed by this result either). Second, the estimated asymptote declines systematically as each new retention interval is added. This is not a sensible outcome, and it is no great stretch of the imagination to assume that the trend would continue if additional (even longer) retention intervals were added to the design. The fact that the asymptote changes systematically as each new retention interval is added suggests that this equation does not accurately characterize the time course of forgetting. An asymptote does not decline; it is the point to which the function itself supposedly declines. The upper panel of Figure 8 shows all four fits on the same graph, and the systematically decreasing asymptote is visually apparent.

The same least squares fitting exercise was repeated with the three-parameter power function, and the results are shown in the lower panel of Figure 8. Like the exponential, this function fits the data extremely well in every case. Unlike the exponential, the power function is reasonably accurate in predicting where the data from each new retention interval will fall. Indeed, the successive fits yield functions that fall virtually atop one another as each new retention interval is introduced. This analysis provides evidence that the forgetting function really is on a trajectory toward an asymptote of zero.[2] The same analysis was also performed by means of maximum likelihood estimation instead of least squares, and the results were visually identical (though the exact parameter estimates differed very slightly).

I do not mean to imply that this analysis proves beyond all doubt that forgetting functions learned in the laboratory descend to an asymptote of zero. The data were averaged over subjects, for example, and it is conceivable that averaging artifacts were responsible for the outcome of this test. Still, the successive fitting test described above provides evidence that weighs in favor of the idea that forgetting functions decline toward an asymptote of zero.

Even if one accepts the claim that forgetting functions based on lists of words learned in the laboratory decline to an asymptote of zero, one might still hold to the idea that more substantial learning (such as the kind of learning that takes place in school) will yield forgetting functions that decline to a true nonzero asymptote. Indeed, Bahrick's (1984) famous concept of "permastore," which is based on tests of memory for Spanish words over the course of a lifetime, is a concept that is sometimes construed as being equivalent to the idea that forgetting functions decline to a nonzero

[2] A similar analysis could have been performed on the recognition data reported by Rubin et al. (1999), but those data are problematic because the false alarm rate increased dramatically throughout the course of the experimental session, from approximately 35% at the beginning to nearly 80% at the end. Given this increasing trend, the longer retention intervals, which were necessarily tested later in the session, were unavoidably associated with a larger false alarm rate than the shorter retention intervals. However, the nature of the design did not allow for an assessment of retention interval-specific false alarm rates. Instead, a single false alarm rate, averaged across the entire session, was used. This has the effect of systematically distorting the retention estimates at the longest retention intervals.

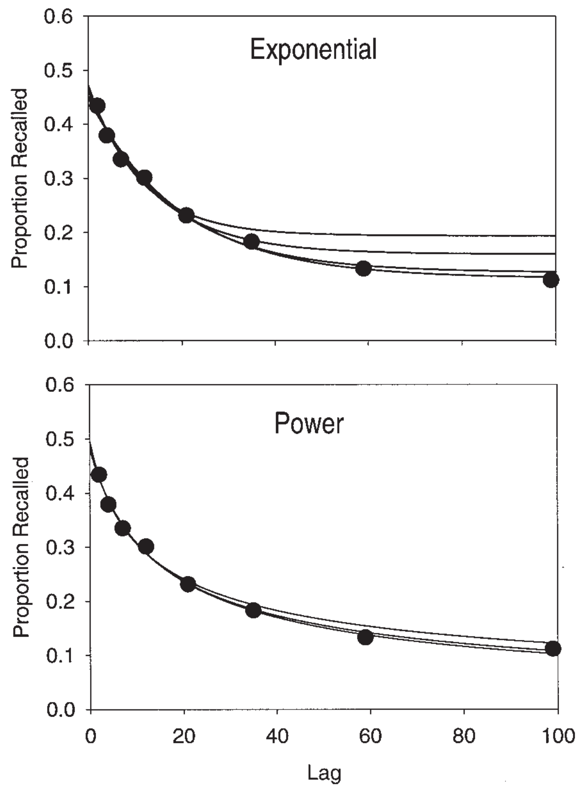

Figure 8. Top: Successive fits of the three-parameter exponential to different subsets of the 8 data points reported by Rubin et al. (1999) that reflect forgetting from long-term memory. These are the same fits shown in Figure 7, now shown on one graph. Bottom: Successive fits to the same data of the three-parameter power function.

asymptote. Does Bahrick's (1984) amazing forgetting function provide strong evidence for that? Not really.

The upper panel of Figure 9 shows the data from Bahrick's (1984) study with the least squares fit of the three-parameter exponential drawn through the data. Bahrick (1984) actually reported separate long-term forgetting functions for 10 different subtests of the Spanish test, but those data were averaged into a single function for purposes of this analysis. Before averaging, the score for each subtest was converted to a percentage scale, with zero representing the performance of control subjects who had never taken a Spanish class and 100 representing the maximum possible score on the subtest (following Hintzman, 1990). The averaged data are somewhat variable, but the three-parameter exponential accounts for 90% of the variance and seems to faithfully convey the permastore story. That is, forgetting occurs at a rapid rate over the first 5 years or so, but then levels off to an asymptote well above zero (which, one might argue, represents permastore).

The lower panel of Figure 9 shows the same data fit with the three-parameter power function. Once again, for the third time, we see that the power function rivals the three-parameter exponential in its ability to describe forgetting from LTM. It was true of the data reported by Wixted and Ebbesen (1991), which spanned a range of only 40 s; it was true of the data reported by Rubin et al. (1999), which spanned a range of approximately 10 min; and is true of the data reported by Bahrick (1984), which spans a range of 50 years.

Adding a fourth parameter to the power function to allow for a nonzero asymptote does not significantly (or even appreciably) improve the fit. Thus, the good fit provided by the power function, coupled with the failure of that fourth parameter to improve the fit at all, suggests that Bahrick's (1984) data may not decline to a nonzero asymptote. Further evidence for this can be found in Figure 10. This figure shows the results of a successive fitting exercise like that performed on the data from Rubin et al. (1999) shown in Figure 8. Specifically, the upper panel shows the results of four separate fits of the three-parameter exponential to Bahrick's (1984) data. The first fit involved the first six points only, and it yielded the highest estimated asymptote (26.6%). The next fit involved the first seven points, and the estimated asymptote dropped slightly (to 25.5%). The next fit involved the first eight points, and the estimated asymptote dropped slightly yet again (to 24.5%). The lowest estimated asymptote was obtained when all nine points were fit (23.5%). Thus, as with Rubin et al.'s (1999) data, the estimated asymptote drops systematically each time a longer retention interval is added (as might be expected if the forgetting function is on a trajectory toward zero). When this successive fitting exercise was repeated using the three-parameter power function (shown in the lower panel of Figure 10), it is clear

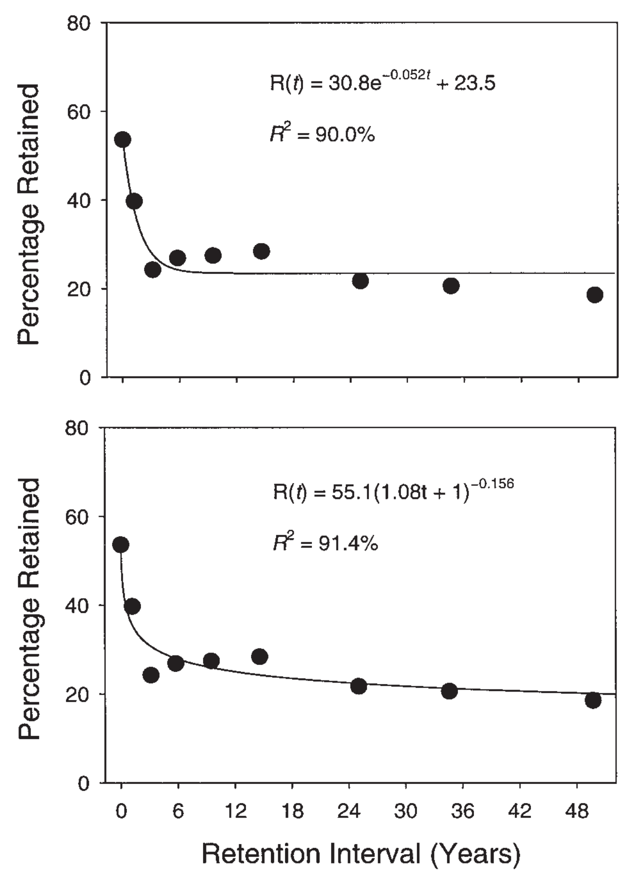

Figure 9. Least-squares fits of the three-parameter exponential (top) and three-parameter power function (bottom) to forgetting data reported by Bahrick (1984).

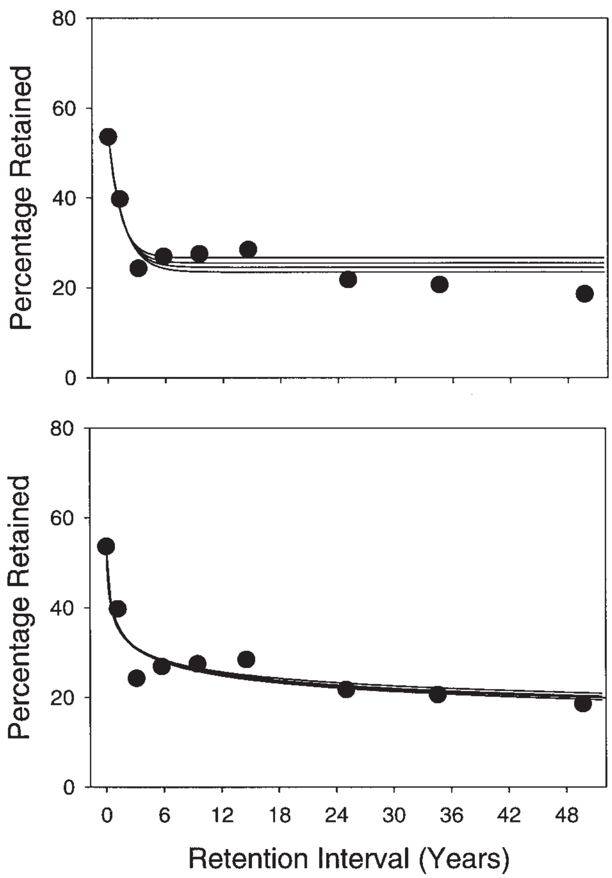

Figure 10. Top: Successive fits of the three-parameter exponential to different subsets of the data reported by Bahrick (1984). Bottom: Successive fits to the same data of the three-parameter power function.

that virtually the same path is predicted each time a longer retention interval is added. Beginning this fitting exercise with even fewer than 6 points leads to inaccurate predictions for both the power function and the exponential, presumably because the data are rather variable, and the more variable the data are, the more points are needed for even the correct function to provide accurate predictions. The notable finding is that the power function begins to make essentially the same (accurate) prediction about the future course of the function when enough points are fit to constrain it, whereas the exponential continues to make systematic errors of the kind that would be expected if the data were on a trajectory toward an asymptote of zero. When these analyses were repeated using maximum likelihood estimation instead of least squares, virtually identical results were obtained. The estimated asymptote of the exponential declines monotonically from 25.9% to 23.6% across the four fits, whereas the four estimated power functions are superimposed and appear to be only one function.

These results weigh against the idea that forgetting functions fall to a nonzero asymptote, but they do not seem to be greatly at odds with the notion of permastore. As indicated earlier, the proportional rate of forgetting associated with the power function declines with the passage of time such that, after many years, the decay rate would be negligible. Indeed, from Years 20 to 50, the predicted decline derived from the best-fitting power function is from 23.1% correct to 20.1% correct (a difference of only 3% over 30 years). Thus, the difference between the notion of permastore and the idea that forgetting proceeds according to a power law is the difference between no further forgetting and a negligible amount of additional forgetting. Although this difference is of great theoretical interest, it is of no practical interest. As Bahrick (1984) said, "[i]t is important to point out that the term permastore has been used here without any intended structural implications. It simply refers to the finding that much of the information in memory has a life span of several decades" (p. 24). The idea that forgetting proceeds according to a power law reinforces rather than challenges this idea. The main point I am making here is that Bahrick's data offer no reason to believe that the data do not decline to an eventual asymptote of zero (though the asymptote would be approached only if people lived for hundreds of years). In fact, those data weigh in favor of the idea that forgetting proceeds toward an asymptote of zero. In a similar vein, Squire (1989) presented data on memory for one-season television programs that had aired during a single year from 1 to 15 years ago, and he concluded that "forgetting in very long-term memory can be gradual and continuous for many years after learning" (p. 241). The analyses presented here reinforce that conclusion as well. Even "flashbulb memories" are no longer regarded as the permanent entities they were once thought to be (e.g., Schmolck, Buffalo, & Squire, 2000). Instead, they decay as other memories do. Thus, on balance, the evidence weighs against the idea that Jost's law holds because an old forgetting function (associated with a relatively high degree of learning) declines to a higher asymptote than the young forgetting function.

Do Memory Traces Become More Robust With the Passage of Time?

The fourth possible explanation for the ever-decreasing rate of decay associated with empirical forgetting functions is that the underlying memory traces themselves individually exhibit an ever-decreasing rate of decay. Jost's law would then be a manifestation of that process.

Why would memories become more resistant to decay with the passage of time? Quite possibly for the same reason that Ribot's law holds. That is, memories become more resistant to the effects of disruptive forces with the passage of time because of the effects of consolidation (Wickelgren, 1974). This is true whether the disruptive forces involve damage to the brain or interference from subsequent learning (i.e., RI). Ribot's law looks backward in time; Jost's law looks forward. But, according to this view, they are both based on the same underlying process: Memories become less vulnerable to the forces of disruption as they consolidate with the passage of time.

If traces do become less vulnerable to interference as they age, then the idea that individual memory traces decay exponentially, which is an idea that all three accounts of Jost's law reviewed above have in common, becomes difficult to sustain. The exponential assumes a constant half-life, and a half-life that remains unchanged even though memory traces become less vulnerable to the forces of interference is something of a contradiction. If traces fall prey to the forces of RI (as they surely do), and if those traces become less vulnerable to RI as they age, then the half-life of a memory trace should increase with age and forgetting would not be exponential in form. Moreover, Jost's law of forgetting would naturally follow.

The idea that memories consolidate and become less vulnerable to the forces of traumatic brain injury is widely accepted, at least by neuroscientists. By contrast, the idea that they become less vulnerable to the forces of ordinary RI as a result of consolidation has been widely and explicitly rejected, at least by psychologists. The probable reason for this rejection is that quite a few studies designed to directly address this issue found that although RI impairs memory, it does not matter whether the RI occurs early or late in the retention interval (the degree of impairment is the same). Wickelgren (1974, 1977), Archer and Underwood (1951), Postman and Alper (1946), McGeoch and Nolen (1933), and Houston (1967) are just some of the researchers who came to this conclusion. Thus, on the basis of studies like these, interference theorists concluded that consolidation, while possibly helping to explain the effects of brain damage, apparently plays no role in the understanding of the time course of forgetting. That verdict has not been seriously challenged in the field of experimental psychology, so no modern cognitive theory of memory includes any provision for a consolidation process. Such a challenge, which seems long overdue, opens the door to a simple explanation for one of psychology's most enduring laws.

In a recent study, Wixted (2004) reviewed a century of research on forgetting and argued that a close look at the relevant literature tells quite a different story from the one summarized above. RI occurring early in the retention interval is more disruptive than RI occurring later in the retention interval, as consolidation theory naturally predicts, but this result has been obscured largely because the RI that matters is not the cue-overload variety (Watkins & Watkins, 1975) that psychologists have diligently studied at least since the 1930s (e.g., an A-B list followed by an A-C list). Instead, what matters is the less specific RI associated with mental exertion and the attendant memory formation, whether or not the new memories are associated with the same retrieval cues as the ones being interfered with.

Oddly enough, this idea corresponds closely to the way of thinking in effect at the dawn of experimental psychology. Müller and Pilzecker (1900) conducted the first study on the effects of RI. In one experiment, which was described by Lechner, Squire, and Byrne (1999) in a partial translation of Müller and Pilzecker's (1900) monograph (which is in German), subjects studied six pairs of syllables and then studied an interfering list either 17 s or 6 min later. The results showed that the retention of the first list was impaired on a cued recall test 1.5 hr later only when the interference occurred 17 s after learning. To explain these results, Müller and Pilzecker (1900) presented what was at the time a novel argument: Memories consolidate over time. If the interfering list is learned before the consolidation of memory for the first list is complete, RI occurs. Müller and Pilzecker were experimental psychologists, and although their work has had virtually no influence on the field of psychology, they are widely credited in the contemporary neuroscience literature with having first advanced the theory of consolidation.

The kind of interference envisioned by Müller and Pilzecker (1900) was relatively nonspecific. That is, mental activity per se was thought to be the force that disrupted recently formed memories. I traced the subsequent history of work on this issue and showed how the field of experimental psychology quickly moved away from the idea that the mental activity impairs recently formed memories and toward the idea interference is primarily a cue-overload retrieval phenomenon (Wixted, 2004). On this view, the mechanism of forgetting was not thought to be trace degradation due to mental activity; instead, forgetting was thought to occur because a retrieval cue that once functioned effectively is rendered less useful by virtue of its association with other memories.

Interference due to cue overload is, of course, a real phenomenon, as the plethora of studies using the A-B, A-C paired-associates paradigm has amply documented. However, Wixted (2004, in press) reviewed a great deal of evidence suggesting that interference due to trace degradation probably accounts for everyday forgetting to a much greater degree than cue overload interference (which is, perhaps, mainly a laboratory phenomenon). And it is interference due to trace degradation that is relevant to consolidation theory and to Jost's law of forgetting. The interfering force is more likely to be the formation of new memories rather than mental activity per se, but the basic idea is otherwise much like what Müller and Pilzecker (1900) had in mind. Traces that are more fully consolidated are less vulnerable to the interfering force of new memory formation, and that may be why forgetting functions exhibit an ever-slowing rate of decay as time elapses. Interference caused by cue overload (a retrieval phenomenon), by contrast, does not decrease as the trace ages.

The idea that memories decay at an ever-slowing proportional rate because they become less vulnerable to the forces of RI with the passage of time, which had been largely abandoned by about 1935, was briefly considered again in the second half of the 20th century by Wickelgren (1974). However, he quickly abandoned the idea as well, mainly because of the results of an experiment that he performed to test it. As with the interference researchers who preceded him, Wickelgren (1974) failed to draw a clear distinction between interference caused by trace degradation and interference caused by cue overload. Instead, he simply took it for granted that cue-overload procedures appropriately model the kind of interference that causes forgetting with the passage of time. His experiment involved a continuous associative recognition memory procedure in which subjects were presented with a long series of paired-associate words to learn. Occasionally during that series, a test pair was presented and subjects were asked to indicate whether the pair was intact (consisting of two words that had appeared together earlier in the series) or rearranged (consisting of two words that had appeared earlier in the series as part of different pairs). An important point to understand about this procedure is that the entire retention interval between study and test was filled with the intentional learning of intervening word pairs, so the temporal point of interpolated interference resulting from memory formation itself was not manipulated. What was manipulated was the temporal point of interpolated interference resulting from cue overload. That is, if an A-B word pair had been studied earlier in the series, an interfering A-C pair was presented either early or late in the retention interval. Compared with a control condition, performance was impaired by the presentation of an A-C pair, but the degree of impairment was the same whether the A-C pair appeared early or late in the retention interval. On the basis of results like these, Wickelgren (1974) rejected his own earlier idea that RI interferes with a young memory trace more than an older one (the very idea I am attempting to resurrect). But the experiment that convinced him to make that move involved an interference procedure that is not relevant to consolidation and may not be very relevant to ordinary forgetting (Wixted, 2004).

The idea that Jost's law reflects a consolidation process that renders memory traces less and less vulnerable to the forces of nonspecific RI is consistent with a great deal of work in the contemporary neuroscience literature (Wixted, 2004). Some of that work is conceptually identical to the work performed by Müller and Pilzecker (1900) long ago. Izquierdo, Schröder, Netto, and Medina (1999), for example, conducted an animal learning study in which they first trained rats on a task called "one-trial step-down inhibitory avoidance" and subsequently interfered with memory for that task by exposing them to a novel environment. The inhibitory avoidance task involved placing the rat on a platform and then delivering a brief shock when it stepped down onto a metal grid. Latency to step down from the platform on subsequent test trials was the measure of memory for the training trial. Exposure to the novel environment (the interfering task) involved placing the rat in an open field with a pink floor adorned with black-lined squares.

After exposure to the original learning trial, the animals were exposed to the interfering task either 1 hr or 6 hr later, and memory for original learning was assessed after a 24-hr retention interval. In agreement with the results reported by Müller and Pilzecker (1900), a temporal gradient of RI was observed: 24-hr memory of the avoidance task was impaired only when the unrelated interfering task was presented 1 hr after learning. This presumably occurred because the memories had not yet had a chance to consolidate when the interfering learning took place. After 6 hr, the memories were more fully consolidated, so exposure to the novel environment had less of an interfering effect. Very similar results had been reported much earlier in the psychological literature in one of the few direct tests of Jost's law of forgetting (Britt & Bunch, 1934).

Similar consolidation effects have been repeatedly observed for a neural phenomenon thought to underlie the initial stages of memory formation in the brain, namely, long-term potentiation (LTP). LTP is an artificially induced enhancement of synaptic transmission that can last for hours or days (and sometimes weeks). It is induced by brief, high-frequency electrical stimulation of a neural pathway, typically in the hippocampus. Xu, Anwyl, and Rowan (1998) conducted an LTP experiment that was conceptually much like the experiment performed by Izquierdo et al. (1999) except that instead of using an actual learning task, these researchers artificially induced hippocampal LTP in freely behaving rats. In essence, LTP was a surrogate memory that would later be exposed to RI after varying delays. Exposure to a novel environment 1 hr after induction completely reversed the previously induced LTP (i.e., interference was complete). However, if exposure to the novel environment was delayed for 24 hr after induction of LTP, no effect of that exposure on LTP (i.e., no interference) was observed. Thus, a temporal gradient was observed yet again, and it suggests that recently established LTP is more vulnerable to the disruptive effects of subsequent interference than more remotely established LTP is (presumably because the latter has had time to consolidate). A similar LTP study performed by Abraham et al. (2002) involved a much more prolonged interference phase and showed that, under such conditions, the LTP temporal gradient can be observed over a period of many weeks (instead of hours, as in Xu et al., 1998). Thus, at both the behavioral level and the neural level, memories appear to become less vulnerable to the forces of nonspecific RI as they age.

This way of thinking is also consistent with the well-known effects of sleep on memory. The classic Jenkins and Dallenbach (1924) study, for example, found that subjects recalled more items after they slept through the retention interval compared with when they remained awake. The intervening activity that is eliminated by sleep does not necessarily involve activities that are captured by A-B, A-C cue-overload methods. Instead, it is ordinary mental activity and memory formation that is prevented by sleep, thereby protecting recently formed memories from that kind of interference. Ekstrand (1972) further showed that recall following a 24-hr retention interval was better when 8 hr of sleep occurred shortly after learning compared with when it occurred just prior to recall. The enhanced performance in the immediate sleep condition presumably arose because, in that condition, memories were protected from interference (of the mental activity and memory formation variety) during the time period when they are the most vulnerable.

Conclusion

If it is true that memory traces become less vulnerable to the effects of subsequent memory formation as a result of the process of consolidation, then forgetting functions would be expected to exhibit an ever-decreasing rate of decay and Jost's law of forgetting would follow naturally. Moreover, the process that accounts for these effects is the same one that is implied by Ribot's law. All may be manifestations of the same underlying process of consolidation.

Why is this explanation for Jost's law more compelling than the alternatives considered earlier? The explanations based on scaling artifacts, "survival of the fittest" memory trace, and asymptotes greater than zero all assume that the temporal properties of a memory trace do not change with the passage of time (i.e., forgetting is exponential in form). Evidence that weighs against each of these ideas is compelling: Forgetting functions based on a measurement scale that, theoretically, is linear with respect to underlying memory strength exhibit the same properties that almost all other forgetting functions do (namely, an ever-decreasing rate of decay). The notion that this property of forgetting functions reflects the survival of the fittest memory trace is challenged by the observation that a great deal of variability in exponential forgetting rates is required, even though the available evidence suggests that very different kinds of items decay at approximately the same rate once they are encoded. And the data purportedly showing that forgetting functions decline to a nonzero asymptote actually provide the most compelling evidence to date that they decline to an asymptote of zero instead.

But most damning for these accounts is independent evidence from the neuroscience literature (and old but forgotten evidence from the psychology literature) suggesting that memory traces become more resistant to interference with the passage of time. A wealth of independent evidence from research on LTP, retrograde amnesia, and the temporal gradient of RI supports the notion that memory traces become less fragile with the passage of time. The idea that traces become less vulnerable to the forces of interference with the passage of time is not easily reconciled with the idea that memory traces have a constant rate of decay. The most parsimonious interpretation, therefore, is that Ribot's law of retrograde amnesia and Jost's law of forgetting have the same theoretical implications: Memories become less fragile to disruptive forces with the passage of time.

If this interpretation of Jost's law is correct, then the dominant view of the time course of forgetting in the field of psychology for 70 years is incorrect. According to the prevailing view, forgetting is largely a cue-overload phenomenon that operates at the time of retrieval and is both proactive and retroactive in direction. In such a view of forgetting (as has long been observed), the notion of consolidation has no role to play. However, the interpretation of Jost's law advanced here implies that forgetting is largely a trace degradation phenomenon and that interference is retroactive in direction. New memories degrade (but do not necessarily overwrite) previously formed memories, more so for recently formed memories than for ones formed longer ago. That way of thinking was in effect at the dawn of experimental psychology, and recent developments suggest that the psychologists who proposed it (viz., Müller & Pilzecker, 1900) were ahead of their time.

References

Abraham, W. C., Logan, B., Greenwood, J. M., & Dragunow, M. (2002). Induction and experience-dependent consolidation of stable long-term potentiation lasting months in the hippocampus. *Journal of Neuroscience, 22,* 9626–9634.

Anderson, J. R. (2000). *Learning and memory: An integrated approach* (2nd ed.). New York: Wiley.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2,* 396–408.

Archer, J. E., & Underwood, B. J. (1951). Retroactive inhibition of verbal associations as a multiple function of temporal point of interpolation and degree of interpolated learning. *Journal of Experimental Psychology, 42,* 283–290.

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

Bogartz, R. S. (1990). Evaluating forgetting curves psychologically. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 138–148.

Britt, S. H., & Bunch, M. E. (1934). Jost's law and retroactive inhibition. *American Journal of Psychology, 46,* 299–308.

Ebbinghaus, H. (1913). *Memory. A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College, Columbia University. (Original work published 1885)

Ekstrand, B. R. (1972). To sleep, perchance to dream (about why we forget). In C. P. Duncan, L. Sechrest, & A. W. Melton (Eds.), *Human memory: Festschrift for Benton J. Underwood* (pp. 59–82). New York: Appelton-Century-Crofts.

Estes, W. K. (1979). Experimental psychology: An overview. In E. Hearst (Ed.), *The first century of experimental psychology* (pp. 623–667). Hillsdale, NJ: Erlbaum.

Hintzman, D. (1990). *Permastore or grade inflation? Adjusting Bahrick's data for changes in academic standards* (Tech. Rep. No. 90-15). Eugene: University of Oregon, Institute of Cognitive and Decision Sciences.

Houston, J. (1967). Retroactive inhibition and point of interpolation. *Journal of Verbal Learning and Verbal Behavior, 6,* 84–88.

Izquierdo, I., Schröder, N., Netto, C. A., & Medina, J. H. (1999). Novelty causes time-dependent retrograde amnesia for one-trial avoidance in rats through NMDA receptor- and CaMKII-dependent mechanisms in the hippocampus. *European Journal of Neuroscience, 11,* 3323–3328.

Jenkins, J. B., & Dallenbach, K. M. (1924). Oblivescence during sleep and waking. *American Journal of Psychology, 35,* 605–612.

Jost, A. (1897). Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen [The strength of associations in their dependence on the distribution of repetitions]. *Zeitschrift fur Psychologie und Physiologie der Sinnesorgane, 16,* 436–472.

Keppel, G. (1968). Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 172–213). Englewood Cliffs, NJ: Prentice-Hall.

Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 years of consolidation—Remembering Müller and Pilzecker. *Learning & Memory, 6,* 77–87.

Loftus, G. (1985a). Consistency and confoundings: Reply to Slamecka. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 817–820.

Loftus, G. (1985b). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 397–406.

Loftus, G., & Bamber, D. (1990). Learning-forgetting independence, unidimensional memory models, and feature models: Comment on Bogartz (1990). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 916–926.

McGaugh, J. L. (2000, January 14). Memory: A century of consolidation. *Science, 287,* 248–251.

McGeoch, J. A., & Nolen, M. E. (1933). Studies in retroactive inhibition: IV. Temporal point of interpolation and degree of retroactive inhibition. *Journal of Comparative Psychology, 15,* 407–417.

McKee, R., & Squire, L. R. (1992). Equivalent forgetting rates in long-term memory for diencephalic and medial temporal lobe amnesia. *Journal of Neuroscience, 12,* 3765–3772.

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

Myung, I. J., & Pitt, M. A. (2002). Mathematical modeling. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology (3rd ed.), Vol. 4: Methodology in experimental psychology* (pp. 429–460). New York: Wiley.

Pitt, M. A., Kim, W., & Myung, I. J. (2003). Flexibility versus generalizability in model selection. *Psychonomic Bulletin & Review, 10,* 29–43.

Postman, L., & Alper, T. (1946). Retroactive inhibition as a function of the time interpolation of the inhibitor between learning and recall. *American Journal of Psychology, 59,* 439–449.

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

Ribot, T. (1882). *The diseases of memory: An essay in the positive psychology* (W. H. Smith, Trans.). London: Kegan Paul, Trench, & Co.

Rubin, D. C., Hinton, S., & Wenzel, A. (1999). The precise time course of retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5,* 1161–1176.

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

Schmolck, H., Buffalo, E. A., & Squire, L. R. (2000). Memory distortions develop over time: Recollections of the O. J. Simpson trial verdict after 15 and 32 months. *Psychological Science, 11,* 39–45.

Sikström, S. (1999). Power function forgetting curves as an emergent property of biologically plausible neural network models. *International Journal of Psychology, 34,* 460–464.

Sikström, S. (2002). Forgetting curves: Implications for connectionist models. *Cognitive Psychology, 45,* 95–152.

Simon, H. A. (1966). A note on Jost's law and exponential forgetting. *Psychometrika, 31,* 505–506.

Slamecka, N. J. (1985). On comparing rates of forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 812–816.

Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 384–397.

Squire, L. R. (1989). On the course of forgetting in very long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 241–245.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99,* 195–231.

Squire, L. R., Clark, R. E., & Knowlton, B. J. (2001). Retrograde amnesia. *Hippocampus, 11,* 50–55.

Underwood, B. J., & Keppel, G. (1963). Retention as a function of degree of learning and letter-sequence interference. *Psychological Monographs: General & Applied, 77*(4, Whole No. 567).

Underwood, B. J., & Postman, L. (1960). Extraexperimental sources of interference in forgetting. *Psychological Review, 67,* 73–95.

Watkins, C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning & Memory, 1,* 442–452.

White, K. G. (2001). Forgetting functions. *Animal Learning & Behavior, 29,* 193–207.

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

Wickelgren, W. A. (1977). *Learning and memory*. Englewood Cliffs, NJ: Prentice Hall.

Wickens, T. D. (1998). On the form of the retention function: Comment on Rubin and Wenzel (1996): A quantitative description of retention. *Psychological Review, 105,* 379–386.

Wixted, J. T. (1990). Analyzing the empirical course of forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 927–935.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55,* 235–269.

Wixted, J. T. (in press). Why do we forget what we once knew? *Current Directions in Psychological Science*.

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

Wixted, J. T., & Ebbesen, E. (1997). Genuine power curves in forgetting. *Memory & Cognition, 25,* 731–739.

Woodworth, R. S. (1938). *Experimental psychology*. New York: Henry Holt and Company.

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology*. New York: Henry Holt and Company.

Xu, L., Anwyl, R., & Rowan, M. J. (1998, August 27). Spatial exploration induces a persistent reversal of long-term potentiation in rat hippocampus. *Nature, 394,* 891–894.

Youtz, A. C. (1941). An experimental evaluation of Jost's laws. *Psychological Monographs, 53*(1, Whole No. 238).

Received July 9, 2003
Revision received March 24, 2004
Accepted March 26, 2004 ■