Absolute vs. Relative Forgetting

John T. Wixted[1]

[1]University of California, San Diego

John T. Wixted, Department of Psychology, University of California, San Diego.

Correspondence concerning this article should be addressed to John T. Wixted (jwixted@ucsd.edu).

Abstract

Slamecka and McElree (1983) and Rivera-Lares et al. (2022), like others before them, factorially manipulated the number of learning trials and the retention interval. The results revealed two unsurprising main effects: (1) the more study trials, the higher the initial degree of learning, and (2) the longer the retention interval, the more items were forgotten. However, across many experiments, the interaction was not significant, a finding that is often interpreted to mean that the degree of learning is independent of the absolute rate of forgetting (i.e., the absolute number of items forgotten per unit time). Yet there is considerable tension between that interpretation and the fact that forgetting has long been characterized by a power law, according to which the absolute rate of forgetting is not a particularly meaningful measure. When the power function is fit to the same data, the results show that a higher degree of learning results in a lower relative (i.e., proportional) rate of forgetting. This raises an interesting question: which of the two definitions of "forgetting rate" (absolute vs. relative) is theoretically relevant? Here, I make the case that it is the relative rate of forgetting. Theoretically, the explanation of why a higher degree of learning is associated with a lower relative rate of forgetting may be related to why, as observed by Jost (1897) long ago, the passage of time itself is associated with a lower relative rate of forgetting.

Absolute vs. Relative Forgetting

Rivera-Lares et al. (2022) recently revisited a classic question: Does the initial degree of learning affect the rate of forgetting? The question was last addressed in earnest when Slamecka and McElree (1983) reported the results of three experiments in which they varied both the number of study trials (i.e., the degree of learning) and the retention interval, the longest of which was five days. The dependent measure was the number of items correctly recalled (n) for each retention interval (t). In each of three experiments, the degree of learning (i.e., n at $t = 0$) increased with the number of study repetitions, and the amount retained decreased with the size of the retention interval. However, the interaction was consistently not significant. Slamecka and McElree (1983) therefore concluded that the rate of forgetting is independent of the degree of learning, noting that this conclusion neither confirmed nor contradicted any existing theory of memory because theorists tend to neglect the problem of normal forgetting.

Given the potential importance of findings like these to theory development, Rivera-Lares et al. (2022) recently replicated one of Slamecka and McElree's (1983) experiments multiple times in four new experiments. In all four experiments, the number of list repetitions at study differed for different subgroups of participants (e.g., 2, 4 or 6 repetitions), and each participant was tested on a different subset of studied items following three different retention intervals (e.g., 30 seconds, 1 day, and one week). The results confirmed Slamecka and McElree's (1983) findings: the degree of learning increased with the number of study repetitions, and the amount retained decreased with the size of the retention interval. However, the interaction was again consistently not significant. Like Slamecka and McElree (1983), Rivera-Lares et al. (2022) concluded that "No current theory of forgetting can account for forgetting that starts from different levels yet results in parallel forgetting slopes that are negatively accelerated" (p. 11).

Testing for an interaction implicitly assumes that the forgetting rate of interest—that is, the measure to be explained at the level of theory—is the absolute loss per unit time. As an example, if performance drops from 10 items recalled to 8 items recalled between $t = 0$ days and $t = 1$ day for the high-degree-of-learning group (a loss of 2 items per day), and if it falls from 6 items recalled to 4 items recalled over the same period of time for the low-degree-of-learning group (also a loss of 2 items per day), then the forgetting rates would be considered equivalent, and a theoretical mechanism would be needed to explain why.

In the years following the publication of Slamecka and McElree (1983), Loftus (1985a) ignited a debate about how to analyze the time course of forgetting by pointing out that the use of absolute loss per unit time does not necessarily provide an interval measurement scale of a latent variable like memory strength (Bogartz, 1990; Loftus, 1985b; Slamecka, 1985; Wixted, 1990). If not, then the apparently similar rates of forgetting across learning conditions could be an illusion.

This is clearly a valid concern, but given how consistent the empirical findings seem to be, it also seems reasonable to proceed as if the dependent measure may have the appropriate measurement-scale properties over much of its range. This might be why the ANOVA-based approach to comparing the rate of forgetting across conditions remains a common practice to this day (e.g., Cohen-Dallal, Fradkin, & Pertzov, 2018; Lombardi et al., 2018; Rivera-Lares et al., 2022; Staugaard & Berntsen, 2019). Yet even if one allows for the possibility of an interval measurement scale under those conditions, the absolute loss per unit time may not be the most theoretically informative way to quantify the rate of forgetting, and that is the issue I focus on in this article.

The alternative to an absolute measure of forgetting rate is a relative measure of forgetting rate, with the loss of information per unit time expressed in proportional (or percentage) terms. Slamecka and McElree (1983) argued strongly in favor of the absolute measure, but I here advance the argument that the relative measure is much more relevant from a theoretical perspective. Before considering Slamecka and McElree's (1983) argument in favor of an absolute measure, it worth considering an issue that has not been factored into the discussion as much as it perhaps should be, namely, the mathematical form of forgetting.

The mathematical form of forgetting

Ebbinghaus (1885/1913) first reported a finding that has been replicated many times since: the time course of forgetting is typically curvilinear, with information lost rapidly at first and more slowly thereafter. That is, the rate of forgetting decreases with t. Critically, that generic observation applies whether the rate of forgetting is conceptualized in absolute or relative terms. However, these two measures need not agree with each other. This point can be illustrated by comparing three candidate forgetting functions, namely, the exponential function, the power function, and the exponential-power function.

The mathematics of absolute vs. relative forgetting

Exponential function. Imagine first that forgetting is characterized by the exponential function, $n(t) = n_0 e^{-bt}$, where n_0 represents the number of items recalled at time $t = 0$ (i.e., the degree of learning) and b represents the decay constant that governs both the absolute and relative rate of forgetting (Loftus, 1985). Beginning with the absolute rate of forgetting, consider two retention intervals that differ by Δt. In that case, the absolute loss from t to $t + \Delta t$ would be $n(t) - n(t + \Delta t)$, and the absolute *rate of forgetting*, represented here by $a(\Delta t)$, would equal the absolute loss divided by Δt. That is, $a(\Delta t) = \frac{n(t)-n(t+\Delta t)}{\Delta t}$. If we set $\Delta t = 1$ day for simplicity, then $a(\Delta t) = \frac{n(t)-n(t+1)}{1} = n(t) - n(t + 1)$. Thus, under the exponential-forgetting scenario with $\Delta t = 1$, $a(\Delta t) = n_0 e^{-bt} - n_0 e^{-b(t+1)} = n_0(1 - e^{-b})e^{-bt}$. This equation can be further simplified to $a(\Delta t) = ke^{-bt}$, where k is a constant given by $k = n_0(1 - e^{-b})$. This version of the equation underscores the fact that only e^{-bt} changes with respect to t. And because e^{-bt} decreases as t increases, exponential forgetting implies that, for any $t \geq 0$, the absolute number of items forgotten from t to $t + 1$ decreases with time.

The negative of $a(\Delta t)$, or $-a(\Delta t)$, represents the absolute rate of *change* in the number of items recalled from t to $t + 1$ (i.e., the change is in the negative direction). Transforming the absolute rate of forgetting into a change score makes it directionally compatible with the first derivative of the exponential, which is the absolute rate of change from t to $t + \Delta t$ as $\Delta t \to 0$. In other words, it is the instantaneous absolute rate of forgetting at time t, or $a(t)$. The first derivative of the exponential is $a(t) = \frac{d}{dt}n(t) = \frac{d}{dt}[-n_0 be^{-bt}] = -bn_0 e^{-bt}$. Note that both n_0 and b in this expression are constant with respect to t. Thus, as with the simpler example using $\Delta t = 1$ day, the operative variable governing the absolute rate of forgetting remains e^{-bt} even as $\Delta t \to 0$. Therefore, according to the exponential, the absolute loss per unit time decreases with t.

Now consider the relative (i.e., proportional) loss from t to $t + \Delta t$, which is simply the absolute loss, $n(t) - n(t + \Delta t)$, divided by the level of performance at time t, namely, $n(t)$. That is, the relative loss is equal to $\frac{n(t)-n(t+\Delta t)}{n(t)}$. The *relative rate of forgetting*, represented here by $r(\Delta t)$, is simply the relative loss divided by Δt. That is, $r(\Delta t) = \frac{n(t)-n(t+\Delta t)}{n(t)\Delta t}$. For the simple case of $\Delta t = 1$ day, this comes to $r(\Delta t) = \frac{n(t)-n(t+1)}{n(t)} = \frac{n_0 e^{-bt} - n_0 e^{-b(t+1)}}{n_0 e^{-bt}} = 1 - e^{-b}$. Note that t does not appear on the right side of this equation, which means that, unlike the absolute rate of forgetting, the relative rate of forgetting from t to $t + 1$ is constant for all t. Thus, for example, if 20% of the items that were recalled at $t = 0$ days were lost by $t = 1$ day, then 20% of items that were recalled at $t = 1$ day would be lost by $t = 2$ days (and so on).

The instantaneous version of the relative rate of forgetting (i.e., as $\Delta t \to 0$), represented here as $r(t)$, is known as a hazard function, and it leads to the same conclusion (Chechile, 2006). The hazard function is given by $r(t) = \frac{f(t)}{S(t)}$, where $S(t)$ is the "survival" function, and $f(t)$ is the probability density function of the survival random variable.

$$n(t) = n_0 e^{-bt}$$

$$a(t) = -bn_0 e^{-bt}$$

$$r(t) = \frac{-a(t)}{n(t)} = \frac{bn_0 e^{-bt}}{n_0 e^{-bt}} = b$$

$S(t)$ is the probability than an encoded memory is still alive after time t, which, for the exponential, is simply $S(t) = e^{-bt}$. The corresponding cumulative density function, $F(t)$, is $F(t) = 1 - S(t) = 1 - e^{-bt}$, and $f(t)$ is the first derivative of $F(t)$. That is, $f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}(1 - e^{-bt}) = be^{-bt}$. Thus, $r(t) = \frac{f(t)}{S(t)} = \frac{be^{-bt}}{e^{-bt}} = b$. In other words, the instantaneous relative rate of forgetting, $r(t)$, is equal to a constant with respect to t. A constant relative rate of decay is the defining—and most theoretically interesting—property of the exponential. It is the reason why the exponential is referred to as being "memoryless" (i.e., the relative rate of decay remains the same no matter how much time has elapsed).

Power function. Now imagine that forgetting is instead characterized by a power function, as empirical forgetting functions have long been thought to be. In a hand-written document cited by Murre and Dros (2015), Ebbinghaus (1880) considered the possibility that his classic savings function could be characterized by a power function, though, in his 1885 monograph, he ended up using a logarithmic function with a very similar shape. Nearly a century later, Wickelgren (1974) noted that forgetting from long-term memory is well-characterized by a power function of the form $n(t) = n_0(1 + \delta t)^{-b}$, where, again, n_0 represents the degree of learning, b governs both the absolute and relative rates of forgetting, and δ is a scaling constant that is henceforth set to 1 for simplicity. Note that this version of the power function, like the exponential, ranges from a maximum of n_0 at $t = 0$ (the degree of learning) to a minimum of 0 as $t \to \infty$. The idea that forgetting follows something close to a power law is consistent with a considerable body of research conducted since Wickelgren proposed it (e.g., Anderson & Schooler, 1991; Donkin & Nosofsky, 2012; Rubin & Wenzel, 1996; Murre & Dros, 2015; Wixted & Ebbesen, 1991; Wixted & Carpenter, 2007).

Setting $\Delta t = 1$ for simplicity and following the same steps as for the exponential above, according to the power law of forgetting, the absolute rate of forgetting from day t to day $t + 1$ is equal to $a(\Delta t) = \frac{n(t)-n(t+1)}{1} = n(t) - n(t+1) = n_0(1+t)^{-b} - n_0(1+t+1)^{-b}$. Because both terms on the right are power functions that approach 0 as $t \to \infty$, the absolute rate of forgetting decreases with time, a property it shares with the exponential. The same point can be made in a more formal way by computing the first derivative of the power function, $a(t)$, which is the instantaneous absolute rate of forgetting (i.e., as $\Delta t \to 0$). For the power function, $a(t) = \frac{d}{dt} n(t) = \frac{d}{dt}[n_0(1+t)^{-b}] = -bn_0(1+t)^{-b-1}$. Thus, the instantaneous absolute rate of forgetting is simply another power function that declines to 0 as $t \to \infty$.

The relative rate of forgetting from day t to day $t + 1$ (i.e., $\Delta t = 1$) is equal to $r(\Delta t) = \left[\frac{n(t)-n(t+1)}{n(t)\Delta t}\right] = \frac{n(t)-n(t+1)}{n(t)} = \frac{n_0(1+t)^{-b}-n_0(1+t+1)^{-b}}{n_0(1+t)^{-b}} = 1 - \left(\frac{t+1}{t+2}\right)^b$. Because $\left(\frac{t+1}{t+2}\right)^b \to 1$ as t increases, it follows that $1 - \left(\frac{t+1}{t+2}\right)^b \to 0$ as t increases. Thus, the relative rate of forgetting also decreases with time, in contrast to the exponential (i.e., the power law is not memoryless).

$$n(t) = n_0(1+t)^{-b}$$

$$a(t) = -bn_0(1+t)^{-b-1}$$

$$r(t) = \frac{-a(t)}{n(t)} = \frac{bn_0(1+t)^{-b-1}}{n_0(1+t)^{-b}} = b(1+t)^{-1} = \frac{b}{1+t}$$

As before, the instantaneous relative rate of forgetting (i.e., as $\Delta t \rightarrow 0$) is given by the hazard function, $r(t) = \frac{f(t)}{S(t)}$. For the power function, $S(t) = (1+t)^{-b}$, which means that $F(t) = 1 - S(t) = 1 - (1+t)^{-b}$, and $f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}[1 - (1+t)^{-b}] = b(1+t)^{-b-1}$. Thus, $r(t) = \frac{f(t)}{S(t)} = \frac{b(1+t)^{-b-1}}{(1+t)^{-b}} = b(1+t)^{-1} = \frac{b}{1+t}$. This equation indicates that the instantaneous relative rate of forgetting is not constant but decreases as t increases. This is a mathematical instantiation of Jost's (1897) famous law of forgetting, which holds that if two memories that are of the same strength but different ages, the older trace will decay more slowly than the younger trace.

Exponential-power function. Finally, consider an exponential-power function, $n(t) = n_0 e^{-bt^c}$, where $c > 1$. For $\Delta t = 1$, the absolute rate of forgetting from day t to day $t+1$ would be $a(\Delta t) = \frac{n(t) - n(t+1)}{1} = n(t) - n(t+1) = n_0 e^{-bt^c} - n_0 e^{-b(t+1)^c} = n_0\left(e^{-bt^c} - e^{-b(t+1)^c}\right)$. For $t = 0$ (i.e., over the first day from $t = 0$ to $t = 1$), the absolute rate of forgetting would be $a(\Delta t) = n_0(e^{-0} - e^{-b}) = n_0(1 - e^{-b})$. As an example, for $b = 1$, $1 - e^{-b} = 0.63$, so the absolute rate of forgetting over the first day would be $0.63n_0$. However, as $t \rightarrow \infty$, both e^{-bt^c}

and $e^{-b(t+1)^c}$ approach 0, so the entire expression approaches 0 as well (implying that the absolute rate of forgetting decreases to 0 over time). Once again, the same point can be made in a more formal way by computing the first derivative of the exponential power function. For the exponential-power function, $a(t) = \frac{d}{dt}n(t) = \frac{d}{dt}[n_0 e^{-bt^c}] = -bct^{c-1}n_0 e^{-bt^c}$. This is a somewhat complicated expression, but it is nevertheless easy to see that with $n_0 > 0$, $b > 0$, and $c > 1$, such that both bt^{c-1} and $n_0 e^{-bt^c}$ are positive for all $t > 0$, the derivative itself is always negative. A negative first derivative for all $t > 0$ means that the instantaneous absolute rate of forgetting always decreases as t increases. Thus, all three functions share this property (as would any curvilinear function that is generally consistent with the curvilinear form of forgetting).

By contrast, the relative rate of forgetting from day t to day $t + 1$ is equal to $r(\Delta t) = \frac{n(t)-n(t+1)}{n(t)\Delta t} = \frac{n(t)-n(t+1)}{n(t)} = \frac{n_0 e^{-bt^C}-n_0 e^{-b(t+1)^C}}{n_0 e^{-bt^C}} = 1 - \frac{e^{bt^C}}{e^{b(t+1)^C}}$. For $t = 0$ (i.e., over the first day), the relative rate of forgetting would be $r(\Delta t) = 1 - \frac{e^{b0^C}}{e^{b(0+1)^C}} = 1 - \frac{e^0}{e^b} = 1 - \frac{1}{e^b}$ (which comes to 0.63 for $b = 1$), but as $t \to \infty$, $\frac{e^{bt^C}}{e^{b(t+1)^C}} \to 0$, so $1 - \frac{e^{bt^C}}{e^{b(t+1)^C}} \to 1$. In other words, for $b = 1$, as t increases from 0 to ∞, the relative rate of forgetting from day t to day $t + 1$ *increases* from 0.63 when $t = 0$ to 1 when $t \to \infty$ (the opposite of the trend exhibited by the power law).

$$n(t) = n_0 e^{-bt^c}$$

$$a(t) = -bct^{c-1}n_0 e^{-bt^c}$$

$$r(t) = \frac{-a(t)}{n(t)} = \frac{bct^{c-1}n_0 e^{-bt^c}}{n_0 e^{-bt^c}} = bct^{c-1}$$

The instantaneous relative rate of forgetting (i.e., as $\Delta t \rightarrow 0$) is given by $r(t) = \frac{f(t)}{S(t)}$, and it leads to the same conclusion. For the exponential-power function, $S(t) = e^{-bt^c}$, which means that $F(t) = 1 - S(t) = 1 - e^{-bt^c}$, and $f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}[1 - e^{-bt^c}] = bct^{c-1}e^{-bt^c}$. Thus, $r(t) = \frac{f(t)}{S(t)} = \frac{bct^{c-1}e^{-bt^c}}{e^{-bt^c}} = bct^{c-1}$. For $c > 1$, this equation indicates that the instantaneous relative rate of forgetting increases with t (whereas it remains constant for the exponential and decreases for the power function).

A graphical illustration of absolute vs. relative forgetting

Basically, any curvilinear function consistent with one's intuition about the general form of forgetting, including the three functions under consideration here, will exhibit an ever-decreasing absolute rate of forgetting. Even so, as explained mathematically above, they differ in what they assume about the relative rate of forgetting. The same point can be illustrated graphically using a concrete example for each function. The left column of Figure 1 illustrates the exponential, power, and exponential-power forgetting functions (top to bottom). Each function represents the amount of information still retained as a function of time, and they all roughly correspond to a typical forgetting function (i.e., forgetting is curvilinear and negatively accelerated with respect to time). As illustrated in the middle column of Figure 1, all three functions also agree that the absolute loss per unit time decreases as a function of time. There is nothing particularly interesting about that fact because it is just another way of saying that the form of forgetting is curvilinear. However, as illustrated in the right column of Figure 1, the relative loss per unit time differs for the three functions. The relative loss per unit time is constant for the exponential, ever decreasing for the power function, and ever increasing for the exponential power function (with $c > 1$).

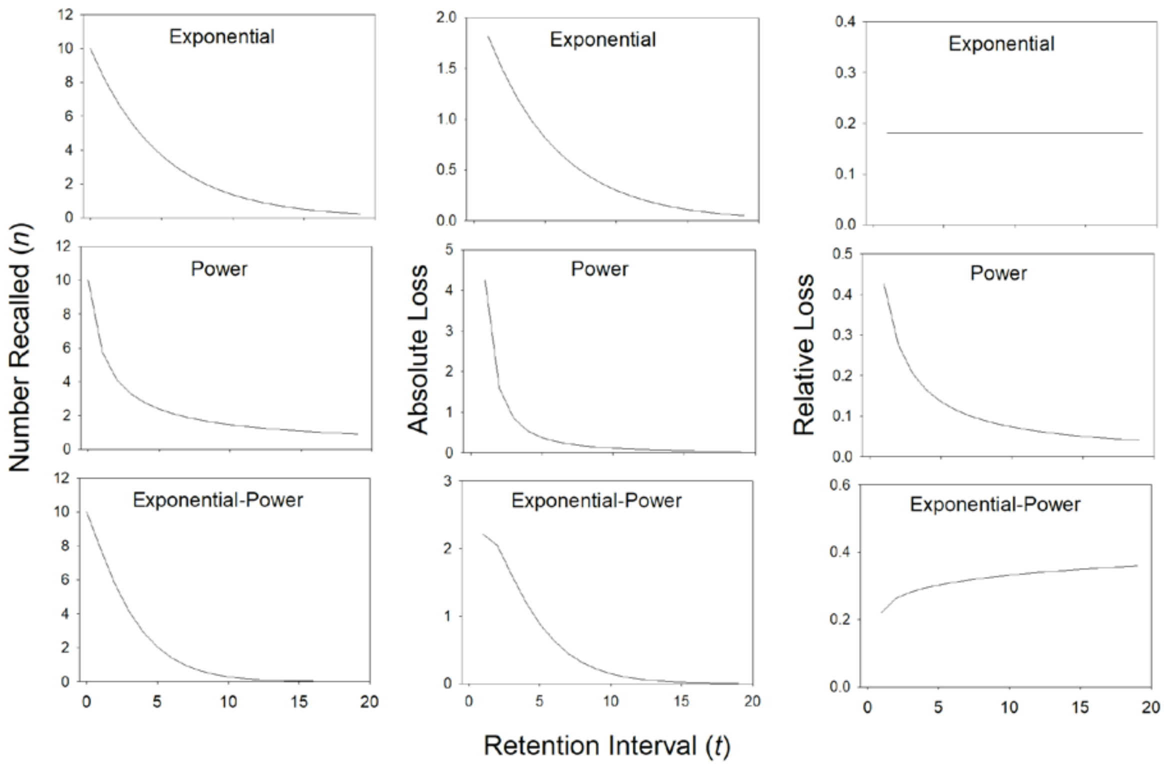

Figure 1. Left Panel: Illustration of three functions that differ in their mathematical forms but nevertheless capture the typical curvilinear (negatively accelerating) shape of most empirical forgetting functions. Middle Panel: A plot of the absolute loss per day, $n_t - n_{t+1}$, for each function shown in the left panel. Right Panel: A plot of the relative loss per day, $(n_t - n_{t+1})/n_t$, for each function shown in the left panel.

Which rate-of-forgetting measure—absolute or relative—is the more theoretically relevant of the two? The next section provides three analogies to illustrate intuitively why it is the relative measure, not the absolute measure, that reflects the underlying theoretical mechanism. The subsequent section then reviews extant theories of forgetting that all predict an ever-decreasing absolute rate of forgetting but make different predictions about the relative rate of forgetting, which again underscores the point that it is the relative measure that is of theoretical interest. This will set the stage for a reconsideration of the main question of interest: how does the degree of learning affect the rate of forgetting?

Analogies

Imagine two car companies. Company A produces and sells 100 cars, and Company B produces and sells 1000 cars. In this analogy, the cars correspond to memory traces, and Company B corresponds to a higher degree of learning. If each car manufactured by either company has a first-year probability-of-failure equal to .01 (i.e., the two companies make cars of equal reliability), then the absolute number of cars lost by Company A in the first year would be $.01 \times 100 = 1$, whereas the absolute number of cars lost by Company B in the first year would be $.01 \times 1000 = 10$. By the absolute measure, the cars manufactured by Company A are superior (only 1 car lost during the first year), but the proportional measure more accurately reveals that the quality of the cars produced by the two companies is the same. Therefore, focusing on the absolute measure to elucidate the underlying mechanism accounting for Company A's apparent superiority would be futile.

Next consider two countries, both of which are suddenly plagued by a deadly COVID-19 pandemic. Country A has a population of 10,000 people, and Country B has a population of 1,000,000 people, and everyone in both countries is infected by the virus. In this analogy, the people of each country correspond to memory traces, and Country B corresponds to a higher degree of learning. In both countries, some people die of the disease. Country A experiences an absolute loss of 100 people to the disease in the first month, whereas Country B experiences a much larger absolute loss of 1000 people over the same period of time. In terms of the absolute fatality rate, the people of Country A (100 lost / month) seem far more resilient to the disease than the people of Country B (1000 lost / month). However, the opposite is true when the fatality rate is measured in relative terms. Whereas Country A loses 100 / 10,000 =.01 of its population in the first month, Country B loses only 1000 / 1,000,000 =.001 of its population.

In a scenario like this, researchers who focus on the absolute fatality rate to understand why the people of Country A are more resilient to the disease than the people of Country B would be wasting their time. Researchers who instead focus on the relative fatality rate would stand a much better chance of uncovering the true underlying mechanism. For example, they might discover that the people of Country A are all older than 65 and have weakened immune systems due to the normal aging process, whereas the people of Country B are all younger than 40 and have stronger immune systems.

These two examples considered a change in the measure of interest over a single Δt (one year in the car example and one month in the contagious disease example) to illustrate why the relative measure is more interesting at the level of theory than the absolute measure. Additional insight into the underlying theoretical mechanism can be obtained by considering how the relative rate of change itself changes over time (if it does). Consider, for example, radioactive decay. Before they decay, radioactive atoms do not change as a function of time (i.e., they do not "age" in any meaningful sense). Instead, an individual atom remains as it always was until a random statistical event causes it to decay. This means that the instantaneous probability of decay (i.e., the hazard function) is constant, which in turn means that radioactive decay is exponential in form. It is, in fact the constant hazard function (i.e., the constant relative rate of decay) that reveals the fact that radioactive decay is a purely random process.

The point is that, with regard to the underlying theoretical mechanism(s) of forgetting, it is the relative rate of forgetting that is of interest. Indeed, as described next, extant theories of forgetting differ not in what they predict about how the absolute rate of forgetting changes as a function of time but instead differ in what they predict about how the relative rate of forgetting changes as a function of time.

Theories of normal forgetting

A constant relative rate of forgetting (associated with the exponential) would be observed if (1) the resistance of memory traces to interference remains constant over time (i.e., traces do not age in any way) and (2) the magnitude of the random disruptive force applied by interference remains constant over time (e.g., retroactive interference that erases the memory is as likely to occur in the moments after the memory was formed as it is to occur years later). In that case, forgetting would be a completely random process, and the forgetting curve would be exponential in form (just as radioactive decay is). Thus, a pure retroactive-interference model of forgetting would naturally predict exponential forgetting. Models that explain forgetting in terms of contextual drift (i.e., the decreasing overlap between the encoding and retrieval contexts with the passage of time) also generally assume exponential forgetting (e.g., Mensink & Raaijmakers, 1988; Murdock, 1997), though perhaps not necessarily. In addition, Sikström (2002) noted that connectionist models often predict exponential decay as well.

By contrast, an increasing relative rate of forgetting (exemplified by the exponential-power function with $c > 1$) would be observed if the magnitude of the disruptive force applied by interference does not remain constant but instead increases over time. Such a theory was once prominent in the interference literature. Keppel and Underwood (1963) argued that forgetting is caused by both retroactive and proactive interference. Although the magnitude of the force exerted by naturally occurring RI would not necessarily change as a function time, the magnitude of the force exerted by PI should theoretically increase over time due to the spontaneous recovery of pre-existing associations that were inhibited during learning. For example, while the cue-target pair “king-bed” was being memorized, it would be necessary to inhibit the pre-existing association between “king” and “queen.” Imagine that preexisting associations remained inhibited for 24 hours, after which they began to reawaken. In that case, during the first 24 hours, each memory would have some probability of failure, and only constantly-applied RI would cause forgetting. If 12 items were initially learned, then, due to RI, after 24 hours, perhaps 6 would be forgotten in an absolute sense and 50% in a relative sense. All else equal, over the second day, another 50% would be lost (3 items) due to the force of constantly-applied RI. However, over the second day, with prior associations now recovering and adding to the theoretical forces that cause forgetting, 4 items might be forgotten instead of only 3. This would still represent a slower absolute rate of forgetting over the second day (4 items lost) compared to the first day (6 items lost). However, in a relative sense, 66.7% (i.e., $100\% \times 4/6$) of still-encoded items in this example would be forgotten on the second day, which represents a faster rate of forgetting compared to the first day (50%). The absolute measure does not reflect the additional underlying mechanism that accelerates forgetting in this example, but the relative measure does.

Keppel and Underwood’s (1963) theory could explain why forgetting is characterized by an exponential-power function with $c > 1$ (as illustrated in the bottom row of Figure 1), if it were. However, empirical forgetting functions rarely, if ever, exhibit an accelerated relative rate of forgetting as a function to time. They do not even exhibit the constant relative rate of forgetting implied by the exponential. Instead, they almost invariably exhibit a decreasing relative rate of forgetting, as noted long ago by Jost (1897).

A variety of other theories correctly predict that the relative rate of forgetting decreases as a function of time, consistent with the power law of forgetting. For example, one proposal is that the temporal distinctiveness between memory traces of a search set is a function of the ratio of the time that has elapsed between encoding and retrieval (Brown et al., 2007; Ecker et al., 2015).

According to their model (known as the SIMPLE model), memory traces are represented as locations along a time continuum receding into the past. Because the continuum is assumed to be logarithmically compressed, recent locations are more easily discriminable from each other than are more remote locations. The model does not require a specific mathematical form of forgetting, but "…forgetting in the SIMPLE model closely follows a power law when range artifacts are avoided" (pp. 556-557).

Alternatively, Anderson and Schooler (1991) proposed that the decreasing proportional rate of forgetting reflects an adaptation to the demands of the environment. For example, the probability of a word occurring in a headline of the New York Times on a given day is a power function of how long it has been since the word previously occurred. Similarly, the probability of given word uttered by a parent is a power function of how many utterances it has been since the word previously occurred. And the probability of receiving an email message from a given source is a power function of how many days it has been since a message was last received from that source. According to this theory, the decreasing relative rate of forgetting as a function of t reflects the fact that, in the environment, the probability that a given memory will be needed again proportionally declines in just that way.

Still another (not necessarily incompatible) proposal is that the decreasing proportional rate of forgetting reflects consolidation (Fusi, Drew, & Abbott, 2005; McClelland, McNaughton, & O'Reilly,1995; Wickelgren, 1979; Wixted, 2004a, 2004b), conceptualized as increasing resistance to interference over time (analogous to glue drying over time). This process may reflect a memory system evolved to possess two desirable properties: (1) plasticity (the ability of neurons to quickly encode information) and (2) stability (the ability of neurons to resist change and therefore retain learned information). According to this view, the ability of neurons to quickly change comes at the expense of vulnerability to interference, which is why new memories exhibit a faster relative rate of forgetting than old memories. If memories consolidate, becoming more resistant to interference, then the time course of forgetting should not be exponential in form (implying a constant proportional rate of forgetting). Instead, it should be (and is) characterized by a function that exhibits a continuously decreasing proportional rate of forgetting, such as the power law.

The argument in favor of an absolute measure of forgetting

Recall that Slamecka and McElree's (1983) noted a lack of theory concerning the normal rate of forgetting. However, as just illustrated, that observation applies to the absolute rate of forgetting, not to the relative rate of forgetting. With regard to the relative rate of forgetting, theories abound. Then again, these theories make predictions about how the relative rate of forgetting changes as a function of time within a single degree-of-learning condition, whereas Slamecka and McElree's (1983) argument in favor of an absolute measure focused on how the rate of forgetting changes as a function of the degree of learning between two different conditions. This was their reasoning:

> At first glance it might seem that if learning affects retention, it would likewise affect forgetting, because the latter is derived from retention scores. But this is logically untenable because intercepts and slopes are a priori completely independent components. However, a gratuitous dependence on intercept levels can be arbitrarily introduced by expressing forgetting as the percentage lost on a later test from the baseline of an earlier one. Because this portmanteau value is determined jointly by both intercept and slope, it is confounded and thereby less analytic than the preferred alternative of evaluating forgetting by way of the slope component alone. This is especially important when degrees of learning are manipulated. Because they will produce intercept effects and because the question is whether they also produce different retention losses, this latter aspect should be free to be estimated without any distortion from shifting baselines. Accordingly, evidence about effects of learning on forgetting rates will come from the interaction terms based on untransformed data or from comparison of absolute-loss scores but not from percentage-of-loss scores (pp. 384-385).

Contrary to this argument, the degree of learning (the *y*-intercept) is independent of the rate of forgetting whether the rate is measured in absolute or relative terms. For example, Figure 2A illustrates what power-law forgetting would look like for different degrees of learning associated with the same proportional rate of forgetting (i.e., different values n_0 but the same value of b). Although the three forgetting functions start out at different levels of performance (i.e., they exhibit different degrees of learning), all three drop by 58% from t = 0 to t = 1 day, by 40% from t = 1 day to t = 2 days, by 30% from t = 2 days to t = 3 days, and so on. By contrast, the absolute rate of forgetting is highest for the high degree of learning condition and lowest for the low degree of learning condition.

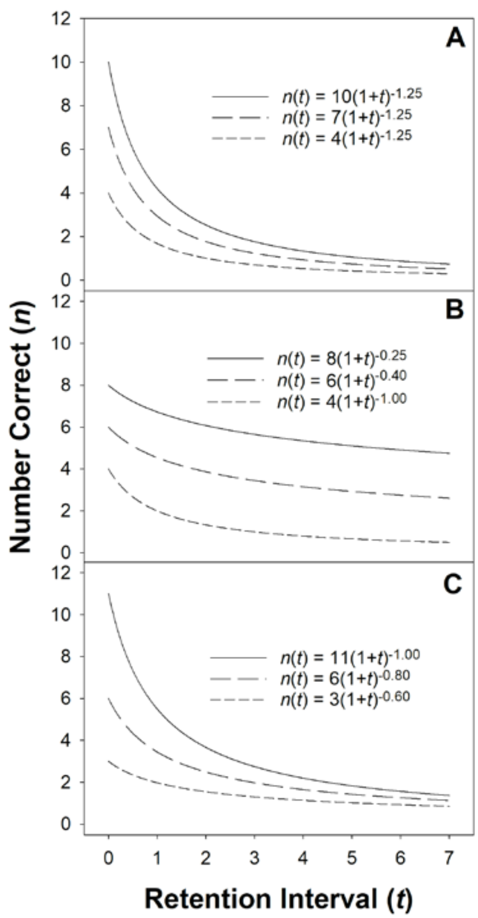

Figure 2. The effect of the degree of learning on the power law of forgetting. Increasing the degree of learning could have no effect on the relative rate of forgetting (A), it could slow the relative rate of forgetting (B), or it could increase the relative rate of forgetting (C).

Empirical forgetting functions typically do not look like the functions shown in Figure 2A because, unlike those functions, the absolute rate of forgetting is typically similar across degrees of learning. Figure 2B illustrates what power-law forgetting would look like if a higher degree of learning yielded a lower relative rate of forgetting (i.e., a smaller value of b). Now, the absolute rate of forgetting is similar across conditions, not unlike the empirical pattern reported by Slamecka and McElree (1983) and Rivera-Lares et al. (2022).

Figure 2C illustrates what power-law forgetting would look like if the relative rate of forgetting *increased* with the degree of learning (i.e., if a higher degree of learning yields a larger value of b). The key point is that the degree of learning does not dictate the proportional rate of forgetting any more than it dictates the absolute rate of forgetting. Using either measure, a higher degree of learning can yield no effect on the rate of forgetting or it can yield an effect in either direction. Empirically, the pattern illustrated in Figure 2B is the one that is most often observed, and for that pattern, approximately equal absolute rates of forgetting necessarily imply unequal proportional rates of forgetting for functions that eventually asymptote to zero (Wixted & Carpenter, 2007).

Power-law fits. To illustrate how reliable the empirical pattern illustrated in Figure 2B is, I fit the 2-parameter version of the Wickelgren (1974) power function, $n(t) = n_0(1 + t)^{-b}$, to several representative data sets by minimizing squared deviations from its predictions. Figure 3A shows the power law fit to the empirical data reported by Krueger (1929) through their 4-day retention interval (similar to the range of more recent studies). Table 1 shows the least-squares estimates of n_0 and b, and the estimates of b are also reproduced in the figure legend for each condition. Although there is no apparent interaction in an absolute sense, those estimates indicate that a higher degree of learning results in a lower relative rate of forgetting (i.e., a smaller estimate of b).

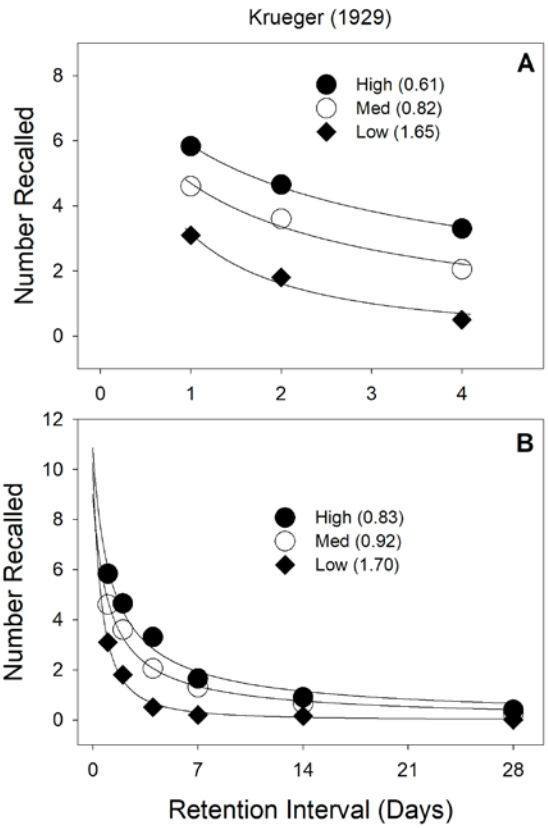

Figure 3. Least squares fits of the power law to data from Krueger (1929) over the first 4 days (A) and over the full 28 days (B). The values shown in the legends indicate the estimated value of the rate parameter, *b*.

Yet the power law implies that the functions will eventually converge, ultimately yielding an interaction for the absolute rate of forgetting. Figure 3B shows the power law fit to the empirical data through the full 28-day retention interval. The parameter estimates change somewhat, but the basic story remains the same, namely, a higher degree of learning results in a lower relative rate of forgetting. Slamecka and McElree (1983) argued that a consideration of the full data set is potentially problematic due to floor effects, but this analysis shows that the interpretation of the relative rate of forgetting remains the same whether the longer retention intervals are included or not. This is how it should be if forgetting is accurately characterized by a power law.

Figure	Panel	Low		Medium		High	
		n_0	b	n_0	b	n_0	b
3	A	9.92	1.65	8.30	0.82	8.96	0.61
	B	10.27	1.70	9.01	0.92	10.87	0.83
4	A	21.37	0.35	-	-	36.59	0.28
	B	18.65	0.48	-	-	27.02	0.24
	C	30.69	0.27	-	-	43.74	0.14
	D	8.26	0.29	-	-	11.12	0.21
	E	11.69	0.37	-	-	15.85	0.25
	F	10.17	0.47	-	-	12.60	0.24
	G	13.99	0.21	-	-	17.16	0.10
5	A	6.53	1.36	9.31	0.74	9.62	0.64
	B	5.31	1.60	7.91	0.71	10.92	0.53
	C	4.38	1.16	6.77	0.58	9.24	0.46
	D	6.08	1.27	9.50	0.63	-	-

Table 1. Least-squares estimates of the power law n_0 and b parameters for the fits shown in Figures 3, 4, and 5.

Figure 4 shows the fits of the power law to the data from Experiments 1 and 2 of Slamecka and McElree (1983). The same story emerges: a higher degree of learning is associated with a slower relative rate of forgetting (Table 1) for all seven experiments. Figure 5 shows the fits of the power law to data from the four experiments recently reported by Rivera-Lares et al. (2022). Once again, in every case, a higher degree of learning is associated with a slower relative rate of forgetting (Table 1). No statistical test is needed to conclude that this pattern did not arise due to chance.

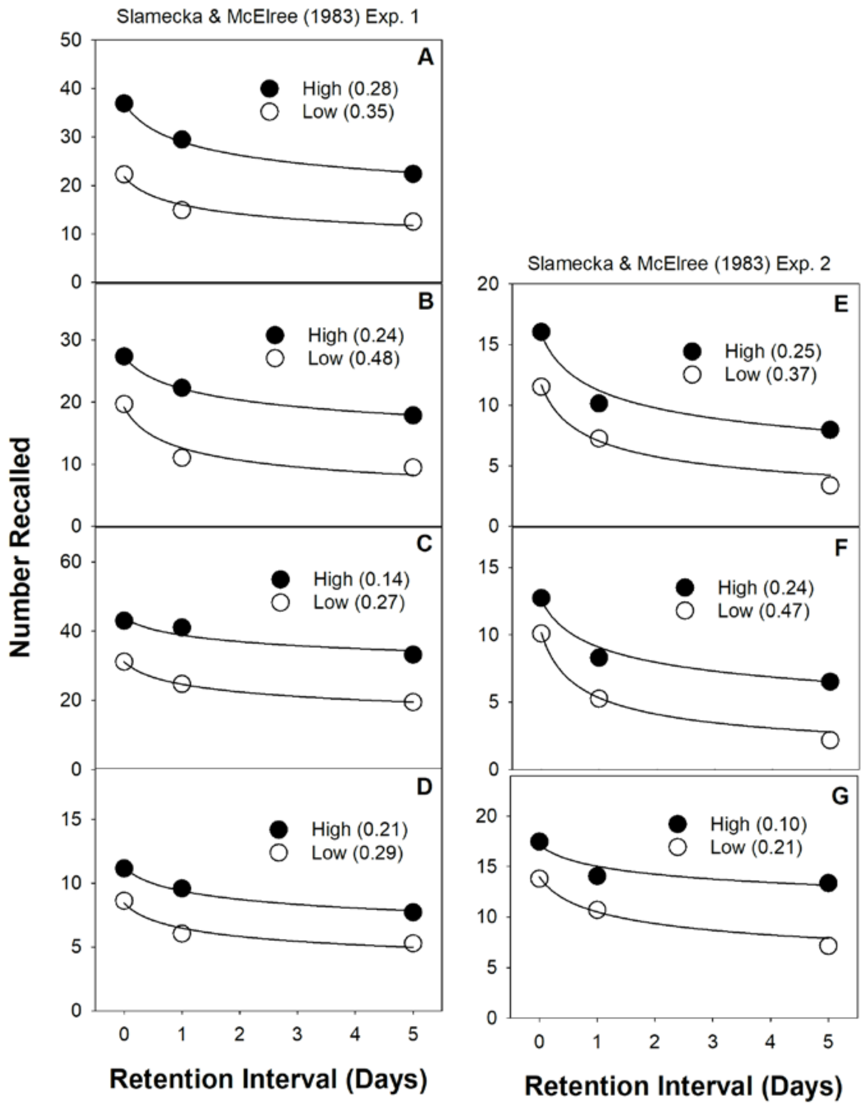

Figure 4. Least squares fits of the power law to data from Slamecka and McElree's (1983) Experiment 1 (left panel, A = free recall, B = easy item recall, C = cued recall, and D = category recall) and Experiment 2 (right panel, E = cued recall, F = easy item recall, and G = matching). The values shown in the legends indicate the estimated value of the rate parameter, *b*.

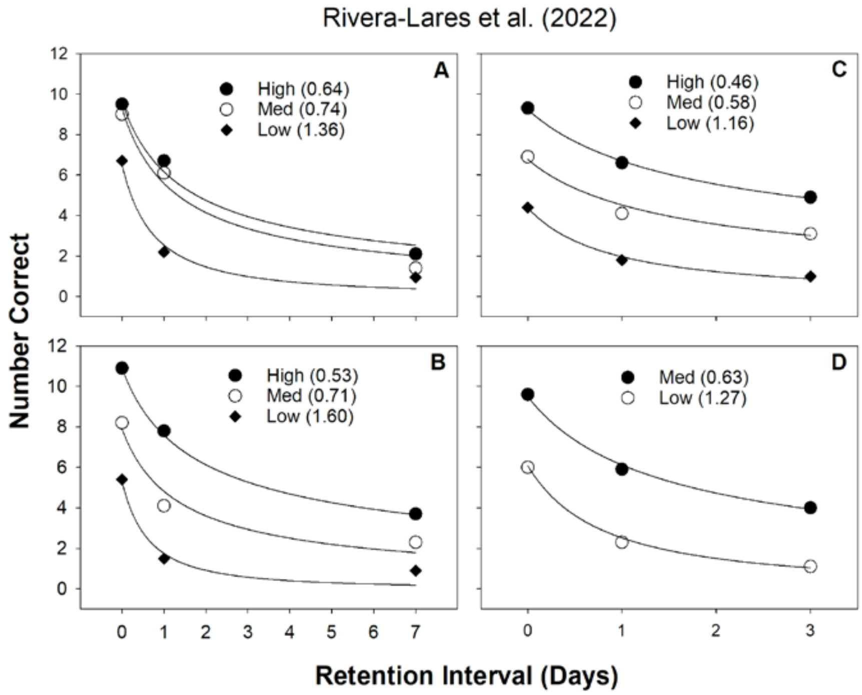

Figure 5. Least squares fits of the power law to data from Experiments 1 through 4 (panels A → D) of Rivera-Lares et al. (2022). The values shown in the legends indicate the estimated value of the rate parameter, *b*.

Checks and Caveats. All of the forgetting data analyzed above were arithmetically averaged over participants. It has long been known that individual forgetting curves that are exponential in form can yield an average curve that is better fit by a power function (e.g., Anderson and Tweney, 1997). However, geometric averaging preserves the shape of both exponential and power functions, and when individual forgetting curves are geometrically averaged, the average forgetting curves are still better fit by a power function (Wixted & Ebbesen, 1997). Here, when the individual participant data reported by Rivera-Lares et al. (2022) were geometrically averaged and fit by the power function,[1] the conclusion remained unchanged: a higher degree of learning was still unambiguously associated with a lower relative rate of forgetting.

[1] Because participants sometimes recalled 0 items at a given retention interval, 1 was added to all values prior to geometric averaging, and then 1 was subtracted from the final result.

For the fits described above, the highest degree of learning was reliably associated with the lowest rate of forgetting. However, potential ceiling effects at the shortest retention interval of that condition could artificially lower the apparent rate of forgetting. Using odds of recall can help to address potential ceiling effects by magnifying small differences that are small only because the values are compressed against the ceiling. The probability of recall (p) is transformed to odds of recall (o) using the equation $o = \frac{p}{1-p}$. When the geometrically averaged data reported by Rivera-Lares et al. (2022) were converted to odds and then fit with the similarly transformed version of the power function, a higher degree of learning was still unambiguously associated with a lower relative rate of forgetting. This makes sense given that the intermediate degree of learning condition (which was not plagued by potential ceiling effects) was reliably associated with a lower rate of forgetting than the low degree of learning condition in the fits reported above. In addition, most of the Slamecka and McElree (1983) data shown in Figure 4 were well below the ceiling and well above the floor, yet the same pattern was evident for those findings.

Finally, it is important to acknowledge that these analyses and conclusions depend on the assumption that the power law is the correct function to use, and it is conceivable that a different function would yield a different conclusion. Still, the power law is at least a reasonable function to use, and it yields a clear interpretation: although the degree of learning may not yield a statistically significant effect on the absolute rate of forgetting over the course of a week, the relative rate of forgetting decreases dramatically as the degree of learning increases. Of these two empirical observations, the former may simply be a theoretically uninteresting byproduct of the latter, which is the opposite of what Slamecka and McElree (1983) once argued.

Why does a higher degree of learning yield a lower relative rate of forgetting?

The relative rate of forgetting is largely determined by the degree of learning and not by many other variables that one might assume would play that role. For example, given equal practice, less meaningful material is usually associated with *both* a lower degree of learning and a higher relative rate of forgetting than more meaningful material. However, as noted by Underwood (1964) long ago, when the degree of learning is equated by means of additional study trials for the less meaningful material, the rate of forgetting is usually equated as well (though exceptions can be found, such as Hockley, 1992, Figure 7, p. 1328). Similarly, given equal practice, individuals who exhibit a lower degree of learning also exhibit a higher relative rate of forgetting. However, when practice is varied to ensure equal degrees of learning, differences in the rate of forgetting across individuals essentially disappear. Thus, he argued, the key to understanding the rate of forgetting is the degree of learning, and for some reason, a higher degree of learning yields a lower (relative) rate of forgetting. Why is that?

As noted earlier, one possibility is that forgetting reflects an adaptation to environmental demands (e.g., Anderson & Hulbert, 2021; Anderson & Schooler, 1991; Schooler & Anderson, 2017). The underlying brain mechanism might involve active processes designed not only to preserve information deemed to be needed but also to erase information deemed to be unneeded (Davis and Zhong, 2018; Hardt, Nader, & Nadel, 2013; Ryan & Frankland, 2022). According to Bean (1912), this basic idea was advanced more than a century ago by Antonio Renda (1875-1959):

> Several more recent investigations of kindred problems have led to conclusions concerning forgetting. The following declaration is made by A. Renda[8] "Forgetting is not merely an accidental characteristic of mental function, but is the result of an active process of dissociation. It is a means by which consciousness gets rid of redundant material." (p. 11).
>
> [8] "L'oblio saggio sull'attivita selettiva della coscienza ," Torino: Fratelli Bocca , 1910 , p . 229.

More recently, Davis and Zhong (2018) observed that “biological processes, in general, operate with separate and dedicated pathways for synthesis and degradation” (p. 494). They argued that it is therefore reasonable to suppose that that there are also biological pathways not only for the encoding and consolidation of memory but also for the selective erasure of memory. As they put it: “Consolidation would play the role of the judge, allowing memories that are deemed important and worthwhile to remain while allowing the irrelevant ones to be removed by intrinsic forgetting” (p. 495).

What might the brain deem to be important? Conceivably, a higher degree of learning signifies the future importance of retaining that information (which is therefore tagged for consolidation) compared to information that is learned less well (which is tagged for selective erasure). Consistent with this idea, repeated learning trials suppress active forgetting mechanisms in drosophila (Zhang et al., 2016).

In humans, learning is not merely a function of repeated presentations; instead, to a large extent, it is a function of making sense of the to-be-learned material (Bransford & Johnson, 1972; Craik & Lockhart, 1972; Craik & Tulving, 1975; Mandler, 1967, 2002). Thus, the degree of learning may often simply be a proxy for how subjectively meaningful the studied material is. From this perspective, active forgetting mechanisms may be suppressed (allowing for newly formed memories to undergo consolidation) to the extent that the to-be-learned material can be processed in terms of existing schematic knowledge. In agreement with the general idea that there is a connection between the prioritized consolidation of newly learned information and pre-existing schematic knowledge, research with rodents suggests that consolidation can occur rapidly if an associative schema into which new information is incorporated already exists in the brain (Tse et al., 2007, 2011). Such findings are consistent with complementary learning systems theory (McClelland et al., 1995; McClelland, 2013).

What becomes of memories that are not preferentially consolidated? Interestingly, memories tagged for selective erasure, when forgotten, are not gone forever. Instead, they may exist in a novel physiological state that is potentially “reactivatable” (Liu et al., 2022). The realization that ostensibly forgotten memories are in fact not gone has long been known to cognitive psychologists (Shiffrin, 1970; Tulving & Pearlsone, 1966). In an interesting new study along these lines, Bäuml & Trißl (2022) analyzed the time course of forgetting beginning shortly after study (as usual) or beginning shortly after a retrieval practice phase implemented 30 minutes after original learning. During the retrieval practice phase, memory for half the studied items was tested several times using cued recall (practiced items), without corrective feedback. The other half was not tested (non-practiced items). Remarkably, when memory was tested shortly thereafter, this selective retrieval boosted recall of not only the practiced items but also the non-practiced items, effectively restoring performance to the level observed shortly after original study for both.

This restoration of performance to the original degree of learning suggests that retrieval practice for half the items reinstated the learning context, making it possible to retrieve apparently “forgotten” items to about the same degree whether or not they had been cued during retrieval practice. From that point on, however, the fates of the practiced items and non-practiced items differed. For the items that were not practiced, the subsequent relative rate of forgetting, measured using parameter b of the power function, was similar to that of the control group (for which testing began shortly after original learning). It was as if these items were simply restored to their original post-learning state. However, for the items that were practiced, the relative rate of forgetting was substantially reduced. It is not clear why, but one possibility is that the tested items were prioritized in such a way that they inherited the state of consolidation that had developed over the 30 minutes after learning (which is when retrieval practice was implemented).

The findings reported by Bäuml & Trißl (2022) constitute an exception to Underwood's (1964) claim the key to understanding the rate of forgetting is the degree of learning. It seemed that way to Underwood because, at the time, the rate of forgetting was usually found to be the same across conditions and across individual participants when the degree of learning was equated. However, in Bäuml & Trißl (2022), the practiced and non-practiced items were restored to their (equivalent) original degrees of learning, yet the practiced items exhibited a lower rate of forgetting from that point forward. Indeed, this is also true of the testing effect in general. That is, when the degree of learning is equated by testing vs. study, the rate of forgetting from that point on is usually lower for the tested items (e.g., Carpenter, et al., 2008).

The same is true of the spacing effect, where massed practice yields the same (or even a higher) degree of learning compared to spaced practice, but the rate of forgetting is usually lower in the spaced condition (Cepeda et al., 2006). Interestingly, Anderson and Schooler (1991) investigated whether this phenomenon might reflect an adaptation to the future need for the practiced material. To do so, they examined email communications received from contacts that were spread out over time (spaced contacts) vs. email communications received from contacts that were clustered together in time (massed contacts). The probability of receiving another email soon after the last contact was higher for the massed contacts, whereas the pattern was reversed for the probability of contact at longer lags. Thus, they concluded that the spacing effect reflects an adaptation to this environmental regularity.

If forgetting does reflect an adaptation to environmental realities, then it seems reasonable to suppose that it would be governed automatically (i.e., outside of conscious awareness). Yet explicit instructions provided to participants about the future need for recently memorized material can have similar effects on the future availability of encoded information. For example, in studies of directed forgetting, participants are instructed to either forget or remember some just-learned items (e.g., a list of words). On a later test of retention, forget-cued material is less likely to be remembered (i.e., those memories are less available) than the remember-cued material (e.g., Basden et al., 1993; Pastötter & Bäuml, 2010). Is this effect yet another example of material being remembered as a function of its future need? The answer is not clear, but it is an intriguing possibility that may be worth exploring.

Conclusion

In a way, Slamecka and McElree (1983) were right that theoreticians have largely neglected the issue of normal forgetting. However, that observation applies only to the absolute rate of forgetting measure that they strongly favored. With regard to the relative rate of forgetting, a variety of theoretical ideas in both experimental psychology and molecular/cellular neuroscience have been proposed. Critically, these theories contain no mechanism that maps on to the absolute rate of forgetting.

It is, of course, possible to specify a theory according to which the absolute loss per unit time is the more theoretically relevant variable. For example, imagine a model in which n_0 items are initially encoded into long-term memory, where they are stored in a buffer. In this model, forgetting is caused by retroactive interference, which is constant force that fills the buffer with 3 irrelevant items per day, evicting 3 of the items from the list that were initially encoded. A higher degree of learning increases the number of items initially encoded (i.e., it increases n_0 and, therefore, the size of the encoded buffer), but it does not change the fact that retroactive interference evicts 3 items per day. In this theory, the absolute rate of forgetting is the relevant measure, and the degree of learning would not affect it (in accordance with the empirical data). This model predicts linear forgetting, which is observed under some conditions (e.g., Fisher & Radvansky, 2019, 2021) but is not usually observed. However, one could accommodate curvilinear forgetting by assuming that it simply reflects a scaling/measurement artifact, in which case the observed mathematical form of forgetting would not be of theoretical interest.

This model is not intuitively plausible, and I am not aware of any research in either psychology or neuroscience suggesting that thinking along such lines will enhance our general theoretical understanding of forgetting. Indeed, no general theory of forgetting has been proposed that denies the curvilinear, negatively accelerating loss of information under typical conditions. I suggest that for any theory of forgetting capturing that general property, it is the relative (not the absolute) rate of forgetting that is of theoretical interest. According to the relative measure, a higher degree of learning yields a lower rate of forgetting—for some theoretically interesting reason.

References

Anderson, M. C., & Hulbert, J. C. (2021). Active forgetting: Adaptation of memory by prefrontal control. *Annual Review of Psychology, 72,* 1–36.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. Psychological Science, 2(6), 396–408.

Anderson, R. B., & Tweney, R. D. (1997). Artifactual power curves in forgetting. *Memory & Cognition, 25(5),* 724–730.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 603–616.

Bäuml, K. T., & Trißl, L. (2022). Selective memory retrieval can revive forgotten memories. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(8), e2114377119.

Bean, C. H. (1912). *The curve of forgetting*. Lancaster, PA: New Era Printing.

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

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior, 11(6),* 717–726.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114(3),* 539–576.

Carpenter, S., Pashler, H., Wixted, J., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36*, 438-448.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T. & Rohrer (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354-380.

Chechile R. A. (2006). Memory hazard functions: a vehicle for theory development and test. *Psychological Review*, *113*(1), 31–56.

Cohen-Dallal, H., Fradkin, I., & Pertzov, Y. (2018). Are stronger memories forgotten more slowly? No evidence that memory strength influences the rate of forgetting. *PloS ONE*, *13*(7), e0200292. https://doi.org/10.1371/journal.pone.0200292

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294.

Davis, R. L., & Zhong, Y. (2018). The biology of forgetting—a perspective. *Neuron 95*, 490–503.

Donkin, C., & Nosofsky, R. M. (2012). A power-law model of psychological memory strength in short- and long-term recognition. *Psychological science*, *23*(6), 625–634.

Ebbinghaus, H. (1880, as cited in Murre & Dros, 2015). Urmanuskript "Ueber das Gedächtniß" [Original manuscript "About the memory"]. Passavia Universitätsverlag.

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

Ecker, U. K., Brown, G. D., & Lewandowsky, S. (2015). Memory without consolidation: Temporal distinctiveness explains retroactive interference. *Cognitive Science, 39(7),* 1570–1593.

Fisher, J. S., & Radvansky, G. A. (2019). Linear forgetting. *Journal of Memory and Language*, *108*, 104035.

Fisher, J. S., & Radvansky, G. A. (2021). Degree of learning and linear forgetting. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218211056464

Fusi, S., Drew, P. J., & Abbott, L. F. (2005). Cascade models of synaptically stored memories. *Neuron, 45(4)*, 599–611.

Hardt, O., Nader, K., and Nadel, L. (2013). Decay happens: the role of active forgetting in memory. *Trends in Cognitive Sciences, 17,* 111–120.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(6),* 1321–1330.

Jost, A. (1897). Die Assoziationsfestigkeit in ihrer Abha¨ngigkeit 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.

Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12*, 71-78.

Liu, H., Wu, T., Canales, X. G., Wu, M., Choi, M. K., Duan, F., Calarco, J. A., & Zhang, Y. (2022). Forgetting generates a novel state that is reactivatable. *Science Advances, 8(6),* eabi9071. https://doi.org/10.1126/sciadv.abi9071

Loftus, G. R. (1985a). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11 (2),* 397–406.

Loftus, G. R. (1985b). Consistency and confoundings: Reply to Slamecka. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 817–820.

Lombardi, M. G., Perri, R., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2018). Forgetting of the recollection and familiarity components of recognition in patients with amnestic mild cognitive impairment. *Journal of Neuropsychology*, *12*(2), 231–247.

Mandler, G. (1967). Organization and memory. In Spence, K.W., Spence, J.T. (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 328–372). New York, NY: Academic Press.

Mandler, G. (2002). Organization: What levels of processing are levels of. *Memory, 10*, 333–338.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, *142(4)*, 1190-1210.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-457.

Mensink, G.-J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review, 95*(4), 434–455.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104*(4), 839–862.

Murre, J. M., & Dros, J. (2015). Replication and Analysis of Ebbinghaus' Forgetting Curve. *PloS one, 10(7),* e0120644. https://doi.org/10.1371/journal.pone.0120644

Pastötter, B., & Bäuml, K.-H. (2010). Amount of postcue encoding predicts amount of directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 54–65.

Renda, A. (1910). L'oblio. Saggio sull'attivita selettiva della coscienza [Oblivion. Essay on the selective activity of consciousness]. Fratelli Bocca, Editori.

Rivera-Lares, K., Logie, R., Baddeley, A., & Della Sala, S. (2022). Rate of forgetting is independent of initial degree of learning. *Memory & Cognition*, 1–13. Advance online publication. https://doi.org/10.3758/s13421-021-01271-1

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*(4), 734–760.

Ryan, T. J., & Frankland, P. W. (2022). Forgetting as a form of adaptive engram cell plasticity. Nature reviews. *Neuroscience, 23(3)*, 173–186.

Schooler, L. J., & Anderson, J. R. (2017). The Adaptive Nature of Memory. In J. H. Byrne (Ed.) *Learning and Memory: A Comprehensive Reference*, 2nd edition. Amsterdam, Elsevier.

Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science*, *168(3939)*, 1601–1603.

Sikström S. (2002). Forgetting curves: implications for connectionist models. *Cognitive Psychology*, *45*(1), 95–152.

Slamecka, N. J. (1985). On comparing rates of forgetting: Comment on Loftus (1985). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 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*(3), 384–397. https://doi.org/10.1037/0278-7393.9.3.384

Staugaard, S. R., & Berntsen, D. (2019). Retrieval intentionality and forgetting: How retention time and cue distinctiveness affect involuntary and voluntary retrieval of episodic memories. *Memory & Cognition*, *47*(5), 893–905.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., & Morris, R. G. (2007). Schemas and memory consolidation. *Science, 316(5821)*, 76–82.

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., & Morris, R. G. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science, 333(6044)*, 891–895.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior, 5(4)*, 381–391.

Underwood, B. J. (1964). Forgetting. *Scientific American, 210(3)*, 91–99. https://doi.org/10.1038/scientificamerican0364-91

Underwood, B. J., & Keppel, G. (1963). Retention as a function of degree of learning and letter-sequence interference. *Psychological Monographs: General and Applied*, *77(4)*, 1–16.

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

Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review, 86(1)*, 44–60.

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

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

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

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.

Wixted, J. T. & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science, 18*, 133-134.

Zhang, X., Li, Q., Wang, L., Liu, Z. J., & Zhong, Y. (2016). Cdc42-Dependent Forgetting Regulates Repetition Effect in Prolonging Memory Retention. *Cell reports, 16(3)*, 817–825.