Original Article

QJEP

Degree of learning and linear forgetting

Quarterly Journal of Experimental Psychology
2022, Vol. 75(8) 1483–1496

DOI: 10.1177/17470218211056464
qjep.sagepub.com
SAGE

Jerry S Fisher and Gabriel A Radvansky

Abstract

The aim of this study was to assess whether the degree of learning influences the observation of memory retention and forgetting that follows a linear pattern. According to our retention accuracy from fragmented traces (RAFT) model, one factor that should increase the likelihood of this is when there is greater learning of the material. Higher levels of learning can increase the number of trace components, making it more likely that reconstruction or partial retrieval can lead to an accurate response on a memory test. Here, we report three new experiments, as well as re-analyses of existing data from the literature, to show that increasing the level of learning in some ways can lead to increases in the likelihood of observing linear forgetting. For Experiment 1, people learned materials to different levels. This learning involved cued recall testing during memorisation. Linear forgetting was observed with increased learning. For Experiment 2, learning did not involve cued recall testing. Linear forgetting was not observed. Although our aim was not to test theories of retrieval practice, for Experiment 3, we showed that when people engage in this process, the pattern of retention and forgetting becomes more linear. Overall, these data are consistent with the RAFT theory and support mechanisms that it suggests can lead to the observation of linear forgetting.

Keywords

Memory retention; RAFT

Received: 31 May 2021; revised: 11 September 2021; accepted: 24 September 2021

The traditional view of forgetting over time is that it is best fit by a negatively accelerating function, such as a logarithmic (e.g., Ebbinghaus, 1885/1913), power (e.g., Wixted & Ebbesen, 1991), or some other function (e.g., Rubin & Wenzel, 1996). However, recent work has shown that under some conditions forgetting is better fit by a linear function (Fisher & Radvansky, 2019). Given the serious implications that this has for theories of retention and forgetting, it is important to better understand the factors that lead to this. The aim of the current study is to explore one theoretically identified factor that may increase the probability of producing linear forgetting, namely, the degree of learning when this leads to a sufficient number of memory trace components.

Linear forgetting

Here, we explore memory retention and forgetting issues initially begun by Ebbinghaus (1885/1913). In the retention and forgetting literature, there is a common consensus that episodic retention follows a negatively accelerating pattern, such as a power function ($m = a{\cdot}t^{-b}$; e.g., Wixted & Ebbesen, 1991, 1997). Given the broad acceptance of this idea of a negatively accelerating pattern of memory loss, we refer to it here as the *default view*.

Although a large-scale analysis by Rubin and Wenzel (1996) failed to identify a single, universal function, several researchers have argued that one of the best fitting functions is a power function (Anderson & Schooler, 1991; Averell & Heathcote, 2011; Wixted & Ebbesen, 1991). This is based on both fitting functions to newly collected data (Wixted & Ebbesen) and to Ebbinghaus's (1885/1913) original data (Anderson & Schooler; Wixted & Ebbesen). Averell and Heathcote argued for a power function using a hierarchical Bayesian analysis of retention data. Although a power function may emerge artefactually from averaging other sorts of functions, such as exponentials (Anderson, 2001; Anderson & Tweney, 1997), it does a good job at capturing many negatively accelerating retention patterns across a range of material and test types.

However, despite all that, Fisher and Radvansky (2019) showed that a linear function ($m = a – b{\cdot}t$) is reliably observed under some conditions and over long retention

Department of Psychology, University of Notre Dame, Notre Dame, IN, USA

Corresponding author:
Jerry S Fisher, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, USA.
Email: jfishe12@nd.edu

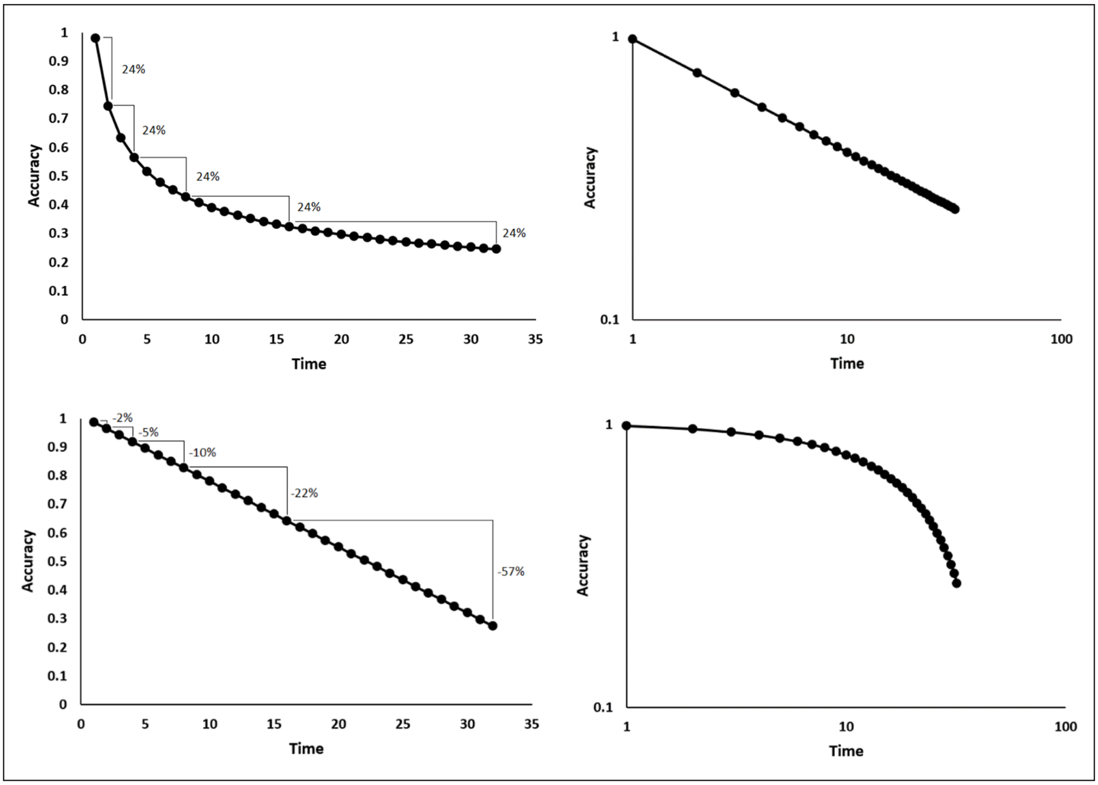

Figure 1. A power function is shown at the top. As each time point doubles, the total proportion lost remains constant at 24%. The left figure is with standard axes and the right is with logarithmic axes. A linear function is shown at the bottom. As each time point doubles, the total proportion lost increases. The left figure is with standard axes and the right is with logarithmic axes.

spans. The difference between negatively accelerating and linear patterns of forgetting has important implications for how we predict and model retention and forgetting. For example, as shown in Figure 1, a power function reflects a constant rate of forgetting across log time, whereas, a linear function reflects an increasing rate. Given other models of decay in science (e.g., radioactive decay), it is readily easy to understand how memory could lose a constant proportion of information over time. It is harder to understand how it could lose a constant amount (increasing proportion) over time.

Prevalence of linear forgetting

Linear forgetting is not an anomaly of a single study. Fisher and Radvansky (2019) also noted patterns of linear forgetting in ten other studies (16 data sets) with 3–52 retention intervals each (Bahrick et al., 1975; Burtt & Dobell, 1925; Cepeda et al., 2008, 2009; Jeunehomme et al., 2018; Kristo et al., 2009; Meeter et al., 2005; Nunoi & Yoshikawa, 2016; Runquist, 1983; Wagenaar, 1986). An outline of these studies is provided by Fisher and Radvansky.

What is notable is that linear forgetting is observed across a range of material and memory test types. In terms of material types, linear forgetting is seen with paired associates (Burtt & Dobell, 1925; Cepeda et al., 2009; Runquist, 1983), classmates names (Bahrick et al., 1975), object–location associations (Nunoi & Yoshikawa, 2016), sentences (Fisher & Radvansky, 2019), autobiographical memory (Kristo et al., 2009; Wagenaar, 1986), news stories (Meeter et al., 2005), trivia facts (Cepeda et al., 2008, 2009), and a walk around town (Jeunehomme et al., 2018). In terms of test types, it is seen with recall (Bahrick et al., 1975; Burtt & Dobell, 1925; Cepeda et al., 2008, 2009; Jeunehomme et al., 2018; Kristo et al., 2009; Runquist, 1983; Wagenaar, 1986), recognition (Fisher & Radvansky, 2019; Meeter et al., 2005; Nunoi & Yoshikawa, 2016), and matching tests (Bahrick et al., 1975).

Note that linear forgetting is also found in other studies not discussed by Fisher and Radvansky (2019), including Linton (1982), and Staats et al. (1970). This is a pervasive phenomenon that requires a cogent explanation given its contrast to the default understanding. In addition, linear forgetting is not simply an artefact of taking a narrow segment of a curvilinear function because it has been observed over very long periods of time, such as month or years (e.g., Fisher and Radvansky, 2019; Linton, 1982).

A theory of linear forgetting

What brings about linear forgetting? Fisher and Radvansky (2019) developed an account, that we call the retrieval accuracy from fragmented traces (RAFT) model.[1] This

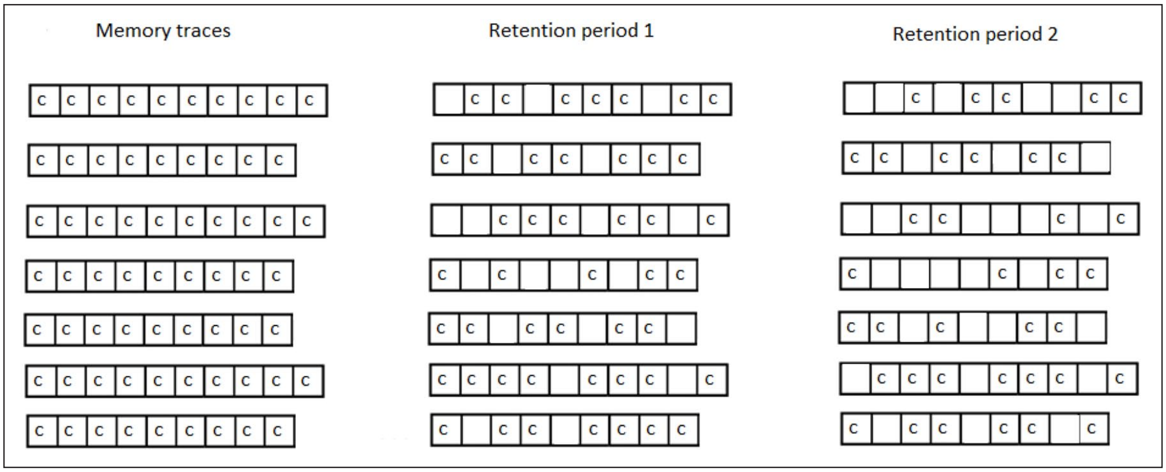

Figure 2. Illustration of how RAFT operates. On the left are seven memory traces that are represented as a vector of components. As each time step increases, these components are forgotten probabilistically (represented by the grey star), such that the proportion loss of all components follows a negatively accelerating pattern. Retrieval is successful at a given retention period if a sufficient number of components are still remembered. For instance, if a trace requires six components for a successful retrieval judgement to be made, six of the seven traces would be accessible for Retention Period 1 and three of the seven traces would be accessible for Retention Period 2.

account combines the premise of negatively accelerating loss of information (Wickens, 1998) with the use of partially degraded information in retrieval judgements. The RAFT model can be specified under the premise that memory trace degrade and information is lost according to a negatively accelerating (e.g., power or exponential) function. RAFT operates according to several widely accepted ideas of memory processing[2]:

1. Memory traces are composed of multiple components.
   a. Memory trace components are forgotten in a negatively accelerating manner.
   b. Different components within a trace may be lost at different rates.
2. People may be able to provide an accurate memory test response using partial traces.
   a. Accurate responses may be made using reconstruction.
   b. Accurate responses may be made using partial matches (e.g., with recognition).
3. It is the combination of these influences that lead to patterns of linear forgetting.
   a. This more likely to be observed when there are more memory trace components.

For RAFT, memory traces are made up of a series of components. These can come from either the features of materials from the world or inferences generated using world knowledge. Thus, traces vary in the number of components they have. This may be influenced by several factors, including the complexity of the information itself and the number of inferences drawn at the time of encoding.

Over time, these trace components are forgotten (i.e., they become either inaccessible or unavailable) at a probability that is captured by a negatively accelerating loss function.[3] Thus, our theory preserves the default Ebbinghaus-based view, albeit at the trace component level, rather than at the entire memory trace level.

At retrieval, memory judgements do not necessarily require complete and intact memory traces. Instead, successful judgements can be made if an adequate proportion of the components are retained. Under such circumstances, people could respond accurately on a recognition memory test using a partial matching process (e.g., Norman & O'Reilly, 2003). Alternatively, through a process of memory reconstruction, prior world knowledge, such as schemas, could be used to fill in the gaps in degraded traces (e.g., Bartlett, 1932). The precise mechanism for retrieval has been left unspecified, and it is likely that in practice some components are more diagnostic for a given memory task than others. Nonetheless, the underlying idea that only part of a memory trace is needed at the time of retrieval is consistent with the well-established processes of partial matching and reconstruction. See Figure 2 for an illustration of this process.

This theoretical framework was implemented in a simulation of the RAFT theory to capture the forgetting patterns reported by Fisher and Radvansky (2018, 2019). For RAFT, even if memory trace components are lost following a negatively accelerating function, so long as, there are a sufficient proportion of the components of a trace available, a correct response may be made. Figure 3 provides the simulation of Fisher and Radvansky (2019) in Experiment 2 assuming that component loss follows a power function. This simulation attempts to simulate the pattern of forgetting observed over a retention interval of

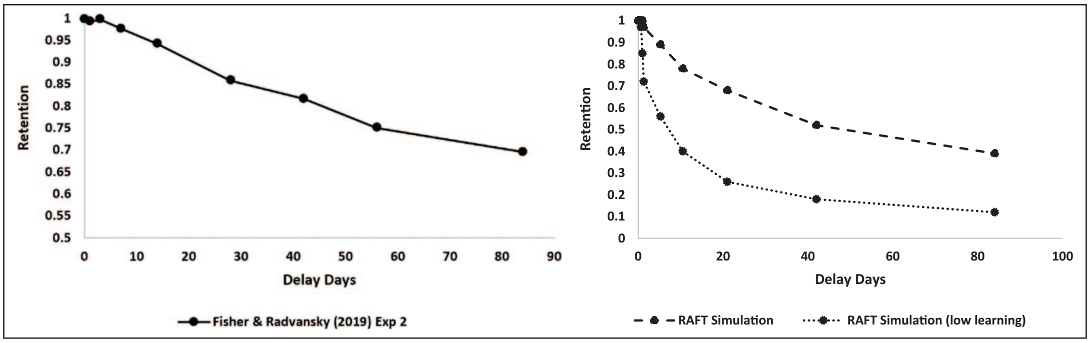

Figure 3. Simulation of Fisher and Radvansky (2019) Experiment 2. The accuracy data are in the left plot where .5 is chance performance. The power fit of these data using R^2 was .42 and the linear fit was .98. The RAFT simulation of these data is on the right (dashed line) where a power fit is .72 and a linear fit is .89. The simulation under low learning (dotted line) where a power fit is .74 and a linear fit is .64. Note that the retention value of 0 in the simulation is equivalent to the chance recognition score of .5.

84 time steps. Note that linear forgetting is only observed when there are enough trace components for reconstruction and/or partial matching processes. When the number of trace components drops below a certain level, the underlying pattern of loss (e.g., exponential or power) for the components should emerge in the behavioural data. This is also shown in Figure 3. The parameters of this simulation are provided as an online supplement on the Open Science Framework (OSF).

As mentioned earlier, our focus here is on the levels of learning. Specifically, the RAFT model predicts that higher levels of learning can produce linear forgetting if the additional learning leads to enough trace components being present to allow for reconstruction and/or partial matching. Higher levels of learning can be achieved in several ways. In some cases, the materials were learned through repeated presentations (Burtt & Dobell, 1925; Nunoi & Yoshikawa, 2016; Runquist, 1983) and/or practice testing (Cepeda et al., 2008, 2009; Fisher & Radvansky, 2019; Nunoi & Yoshikawa, 2016; Runquist, 1983).

The question of whether retention patterns differ as a function of the level of learning has been explored previously (Bogartz, 1990a, 1990b; Loftus, 1985a, 1985b; Loftus & Bamber, 1990; Slamecka, 1985; Slamecka & McElree, 1983). However, this discussion was largely oriented around interpretations of Slamecka and McElree's study that assessed whether the degree of original learning influenced the speed of forgetting over an interval of 5 days. The shape of the forgetting function was not considered.

It is important to note that higher levels of learning alone will not result in linear forgetting. As a classic example of this, Ebbinghaus (1885/1913) learned lists of nonsense syllables to a high degree using rote repetition, and linear forgetting is clearly not observed in his retention data. For RAFT, there needs to be a sufficient number of trace components to allow for reconstruction and/or partial matching. Lists of nonsense syllables, as well as other impoverished material sets, are less likely to lead to this state of affairs even after substantial learning of the type Ebbinghaus did. What does seem to encourage the creation of traces that have a sufficient number of components is when the materials are inherently elaborate or interrelated with one another (as with a narrative) or when simpler materials (such as sentences) are used, and there is an opportunity to elaborate upon that information during learning. Here, we examine the role of degree of learning on the shape of the pattern of forgetting in three experiments that examine retention for periods of up to 2 weeks. We predict that as learning increases, then the shape of the pattern can change.

Experiment 1

For Experiment 1, people learned lists of sentences like those used by Fisher and Radvansky (2019). The aim was to examine the influence of the degree of learning on the observation of linear forgetting. According to the RAFT theory, higher degrees of learning should allow for more opportunities for the materials to be encoded and/or elaborated upon. This contributes to the observation of linear retention patterns because of the better encoding of trace features, and a greater probability of inference making. The prediction is that as the level of learning increased, there would be a concomitant increase in the linearity of the pattern of retention and forgetting.

Method

Participants. Four hundred thirty-two native English-speaking participants (281 females), ranging from 18 to 76 years of age ($M = 38.2$; $SE = .58$), were recruited through Amazon Mechanical Turk (AMT) in exchange for Amazon credit. Twenty-four participants were randomly assigned to each retention/learning degree group. Three hundred forty-five participants were replaced for failing to return for the second half (memory testing). Experiment 1 was a 3 (learning) × 6 (delay) design. For each group, learning level was either study (S), study–test (ST), or study–test–study–test (STST). Similarly, retention interval was immediate, 1, 3, 7, 10, and 14 days. Participants in the immediate groups were reimbursed US$.60, whereas participants in all other groups were reimbursed US$.40 for learning section and another US$.40 for testing. This and all other experiments had *Institutional Review Board* (IRB) approval from the University of Notre Dame.

Materials. There were 18 study sentences, which were created from random pairings of 18 people and 18 activities/traits. Specifically, nine of the sentences were combinations of people and activities of the form "The *person* is *activity*" (e.g., "The student is eating."), and nine of the sentences were combinations of people and traits of the form "The *person* is *trait*" (e.g., "The clerk is skillful"). Each participant had their own randomised pairings. In addition, to broaden the set of materials, we used two sets (A and B) of the 18 activities/traits that were synonymous with one another. Half of the participants in each group were randomly assigned to Set A and half to Set B. These are provided at the OSF website (https://osf.io/7mr64).

In addition to the 18 study sentences, there were 18 nonstudied recognition foils made up of recombination of these pairings. This was done so that participants could not use familiarity judgements based on the words in the sentences, but needed to evaluate the particular combination of elements to make an appropriate response.

Procedure. Both Experiments 1 and 2 were coded in JavaScript using the jsPsych library (De Leeuw, 2015). After providing informed consent, learning took place online. Learning for the S group consisted of one presentation of the list of sentences. These sentences were presented one at a time for 7 s each. For the ST group, they studied the sentences like the S group, and were then given a 36-item cued recall test with feedback, as is done in other sentence memory studies (e.g., Fisher & Radvansky, 2019). Cued recall consisted of sentence blanks of the form "The ______ is eating" or "The student is ______." Participants were asked to fill in the correct word and received feedback for incorrect responses. The feedback was in the form, "Sorry, [provided answer] is incorrect. The correct answer is [correct answer]." In this way, memory for each person and activity/trait was assessed. These questions were presented randomly, one at a time, in a self-paced manner. Finally, the STST group went through the same process as the ST group, except that this procedure was done twice. Note that for the STST group, a different random order was used on each ST cycle.

After learning, the immediate group completed a sentence sensibility task. This consisted of 20 sentences, ten of which were sensible and ten of which were not. This served as a distractor task. The immediate group followed the sensibility task with the recognition test. The delay groups were contacted by email after the designated retention interval. For the 1-day group, the email was sent 2 hr prior to testing time. For the 3-day group, it was sent 5 hr prior to the testing time. All other delay groups received the email the morning of the testing day.

For the recognition test, participants responded to 36 probe sentences (18 studied and 18 nonstudied). Probes were presented one at a time with the task of indicating whether each sentence was studied or not by clicking a button on the screen. No feedback was provided.

Results and discussion

The deviations from the assigned participant retention intervals are provided at the OSF website (https://osf.io/7mr64). Recognition accuracy (proportion of hits and correct rejections) is shown in Figure 4.[4] These data were fit to power and linear functions, which are provided in Table 1.[5] As can be seen, the S condition was better fit by a power function, whereas the ST and STST conditions were better fit by linear functions. Moreover, the linear fit was greater for the STST than the ST condition. These results support our prediction that as the level of learning increased, there would be a concomitant increase in linear forgetting. This is consistent with our theory that increases in the degree of learning would make it more likely that trace elements will be encoded, and perhaps elaborated upon.

The results of Experiment 1 are in line with other results reported in the literature. The value of degree of original learning on the observation of linear forgetting is hinted at in a study by Craig et al. (1972; see also Hellyer, 1962; Postman & Riley, 1959; Youtz, 1941). This study used printed advertisement slides with brand names as materials and manipulated learning to involve either 7 (100% overlearning), 14 (200% overlearning), or 21 exposures (300% overlearning). They tested free recall of the brand names either immediately, 1 day, 1 week, or 4 weeks later. As the number of repetitions increase, the fit of a power function declined ($R^2 = .84$, .73, and .64) and the fit of the linear function increased ($R^2 = .63$, .70, and .86). This was not noted by Craig et al.

However, it is important to note that there was retrieval practice during memorisation in Experiment 1. This may be linked to the observation of linear forgetting here. For example, consider a study by Runquist (1983). In this study, people were tested for memory for word pairs. These were presented one or three times, and there was retrieval practice during memorisation or not. The results revealed that memory was better with three presentations compared with one, and better with retrieval practice. Of interest, here is how well the various conditions were fit by power and

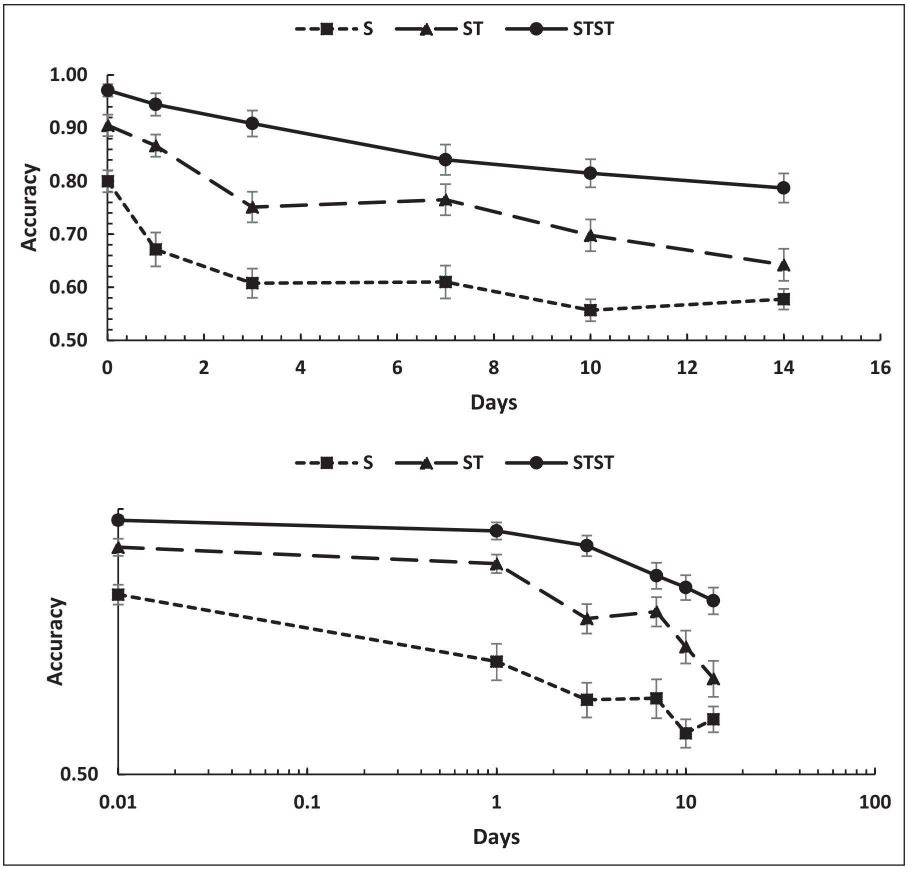

Figure 4. Recognition accuracy for Experiment 1 with linear axes above and logarithmic axes below.

Table 1. Fit of the Experiment 1 data to power and linear functions fits in terms of R^2.

Group	Power function fit	Linear function fit
S	.95	.63
ST	.71	.87
STST	.70	.96

Table 2. Fit of the Runquist's (1983) data to power and linear functions fits in terms of R^2.

Group	Power function fit	Linear function fit
S	.85	.40
ST	.68	.65
SSS	.93	.63
STSTST	.70	.93

linear functions. This is shown in Table 2. As can be seen, linear forgetting was observed with a higher level of learning, but only when there was retrieval practice.

Experiment 2

The aim of Experiment 2 was to assess whether the results of Experiment 1 were due to repeated study alone, or to repeated study along with retrieval practice. Experiment 2 replicated much of Experiment 1 except that the different levels of learning were defined in terms of the number of study cycles alone. No cued recall testing was done during memorisation. If the results of Experiment 1 are due to the number of exposures, then the pattern of results for Experiment 2 should resemble those of Experiment 1. However, if the results are brought about by the additional elaboration and processing brought on by testing, then linear forgetting should not be observed.

Method

Participants. For Experiment 2, 432 native English-speaking participants (225 females), ranging from 18 to 81 years of age (M=37.4; SE=.55), were recruited through AMT in exchange for Amazon credit like Experiment 1. Each was randomly assigned to a group of 24 participants assigned to each retention/learning degree group. Experiment 2 was a 3 (learning) × 6 (delay) design. Four hundred sixty-four participants were replaced for failing to return for the second half (memory testing). For each group, learning level was either study (S), study–study

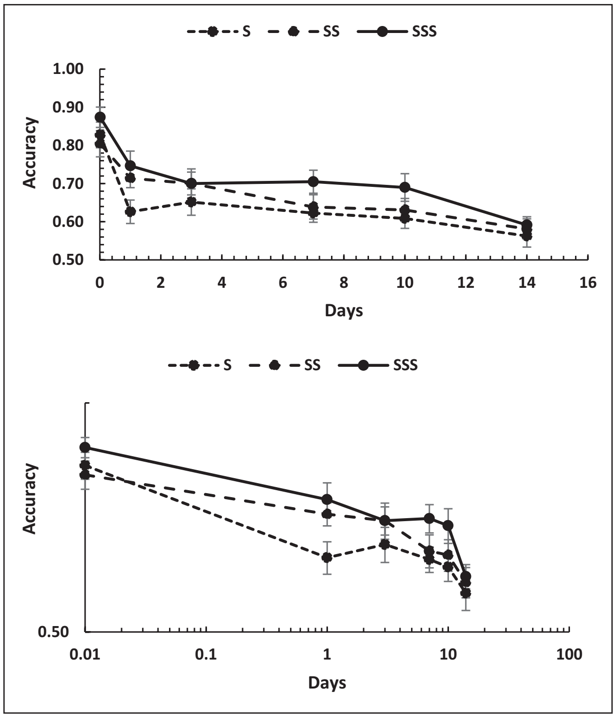

Figure 5. Recognition accuracy for Experiment 2 with linear axes above and logarithmic axes below.

(SS), or study–study–study (SSS). The same retention intervals were used as in Experiment 1.

Materials and procedure. The same materials and procedure were used as in Experiment 1, with the exception that there was no cued recall testing during learning for any of the groups. After the 18 sentences are presented, participants in the SS and SSS groups were told to memorise the sentences once or twice more. The order of sentences was randomised for each study cycle.

Results and discussion

The deviations from the assigned participant retention intervals (in terms of minutes) are provided at the OSF website (https://osf.io/7mr64). The accuracy data are shown in Figure 5. These data were better fit by a power function in all conditions. These results do not replicate Experiment 1. Thus, the results of Experiment 2 suggest that learning by study alone is insufficient to produce linear forgetting, at least insofar as the repetitions are two or fewer. All conditions were better fit by power functions with similar slope parameters. Moreover, this is consistent with prior work, such as that by Slamecka and McElree (1983). Interpreted under the RAFT framework, mere repetition of presentation alone did not result in a sufficient number of encoded trace components, such that partial matching and reconstruction would result in linear forgetting. This may be due to lack of inferences being generated during the repeated presentations. After all, it is well-known that rote repetition alone may be insufficient to lead to more complex memory traces (e.g., Glenberg et al., 1977; Nickerson & Adams, 1979).

Experiments 1 and 2 comparison

The comparison of Experiments 1 and 2 parallels the idea that retrieval practice during memorisation is important for observing linear forgetting with these types of materials. This is consistent with other data in the literature, such as the study by Runquist (1983). Thus, there is some benefit to directly comparing conditions that do and do not involve retrieval practice during memorisation.

To do this, we combined the accuracy data from Experiments 1 and 2 into a 2 (experiment) × 3 (degree of l) × 6 (delay) analysis of variance (ANOVA). This analysis revealed main effects of experiment, $F(1, 848) = 74.50$, $MSE = 1.43$, $p < .001$, $\eta_p^2 = .08$; degree of learning, $F(2, 848) = 89.46$, $MSE = 1.71$, $p < .001$, $\eta_p^2 = .17$; and delay, $F(5, 848) = 53.15$, $MSE = 1.02$, $p < .001$, $\eta_p^2 = .24$, along with a significant experiment × degree of learning interaction, $F(2, 848) = 28.64$, $MSE = .55$, $p < .001$, $\eta_p^2 = .06$. No other interactions were significant. Simple effects tests at each degree of learning showed no difference between the S groups for Experiment 1 ($M = .64$; $SE = .01$) and Experiment 2 ($M = .64$; $SE = .01$), as expected, $F(1, 276) = .64$, $MSE = .02$, $p = .43$, $\eta_p^2 = .002$. However, there was a significant difference between the ST ($M = .77$; $SE = .01$) versus SS ($M = .67$; $SE = .01$) groups, $F(1, 276) = 35.35$, $MSE = .02$, $p < .001$, $\eta_p^2 = .11$, as well as between the STST ($M = .88$; $SE = .01$) versus SSS ($M = .72$; $SE = .01$) groups, $F(1, 276) = 93.64$, $MSE = .02$, $p < .001$, $\eta_p^2 = .25$.

This finding is consistent with a testing effect (Bjork, 1975; Roediger & Karpicke, 2006), although Experiments 1 and 2 were not designed to target the testing effect per se. Overall, the comparison of these two experiments suggests that retrieval practice influences the shape of the retention function. In addition, the better performance of the groups that involved practice testing compared with those that only involved study sessions mirrors the results of Roediger and Smith (2012) who found that testing resulted in better learning compared with study alone.

Thinking about the testing effect more specifically, we can look at Roediger and Karpicke's (2006) highly cited study. Of particular importance here, in their Experiment 1, they assessed memory at three retention intervals, and reported an interaction between learning task and delay. These data are shown in Figure 6. If we fit these data to power and linear functions, we find that the SS condition was better fit by a power function ($R^2 = .95$) than a linear function ($R^2 = .80$); however, the ST condition was more poorly fit by power function fit ($R^2 = .72$) than a linear function ($R^2 = .99$). This is in line with our findings that

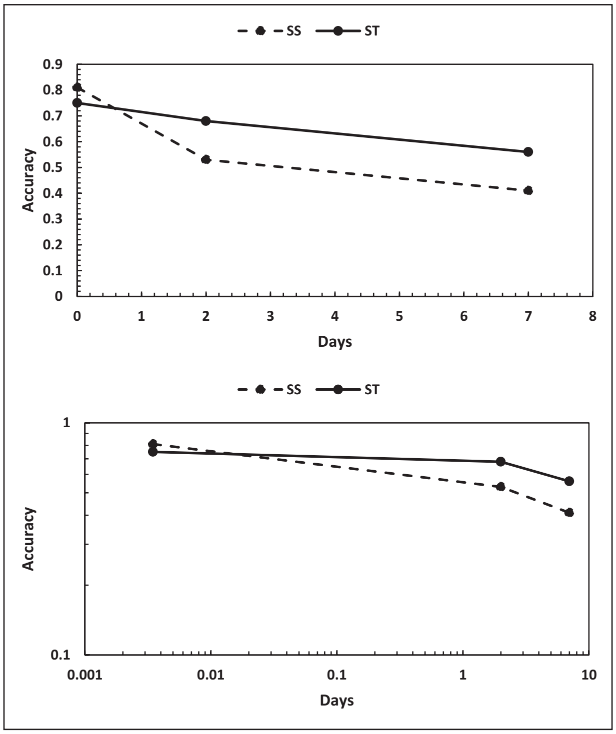

Figure 6. Data from Roediger and Karpicke (2006).

greater learning is more likely to lead to linear forgetting (Table 3).

Experiment 3

Why would retrieval practice lead to more linear forgetting? For the RAFT model, an important aspect of the observation of linear forgetting is whether the traces have a large enough number of trace components to allow for partial matching and/or reconstruction. How would retrieval practice during memorisation affect this? Some retrieval practice theories suggest how this might happen. For example, according to the *elaborative retrieval hypothesis* (Carpenter, 2009), there is a spreading activation process that results in elaborative associations during retrieval practice. This elaboration could be interpreted as resulting in making inferences, and, as a result, from our view, would increase the number of trace components. Alternatively, according to the *gist-trace processing account* (Bouwmeester & Verkoeijen, 2011), retrieval practice enhances the strength of a memory by bringing to bear schematic information. People use schematic knowledge to elaborate on a memory, again, according to our view, increasing the number of trace components. Finally, for the *dual memory framework* (Rickard & Pan, 2018), retrieval practice results in the creation of two memory traces, one each for the learning and retrieval experiences. The result would be multiple traces containing the target information, both of which contribute to retrieval. Although our simulation does not involve multiple traces, per se, with multiple traces, each with different features, then the outcome would be the same as having a single trace with multiple features.

Table 3. Fit of the Experiment 2 data to power and linear functions in terms of R^2.

Group	Power function fit	Linear function fit
S	.91	.53
SS	.85	.86
SSS	.82	.71

Overall, while our work does not distinguish between these theories, they are all consistent with the idea that retrieval practice can result in the creation of memory traces with more components, which would increase the likelihood of observing linear forgetting. Experiment 3 was a study that we had done years before, but had not yet done anything with the data, that had involved an explicit assessment of the testing effect with the aim of assessing whether linear forgetting is observed more under retrieval practice than under repeated study conditions. It was modelled after a study by Roediger and Karpicke (2006) and used the longer texts of Chan et al. (2006). Experiment 3 also examined the testing effect across four retention intervals of up to 4 weeks later.

Method

Participants. Ninety-five students were recruited from the University of Notre Dame participant pool in exchange for partial course credit. They were assigned to one of four retention groups (immediate, 1 day, 1 week, and 4 weeks), each made up of 24 participants, except for the 1-day condition in which there were only 23 participants.

Materials and procedure. The materials used in this study were the same as those used by Chan et al. (2006). They consisted of two texts printed on paper. The topics were the big bang theory and the Shaolin temple, which were 1948 and 1890 words long, respectively. Learning consisted of initial reading followed by either a restudy or a test phase. Within each text, there were 24 critical sentences identified by Chan et al. (2006). These sentences were either restudied or tested for after the initial reading phase.

For the initial reading period, people were given 15 min to read each of the two texts. They were told that to continue reading until the 15 min were up if they finished early. The order of the texts was counterbalanced.

For the second phase, people again studied parts of one of the texts (SS condition) and were given a test for parts of the other (ST condition). Following Roediger and Karpicke (2006), for the SS text, people were given the 24 critical sentences from the text to restudy, one at a time. In comparison, for the ST text, people were given the critical sentences from the text with a target answer left blank. For example, "After the Big Bang, gravity condensed clumps of matter together and these clumps eventually formed ____________." The task was to type the answer to the blank. No feedback was provided.

After the second phase, a final cued recall test was given. For this test, people were given 24 fill-in-the-blank questions for each text (48 total). These items were of the same format as the ST probes that were given during the second phase of additional learning. There was no time limit. The order of texts was randomised for each person, as was the order of the items within each one.

For the immediate group, they were given the test instructions right after the end of the second phase, and then took the final memory test on the lab computers. In comparison, participants in the other retention groups were told to expect an email with a weblink. They then took the final memory test on their own devices.

Results

Responses were scored as correct if they clearly conveyed the propositional idea in the text. For example, if the text was "According to the flat and open models, the universe will continue to _________.," and the answer was "expand indefinitely," a response of "expand for eternity" was scored as correct. In addition, some answers were given half credit if they captured an element of the propositional idea. For example, the answer of "expand" was counted as half correct.

The recall accuracy data are shown in Figure 7. As can be seen, while memory was initially better in the Study-Study conditions, the rate of forgetting was slower for the ST condition and was superior to it by 7 days. This is consistent with Roediger and Karpicke's (2006) basic finding.

These data were submitted to a 4 (retention interval) × 2 (condition: test vs study) mixed ANOVA, with retention interval as a between-participants variable and condition within. This analysis revealed main effects of retention interval, $F(3, 91)=28.26$, $MSE=.047$, $p<.001$, $\eta_p^2=.48$, with memory becoming worse over longer retention intervals, and condition, $F(1, 91)=4.91$, $MSE=.01$, $p=.03$, $\eta_p^2=.05$, with performance being better, overall in the SS condition. Importantly, the interaction was significant, $F(3, 91)=13.96$, $MSE=.01$, $p<.001$, $\eta_p^2=.32$. Simple effects tests revealed that the effect of condition was significant for the immediate group, $F(1, 23)=33.15$, $MSE=.01$, $p<.001$, $\eta_p^2=.59$, and the 1-day group, $F(1, 22)=9.73$, $MSE=.008$, $p=.005$, $\eta_p^2=.31$, with the SS condition doing better than the ST condition. The difference was not significant for the 7-day group, $F(1, 23)=2.54$, $MSE=.013$, $p=.13$, $\eta_p^2=.10$, but was significant for the 28-day group, $F(1, 23)=4.44$, $MSE=.011$, $p=.046$, $\eta_p^2=.16$. Note that

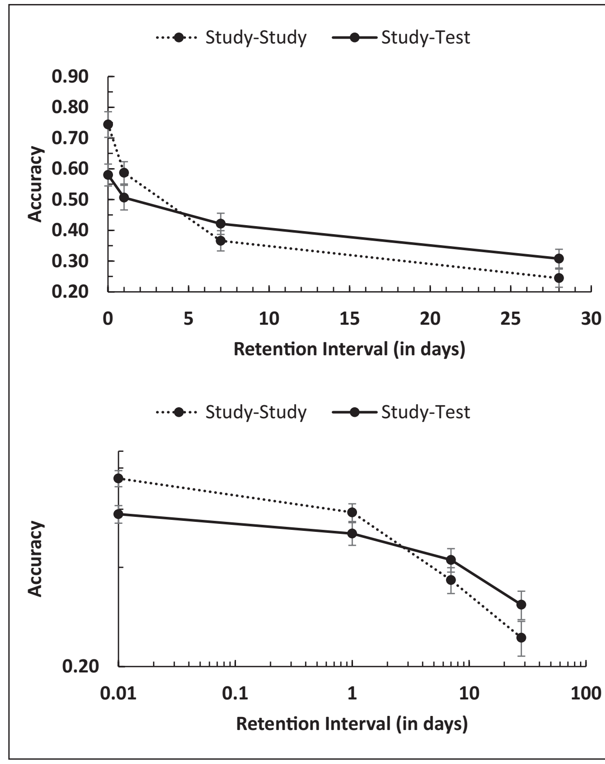

Figure 7. *Recognition accuracy for Experiment 3* with linear axes above and logarithmic axes below.

the testing effect is relatively small here. This is consistent with other research showing that the testing effect may be reduced or absent for more complex materials (van Gog & Sweller, 2015; but see Karpicke & Aue, 2015).

While replicating the basic testing effect is nice, our primary concern here is the best fitting function for the SS and ST conditions. As can be seen in Table 4, these data are in line with all the other studies. When there was less learning (SS), the data were better fit by a power function than a linear function, but with greater learning (ST), the data were better fit by a linear function than a power function. This further supports our theory.

General discussion

The aim of the current study was to explore whether differences in the degree of learning influences the observation of linear forgetting. Our results clearly show that it can. However, at least for the kinds of materials that we used, this additional learning seems to require something more than simply repeated exposure. In our case, additional retrieval practice was sufficient to allow linear forgetting to be observed. This is consistent with other data sets that have been reported in the literature, but which have not been evaluated in this way.

The RAFT simulations for Experiments 1, 2, and 3 are provided in Figures 8 to 10, respectively. All the parameters are constant in these simulations, except for

Table 4. Fit of the Experiment 3 data to power and linear functions in terms of R^2.

Group	Power function fit	Linear function fit
SS	.85	.73
ST	.80	.87

the theoretically defined trace learning and inference parameters. A full listing of the parameter values is provided at the OSF website (https://osf.io/7mr64).

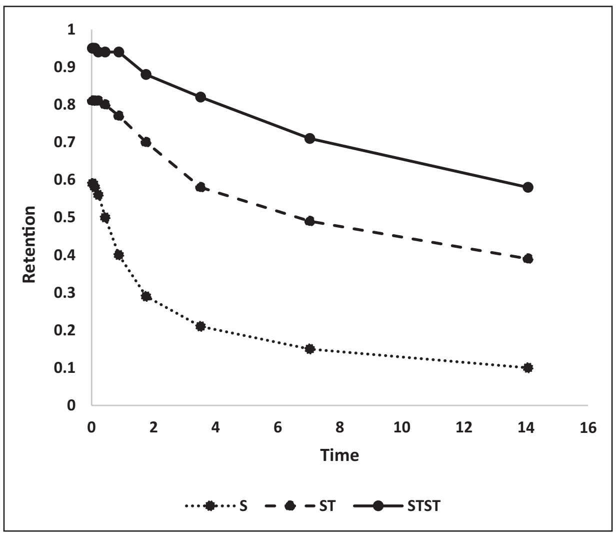

Figure 8. RAFT simulation of Experiment 1.

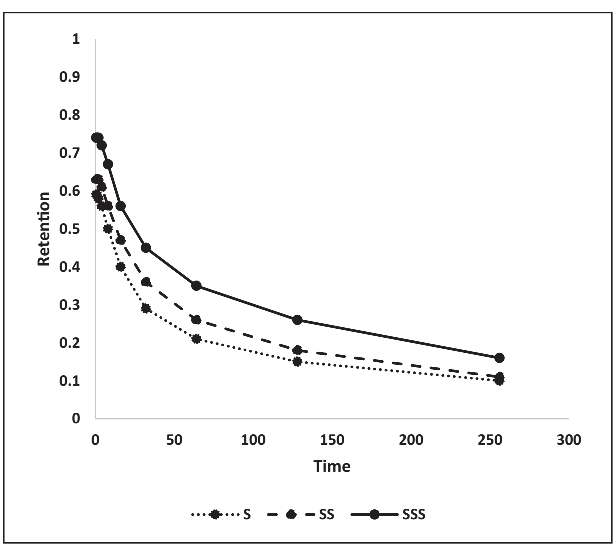

Figure 9. RAFT simulation of Experiment 2.

From the perspective of the RAFT theory, the reason that increased degrees of learning can sometimes produce linear forgetting patterns is because of an increase in the number of trace components present. This increase under the conditions explored here are consistent with some theories of retrieval practice and testing effects (Bouwmeester & Verkoeijen, 2011; Carpenter, 2009; Rickard & Pan, 2018). Essentially, the act of testing during learning results in the generation of more inferences, which then become part of the memory for the content information. This said, it is important to note that there are likely also other ways of increasing trace components besides retrieval practice. For example, Doolen and Radvansky (2021) examined memory retention for novels and found that retention followed a linear pattern for elements that had a high causal connectivity. In this case, number of repetitions was kept the same. What varied was how much information was in a memory trace for a given story event.

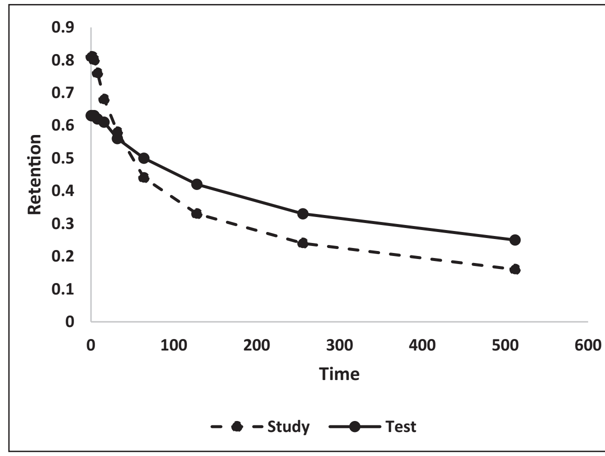

Figure 10. RAFT simulation of Experiment 3.

Consistent with this, in our Experiment 1, linear forgetting was observed after people learned a list of sentences, but only when there was cued recall testing during the memorisation phase. Moreover, two cycles through this process magnified this pattern. In comparison, for Experiment 2, which was essentially the same as Experiment 1, except that there was no cued recall testing during learning, there was no evidence of a pattern of linear forgetting. That is, restudy alone was not sufficient to induce enough trace components. Repetition alone may not encourage inference generation (at least to the extent testing would).

The comparison of these two experiments is analogous to studies of the testing effect. Experiment 3 allowed us to directly assess the issue of whether a testing effect paradigm is more likely to produce linear forgetting. Experiment 3 showed evidence of linear forgetting in the testing condition, even though this study was not initially designed to address such questions. Moreover, this study used more complex texts, rather than just a list of sentences to learn. This provides more support for the RAFT theory suggestion that higher levels of learning are more likely to lead to the observation of linear forgetting.

Note that the observation of linear forgetting is not likely due to an influence of a ceiling effect given that linear forgetting is also observed in the Roediger and Karpicke (2006) data, as well as our Experiment 3 (see also Cepeda et al., 2009; Linton, 1982), when initially performance in the test condition is actually less than in the study condition.

Prior studies

These conclusions are supported not only by our own studies, but also re-evaluations of prior research. As noted earlier, Fisher and Radvansky (2019) reported that the observation of linear forgetting is also present in several other studies (Bahrick et al., 1975; Burtt & Dobell, 1925; Cepeda et al., 2009, 2008; Jeunehomme et al., 2018; Kristo et al., 2009; Linton, 1982; Meeter et al., 2005; Nunoi & Yoshikawa, 2016; Runquist, 1983; Staats et al., 1970; Wagenaar, 1986). This occurs even though it was often not noted by the researchers themselves and was not the aim of those studies. This observation is made even in the face of a wide variety of materials and memory test types. Thus, this is a robust phenomenon.

The idea that different levels of practice influence the likelihood of observing linear forgetting is also supported by other work in the literature. For example, if we re-analyse the data reported by Craig et al. (1972), we see that increased levels of learning are accompanied by more of a linear forgetting function. This may have occurred for two reasons. First, the number of repetitions was quite large (21). Thus, there was more opportunity for people to spontaneously elaborate on the material that was seen repeatedly. Second, the materials were pictures and product names. The act of understanding why a name went with a certain product is a form of elaborative rehearsal, which would increase the number of trace components. The idea that mental processing beyond simple study increases the probability of observing linear forgetting is support by a reassessment of work by Runquist (1983) and Roediger and Karpicke (2006). Consistent with theories of the testing effect, this seems to be related to increased elaborative processing that accompanies retrieval practice.

Implications for retrieval practice studies

Although not the focus of our work, the experiments reported here have implications for theories of retrieval practice. As already noted, these data support the idea that retrieval practice involves, in some form or another, the storage of a larger amount of information in memory, either in the form of elaborative inferences, multiple memory traces (Bouwmeester & Verkoeijen, 2011; Carpenter, 2009; Rickard & Pan, 2018). Of the studies showing a linear retention pattern mentioned previously, Cepeda et al., Chan, and Runquist all used explicit practice retrieval during learning.

One of the features of the testing effect is the acknowledgement that testing during learning does not always provide a benefit immediately and may even show the opposite pattern at that time particularly when testing performance is poor and no feedback is given (Kang et al., 2007). The testing effect is a phenomenon that is more likely to be observed after a delay (e.g., Roediger & Karpicke, 2006). Thus, there is some acknowledgement among researchers exploring the testing effect that something is changing over time. What is absent is any consideration of how the fact that the pattern becomes more linear informs the mechanisms involved in producing the effect.

Consistent with views, such as the elaborative retrieval hypothesis, it has also been noted that testing increases the degree of organisation of the information in memory (Zaromb & Roediger, 2010). Our results suggest that this increased organisation could involve the creation of additional trace components, perhaps in the form of inferences. It has also been found that testing reduces proactive interference effects (Szpunar et al., 2008). Taking the idea that greater trace overlap produces greater interference, having a larger number of trace components in memory following retrieval practice, as suggested by the RAFT theory, could be what reduces interference.

Another theory of retrieval practice is the *episodic context account* (Lehman et al., 2014). For this view, people encode both the study context and the retrieval context. People then use this context as part of their memory search. If retrieval is successful, then there is also a memory with the new context that can be used to help later retrieval (Akan et al., 2018). While this mechanism of retrieval practice seems reasonable, and is supported by some data, it is not immediately clear how this would result in a more linear pattern of forgetting according to our theory given that the focus is on context, whereas the RAFT model uses additional content components to achieve linear forgetting patterns. One would need to assume that the different contexts are memory trace components that can be used to make correct responses, and that more of them allow for more accurate responses with otherwise degraded traces.

One other theory of the testing effect is the *relational processing hypothesis* (Rawson & Zamary, 2019). This idea is grounded in the finding that the testing effect is more likely, and is larger, for recall than for recognition during learning. Recall tests are more likely to emphasise relational processing, whereas recognition tests are more likely to emphasise item-specific processing. Relational processing involves making connections, helping organise the material, and improving performance. Thus, this theory could be consistent with linear forgetting if one assumes that the relational processing amounts to having additional components to the memory traces.

In our studies, although we were manipulating the degree of learning, this was also likely influencing the number of components in the resulting memory traces through spontaneous elaboration and inference making on the part of the participants. This is important given that another way to increase the probability of observing linear forgetting is to increase the complexity of the memory traces. In prior studies, the number of presentations was not explicitly manipulated, but likely occurred. For example, for Bahrick et al.'s (1975) study, participants certainly had repeated exposure to classmate's names, along with a wide range of other experiences involving those people. For autobiographical experience studies (e.g., Jeunehomme et al., 2018; Kristo et al., 2009; Linton, 1982; Wagenaar, 1986), these likely involved extensive inference generation as part of understating the events of one's own life, thus rending the memory traces much more complex. A similar case can be made for the people hearing about news events in Meeter et al.'s (2005) study.

The progress of forgetting

The finding of linear forgetting is important for understanding the progress of memory over time. Science should be able to provide reasonably accurate predictions. One of the most important things that we can predict as memory researchers is how much information from a given event is likely to be remembered after a given period. To have any hope of doing so, we need to understand the nature of the loss of information over time. If we simply assume that it is an Ebbinghaus-like negatively accelerating function, as most researchers do, then we will be quite right on some occasions, and quite wrong on others. However, if we acknowledge that different circumstance can alter the progress of memory over time, then we can provide more accurate predictions.

This ability to predict more accurately can provide useful information for a wide range of applications. For example, in education, knowing what the future of memory for materials learned in the classroom can better guide when memories for materials are likely to be retained, and when they may need to be refreshed. In the field of eyewitness memory, knowing how long different aspects of a witnessed event are likely to persist can help guide an understanding of how memory reports may be more or less reliable. In the area of clinical practice, understanding the progress of forgetting can be helpful in understanding how long clients will remember information that have been told regarding their treatments.

Finally, knowing when different patterns of forgetting may emerge can be additional information that can be used to distinguish different theories of memory. Change over time is a dimension along which we can gather data that will either support or refute various theories of memory. This is certainly a more difficult way to study memory given that more time and data are needed to reach a conclusion compared with most studies of memory that test performance after a single retention interval. While this approach may be valuable under some circumstances, we would argue that an approach that also considers memory change will provide a greater depth of understanding.

Conclusion

Overall, our aim was to assess whether higher levels of learning would be a factor that could lead to an increase in linear forgetting. This is what we found, but only when there was additional processing involved beyond study alone. That is, this was more likely to occur in the presence of retrieval practice during learning. Thus, secondarily, these data also make plain heretofore unnoticed qualities of the testing effect. In addition to providing yet further evidence of linear forgetting, this finding is consistent with our theory of the emergence of linear forgetting following component loss that follows a negatively accelerating function and trace fragment usage. Thus, to be better able to predict memory over time, it needs to be understood when this process will be more linear than the default negatively accelerating function.

Author Note

Jerry S Fisher is now affiliated to University of Grand Valley.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Jerry S Fisher https://orcid.org/0000-0002-0435-9251
Gabriel A Radvansky https://orcid.org/0000-0001-7846-839X

Data accessibility statement

The data and materials from the present experiment are publicly available at the Open Science Framework website: https://osf.io/7mr64

Notes

1. RAFT is publicly available online here: http://ec2-52-204-56-150.compute-1.amazonaws.com/pilot/RAFT4.0.html
2. The details of this theory, as well as a simulation of it, are available elsewhere (Fisher & Radvansky, 2018, 2019).
3. See Wickens (1998). Of note, under exponential forgetting, if components are each lost at different (rather than constant) probabilities, the aggregate pattern may resemble a power function (e.g., Anderson & Tweney, 1997).
4. We chose accuracy as our measurement over other measures such as d'. The concern for us is the *amount* of information remembered, not the ability to *discriminate* old from new items, which is what d' measures.

5. Note that we are assessing the degree to which the patterns of data conform to power or linear mathematical functions. We are not varying free parameters of the RAFT model to fit it to the data. Instead, the model makes predictions of when different alternative patterns would emerge, and we are testing whether model specified manipulations result in that qualitative difference.

References

Akan, M., Stanley, S. E., & Benjamin, A. S. (2018). Testing enhances memory for context. *Journal of Memory and Language*, *103*, 19–27.

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

Anderson, R. B. (2001). The power law as an emergent property. *Memory & Cognition*, *29*(7), 1061–1068.

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

Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*, *55*(1), 25–35.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, *104*(1), 54–75.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), Information processing and cognition: The Loyola symposium (pp. 123–144). Lawrence Erlbaum.

Bogartz, R. S. (1990a). Evaluating forgetting curves psychologically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 138–148.

Bogartz, R. S. (1990b). Learning-forgetting rate independence defined by forgetting function parameters or forgetting function form: Reply to Loftus and Bamber and to Wixted. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 936–945.

Bouwmeester, S., & Verkoeijen, P. P. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language*, *65*(1), 32–41.

Burtt, H. E., & Dobell, E. M. (1925). The curve of forgetting for advertising material. *Journal of Applied Psychology*, *9*(1), 5–21.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563–1569.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236–246.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102.

Chan, J. C., McDermott, K. B., & Roediger III, H. L. (2006). Retrieval-induced facilitation: initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*(4), 553.

Craig, C. S., Sternthal, B., & Olshan, K. (1972). The effect of overlearning on retention. *Journal of General Psychology*, *87*, 85–94.

De Leeuw, J. R. (2015). JsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12.

Doolen, A. C., & Radvansky, G. A. (2021). A novel study: long-lasting event memory. *Memory*, 1–20.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). Columbia University, Teacher's College. (Original work published 1885; Reprinted 1964, Dover).

Fisher, J. S., & Radvansky, G. A. (2018). Patterns of forgetting. *Journal of Memory and Language*, *102*, 130–141.

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

Glenberg, A., Smith, S. M., & Green, C. (1977). Type I rehearsal: Maintenance and more. Journal of. *Verbal Learning and Verbal Behavior*, *16*(3), 339–352.

Hellyer, S. (1962). Supplementary report: Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental Psychology*, *64*(6), 650.

Jeunehomme, O., Folville, A., Stawarczyk, D., Van der Linden, M., & D'Argembeau, A. (2018). Temporal compression in episodic memory for real-life events. *Memory*, *26*(6), 759–770.

Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4–5), 528–558.

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, *27*(2), 317–326.

Kristo, G., Janssen, S. M., & Murre, J. M. (2009). Retention of autobiographical memories: An Internet-based diary study. *Memory*, *17*(8), 816–829.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1787–1794.

Linton, M. (1982). Transformations of memory in every life. In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (pp. 77–91). Freeman.

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.

Loftus, G. R., & 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*(5), 916–926.

Meeter, M., Murre, J. M., & Janssen, S. M. (2005). Remembering the news: Modeling retention data from a study with 14,000 participants. *Memory & Cognition*, *33*(5), 793–810.

Nickerson, R. S., & Adams, M. J. (1979). Long-term memory for a common object. *Cognitive Psychology*, *11*(3), 287–307.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646.

Nunoi, M., & Yoshikawa, S. (2016). Deep processing makes stimuli more preferable over long durations. *Journal of Cognitive Psychology*, *28*(6), 756–763.

Postman, L., & Riley, D. A. (1959). *Degree of learning and interserial interference in retention: A review of the literature and an experimental analysis*. University of California Publications in Psychology.

Rawson, K. A., & Zamary, A. (2019). Why is free recall practice more effective than recognition practice for enhancing memory? Evaluating the relational processing hypothesis. *Journal of Memory and Language*, *105*, 141–152.

Rickard, T. C., & Pan, S. C. (2018). A dual memory theory of the testing effect. *Psychonomic Bulletin & Review*, *25*(3), 847–869.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Roediger, H. L., & Smith, M. A. (2012). The "pure-study" learning curve: The learning curve without cumulative testing. *Memory & Cognition*, *40*(7), 989–1002.

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

Runquist, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, *11*(6), 641–650.

Slamecka, N. J. (1985). On comparing rates of forgetting: Comment on Loftus. *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.

Staats, A. W., Minke, K. A., & Stalling, R. B. (1970). *Long-term Retention of Conditioned Attitudes* (No. TR-6). Hawaii University Honolulu, Department of Psychology.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1392–1399.

van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*(2), 247–264.

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology*, *18*(2), 225–252.

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*(2), 379–386.

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, *2*(6), 409–415.

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, *25*(5), 731–739.

Youtz, A. C. (1941). An experimental evaluation of Jost's laws. *Psychological Monographs*, *53*(1), i–54.

Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, *38*(8), 995–1008.