Journal of Memory and Language 108 (2019) 104035

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Linear forgetting☆

Jerry S. Fisher*, Gabriel A. Radvansky

Department of Psychology, University of Notre Dame, United States

ABSTRACT

Memory retention and forgetting is typically captured by an Ebbinghaus curve in which there is a sharp initial decrease that follows a negatively accelerated function. This pattern, typically well fit by a power function and poorly fit by a linear function, has been observed across a variety of materials, tasks, and retention lengths. However, here we demonstrate, across three experiments, a set of retention patterns that are better fit by a linear function, which is not accounted for by any existing theory of memory retention and forgetting. This linear pattern was also observed, but not noted, in existing studies from the literature. Our assessment suggests that higher degrees of learning and meaningfully complex materials may be jointly needed to observe linear forgetting. A simulation is provided as a proof of concept that linear forgetting may emerge when there are (a) the multiple components of memory traces are lost at different rates, with each following a negatively accelerating function, and (b) memory responses may be made using degraded memory traces, such as through partial matching and/or reconstructive processes. Linear forgetting is important to the study of memory retention and forgetting, and it may apply to the bulk of everyday event memories that concern people over long lasting periods of time.

Introduction

The standard view among memory scientists is that memory retention and forgetting follow a pattern of a large initial rate of forgetting that slows in a negatively accelerated manner. This is the classic Ebbinghaus forgetting curve (Ebbinghaus, 1885), as shown in Fig. 1. Numerous studies have found this to be a reliable pattern that appears across a variety of experimental designs and materials (Rubin & Wenzel, 1996). Moreover, it is generally acknowledged that these retention and forgetting curves are well fit by a power function (Averell & Heathcote, 2011; Wixted & Carpenter, 2007; Wixted & Ebbesen, 1991). The power function that is used to account for patterns of retention and forgetting is $m = at^r$ (e.g., Anderson, 2001; Anderson & Tweney, 1997; Myung, Kim, & Pitt, 2000; Rubin, 1982; Rubin & Wenzel, 1996; Rubin, Hinton, & Wenzel, 1999; Sikström, 1991; Wixted & Carpenter, 2007; Wixted & Ebbesen, 1991, 1997) with *m* being memory, *t* being the delay time, *r* capturing the rate of forgetting, and *a* being a scaling parameter. The pattern of data in Ebbinghaus's original study is well fit by a power function, $r^2 = .97$, but is poorly fit by a linear function, $r^2 = .19$. According to a power function account, there is a consistent proportional loss of information in memory across log time. Thus, this version of a power function serves as the *default hypothesis* in research on human memory, and it conveys the idea that, over time, memory retention and forgetting will show a negatively accelerating pattern.

It is important to note that the observed retention and forgetting functions reflect some degree of averaging across memory traces, and, often, several people. This process of averaging may be what produces the power function. For example, when several exponential functions with differing loss rates are averaged, the composite function is often better fit by a power function (Anderson, 2001; Anderson & Tweney, 1997; Murre & Chessa, 2011; Myung et al., 2000; but see Wixted & Ebbesen, 1997). That said, regardless of whether the retention and forgetting pattern shown by individual memory traces follows an exponential or a power function, the basic pattern is negatively accelerating, consistent with Ebbinghaus (1885). Thus, the default hypothesis is that memory retention and forgetting, in general, will show a negatively accelerating pattern, consistent with a power function.

With this well-established regularity in mind, in a recent study, Radvansky, O'Rear, and Fisher (2017) assessed memory across five intervals of up to two weeks after learning. The primary aim of this work was to assess any changes in the differential fan effect (e.g., Radvansky & Zacks, 1991) over time using materials that were sentences that described objects in locations (e.g., "The potted palm is in the hotel"). These materials were memorized to a high criterion, and memory was tested using timed recognition (in eight separate blocks). The primary finding was that the recognition test data from this study revealed that the basic differential fan effect pattern remained largely intact over this time. In addition to the primary finding, and what is our focus here, is that the pattern of recognition accuracy did not show a typical negatively accelerating function. Instead, the retention function

☆ This work was done, in part, as fulfillment of the master's thesis for the first author. The authors thank Daniel Hochee for programming assistance, and Jack Feger, Laura McKernan, and Jinelfry Rodriguez for their assistance in collecting the data.

* Corresponding author at: Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, United States.
E-mail address: jfishe12@nd.edu (J.S. Fisher).

https://doi.org/10.1016/j.jml.2019.104035
Received 26 July 2018; Received in revised form 20 June 2019; Accepted 26 June 2019
Available online 09 July 2019
0749-596X/

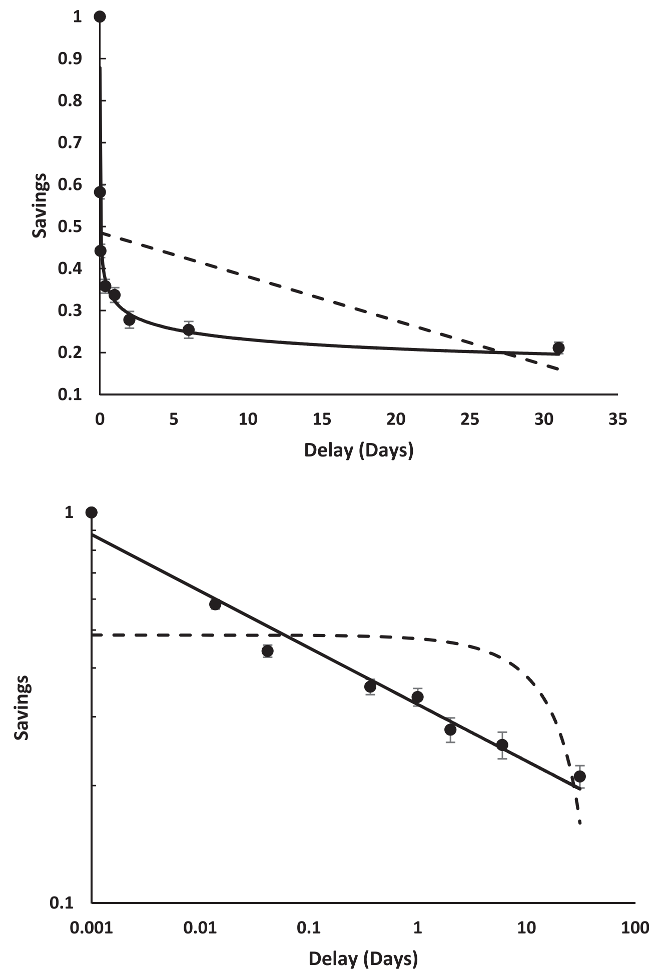

Fig. 1. Plots of the Ebbinghaus (1885) retention curve with linear axes on the top, and logarithmic axes on the bottom.

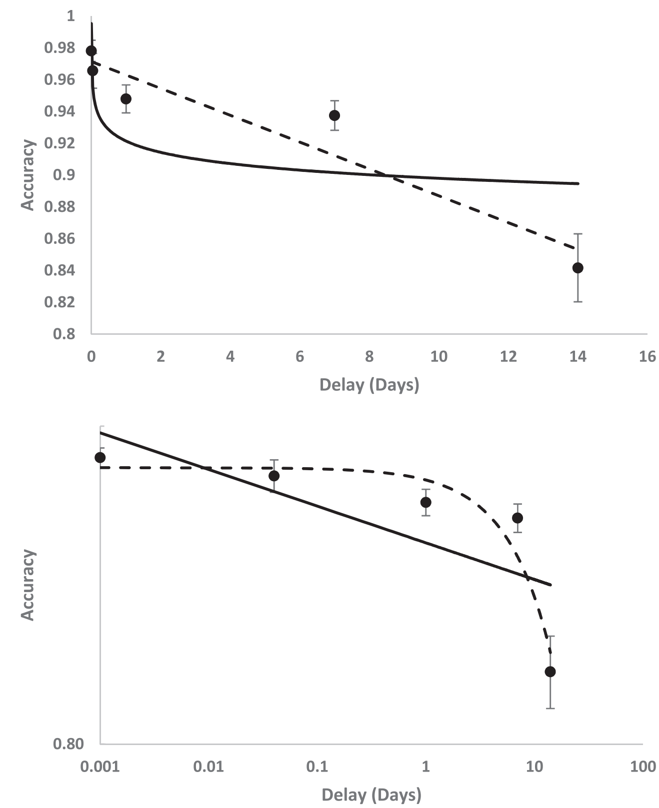

Fig. 2. Plots of the Radvansky et al. (2017) accuracy data for Block 1 only with linear axes above and logarithmic axes below.

appeared to be better fit by a linear function, namely $m = a - t * r$, as is shown in Fig. 2.

Clearly, the retention and forgetting pattern of the Radvansky et al. (2017) study differs from the common Ebbinghaus pattern of a negatively accelerating function. Rather, the pattern of forgetting here better resembles a linear function than a power function. This distinction is important because with a linear pattern of forgetting, there is not a constant loss in the *proportion* of information in memory over log time, but a constant loss in the *amount* of information in memory. In such a case, the proportion of loss actually increases over time. This is important because, if such a pattern were found to be regular and consistent, it is not currently accounted for by any existent theories of memory retention and forgetting, all of which assume a negatively accelerating function. Given the potential importance of such a retention pattern, this issue merits a closer examination.

In this paper, we present three experiments. The first is a reanalysis of the accuracy data reported by Radvansky et al. (2017) to examine the linearity of the retention and forgetting pattern. The second experiment was directly aimed at assessing linear forgetting under a similar design for a period of up to twelve weeks. The third experiment was a further replication using different materials and learning criterion. After this, we report the results of a survey of the literature aimed at exploring whether there are any other sets of data that exhibit linear forgetting, but which were not noticed or reported by the original investigators. That is, is this pattern of data robust enough to be observed in studies done by other researchers, at other times, with different materials and methods, and for different aims? From this assessment, we derive factors that need to be present for linear forgetting to be observed, namely the presence of more meaningfully complex materials that allow for meaningful association and elaboration, and that these materials be learned to a relatively high degree. Finally, we report a simulation that captures these qualities, along with the idea that a successful memory response may be made using partial information via processes of partial matching and reconstruction, to provide a proof of concept for our theoretical account of linear forgetting.

Experiment 1

Experiment 1 is a reanalysis of the Radvansky et al. (2017) dataset that more explicitly explores the nature of the retention and forgetting function. Specifically, we assessed the fit of power and linear functions for the overall accuracy data. According to the default hypothesis, performance is better fit by a power function, whereas according to an alternative hypothesis, performance is better fit by a linear function.

Method

Participants: In this study, 200 undergraduate students from the Department of Psychology at the University of Notre Dame were recruited from the participant pool. There were 40 participants in each of the five retention interval groups. They received partial course credit for their participation. This research was approved by the Institutional Review Board at the University of Notre Dame.

Materials: The study materials consisted of 18 sentences that described objects in locations (e.g., “The potted palm is in the hotel”) that were created in the same manner as previous studies (e.g., Radvansky, Spieler, & Zacks, 1993). These sentences were generated using a set of 12 objects and 12 locations, and they are listed in Appendix A. Because the focus of the Radvansky et al. (2017) study was the differential fan effect, the object and location concepts for this material set each had

1–3 associations. The details of this aspect of the materials and design is reported in Radvansky et al. (2017).

For the recognition test, the studied probes were the sentences learned during the first part of the study. Nonstudied probes were created by re-pairing the objects and locations to form new, nonstudied sentences. By creating nonstudied sentences in this way, people will be much less likely to use plausibility judgments (Reder & Anderson, 1980) to make their recognition decisions.

Procedure: During learning, participants were presented with the study sentences on a computer screen, one at a time, for seven seconds each. The materials were presented in black letters on a white background using 20-point Courier font. The instruction was to memorize the sentences as efficiently as possible. After viewing all 18 sentences, participants were then given a cued recall test with questions of the form "Where is the *object*?" and "What is in the *location*?" Participants responded by typing in their answer into a textbox. After submitting their responses, they were provided with feedback about the correctness of their responses. If any of their responses were incorrect, they were provided with the correct answer(s). At the completion of the cued recall test portion of memorization, the participants then restudied all 18 of the sentences and were tested again. This study-test procedure continued until participants reached a criterion of two perfect cued recall tests.

After memorization, and the designated retention interval, participants were given a recognition test. The retention intervals used in this study were immediate, one hour, one day, one week, or two weeks later. For the recognition test, the memory probes were presented one at a time on the screen with the task of indicating whether the probe item was studied or not. Participants indicated that a probe sentence was studied by pressing the left button on a computer mouse, which was marked with a "Y" for "yes, this sentence was studied." Alternatively, participants indicated nonstudied sentences by pressing the right button, which was marked with an "N" for "no, this sentence was not studied." If a person responded incorrectly to any of these probes, the participant was given feedback in the form "**Error! Sentence true!**" or "Error! Sentence not true!**.

The recognition test consisted of 8 blocks of 36 trials, composed of the 18 studied and 18 non-studied sentences, for a total of 288 trials. To familiarize people with the procedure of the recognition test procedure, a practice period of 18 trials was provided. For practice, the computer displayed items that read either "sentence studied" or "sentence not studied," and the participants responded accordingly. Because of concerns about additional learning for the individual probe items that may occur as a result of feedback during recognition testing, only the first block of trials is analyzed here.

Analysis: The re-analysis of the data for Experiment 1 was aimed at the pattern of retention and forgetting. Specifically, whether it was better captured by a power function or a linear function. This was assessed using a Pearson's coefficient (r^2) to convey the proportion of variance accounted for by the different functions. This common curve-fitting measurement has the advantage of being simple while still providing the same amount of information as would an ANOVA (see Rubin & Wenzel, 1996). Specifically, we compare the power r^2 with the linear r^2. In addition, we also report AIC and BIC measures of fit assessment.[1] These assessments were done for both overall recognition averaged across all blocks and then for the first block only.

[1] AIC and BIC measures are typically used when different models vary in the number of parameters, and these measures take these differences into account. Here, linear and power functions both have the same number of parameters, thus the AIC and BIC measures are unlikely to provide any new and useful information. However, these values are included here to emphasize the difference in r^2 value.

Results and discussion

Learning Rates. As reported by Radvansky et al. (2017), participants took 3–10 study-test learning cycles to memorize the sentences (M = 4.8, SE = .08). The number of learning cycles needed did not vary among the various retention groups, $F < 1$.

Recognition Data: The data shown in Fig. 2 is the retention pattern for the first block of trials for the Radvansky et al. (2017) study. Here the data were poorly fit by a power function, r^2 = .55, $AIC = -13.78$, $BIC = -14.95$, but well fit by a linear function, r^2 = .91, AIC = –22.06, BIC = –23.23. Thus, the pattern of retention and forgetting is far more consistent with a linear pattern than the prediction of a power function by the default hypothesis.

Experiment 2

Our re-analysis of the Radvansky et al. (2017) data in Experiment 1 showed evidence of a linear pattern. The aim of Experiment 2 was to assess whether this pattern of retention and forgetting can be replicated using different materials. While we were able to address some possible reasons for this in our reanalysis of that data, there are other possibilities that have yet to be considered. One of these, consistent with the default hypothesis, is that because in the Radvansky et al. study retrieval performance was near ceiling initially, it may not have allowed for enough time to pass for a drop to a level for the typical negatively accelerating pattern to be observed. Alternatively, the linear pattern of forgetting may be a result of the fact that many of the objects and locations in Experiment 1 had multiple associations because of the assessment of the fan effect. It could be that these associations served as a basis for priming various elements of other study sentences, thereby leading to an unusual pattern of retention and forgetting data.

To address this possibility, Experiment 2 was similar to Experiment 1 in that it (a) had a high memorization criterion, (b) used sentences as materials, (c) included multiple retention intervals, and (d) used accuracy in recognition testing to measure retention. However, there were several important differences as well. First, Experiment 2 used sentences about people doing activities (e.g., "The student is eating.") rather than objects in locations (e.g., "The potted palm is in the hotel.") to generalize the findings to different materials. Second, none of the people or activities in the study sentences had multiple associations across multiple sentences, as was the case in Experiment 1. Also, because accuracy performance had not dropped much after two weeks in the Experiment 1, allowing for the possibility that performance had not moved off of ceiling performance sufficiently to allow for an observation of a normal forgetting curve, the retention intervals in Experiment 2 were expanded up to 12 weeks. Finally, because the linear pattern of forgetting was observed even on the first block of trials for Experiment 1, only one recognition block was used rather than the eight. This also reduces any influences of additional learning that could occur during testing.

For this study, there are two likely outcomes. First, consistent with the default hypothesis, a standard negatively accelerating function will be observed. This would be in line with traditional accounts of memory. Alternatively, a linear forgetting function would be observed, even with longer retention intervals. This would be inconsistent with a proportional loss of information in memory over time but would be consistent with consistent loss in the amount of information lost over time. That is, there would be an increase in the proportion of memory loss as retention intervals grew longer. In other words, the rate of forgetting would be speeding up. There is no prior theory of memory to account for such a pattern.

Method

Participants. One hundred forty-four students (99 female; age 17–22, M = 19.3, SE = .08) native-English speaking were recruited from the University of Notre Dame participant pool in exchange for partial course credit. They were assigned (16 each) into nine retention interval groups: immediate testing, 1-day, 3-day, 1-week, 2-week, 4-week, 6-week, 8-week, and 12-week retention intervals. This research was approved by the Institutional Review Board at the University of Notre Dame.

Materials. The study materials were 18 sentences of people doing activities in the form "The *person* is *activity*", such as "The student is eating" or "The mailman is swimming". These sentences were generated using a set of 18 occupations and 18 activities, and they are listed in Appendix A. The occupations and activities were randomly paired without replacement for each person. These sentences were similar to the object-location sentences of Experiment 1 insofar that they refer to individual events.

For the recognition test, there were 36 probe sentences. Eighteen of these were the sentences that the people studied. The other 18 probes were negatives created by recombining the occupations and activities. Taking the two sample sentences from the previous paragraph, possible negative probe sentences could be "The student is swimming." or "The mailman is eating." For the negatives, each person and activity only occurred once. Both the learning and the testing programs were written in JavaScript using the jsPsych library (De Leeuw, 2015).

Procedure. Learning took place in the Memory Laboratory at the University of Notre Dame, and all people followed the same memorization procedure. After providing informed consent, they were led to individual computers with 22-inch monitors. They were then presented with the 18 study sentences, one at a time, in random order, for four seconds each.

After presentation of all 18 study sentences, participants were given a 36-item cued recall test in which they filled in a sentence blank with the correct person or activity. For example, a test question could be, "The ____ is eating." or "The student is ____.". In this way, memory for each person and activity was assessed. These questions were presented randomly, one at a time, and were self-paced. If people typed in the wrong answer, they received feedback in the form, "Sorry, [*provided answer*] is incorrect. The correct answer is [*correct answer*].", such as "Sorry, 'baker' was incorrect. The correct answer was 'student'." After the cued recall test, if any of the questions were incorrect, people returned to the study portion. This study-test cycle continued until people twice scored perfectly.

Following the appropriate retention interval, people were given a yes/no recognition test. For this test, people were presented with the studied and non-studied memory probes. These probes were presented one a time and in a random order. For each probe, participants responded by clicking the appropriate button on the screen, and there was no time limit to respond. Unlike Experiment 1, there was no feedback during the recognition test.

People in the immediate group took this test after a brief distractor task involving memorizing seven numbers (from 1 to 10) and recalling them in the correct order. This was done twice to encourage some forgetting of the materials. Testing in this group was done on the same computers as the memorization phase. People in any of the delayed testing groups received an email with a link to the recognition test. For the people in the one-day group, this email was sent one hour before their target testing time. For people in the three-day group, this email was sent six hours before their target testing time. All other retention groups received the email 24 h before their target testing time. For people in all of the delayed retention groups, testing was done on a personal computer, tablet, or smartphone.

People were asked to take the recognition test in a quiet place and with enough time. The retention intervals were as follows: 1-day group (*Median deviation* = 27 min), 3-day group (*Median deviation* = 46 min), 1-week group (*Median deviation* = 51 min), 2-week group (*Median deviation* = 44 min), 4-week group (*Median deviation* = 66 min), 6-week group (*Median deviation* = 35 min), 8-week group (*Median deviation* = 403 min), and 12-week group (*Median deviation* = 78 min).

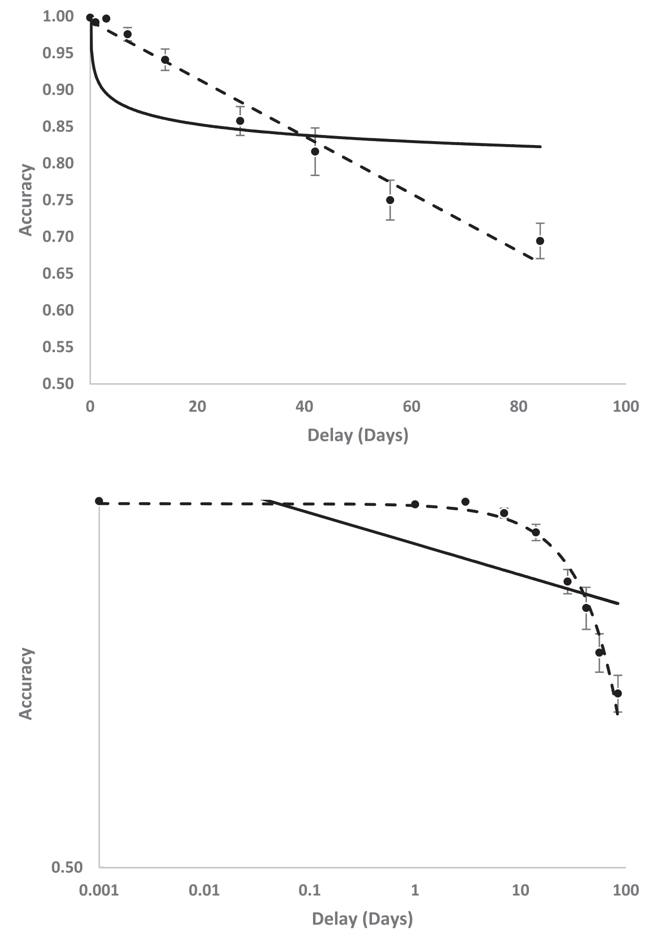

Fig. 3. Recognition accuracy in Experiment 2 as a function of time with linear axes above and logarithmic axes below.

Results and discussion

Learning Rates. People took 2 to 8 study-test learning cycles to memorize the sentences (M = 3.6, SE = .09). The number of learning cycles needed did not vary among the various retention groups, $F < 1$.

Overall Data Fits. Recognition performance was measured by overall proportion of hits and correct rejections. As can be seen in Fig. 3, accuracy remained very high through the first three days, whereupon it began falling through the remaining retention intervals. The data were poorly fit by a power function, r^2 = .42, $AIC = -12.29$, $BIC = -11.7$, but well fit by a linear function, r^2 = .98, $AIC = -40.71$, $BIC = -40.12$. Thus, similar to Experiment 1 and inconsistent with the default hypothesis, the data are better described by a linear function.

Experiment 3

The aim of Experiment 3 was to further replicate the linear pattern of retention and forgetting reported for Experiments 1 and 2 with yet a different set of materials. Specifically, rather than sentences, this study used paired associates. The materials were animal-location pairs, such as Ant-Grotto. Note that although they are word pairs, people are likely to refer to such materials as referring to events (Radvansky, 2005). Also, rather than using two study-test learning cycles in which all of the study list items were repeated until all test questions were answered correctly within a cycle, a drop-out procedure was used. Specifically, if a cue resulted in a correct response, then it was dropped from the study set. However, if a cue was incorrectly responded to, feedback was provided in the form of presenting the correct pair, and that pair remained in the study set.

Method

Participants. One hundred forty four native-English speaking adults (89 female; age 19–69, $M = 40.2$, $SE = 1.00$) were recruited from Amazon's Mechanical Turk (AMT) service. This study was restricted to participants living in the United States. Participants were assigned (24 each) into six retention interval groups: immediate testing, 1-day, 3-day, 1-week, 10-day, and 2-week delay. Participant age did not vary between groups, $F(5, 138) = 1.19$, $MSE = 145$, $p = .32$, $\eta_p^2 = .04$. Participants whose A' performance score was below .8 were removed on the assumption that they were not doing the task carefully. This resulted in dropping four participants in the immediate group, two in the three-day group, four in the one-week group, four in the ten-day group, and three in the two-week group. No one-day participants needed to be replaced. This research was approved by the Institutional Review Board at the University of Notre Dame.

Materials. The study materials were 16 animal-location pairs. To increase the number of possible combinations, each of the 16 animals were randomly selected from a larger set of 32 animals, and each of the 16 locations were randomly selected from a larger set of 32 locations for each participant. The animal and locations concepts used are provided in Appendix C. The pairs were separated by a hyphen, such as "Ant – Grotto" or "Deer – Beach".

For the recognition test, there were 32 probe pairs. Sixteen were the studied pairs. The other 16 were negatives created by recombining the animals and locations, similar to Experiments 1 and 2. For the negatives, each animal and location only occurred once. Both the learning and the testing programs were written in JavaScript using the jsPsych library (De Leeuw, 2015).

Procedure. Learning took place online. Participants were presented with the 16 pairs, one at a time, in random order, for seven seconds each. After presentation of all 16 study pairs, participants were given two 32-item cued recall tests with a drop-out procedure. In these tests, each study list item (animal or location) was used as a cue for the participant to type its corresponding pair. For example, a test question could be, "Ant - ____" or "____ - Beach". These questions were presented randomly, one at a time, and were self-paced. If people typed in the wrong answer, they received feedback in the form, "Sorry, [*provided answer*] was incorrect. The correct pair is [*correct pair*]", such as "Sorry, Hippo was incorrect. The correct pair is Ant - Grotto." This memorization protocol used a drop-out testing procedure. Specifically, if the cued-recall answers for a given cue was correct, then that item was removed from the study set. However, cues that had incorrect answers remained in the study list. In this way, each of the cued-recall probes had to be answered correctly once for memorization.

Following the appropriate retention interval, participants were given a yes/no recognition test. For this test, people were presented with the studied and non-studied memory probes. These probes were presented one a time and in a random order. For each one, participants responded by clicking the appropriate button on the screen. Again, like Experiment 2, there was no feedback or time limit.

People in the immediate group took this test after a brief distractor task in which they judged the semantic sensibility of 10 sentences, presented one at a time. For example, the sentence "It was the stereo that the fraternity played loudly." was a semantically sensible item, whereas "It was the necklace that stole the thief." was not.

People in any of the delayed testing groups received an email through AMT's server with a link to the recognition test. For the people in the one-day group, this email was sent two hours before their target testing time. For people in the three-day group, this email was sent six hours before their target testing time. All other retention groups received the email 12 h before their target testing time. For people in all of the delayed retention groups, testing was done on a personal computer, tablet, or smartphone. Following recognition testing, the semantic sensibility task was given.

The retention intervals were as follows: 1-day group (*Median deviation* = 1 min), 3-day group (*Median deviation* = 159 min), 1-week group (*Median deviation* = 243 min), 10-day group (*Median deviation* = 114 min), 2-week group (*Median deviation* = 54 min).

Results and discussion

Recognition performance was measured by overall proportion of hits and correct rejections. As can be seen in Fig. 4 and similar to Experiment 2, accuracy remained high through the first three days before falling through the remaining retention intervals. The data were poorly fit by a power function, $r^2 = .49$, $AIC = -12.05$, $BIC = -12.68$, but well fit by a linear function, $r^2 = .93$, $AIC = -23.57$, $BIC = -24.2$. Thus, similar to Experiments 1 and 2, and inconsistent with the default hypothesis, the data are better described by a linear function.

Prior studies with linear forgetting

The results of the three experiments reported here suggest that a reliable linear retention and forgetting pattern can be repeatedly observed. Rather than doing further new data collection, at this point we chose to assess whether there are any other studies in the literature that show clear evidence of linear forgetting. This was done because we felt that it would be more convincing to observe linear forgetting in data from other labs, at different times, using different materials and procedures, and for studies that were not done with the aim of testing for linear forgetting. The primary aim was to use studies in which linear forgetting is observed to help identify which characteristics they share

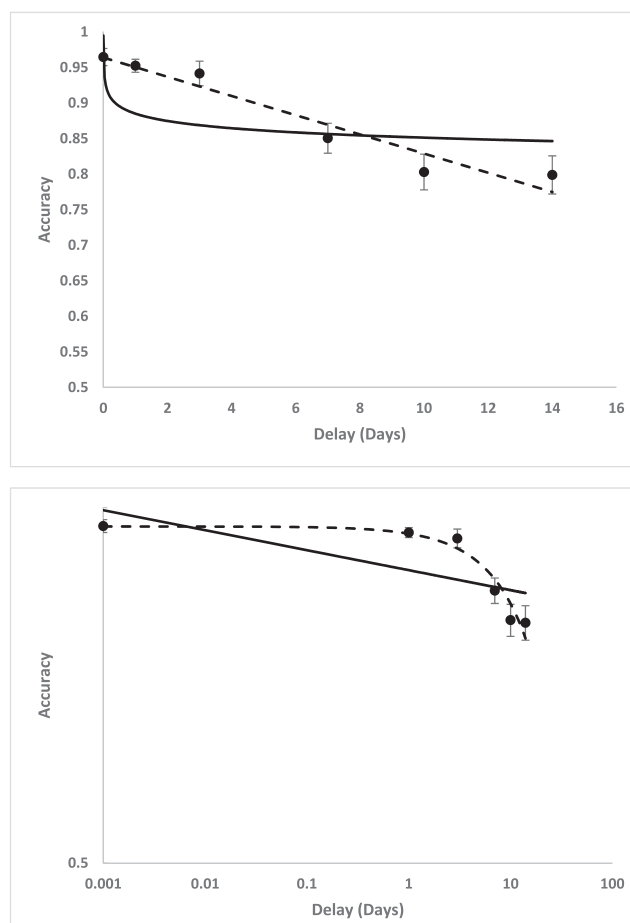

Fig. 4. Recognition accuracy in Experiment 3 as a function of time with linear axes above and logarithmic axes below.

with our experiments to help describe when linear forgetting would be more likely to be observed.

To assess whether linear forgetting is observed we surveyed our database of 97 studies reporting 392 experiments with three or more retention intervals. We used the following criterion for inclusion. First, the study needed to have four or more retention intervals. This was done to ensure that there was enough data to accurately fit a function.[2] Second, the study needed retention intervals of at least one day because we are primarily interested in retention over long periods of time. Third, a study needed to have a greater linear than power function fit because our aim here was to find studies in which forgetting that conforms more to a linear function. Finally, the r^2 of the fit for at least one of the assessed linear functions needed to be at least .75. This criterion was used to assess when a good linear pattern was observed.

We found several data sets that meet these criteria (Bahrick, Bahrick, & Wittlinger, 1975; Burtt & Dobell, 1925; Carpenter, Pashler, Wixted, & Vul, 2008; Cepeda et al., 2009; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Jeunehomme, Folville, Stawarczyk, Van der Linden, & D'Argembeau, 2018; Kristo, Janssen, & Murre, 2009; Meeter, Murre, & Janssen, 2005; Nunoi & Yoshikawa, 2016; Runquist, 1983; Thompson, Skowronski, Larsen, & Betz, 1996; Wagenaar, 1986). The details of these studies are provided below. Across these studies, we explored three methodological characteristics that we considered may lead to linear retention and forgetting. These are (a) the type of memory test used, (b) the degree of original learning, and (c) the meaningful complexity of the materials used.

Burtt and Dobell (1925)

Following up on Ebbinghaus's work, Burtt and Dobell (1925) examined retention for paired associates for up to four weeks. They had three experiments, where each involved learning 100 commodities (e.g., perfume) paired with a fictitious brand name (e.g., Pettal). In all three experiments, each pair was presented twice (in succession) and memory was tested using recall and recognition after a designated retention interval. Participants returned at each retention interval (thus, everyone had all retention intervals).

Experiment 1 had 58 participants who were tested at five intervals: immediate, one-week, two-weeks, three-weeks, and four-weeks. Each cued recall session consisted of people being given a blank with a set of 20 cues (the commodity) and the task of recalling the product name. This was followed by a recognition test in which each of the 20 probed commodities was provided along with five possible pairs: the valid brand and four alternative brands (it is not stated whether the foil brands were studied before and paired with other commodities, but this seems to have been the case). The dependent variable was proportion correct for both recall and recognition.

For Experiments 2 and 3, they used the same materials as Experiment 1, and also included conditions in which a subset of the materials was given an additional learning at a later time. This reduced the original 100 paired associates from Experiment 1 to 40 that had no subsequent learning. The retention of these 40 associates with no additional learning is of primary concern here. Experiments 2 and 3 also differed from Experiment 1 in that they had fewer participants (41 and 47, respectively) and only retention intervals that spanned 16 (Experiment 2) and 17 days (Experiment 3).

For these experiments, while the cued recall was better fit by a power function, recognition showed a more linear pattern. These results are shown in Table 1, and plots of the recognition data is shown in Fig. 5. Specifically, for Experiment 1, the recognition data was better fit with a linear than a power function. The opposite was true for the recall data.[3] For Experiment 2, recognition and recall data were both better fit by a power function than a linear function. Note that this experiment also had the smallest sample size. Finally, for Experiment 3, the recognition data was better fit with a linear than a power function, and the opposite was true for the recall data.

Considering Burtt and Dobell (1925) experiments as a whole, recognition was better fit by a linear than a power function in two of the three experiments, but cued recall performance was better fit by a power function. They used meaningfully complex materials of paired associates of words with related but different meanings. Finally, they presented these materials more than once during learning.

Bahrick et al. (1975)

In another study, Bahrick et al. (1975) examined the retention of names and faces of high school classmates over the course of up to 57 years. They used a cross sectional design involving 392 high school graduates from nine different time period groups since graduation. Participants had six memory tasks: free recall, name recognition, picture recognition, picture matching, name matching, and picture cuing. Although the pattern of retention across tasks was somewhat inconsistent, they were all better fit by a linear than a power function, as seen in Table 1, with name recall and name matching tasks having good linear fits. A plot of these tasks is provided in Fig. 6. Free recall was measured by an adjusted mean of the number of correct recalls, and recognition by an adjusted percentage correct.

This study used multiple memory test types, all of which resulted in better linear than power function fits, although the linear fits were best for name recall and name matching. The materials were meaningfully complex in that they involved more than one component (i.e., pairing of first and last names). Finally, because these names were for high school classmates, it is presumed that they had been overlearned during the participants' high school years.

Runquist (1983)

Runquist (1983) examined the effect of initial testing upon subsequent retention using paired associates and cued recall testing. In one experiment, 288 people learned 24 paired associates (presented once or three times) and then were given a cued recall test for 12 of these pairs. After a designated retention interval of either immediate, one-hour, six-hours, two-days, seven-days, or 21-days, participants were given another cued recall test for all 24 pairs. The results, shown in Fig. 7. Retention of tested items that were presented once or three times, were better fit by a linear than a power function. In contrast, the retention of the untested items presented once or three times were better fit with a power function than a linear function.

This study used cued recall for paired associates of items with different meanings. Moreover, this study showed linear forgetting for the paired associates that were learned with additional testing, but power forgetting for the associates that did not receive the learning benefit of additional testing. This suggests that a greater degree of learning influences the linearity of the retention function.

Wagenaar (1986)

Wagenaar (1986) examined his own autobiographical memory for

[2] That said, some studies with three retention intervals can show stronger linear than power function forgetting patterns (e.g., Goetschalckx, Moors, & Wagemans, 2018).

[3] Note that we removed the last time point of Experiment 1 in Table 1 because performance fell to chance by three weeks. If these data are included, the pattern is not altered much. When this extra time point is included, for recognition the data are still better fit by a linear function, $m = .83 * t - .01$, $r^2 = .90$, than a power function, $m = .65 * t^{-.05}$, $r^2 = .78$, whereas for cued recall the data are still better fit by a power function, $m = .27 * t^{-.03}$, $r^2 = .90$, than a linear function, $m = .27 * t - .087$, $r^2 = .59$.

Table 1
Linear and power function fits (in terms of r^2) of previously published data, along with values for other components of the corresponding functions.

Publication	Data	Memory test	Meaningfully complex	Degree of learning	Power fit	Exponent	Constant	Linear Fit	Slope	Intercept
Ebbinghaus										
Ebbinghaus (1885)		Savings	No	High	.97	−.15	.32	.19	−.011	.49
Current Study										
Current Study	Experiment 1 (all)	Recognition	Yes	High	.40	−.01	.94	.82	−.004	.97
	Experiment 1 (block 1)	Recognition	Yes	High	.55	−.01	.92	.91	−.008	.97
	Experiment 2	Recognition	Yes	High	.42	−.03	.92	.98	−.004	.99
	Experiment 3	Recognition	Yes	High	.49	−.02	.89	.93	−.01	.96
Other Linear Fits										
Burtt and Dobell (1925)	Experiment 1	Cued recall	Yes	High	.89	−.06	.29	.70	−.018	.33
	Experiment 1	Recognition	Yes	High	.77	−.04	.66	.97	−.017	.85
	Experiment 2	Cued recall	Yes	High	.97	−.23	.11	.56	−.024	.36
	Experiment 2	Recognition	Yes	High	.92	−.04	.73	.76	−.017	.86
	Experiment 3	Cued recall	Yes	High	.84	−.19	.17	.69	−.026	.44
	Experiment 3	Recognition	Yes	High	.78	−.03	.79	.92	−.015	.91
Bahrick et al. (1975)	Name	Recall	Yes	High	.69	−.06	.93	.90	−.004	.32
	Name	Recognition	Yes	High	.37	−.03	.89	.51	−.003	.90
	Picture	Recognition	Yes	High	.03	−.01	.89	.18	−.002	.91
	Picture	Matching	Yes	High	.16	−.02	.89	.39	−.003	.90
	Name	Matching	Yes	High	.53	−.06	.93	.90	−.007	.94
	Picture	Cued Recall	Yes	High	.33	−.13	.58	.48	−.006	.59
Runquist (1983)	Tested Items (3 presentation)	Cued Recall	Yes	High	.58	−.09	.67	.93	−.032	.95
	Tested Items (1 presentations)	Cued Recall	Yes	High	.55	−.06	.76	.65	−.022	.94
	Untested Items (3 presentation)	Cued Recall	Yes	High	.82	−.17	.38	.63	−.029	.68
	Untested Items (1 presentation)	Cued Recall	Yes	Low	.77	−.16	.31	.4	−.020	.53
Wagenaar (1986)	One Cue	Cued Recall	Yes	High	.97	−.46	.35	.77	−.058	.41
	Two Cues	Cued Recall	Yes	High	.95	−.33	.60	.81	−.075	.68
	Three Cues	Cued Recall	Yes	High	.86	−.26	.75	.63	−.075	.83
	Critical Detail	Cued Recall	Yes	High	.90	−.34	.78	.90	−.102	.89
Thompson et al. (1996)	Location	Cued Recall	Yes	High	.79	−.10	1.5	.94	−.0002	.94
	Who With	Cued Recall	Yes	High	.58	−.14	1.78	.79	−.0003	.94
Meeter et al. (2005)	Experiment 1	Cued Recall	Yes	Unclear	.87	−.16	.79	.75	−.001	.51
	Experiment 1	Recognition	Yes	Unclear	.90	−.07	.85	.77	−.001	.70
	Experiment 2	Cued Recall	Yes	Unclear	.63	−.08	.43	.35	−.001	.33
	Experiment 2	Recognition	Yes	Unclear	.76	−.05	.66	.79	−.001	.57
	Experiment 3	Cued Recall	Yes	Unclear	.80	−.14	.67	.56	−.001	.43
	Experiment 3	Recognition	Yes	Unclear	.78	−.07	.85	.63	−.001	.68
Cepeda et al. (2008)	Experiment	Cued Recall	Yes	High	.52	−.17	.66	.76	−.007	.78
Carpenter et al. (2008)	Experiment 1- Study-Test	Cued Recall	Yes	High	.56	−.09	.68	.80	−.014	.84
	Experiment 1- Study Only	Cued Recall	Yes	High	.61	−.10	.63	.78	−.013	.78
	Experiment 2- Study-Test (x 3)	Cued Recall	Yes	High	.44	−.07	.78	.98	−.013	.93
	Experiment 2- Study Only (x 3)	Cued Recall	Yes	High	.56	−.08	.70	.85	−.013	.84
Cepeda et al. (2009)	Experiment 1	Cued Recall	Yes	High	.57	−.11	.58	.86	−.044	.82
	Experiment 2a	Cued Recall	Yes	High	.64	−.12	.63	.76	−.004	.81
	Experiment 2b	Cued Recall	Yes	High	.61	−.19	.46	.76	−.009	.75
Kristo et al. (2009)	Content	Recall	Yes	High	.91	−.11	.96	.95	−.006	.85
	Time	Recall	Yes	High	.96	−.13	1.01	.90	−.006	.87
	Detail	Recall	Yes	High	.98	−.25	.78	.73	−.007	.58
Nunoi and Yoshikawa (2016)	Reminder learning (5 presentations)	Recognition	Yes	High	.34	−.01	.86	.98	−.002	.88
	Reminder learning (1 presentation)	Recognition	Yes	Low	.35	−.01	.75	.96	−.003	.79
	Spatial learning (5 presentations)	Recognition	Yes	High	.54	−.01	.70	.86	−.003	.74
	Spatial learning (1 presentation)	Recognition	Yes	Low	.40	−.01	.61	.98	−.003	.65
Jeunehomme, Filville, Stawarczyk, Van der Linden, & D'Argembeau (2018)	Experiment	Recall	Yes	High	.56	−.04	39[a]	.91	−.604	46[a]
The Influence of Degree of Learning										
Craig et al. (1972)	7 presentations	Recall	Yes	Some	.84	−.10	.35	.63	−.114	.65
	14 presentations	Recall	Yes	More	.73	−.06	.50	.70	−.098	.74
	21 presentations	Recall	Yes	Most	.64	−.07	.46	.86	−.107	.72

(*continued on next page*)

Table 1 (*continued*)

Publication	Data	Memory test	Meaningfully complex	Degree of learning	Power fit	Exponent	Constant	Linear Fit	Slope	Intercept
Krueger (1929)	100% Overlearning	Recall	No	Some	.87	−.87	.17	.38	−.005	.12
	150% Overlearning	Recall	No	More	.94	−.20	.37	.65	−.005	.32
	200% Overlearning	Recall	No	Most	.92	−.21	.46	.54	−.007	.39

[a] These values reflect the use of “experience units” rather than proportions.

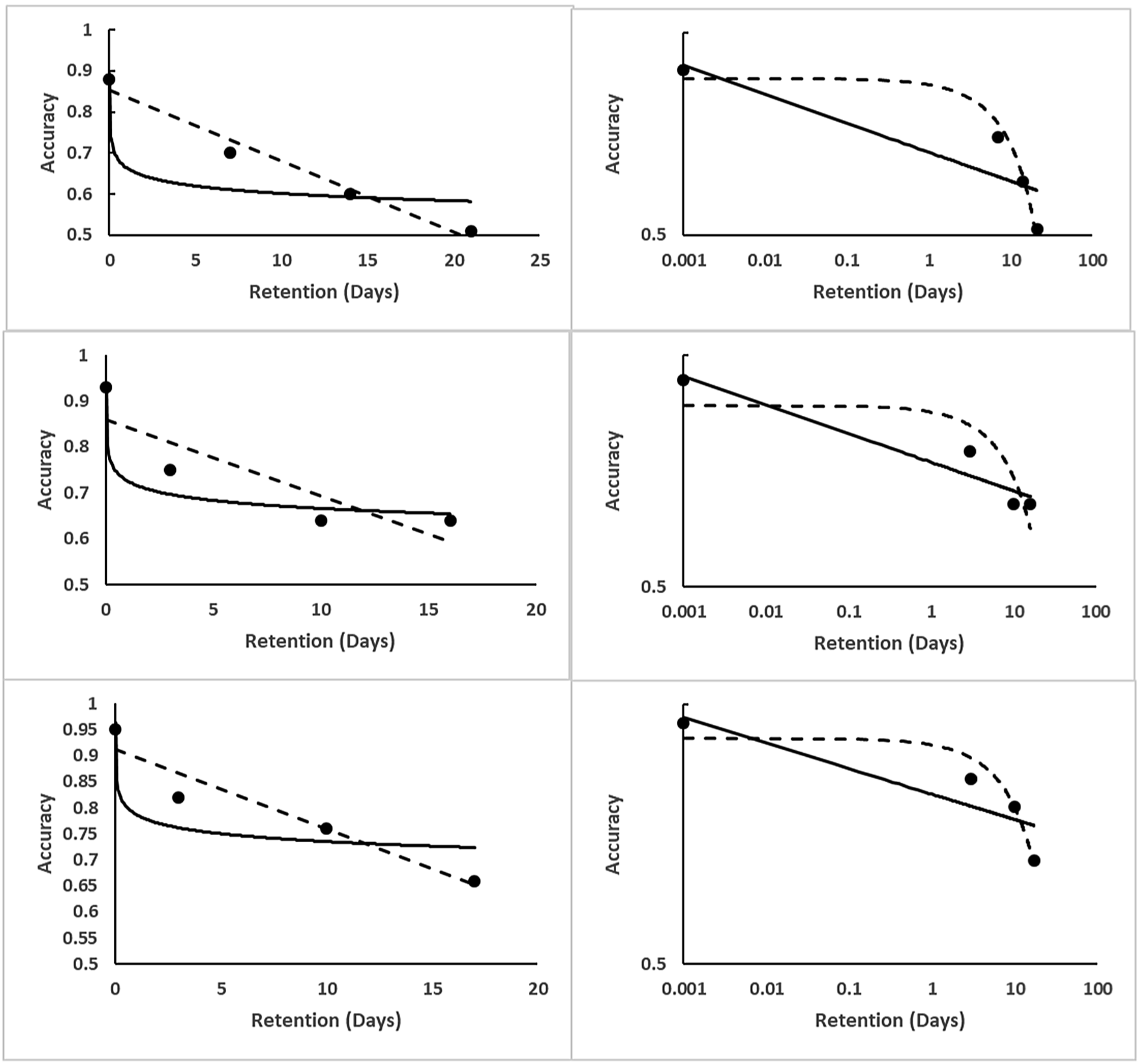

Fig. 5. Plots of Burtt and Dobell (1925) three recognition experiments with linear axes on the left, and logarithmic axes on the right. Experiment 1 is at the top, followed by Experiment 2, and Experiment 3 at the bottom. The last time point of Experiment 1 was removed because performance fell to chance by three weeks. Experiments 2 and 3 only involve the materials with no subsequent relearning at later time points.

2402 events over a period of five years. He recorded these events by providing details as to *who* was present, *what* happened, *where* did it take place, and *when* did it occur. Additionally, he recorded a critical detail to each event, the retrieval of which allowed him to measure of how well he retained the gist of the event. He probed his memory using cued recall at half-year intervals. The results for the one, two, and three cue and critical detail conditions are shown in Fig. 8. Although memory for units of information in the one, two, and three cue conditions were better fit by a power function, memory for the critical detail was not. Wagenaar ascribed memory for the critical detail as memory for the gist of the event. Consistent with our later account that linear forgetting is more likely to be observed with more meaningfully complex information, this was better fit by a linear function.

This study used cued recall for memories that were likely well-learned. Moreover, they all seemingly had a high level of meaningful complexity, although it is unclear how well the details being recalled in the one, two, and three conditions were integrated and centered in the event as a whole. In comparison, the critical detail was central to the event, and better captured the event as a whole.

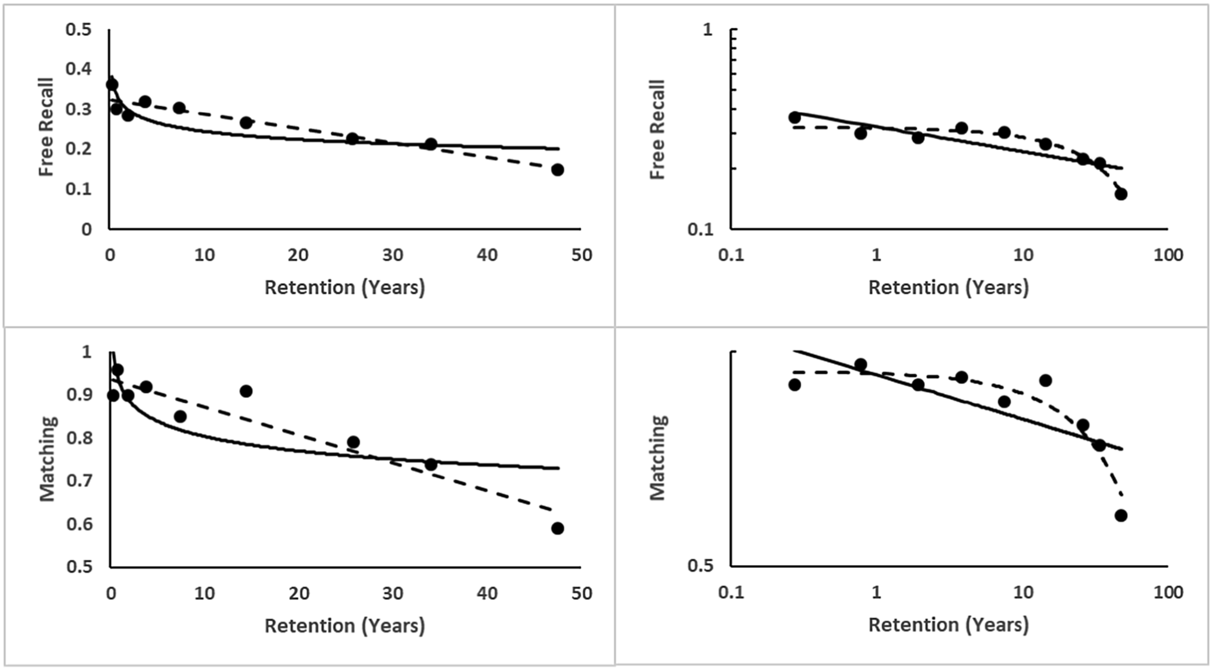

Fig. 6. Plots of the name free recall (above) and name matching (below) tasks in Bahrick et al. (1975) with linear axes on the left and logarithmic axes on the right.

Thompson et al. (1996)

In their book on autobiographical memory, Thompson et al. (1996) report a study where six participants kept a diary over a span of 900 days. At the end of this retention period, these participants were probed on specific location information (where did the event take place) or person information (who was with you). The retention patterns for these two sources of information are given in Fig. 9. For both sources, retention was better fit by a linear function.

Meeter et al. (2005)

Meeter et al. (2005) studied memory for news events for thousands of participants for a period of up to either one or two years. They did this via an Internet test in which they gave either cued recall or four-alternative forced choice recognition tests to news items from different dates. Their Experiments 1 and 2 tested memory for up to a year later. Experiment 1 had Dutch participants, and Experiment 2 had a more international sample. Experiment 3 assessed memory for news items two years later. The results are shown in Figs. 10 and 11 for recognition and cued recall tests, respectively. As can be seen, while the data are better fit by a linear function for the recognition data of Experiment 2, the rest of the data are better fit by power functions.

This study used both recall and recognition, with a better linear pattern only emerging when for recognition in only one of three experiments was used. Moreover, while the materials were complex, namely news events, many of the retrieval questions only probed for information that was not well-tied to describing the structure of the event, per se, but were bits of information such as proper names (e.g., “Which famous singer died on September, 12, 2003?”). Finally, while performance was reasonably good, the degree of learning for the news stories was not controlled, so it is unclear how well some of them were encoded by participants. Thus, these data may reflect a mixture of well-learned and poorly-learned information.

Carpenter et al. (2008)

Carpenter et al. (2008) examined the effect of memory tests on the rate of forgetting. They report three experiments, each of which span a retention period of up to 42 days. Of interest here are the first two experiments where participants learned obscure facts under differing learning conditions that included either repeated study or practical retrieval after an initial presentation of the facts. Performance was assessed using cued recall where the fact was presented in a question format with a one-word answer as the target information. The results for Experiments 1 and 2 are shown in Fig. 12. In all learning and testing conditions, these data were better fit by a linear function than a power function.

Cepeda et al. (2008)

A study by Cepeda et al. (2008) examined the effect of spaced practice on retention by having 1350 participants learn a set of 32 trivia facts. These facts were presented in a question–answer format (Phase I). This phase used a drop-out testing procedure where each fact was presented in question form (e.g., “What European nation consumes the most spicy Mexican food?”) and incorrect answers were given correct feedback (e.g., “Norway”) and recycled in the learning set. After all the facts were answered correctly, Phase I was complete. People then received a restudy session after a delay of up to 3.5 months (Phase II). Finally, they took a recall and recognition test one year later (Phase III). What is of interest here is not the final test performance of Phase III that was reported by Cepeda et al., but the unpublished data from the restudy session of Phase II. The delay between Phases I and II was either immediate, one-day, two-days, seven-days, 21-days, 35-days, 70-days, or 105-days. During Phase II, the same 32 questions were given twice through with feedback. Therefore, this performance was a cued-recall test for the answers to the question probes.

The results, shown in Fig. 13, are the accuracy proportions. These data were better fit by a linear function than a power function.

This study used cued recall for meaningful information (facts), and there was a high degree of learning.

Cepeda et al. (2009)

Similar to Cepeda et al. (2008), Cepeda et al. (2009) examined the effect of spaced practice on retention. What is of concern to us here is not the result of the final test of Phase III, but the results of the Phase II

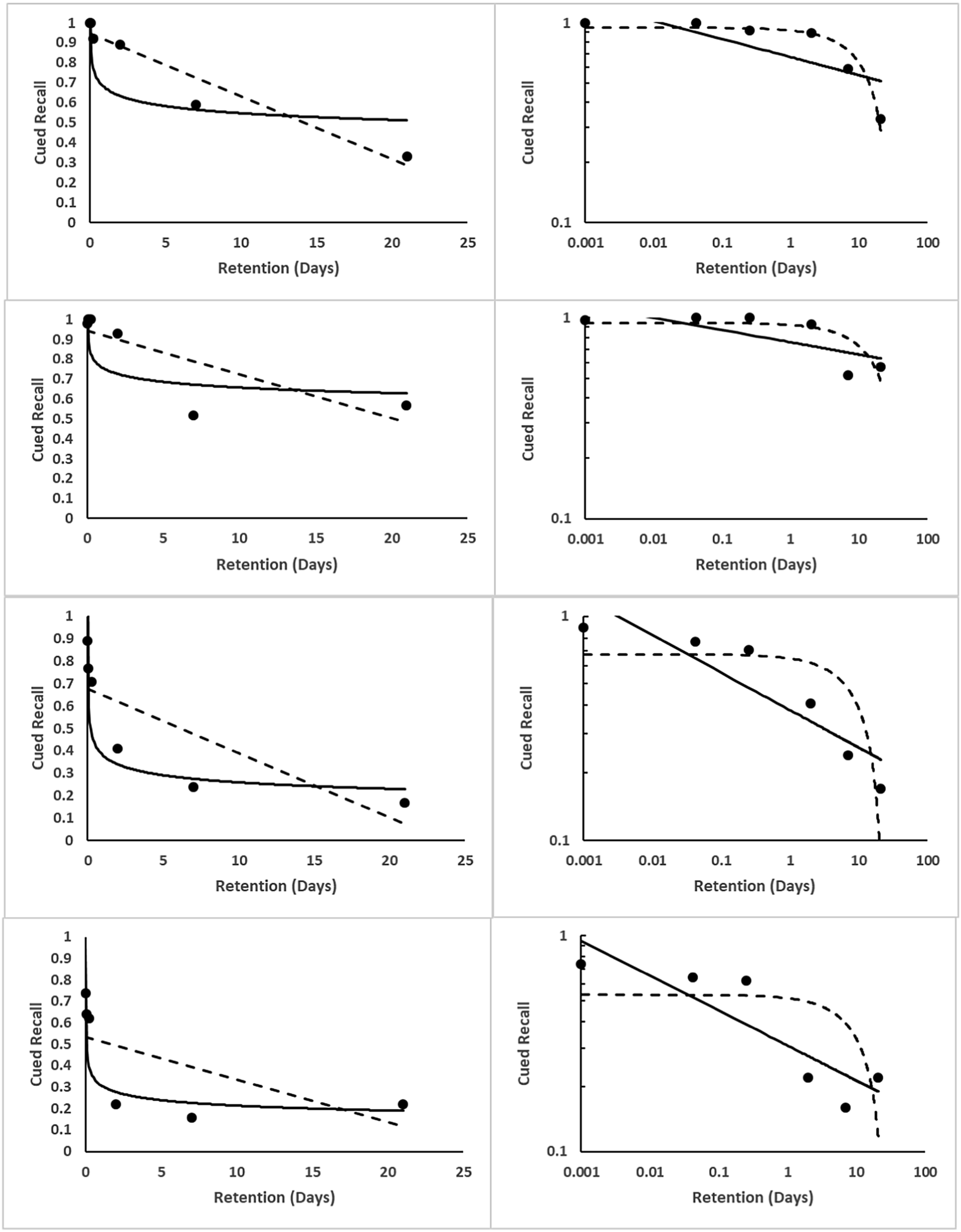

Fig. 7. Plots based on the data from Runquist (1983) Experiment 2 with linear axes on the left and logarithmic axes on the right. The top plot is of retention for items that received practice testing after three presentations. The second is for items that received practice testing after one presentation. The third is for untested items that were presented three times. The last is for untested items presented once.

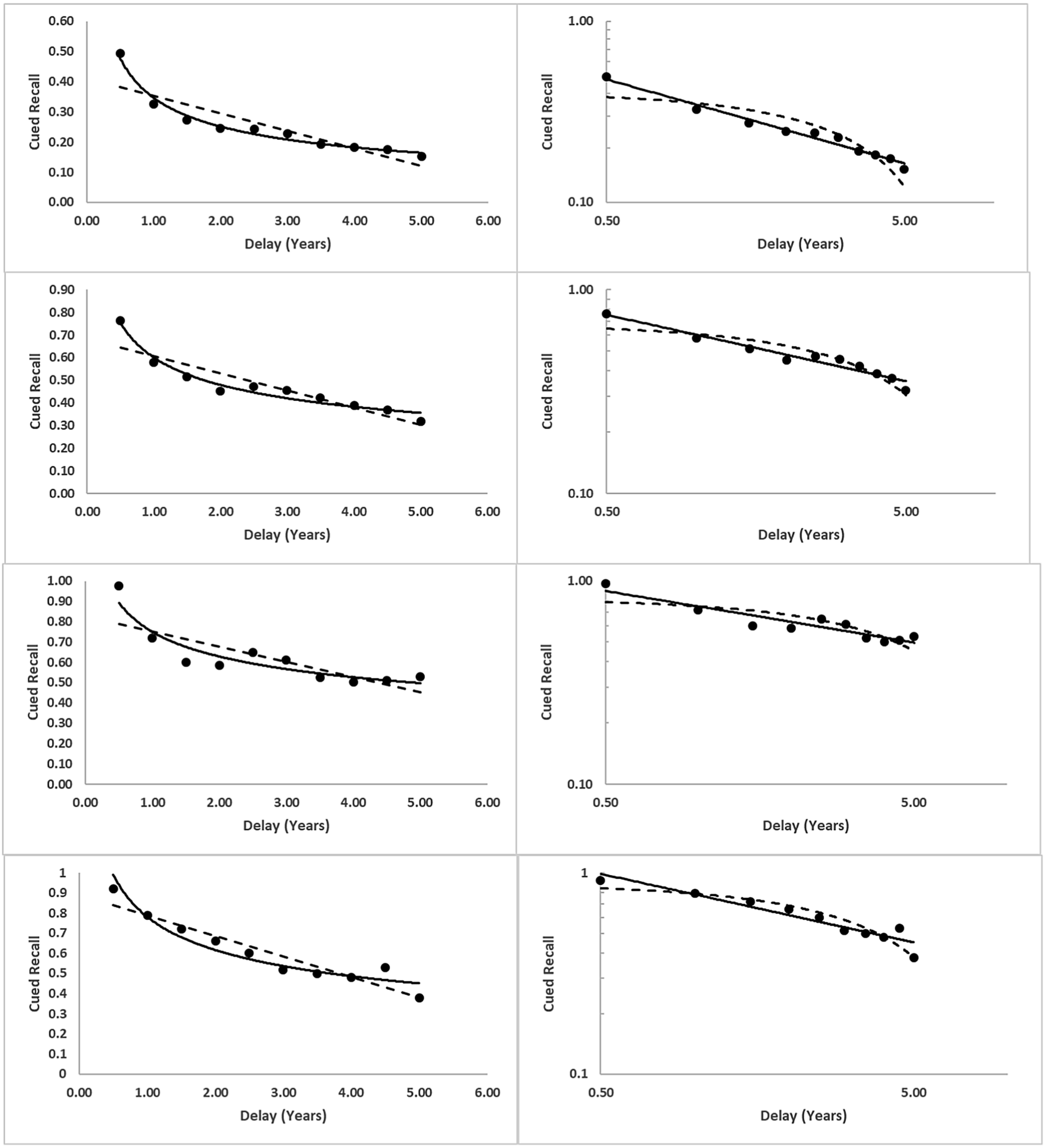

Fig. 8. Plots based on the data from Wagenaar (1986) with linear axes on the left and logarithmic axes on the right. The top plot is of retention after one cue. The second is after two cues. The third is after all three cues. The last is for the critical detail.

restudy session that occurred at varying delays after the original learning of Phase I. In their Experiment 1, Cepeda et al. had 215 40 Swahili-English word pairs. During Phase I, this involved a single presentation of each pair for seven seconds each, followed by a drop-out testing procedure in which participants were cued with a Swahili word and were asked to respond with the English equivalent. Feedback was provided for all responses. Pairs that were correctly answered twice were removed from the learning set. For Phase II, participants returned for a restudy session after 5-minutes, one-day, two-days, four-days, seven-days, or two-weeks. During this session, they were again given a cued-recall test with feedback.

For their Experiments 2A and 2B, Cepeda et al. (2009) had two sets of materials for two different parts. Part A used 23 facts in a question-answer format, and Part B used 23 photographs of uncommon objects associated with a fact for each. The Phase I learning session consisted of a pretest, an initial exposure, and three blocks of tests with feedback. The Phase II restudy session consisted of two blocks of tests with feedback. The intervals between Phases I and II were five-minutes, one-day, seven-days, 28-days, 84-days, or 169-days.

The results of these experiments are shown in Fig. 14. Experiments

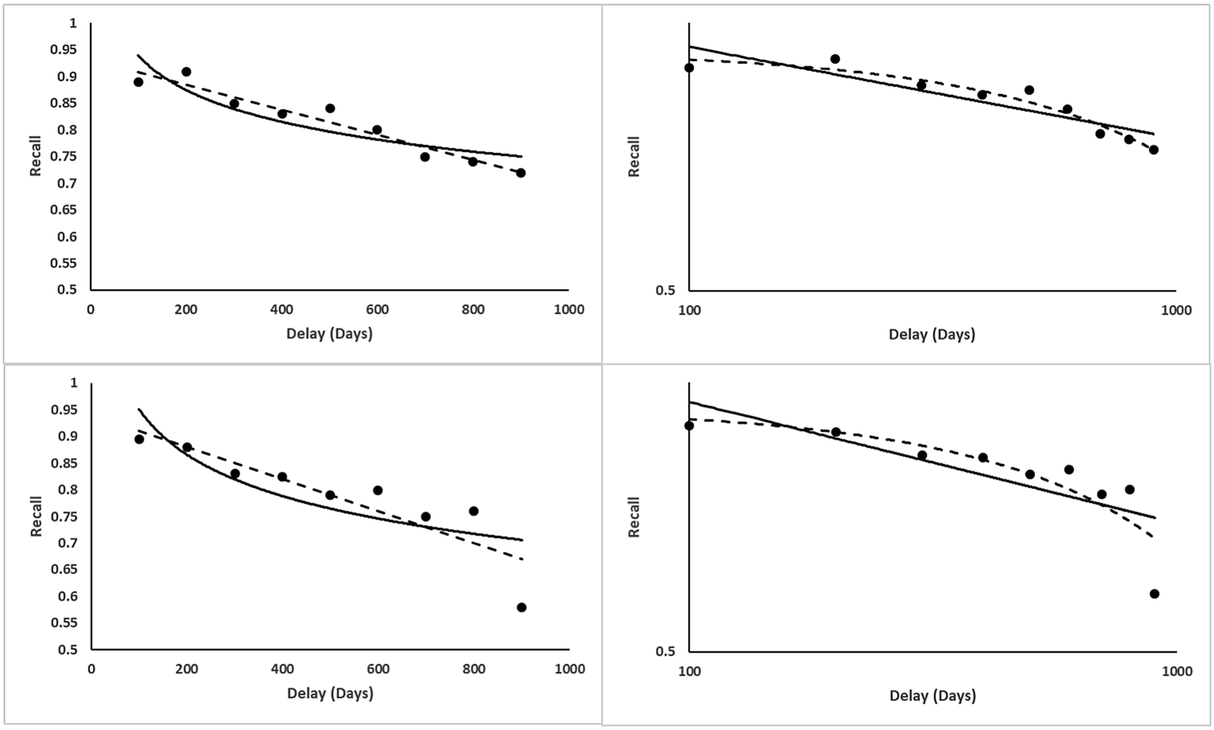

Fig. 9. Plots based on Thompson et al. (1996) with linear axes on the left and logarithmic axes on the right. The top plot is for the *location* information. The bottom plot is for the *who with* information.

1, 2A, and 2B were better fit by a linear function than a power function.[4] This study used cued recall for paired associates of a word and its translation. Thus, both words referred to the same meaning. Moreover, these pairs were learned to a very high level.

Kristo et al. (2009)

Kristo et al. (2009) examined long-term memory for various types of autobiographical information. They used an online population of 878 people assigned to one of five retention groups: two, seven, 15, 31, or 46 day retention intervals. Participants first logged the details of a single recent personal event. Then, after a retention interval, they were asked to recall the details of that one event.

The results of this study are shown in Fig. 15. The qualities of the autobiographical memories that were better fit by a linear function were memory for the content of the event (who, where, and what). In comparison, memory for time (time of day, day of week, day of month, and month) and other details are better fit by a power function. This may be due to a better integration of the content information into a representation of the meaning of the whole event compared to the time or detail information. Information that is better integrated is better fit by a linear function, while information that is less well integrated is better fit by a power function.

This study used recall for a meaningfully complex set of information (an autobiographical event). Moreover, given that this was an autobiographical event selected by the participants, it was likely well-learned.

[4] Note that retention reached floor in Experiment 2B by 84 days and remained at floor at the 168-day time point, thus, we removed the data from the 168-day point from Table 1. When we include the 168-day data, the pattern of retention and forgetting artificially appears more curvilinear than linear, simply because the data had reached floor at that point. When this extra time point is included for Experiment 2B, the data are similarly well-fit by linear, $m = -.005*t + .68$, $r^2 = .65$, and power functions, $m = .41*t^{-.23}$, $r^2 = .66$.

Nunoi and Yoshikawa (2016)

Nunoi and Yoshikawa (2016) examined the retention of pictures of 80 novel objects for a period of up to six weeks. During learning, people were presented with the objects individually for one second each, and were tasked with judging either the spatial position of the object or coming up with an idea of what that object reminded them of. Some of these objects were given five presentation-judgment tasks in a row, whereas others were just given just one. Although the main aim of this study was an effect of levels of processing on the preference ratings, it also tested recognition memory at time intervals of immediate, one-day, one-week, and six-weeks. The results, shown in Fig. 16, showed that the data were better fit by a linear than a power function.

For this study, used recognition memory testing for knowledge that was meaningfully complex by combining an image with either spatial location knowledge or an event that a person was reminded of. Furthermore, items with five-presentation conditions had a greater level of learning.

Jeunehomme et al. (2018)

Finally, a study is by Jeunehomme et al. (2018) investigated the rate of temporal compression in episodic memory over time as well as other factors (e.g., goal processing). What is of concern here is that they also measured the number of episodic details recalled across time. Their learning phase consisted of 32 participants in four delay groups (128 total) walking around a university campus and doing various activities at various locations while wearing a small camera. They assessed participants' free recall of the components for these events (e.g., person, object, thought) either immediately or after a one-day, one-week, or one-month delays. A plot of the mean number of these components remembered at each delay is shown in Fig. 17. These data were better fit by a linear function than a power function.

This study used recall to assess memory for meaningfully complex information (encountered with real world events) in which the learning level was higher than a single brief presentation.

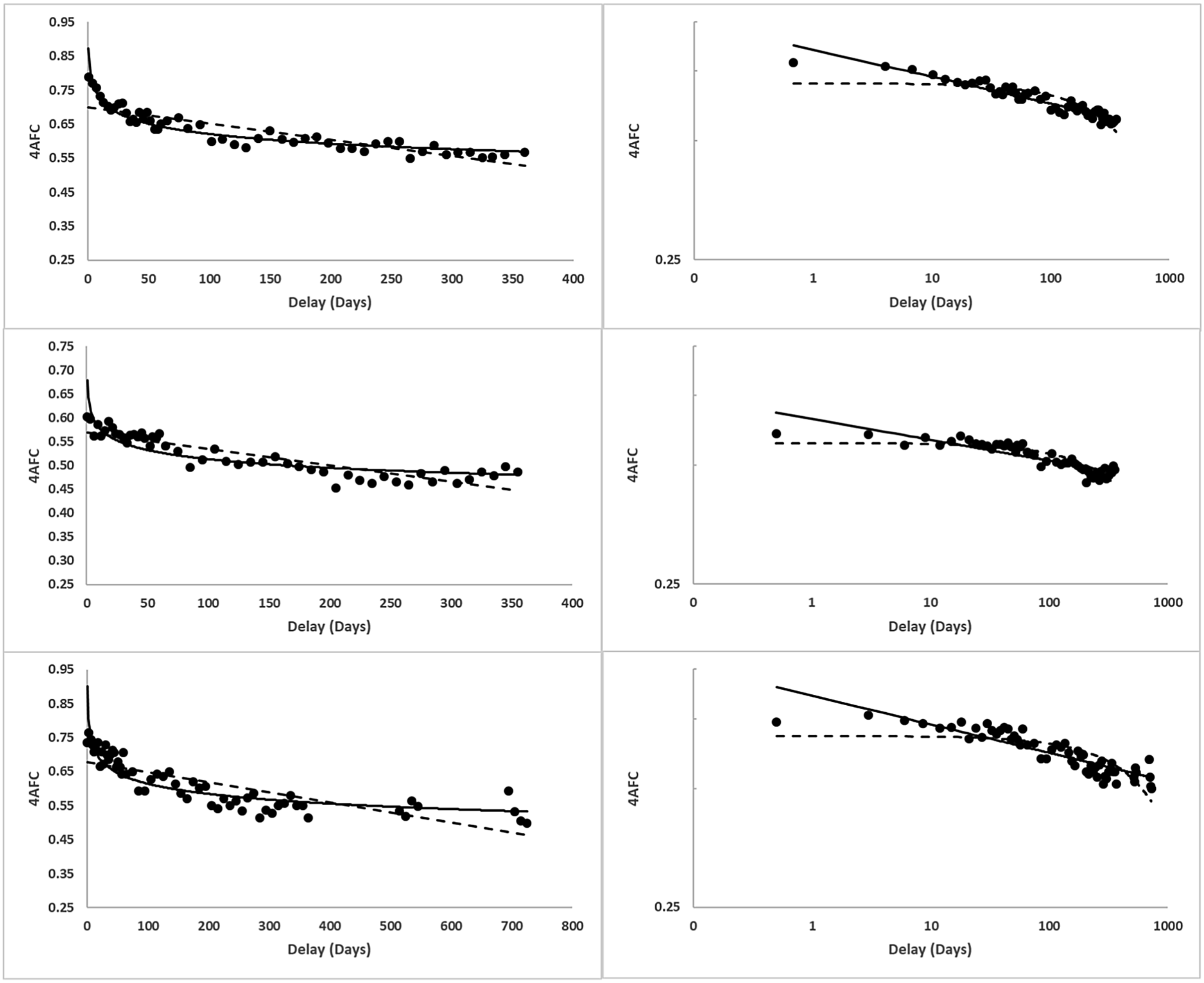

Fig. 10. A graph based on the four alternative forced-choice recognition data from Meeter et al. (2005) with linear axes on the left and logarithmic axes on the right. The top plots are from Experiment 1. The middle plots are of Experiment 2. The bottom plots are of Experiment 3.

Evaluation

In sum, the studies reviewed above all show linear forgetting in some or all of their experiments. It should be noted that the studies that meet our criterion for linear forgetting make up a minority within the retention literature. Moreover, these studies differ in various respects. The question at this point is whether there are methodological factors that create this retention pattern. The factors of memory test type, degree of learning, and meaningful complexity, raised earlier, and each considered in turn.

Memory Test Type. One issue that we aimed to address was whether the observation of linear forgetting was limited to a certain type of memory retrieval task. In general, this does not seem to be the case. Specifically, linear forgetting was observed in studies that used free recall (Bahrick et al., 1975; Jeunehomme et al., 2018), cued recall (Carpenter et al., 2008; Cepeda et al., 2008; Cepeda et al., 2009; Kristo et al., 2009; Runquist, 1983; Thompson et al., 1996; Wagenaar, 1986), and recognition (current study; Burtt & Dobell, 1925; Meeter et al., 2005; Nunoi & Yoshikawa, 2016). Thus, this finding is not limited to a particular type of retrieval task.

That said, it also appears to be the case that it is more likely to be observed with some tasks compare to others. For example, with the data from the Burtt and Dobell (1925) and Meeter et al. (2005) studies, a linear forgetting function was observed with recognition data, but not for with cued recall data. In comparison, for Bahrick et al. (1975), although linear fits were better than power fits throughout, there are better linear fits with matching and recall tasks than with recognition. Thus, while the finding is not limited to certain types of memory tasks, it does appear to be the case that the demands of the task at retrieval can influence the degree to which a linear forgetting pattern is observed.

Degree of Original Learning. Another factor that could influence whether linear forgetting is observed is the degree of original learning. Specifically, in the current experiments, people needed to memorize the materials to a criterion, rather than a single exposure. Likewise, in many of the studies reviewed above, learning was greater than a single exposure and often involved retrieval practice. The exceptions are the Wagenaar (1986), Thompson et al. (1996), Kristo et al. (2009), and Jeunehomme et al. (2018) studies in which the memoranda involved autobiographical experience outside a laboratory which are clearly better remembered than laboratory materials. In these cases, the richness of the real-life events constitutes high learning. Meeter et al. (2005) probed for information that participants had encountered in news stories at some time in the past, although it is unclear how many times their participant encountered a given story.

In a laboratory setting, there are a number of ways that the degree of original learning could be increased, including (a) the duration and number of exposures to the materials, (b) retrieval practice, and (c) overlearning after reaching a criterion level of memorization. All of the laboratory learning studies covered included at least one these ways of increasing the degree of original learning.

The value of degree of original learning to observing linear

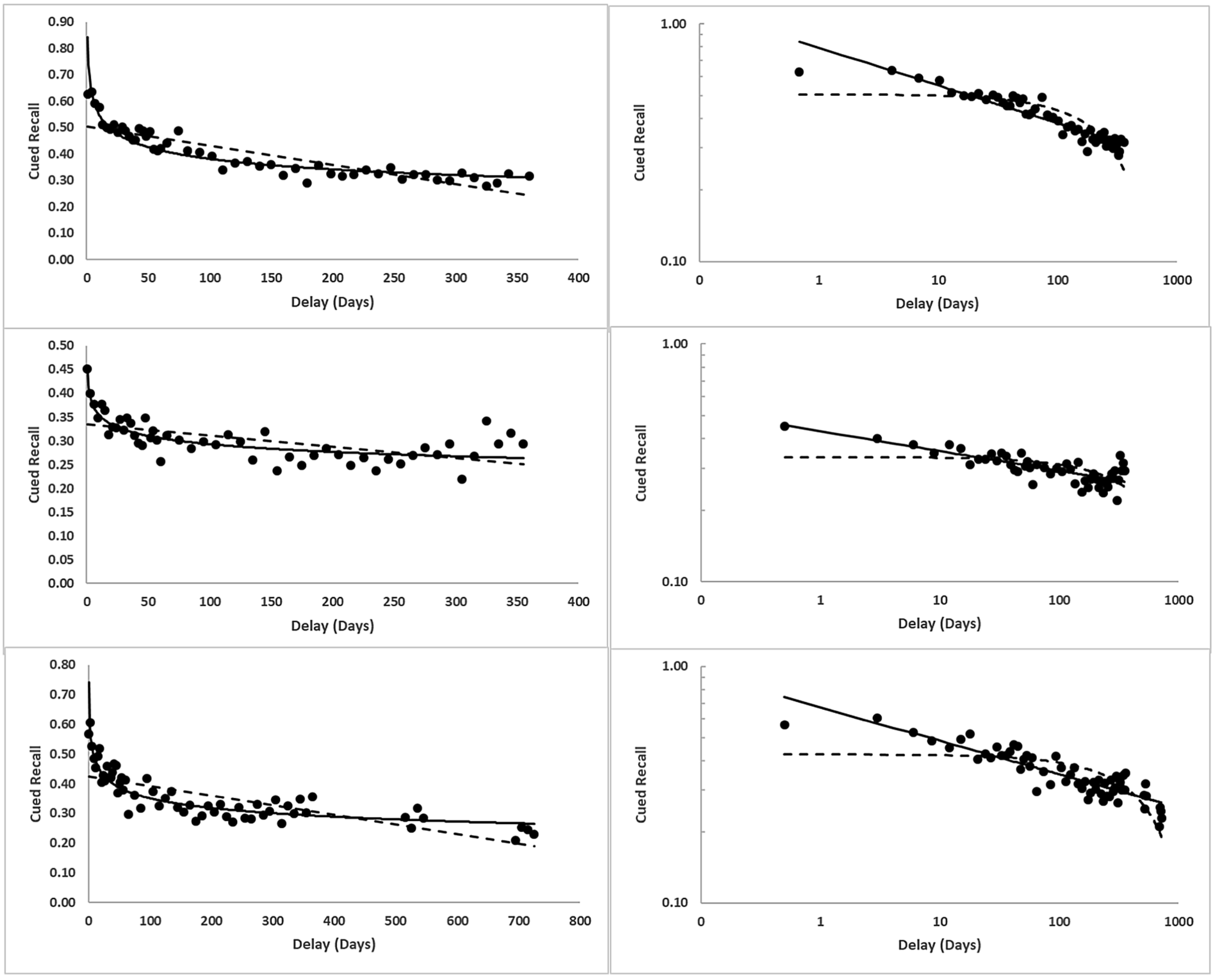

Fig. 11. A plot based on the cued recall data from Meeter et al. (2005) with linear axes on the left and logarithmic axes on the right. The top plots are from Experiment 1. The middle plots are of Experiment 2. The bottom plots are of Experiment 3.

forgetting is clearly illustrated in a study by Craig, Sternthal, and Olshan (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 then tested free recall of the brand names either immediately, one-day, one-week, or four-weeks later. As shown in Table 1 and Fig. 18, the retention pattern became more linear as the number of repetitions increases.

While an increased degree of learning appears to be important for linear forgetting to occur, is it sufficient? The answer seems to be "no". For example, a study by Krueger (1929) used word lists and had levels of overlearning of 100%, 150%, and 200%. He tested recall at one-day, two-day, four-day, one-week, two-week, and four-week intervals. As Table 1 and Fig. 19 show, these data were better fit by a power function than a linear function regardless of the degree of learning. What may matter is when a higher degree of learning is coupled with materials that are more meaningfully complex materials that involve associations of some type to more readily allow for elaborative processing.

Meaningful Complexity. The final factor that we consider here that could lead to a linear pattern of forgetting is the meaningful complexity of the materials. In our work, as well as many of the studies reviewed here, the materials and retrieval task involved some meaningful complexity (e.g., paired associates or sentences). In the current Experiments 1 and 2, we used sentences involving a meaningful combination of two concepts (ether objects and locations or people and activities), whereas for Experiment 3, we used paired associates that could be readily meaningfully combined (animals and locations). Burtt and Dobell (1925) and Runquist (1983) used paired associates in which the two meaning of the words needed to be combined. Similarly, the classmate names from Bahrick et al. (1975) can also be viewed as paired associates (a first name paired with a last name). Carpenter et al. (2008), Cepeda et al. (2008), and Cepeda et al. (2009) also used items that were questions and answers, associated facts and objects, or English-Swahili pairs.[5] The Nunoi and Yoshikawa (2016) study involved elaborating on items with either a spatial position or a prior memory. Wagenaar (1986), Thompson et al. (1996), Kristo et al. (2009), and Jeunehomme et al. (2018) assessed memory for real-life events. Thus, when participants can more easily elaborate on and integrate meaningful and complex materials, rather than just simple items (such as a nonsense syllable or word), then linear forgetting is more likely to be observed.

[5] We would not normally expect a pattern of linear forgetting for word-translation pairings, such as English-Spanish pairings. This is because although there are two words, there is a single meaning. As such, the likelihood of meaningful complexity is reduced. In fact, data from studies that use word-translation pairs are better fit by power than linear functions (e.g., Bahrick, 1984; Wickelgren, 1972). The reason for the observation for linear forgetting for the English-Swahili pairs may involve the very high level of intense learning involved, which may have led some participants to engage in elaborative inference making based on the phonological structure of the Swahili words.

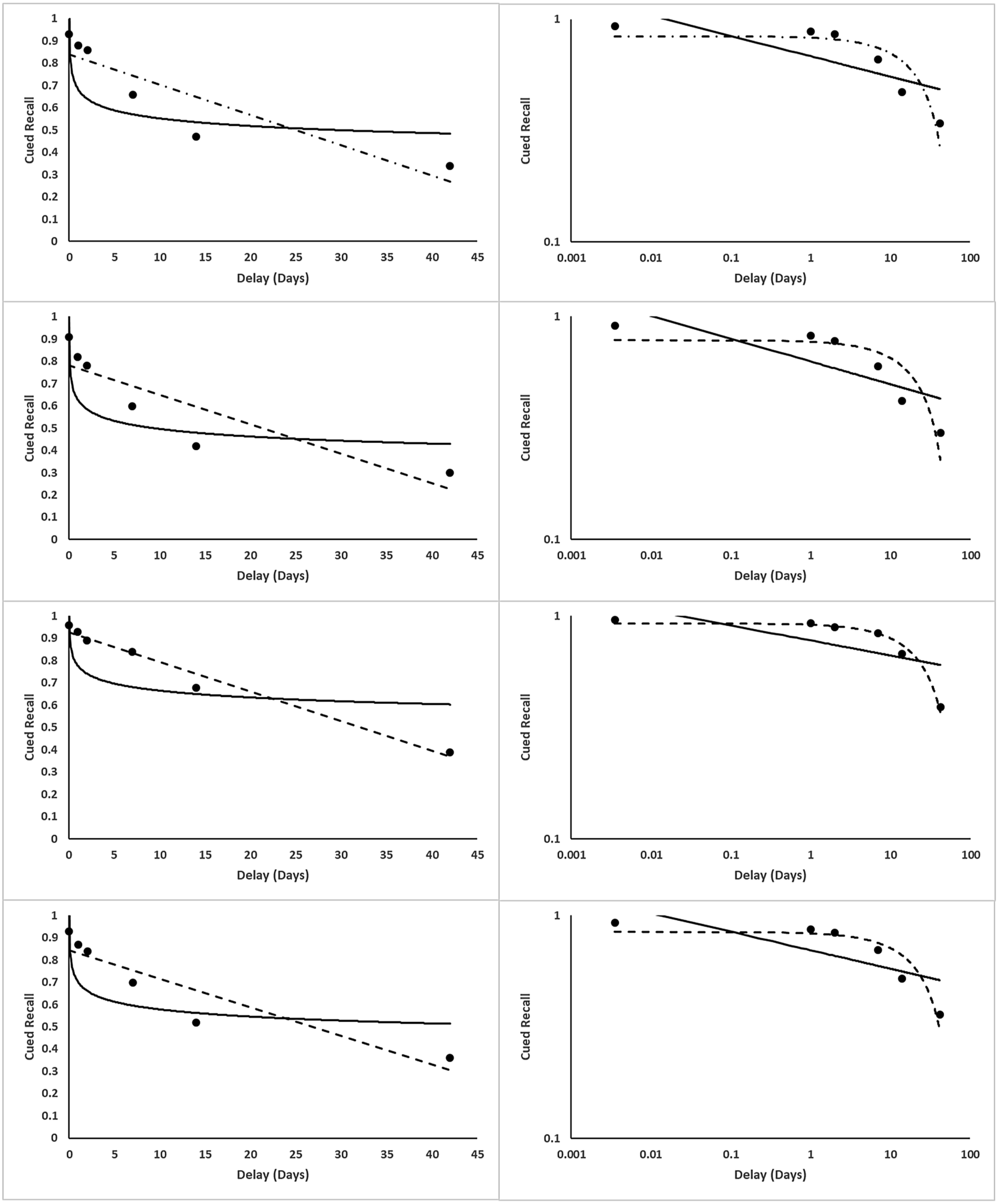

Fig. 12. Plots based on the cued recall data of Experiments 1 and 2 of Carpenter et al. (2008) with linear axes on the left and logarithmic axes on the right. The top plot is Experiment 1 study-test. The second is Experiment 1 study only. The third is Experiment 2 study-test (3 times). The bottom is Experiment 2 study only (3 times).

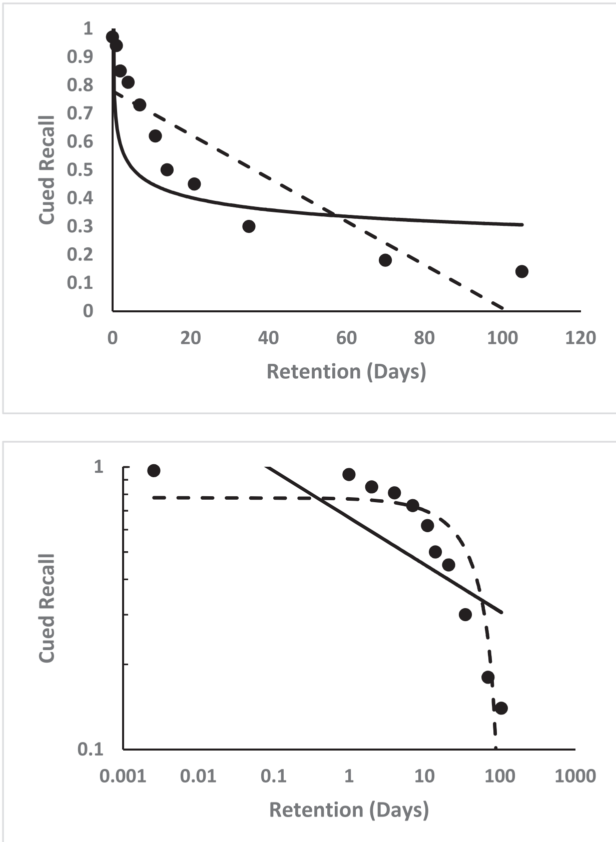

Fig. 13. A plot based on the cued recall data of the restudy session in Cepeda et al. (2008) with linear axes above and logarithmic axes below.

Assessment of factors that influence the observance of linear forgetting

Overall, it appears that the observation of linear forgetting can be influenced by the type of memory test used, although it is possible to observe it with different kinds of tests. It also appears to require well-learned information, and the information needs to be meaningfully complex and the retrieval task need to tap into the integrated larger representation. When some or all of these factors are not at play, linear forgetting may not be observed. For example, when recall but not recognition was used, linear forgetting was not observed in the Burtt and Dobell (1925) and Meeter et al. (2005) studies. Linear forgetting is also not observed when the materials were not well-learned, as is clearly illustrated by the Krueger (1929) data. This may also contribute to the absence of linear forgetting in two of the three Meeter et al. studies given that it is unclear how well learned the news stories were to begin with. Finally, the information needs to be meaningfully complex and this needs to be assessed by the retrieval task. When some of all of these elements are no present, then linear forgetting patterns may not be observed.

For example, a flashbulb memory study by Weaver and Krug (2004) assessed memory two-days, one-week, one-month, three-months, and one-year after the attacks on the Pentagon and World Trade Center on September 11, 2001, and their data are better fit by a power function (r^2 = .94) than a linear function (r^2 = .64). While the memory of the event was certainly meaningfully complex, (a) they used free recall, which seemed to reduce the probability of observing linear forgetting, (b) the experience tested was the context of learning the news rather than the critical event itself, and so was likely less well learned at first, and (c) they probed for information that may not have been well integrated with memory for the event (e.g., "What was the exact time when you heard about the attacks on the World Trade Center?" and "Describe in as much detail as possible the clothes you were wearing at the time you heard the news."). Overall, while there are some elements for the observation of liner forgetting here, there are some aspects of the study that work against it.

Another point that needs to be address is why there are cases where linear and power function fits are both relatively high. Is there actually just one type of function that fits all data? Probably not. One possibility is that there is a mix of the functions in a given data set arising from variations in the nature of the materials and the participants. Another important one is the relationship between linear and power function fits differs when the rate of forgetting are steep or shallow. Fig. 20 shows linear and power function data and how well they are fit by power and linear functions, respectively, when the rate of loss is steep or shallow. When the rate of forgetting is steep, the distinction between the two is more pronounced. However, as the rate of loss becomes shallower, the fit of the inappropriate function becomes greater. Thus, when the rate of forgetting is shallow, it becomes increasingly likely that the pattern could be well-fit by both types of functions.

For the large part, the evidence from our own work, as well as our survey of the literature, generally supports our idea that well-learned, meaningfully complex materials yield linear forgetting over long lasting periods of time. If we consider the degrees of fit for linear and power functions, for well-learned and more meaningfully complex materials, we observe either a good linear fit and a poor power function fit, or a high fit for both. Thus, we have not found any studies that have higher degrees of learning and meaningful complexity that are not discussed in this manuscript. We rarely observe a good power function fit and a poor linear fit. However, good power function and poor linear fits are observed when the conditions for observing linear forgetting are not met. Thus, we are confident that linear forgetting patterns are stable and a reflection of important memory processes.

Theory and simulation

At this point, two factors seem to be jointly important to produce linear forgetting: a sufficient degree of learning and memory complexity. Now the questions is why is this the case? To address this, we developed a theoretical account that can explain such a pattern of data, and we developed a simulation as a proof of concept of this theoretical account. This approach incorporates the ideas that this pattern of retention requires a higher degree of learning and that the memory traces be at least somewhat complex.[6] This same simulation is reported in Fisher and Radvansky (2018) to account for different phases of memory retention in which there is an initial long-term memory phase in which memory stabilizes somewhat, followed by an increase in the rate of forgetting somewhere around seven days. This is why a shift in the retention and forgetting function appears in some of the simulated data patterns.

Retention assumptions

As noted in the introduction, a linear pattern is not immediately explained by current theories of retention and forgetting. By itself, the idea that information is forgotten at a linear pattern is unsatisfactory because (a) it goes against over a century of forgetting data that largely show negatively accelerating functions, and (b) there is no explanation for why the rate of forgetting would increase over time. Therefore, rather than simply assuming that memory trace information is lost at a linear rate, our theory and simulation (a) incorporates standard negatively accelerating forgetting functions for information and (b) does not have increased rates of forgetting over time. In other words, the loss of information in our theory and simulation is entirely consistent with the common assumption of retention.

[6] This simulation can be found at http://ec2-52-204-56-150.compute-1.amazonaws.com/pilot/RetentionSimulationAWS.html.

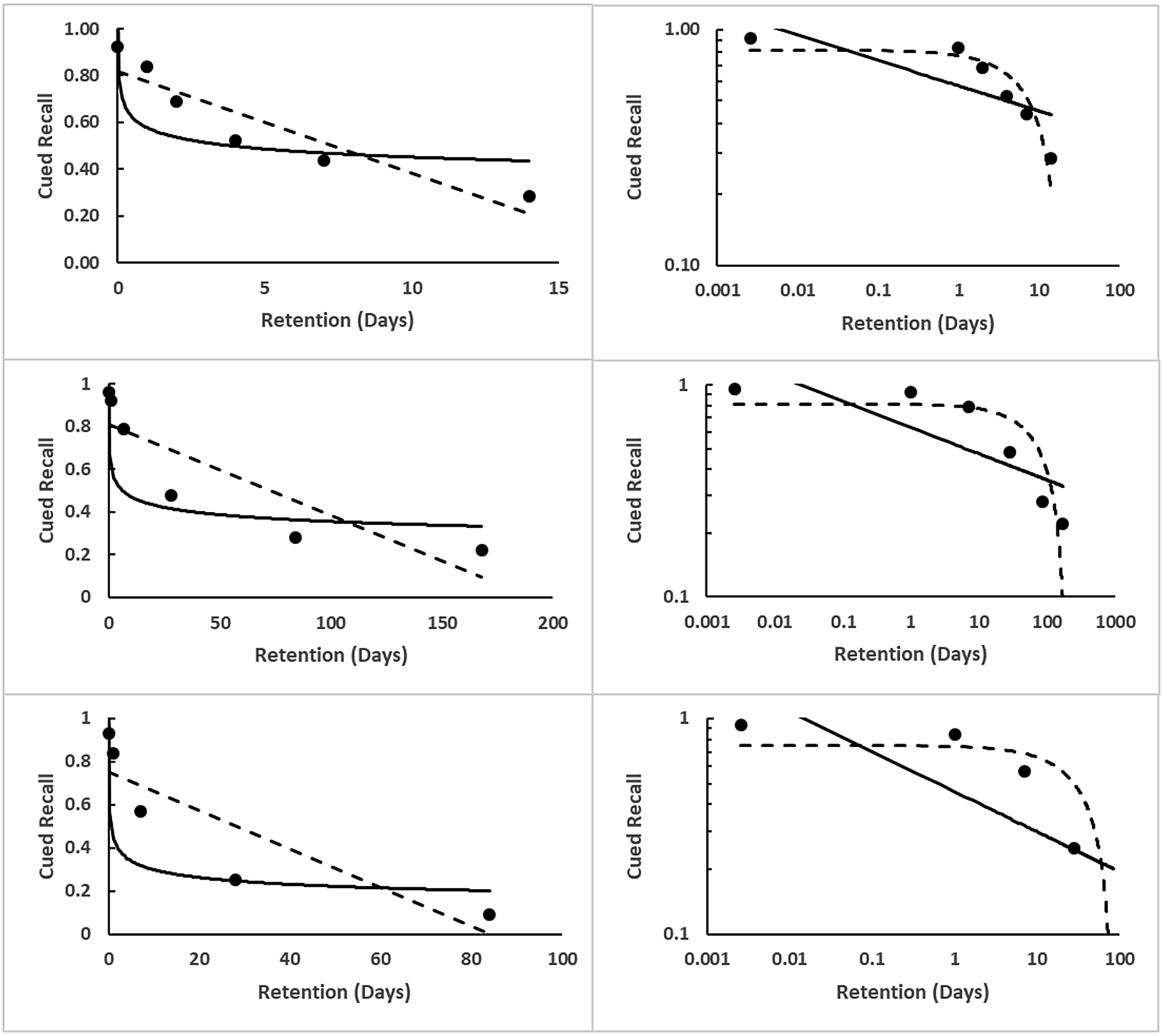

Fig. 14. Plots based on the cued recall data of the restudy session in Cepeda et al. (2009) with linear axes on the left and logarithmic axes on the right. The top plot is for their Experiment 1. The middle plot is Experiment 2a. The bottom plot is Experiment 2b.

For our theory and simulation, what brings about linear forgetting is this standard forgetting of memory trace components along with an assumption that memory responses can be made using a partially degraded memory trace. This can be done through some sort of partial matching or reconstructive process.

Simulation characteristics

A brief summary of the simulation is provided here. More details are provided in Appendix B.

Memory Traces. The simulation assumes that each item was stored as a vector of individual components (e.g., Hintzman, 1984). These components may be either (a) encoded from the environment or (b) inferences drawn from general world knowledge. The precise nature and organization of these components is not important here.[7] What is important is that any target memory trace can be decomposed into individual components. For the sake of simplicity for the simulation, these components are assumed to all have equal strength. That said, the theory itself is neutral with regard to the precise nature of the components that make up a memory trace.

Component Forgetting. For individual components, the theory and simulation assumes that their decay pattern follows a standard exponential function, consistent with the idea that the loss of individual pieces of information may follow an exponential function, but their aggregate pattern may reflect a power function given enough variance among the items (Anderson, 2001; Murre & Chessa, 2011). However, the rate at which each of these components is forgotten can differ. Thus, there is a constant rate of forgetting for each component. In this way, the simulation encapsulates a negatively accelerating forgetting function that is the norm in theories of forgetting. This is avoided in the simulation by allowing for the occurrence of partial matching or reconstruction to occur during retrieval.

Partial Matching/Reconstruction. A critical aspect of this theory and simulation is the idea that a memory is not completely forgotten if some components are not accessible. Instead, we assume that if a sufficient number of components are present, then an accurate response can be made. This is done either through a sufficient partial match and/or a sufficient reconstruction of the memory trace to the point where a

[7] We cannot match the number of components in the simulation to the number of components in the memory traces for the sentences used in the experiment. How does one accurately count the number of components? As the number of critical words? The number of words? The number of syllables? The number of letters? The number of evoked semantic concepts? The number of evoked sensory/motor experiences? And so on.

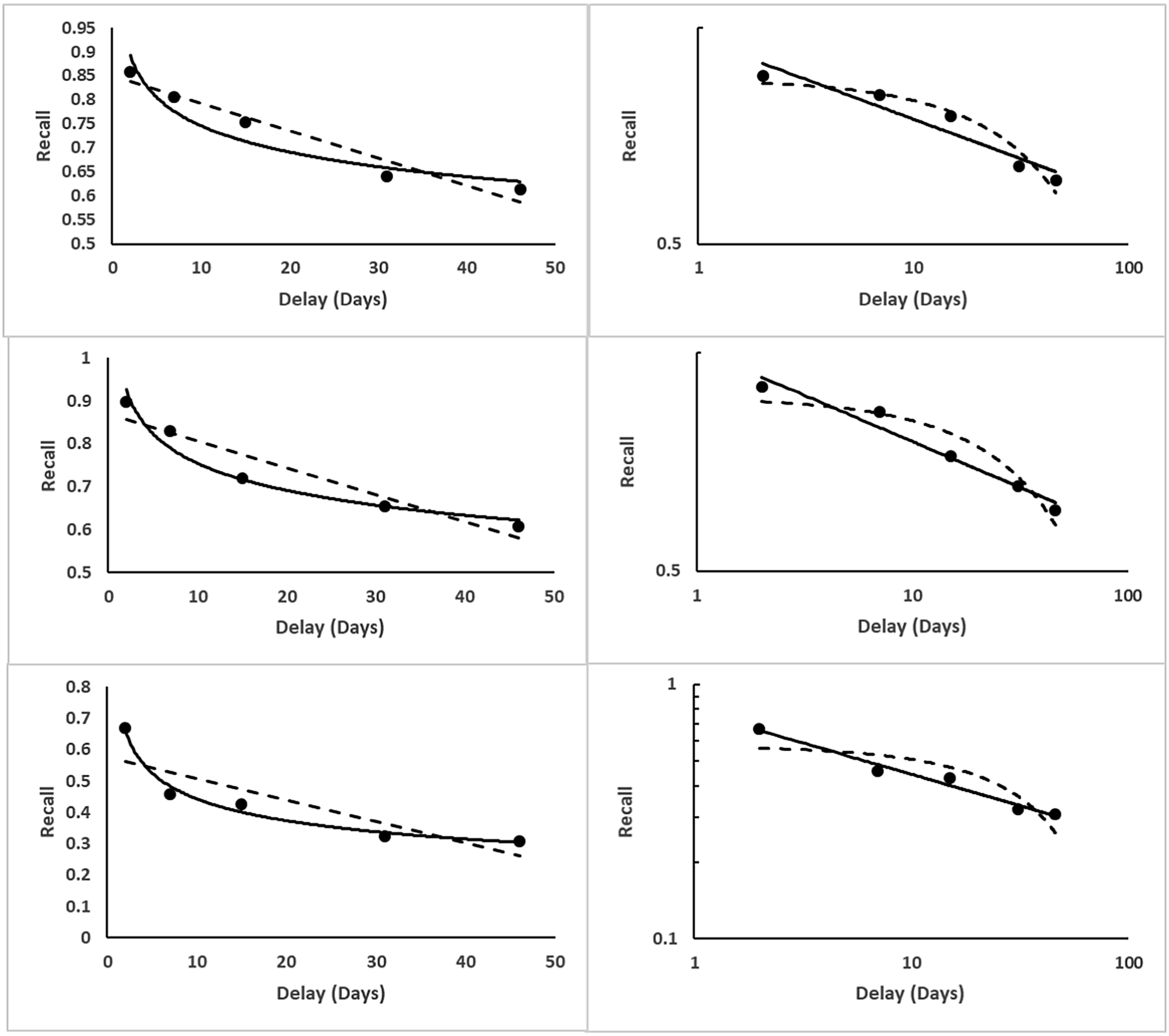

Fig. 15. Plots based on the data from Kristo et al. (2009) with linear axes on the left and logarithmic axes on the right. The top plots are based on the content of the memory (who, what, where). The middle plots are based on the time of the memory (day of week, time of day, day of month, month). The bottom plots are based on the details (important, unimportant).

person is able to produce an accurate response. The specific process of this partial matching or reconstruction is beyond the concern of this theory and simulation. What is of concern here is only that there is a process that allows an accurate memory response to be made based on a memory containing only a portion of its original constituent components.

Memory Storage. Although not specifically relevant here, as reported by Fisher and Radvansky (2018), our theory and simulation also assume that there are two general memory stores. Following McGaugh (2000), these are long-term memory (LTM) and long-lasting memory (LLM). LTM allows for the temporary consolidation of knowledge for a period of time (i.e., phase I), whereas LLM is for the extended storage of memories after the retention of information in LTM begins to wane (i.e., phase II). In the simulation, memory traces are first encoded into LTM. During phase I, there is a probability that LTM trace components will be consolidated and therefore be immune to forgetting while the knowledge is in LTM. Over time, LTM trace components may be transcribed to LLM, or forgotten. One in LLM, these components may also be forgotten.

This characteristic of memory having two retention stores is not particularly relevant for the understanding of linear forgetting. However, it is important to note that this characteristic of the simulation influences its overall pattern. Specifically, although average retention may show a strong power fit within both phase I and phase II, the overall pattern may have a weaker power fit due to differential forgetting rates between the two phases. This idea that the rate of forgetting can differ due to a transition between two phases is argued elsewhere (Fisher & Radvansky, 2018). Of note, the majority of retention studies to date are restricted in their range so that they only can capture the retention either within phase I (i.e., these studies do not have sufficient time points beyond a week) or phase II (i.e., these studies do not have sufficient time points before a week).

Next, we report the outcome of three simulations. The first is a demonstration of linear forgetting when there is exponential forgetting of components and the possibility of partial matching and/or reconstruction from a degraded memory trace. After this, we report two simulations in which there is either low levels of learning or low complexity, which, consistent with what we have found in the

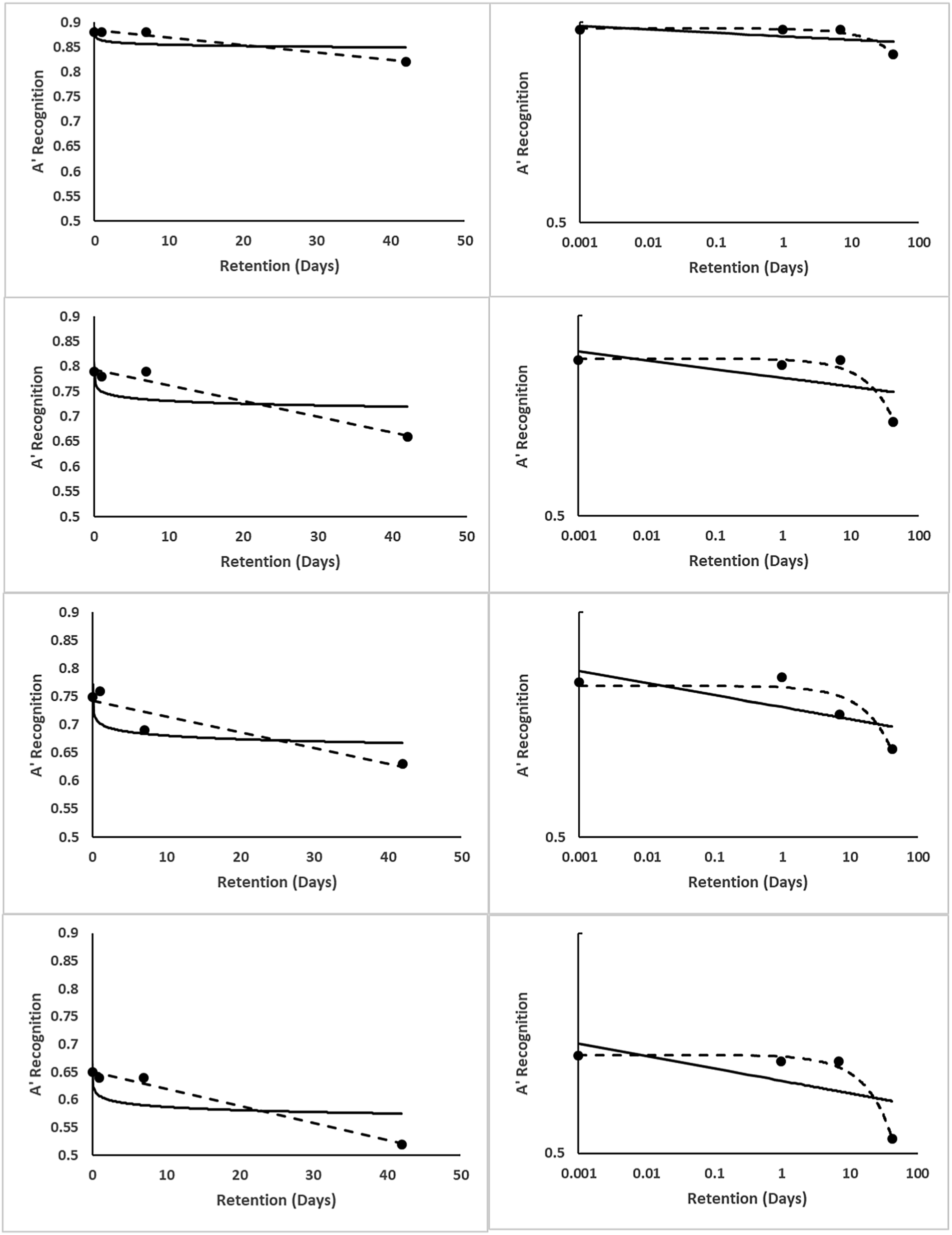

Fig. 16. Plots of Nunoi and Yoshikawa (2016) data with linear axes on the left and logarithmic axes on the right. The top plot is of the recognition score for the group who leaned the items with deep processing and five presentations. The next plot is of deep processing with 1 presentation. The third plot is of shallow processing with five presentations. The final plot is of shallow processing with one presentation.

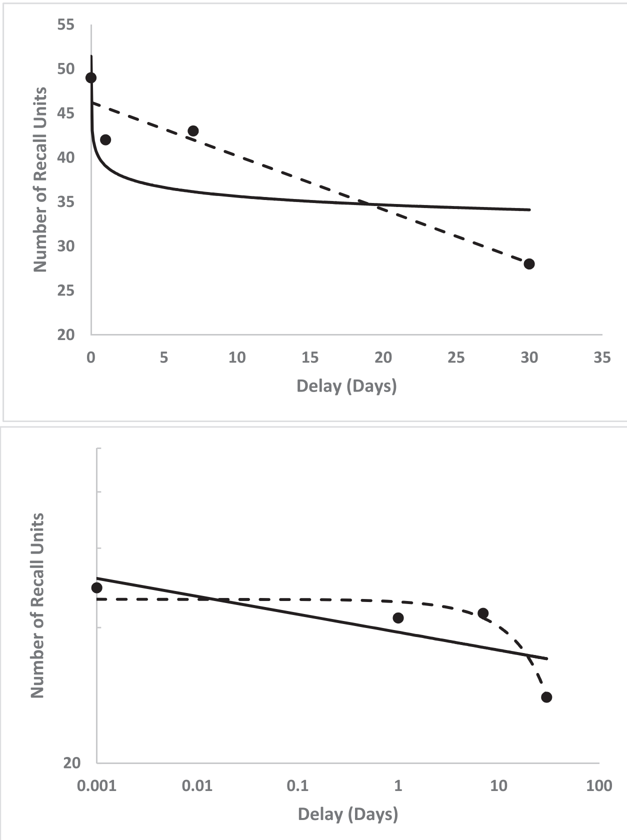

Fig. 17. A plot of the data from Jeunehomme et al. (2018) with linear axes above and logarithmic axes below.

literature, should show more evidence of negatively accelerating forgetting.

Linear forgetting simulation

We first simulated a linear pattern of forgetting. This was done by simulating 48 participants, each retaining 20 memory items. We specified a .98 probability that each item would be encoded into LTM. The idea was that the study sentences in our experiments were very well learned, although small errors may have occurred. Moreover, we assumed a set of ten environmental components for each memory along with another ten potential inference components. The idea was that there were likely a number of different components that could be represented from study sentences, and because these referred to an event, they allowed for the possibility of drawing of a number of inferences. We set a .50 ± .05 probability for each of these potential inference components to be drawn. This value was arbitrarily set to convey the idea inference generation was likely to happen at some chance level.

During retention, we had 30 overall time steps. The probability that a LTM trace would be consolidated was .2 ± .05 with the idea that consolidation takes time to occur. The probability for a LTM component to be transcribed into LLM was set at .1 ± .05 per time step to capture the idea that consolidation in LTM allows for a slower encoding process of information into LLM. The loss probability for LTM components was set .3 ± .1 per unit of time (i.e., there was a 30% probability that a component, experienced or inferred, would be lost at each time step). This value captures the idea that knowledge in LTM can be lost quickly if it is not consolidated. Similarly, the loss probability for a component in LLM was set at .05 ± .03 per unit of time for the idea that information in LLM is more durable over time. We set the duration of consolidation (i.e., phase I) for each LTM trace to be 10 ± 2 time steps since the onset of retention.

For retrieval we assumed that if at least .40 of a memory trace's components were present, then the information could be successfully retrieved; otherwise, it was counted as forgotten to the point that it was not sufficiently recoverable. This value was chosen using the idea that, if part of a sentence of event memory were accessible, then this could serve as a sufficient basis for a partial match or to reconstruct the rest. The results of this simulation are shown in Fig. 21. As can be seen, there was a linear pattern of forgetting. The overall pattern showed a strong linear fit (r^2 = 1.00) compared to a power fit (r^2 = .54). Moreover, if the data are divided into phase I and II segments, as was done by Fisher and Radvansky (2018), then the pattern in phase I (i.e., the first 10 time steps) showed a better linear fit (r^2 = .98) compared to a power fit (r^2 = .63). Likewise, the pattern in phase II (i.e., the last 20 time steps) showed a better linear fit (r^2 = 1.00) compared to a power fit (r^2 = .93).

Low learning

To show that low levels of learning can move performance away from linear forgetting toward more negatively accelerating forgetting, we reran the simulation with all of the same parameters, except that the initial learning parameter was set at .50 (rather than .98). This was done to capture the idea that the information was not well learned. The complexity of the memory trace was unchanged. The results of the simulation are shown in Fig. 22. Although the overall pattern was still better fit by linear (r^2 = .92) compared to a power (r^2 = .61) function, the difference was less than in the original run with high learning. More importantly, the phase I data were better fit by a power function (r^2 = .87) than a linear function (r^2 = .84). Similarly, the phase II data were also better fit by a power function (r^2 = .97) than a linear function (r^2 = .92). It is also notable that, as reported by Fisher and Radvansky (2018), an indication of two retention phases becomes apparent under these conditions.

Low complexity

To show that low levels of complexity can also move performance away from linear forgetting toward more negatively accelerating forgetting, we reran the simulation with all of the same parameters as the basic simulation, except that the number of environment components was set to 4 and the number of possible inference components was set to 1. This was done to capture the idea that the information was not as complex. Additionally, because the materials are less complex, the threshold for trace retrieval was increased to .70. This was done to capture the idea that the lack of complex would require a more complete memory trace for retrieval. That is, we assumed that less complex materials will have fewer interrelations, and so will be less open to pattern matching or reconstructive processes. Thus, the threshold is higher.[8] The results of the simulation are shown in Fig. 23. Again, the overall pattern was still better fit by linear (r^2 = .85) compared to a power (r^2 = .63) function, but the difference was smaller here. More importantly, the phase I data were better fit by a power function (r^2 = .92) than a linear function (r^2 = .73). Likewise, the phase II data were also better fit by a power function (r^2 = .99) than a linear function (r^2 = .81).

Approaching floor performance

One concern that can be raised about the simulation results shown so far is that they appear to imply that a linear pattern will hold

[8] While both of these changes can alter the fit of the different functions, both are done here to capture both changes that theoretically occur when there is low complexity.

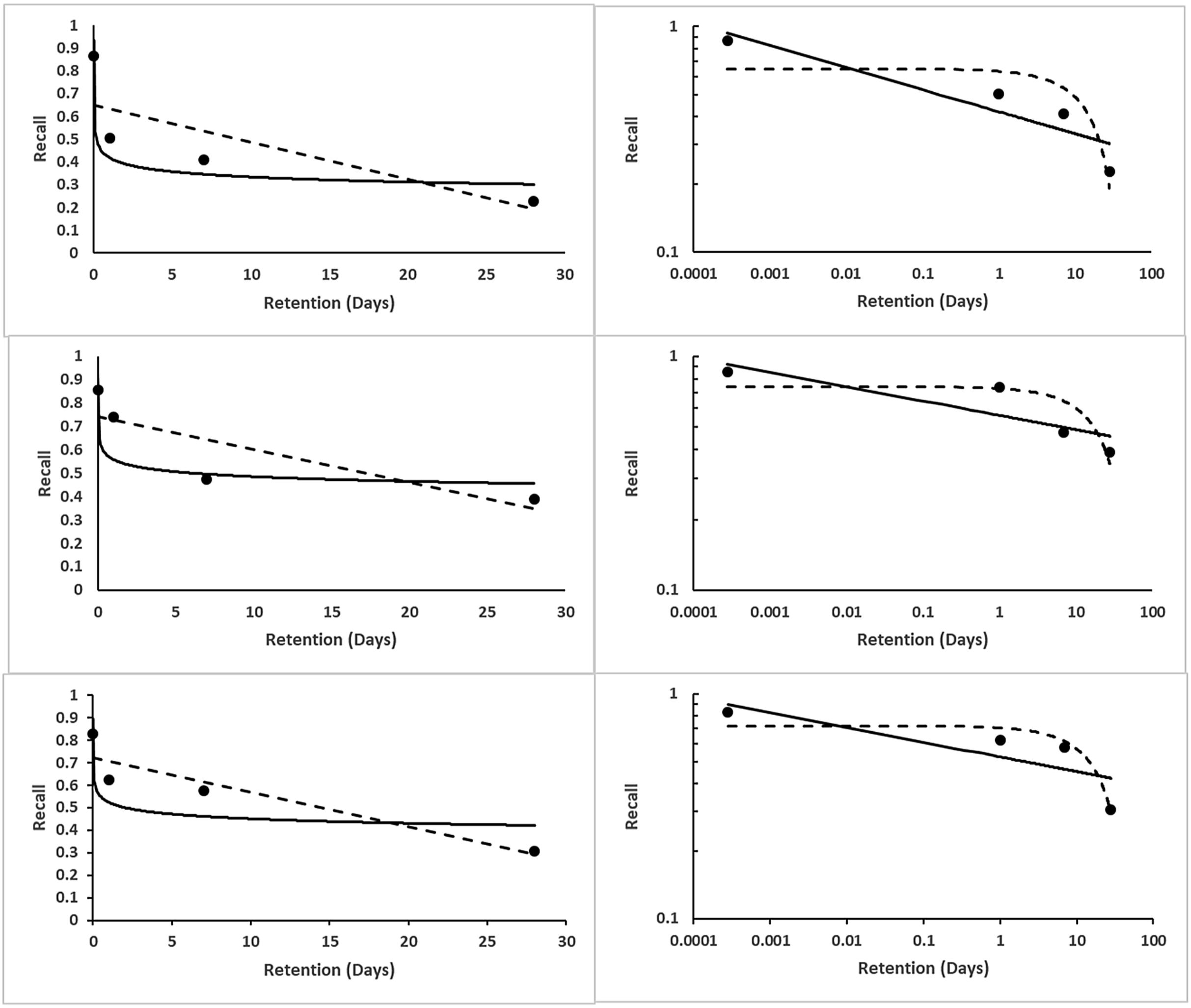

Fig. 18. Plots of the data from Craig et al. (1972) with linear axes on the left and logarithmic axes on the right. The top plot is the group that received 7 exposures, the middle plot is of 14 exposures, and the bottom is of 21 exposures.

throughout the retention period, even to the point when floor performance is approached. However, this is not the case. Instead, as more and more components are lost from memory traces, the ability to engage in partial matching and/or reconstruction becomes lost, and the curvilinear loss pattern of the individual components exerts itself in the pattern of performance of the observed data.

To demonstrate this, we reran the simulation with all of the same parameters as the basic simulation, except that the number of timesteps now went out to 50. The results of the simulation are shown in Fig. 24. Again, while the overall pattern was still better fit by linear (r^2 = .89) compared to a power (r^2 = .34) function, it can clearly be seen that as performance approaches floor, the pattern of data become more curvilinear. Thus, patterns of linear forgetting are more likely to observed at higher levels of performance, consistent with what we have noted in our section on higher levels of learning.

General discussion

In this paper, we have explored the idea that long-term memory retention and forgetting may not always follow the default outcome of a negatively accelerating function, such as power function, as has been noted since the beginning of the scientific study of human memory (Ebbinghaus, 1885) in which there is a constant proportional loss of information over time. In three experiments with data that we collected, we were able to show that, with retention intervals up to 12 weeks, we observed a linear pattern of forgetting in which there is a constant loss in the amount of information, but an increasing proportional loss of information over time.

This finding is not limited to our own work with the research materials and procedure that we employed. A survey of the literature revealed that linear retention and forgetting patterns appear in the studies different in other labs, at different times, with different materials,

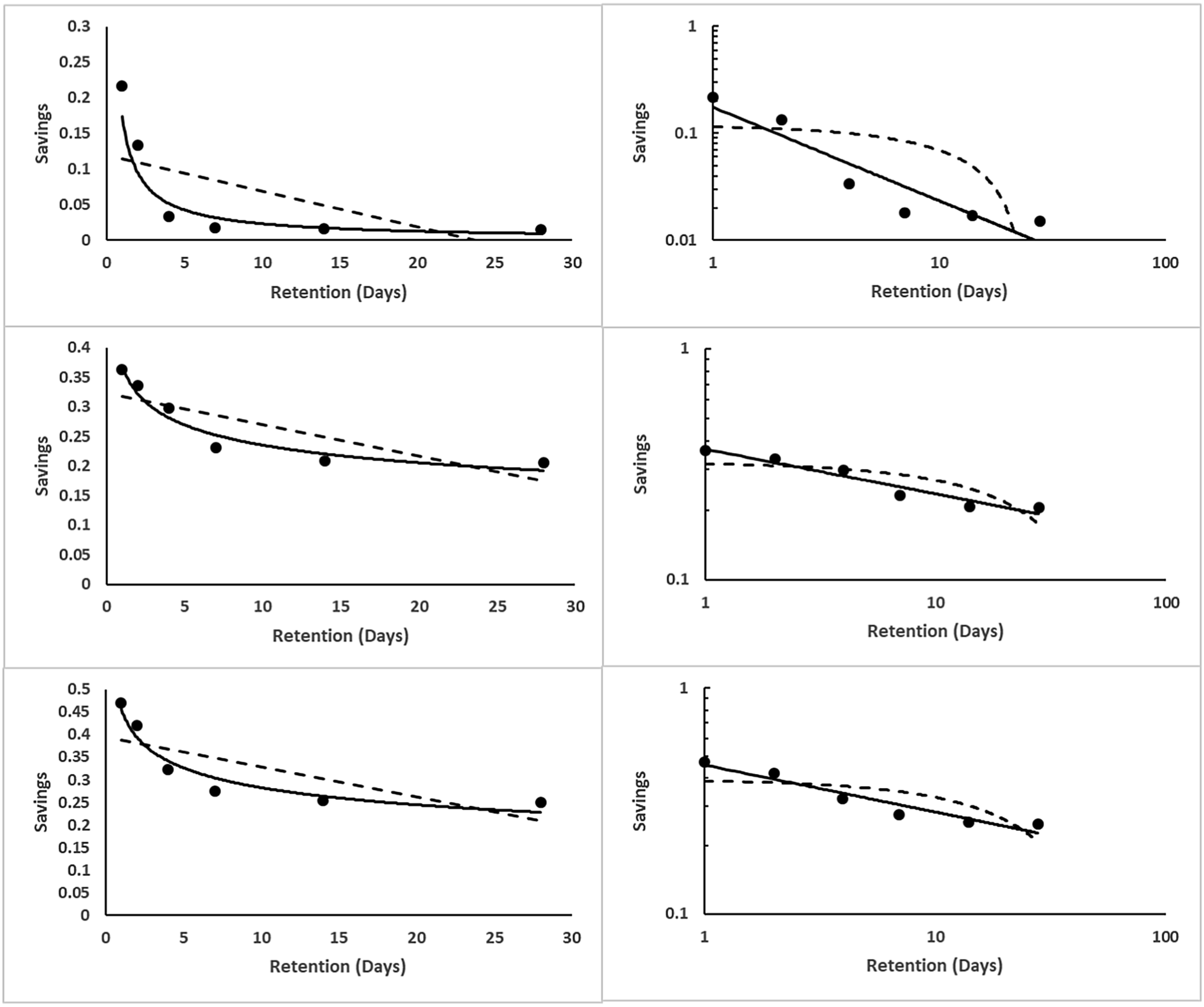

Fig. 19. A plot of the data from Krueger (1929) with linear axes on the left and logarithmic axes on the right. The top plot is the group that received 100% overlearning, the middle plot is 150% overlearning, and the bottom is 200% overlearning.

and for purposes other than the assessment of whether retention and forgetting can show a linear pattern. A common thread emerging looking at this work along with our own is that a pattern of linear retention and forgetting is more likely to be observed when the materials used are more meaningfully complex and allow for meaningful elaboration, and that they were well learned. Moreover, the study by Craig et al. (1972; see also Hellyer, 1962; Postman & Riley, 1959; Youtz, 1941) clearly suggests that as the degree of learning increases, the more linear the pattern of retention and forgetting became (see Fig. 24).

Resting on these findings, we advocate for theories of memory that account for different patterns of retention and forgetting by considering a number of, hopefully noncontroversial, factors. First, memory traces are made up of multiple components, and that memory traces of more complex sets of information will have more components. These components may originate externally from information in the environment or may be generated internally, as with any inferences that people may make. Second, information is not lost equally in a memory trace. Different components are lost at different rates. This is clearly evident in the fact that people can often remember some aspects of an event but not others over time. Third, consistent with the bulk of the memory retention and forgetting literature, we assume that the forgetting of individual components follows a negatively accelerating function. This keeps the basic Ebbinghaus (1885) retention and forgetting curve. Fourth, when people engage in a memory retrieval task, they do not need to have every component of the original set of information intact. Instead, accurate responses can be generated using partial information. During recognition, this may be done using a partial match of information in memory with that in a probe. Similarly, for both recognition and recall, people use some sort of reconstruction process to recover otherwise forgotten elements of a memory. With all of these elements, linear forgetting can be observed. Note well that we are not suggesting that the underlying memory process has a linear character, but that we believe that there are other noncontroversial factors that, working together, can produce an observed pattern of data that is more linear.

To provide further support for this theoretical view, we appealed to

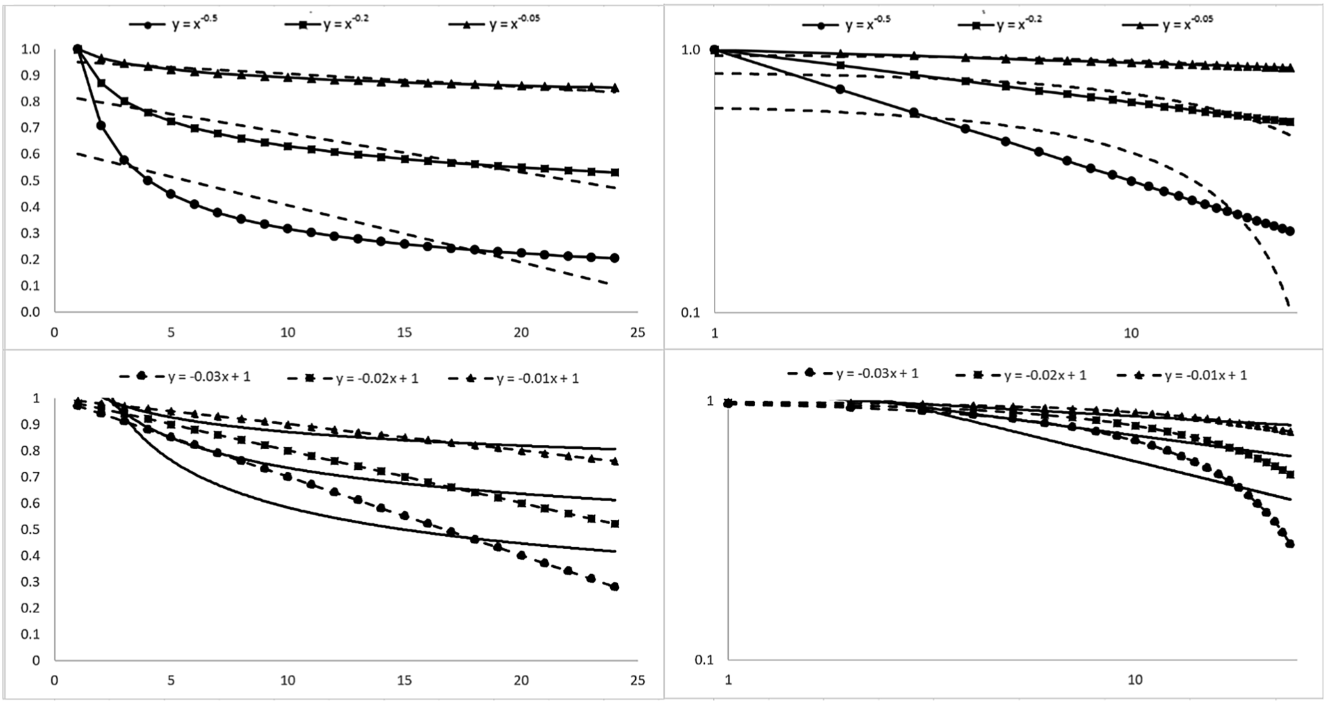

Fig. 20. These plots illustrate how the degree of loss in a function can influence its fit. Note again that power functions are represented by the solid line, and linear functions are represented by a dashed line. The top plots are of three power functions with a linear x-axis on the left and a logarithmic x-axis on the right. The exponents are −.5, −.2, and −.05, and the linear fits (r^2) are .66, .79, and .84, respectively. The bottom plots are of three linear functions with a linear x-axis on the left and a logarithmic x-axis on the right. The slopes are −.03, −.02, and −.01, and the power fits (r^2) are .75, .80, and .84, respectively. Thus, as the rate of loss becomes shallower, the fit of the other type of function increases.

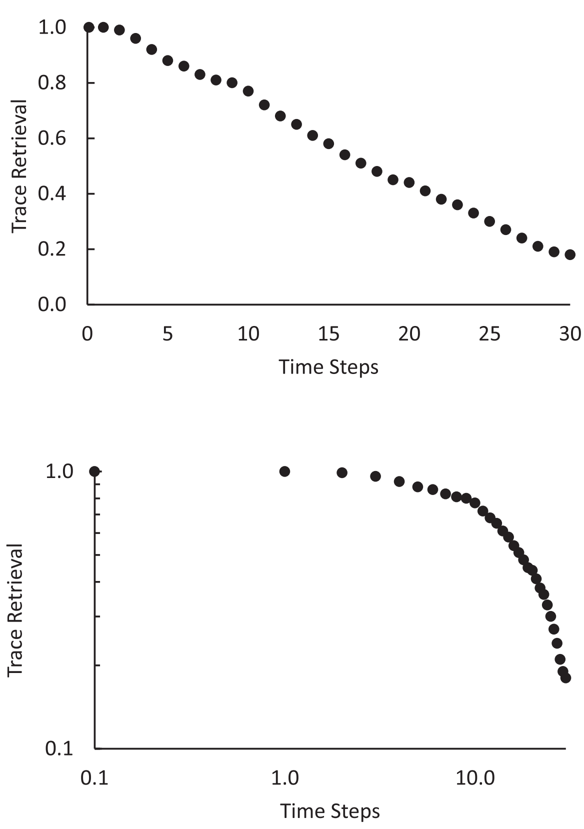

Fig. 21. Simulation with high trace complexity and high learning with linear axes above and logarithmic axes below.

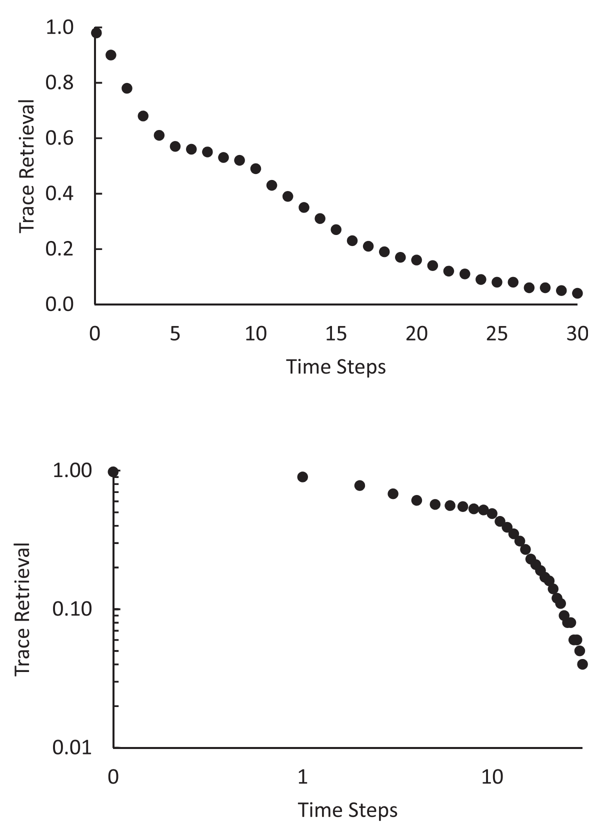

Fig. 22. Simulation with high trace complexity and low learning with linear axes above and logarithmic axes below.

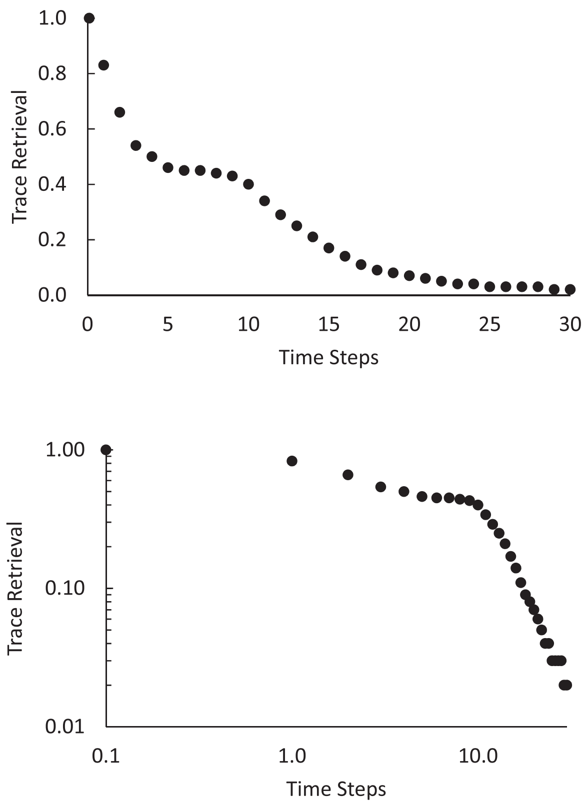

Fig. 23. Simulation with low trace complexity and high learning with linear axes above and logarithmic axes below.

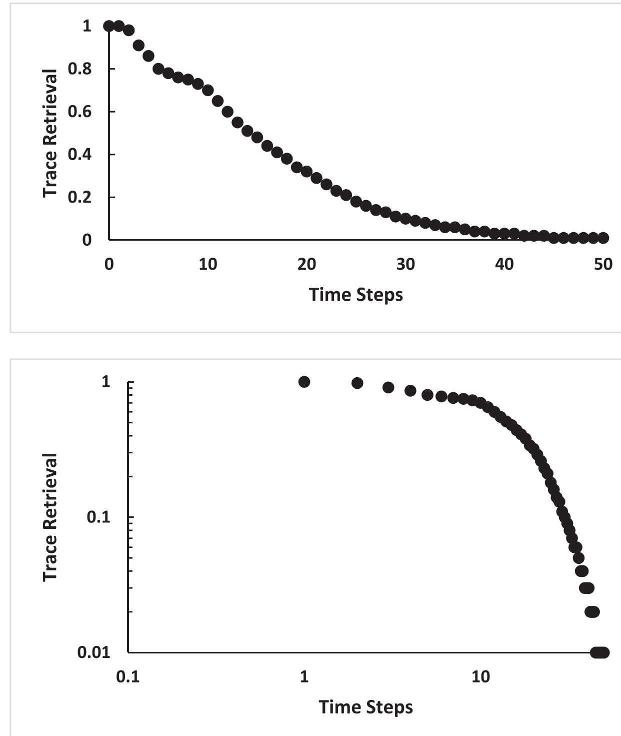

Fig. 24. Simulation with all of the same parameters as the basic simulation, except that the number of timesteps now went out to 50. Linear axes are above, and logarithmic axes are below.

a computer simulation as a proof of concept. Using this simulation, we were able to produce a linear pattern of retention and forgetting, despite the fact that the components of the memory traces are lost following a negatively accelerating function. Moreover, we were also able to highlight the importance of the degree of learning and the complexity of the memory traces by showing that if these are changed so that there is either a low level of learning, or the memory traces are made of only a small number of components, then linear forgetting is not observed. While the simulation is important, further empirical evidence is needed to assess ideas about the importance of factors such as the degree of learning, meaningful complexity, and the use of partial memory traces. We currently have multiple projects in our laboratory directly assessing these factors.

An important aim of science is to be predictive. One type of prediction that is absent from much of the research on memory is our ability to predict how long people will be able to remember sets of information into the future. This is the promise of retention and forgetting curves. If all memories were retained in a manner consistent with a power function, then we could easily set out on an effort of defining the rates of forgetting of different kinds of information under different circumstances, and be able to reliably predict future memory for just about anything. However, the current work shows that not all information is lost in a manner consistent with such a negatively accelerating function. In fact, if we were to use such a function to predict all future memory performance, we would wildly miss the mark. Some types of information follow a linear pattern of retention and forgetting. Finally, given that so many types of information that are important in our everyday experiences are meaningfully complex and persist for long periods of time, such as memories for eyewitnessed events, novels, films, social experiences, and so on, our ability to predict the fate of those forms of knowledge critically depends on our first being able to identify the different patterns of retention that may emerge, and when they are applicable.

Conclusion

Considering that memory retention has long been thought to consistently conform to a negatively accelerating function (e.g., Ebbinghaus, 1885), such as a power function, it is noteworthy that we were able to show a pattern of retention much better fit by a linear function. Furthermore, this pattern has appeared in other published studies, although its presence and significance was overlooked at the time. This pattern of retention appears to require a combination of meaningfully complex materials and a sufficiently high learning (to ensure that the various components of a memory are stored). For a simulation that we created as a proof of concept, linear forgetting may be a result of a loss of individual trace components at negatively accelerating rate (typical forgetting), along with the allowance for the operation of partial matching and/or reconstruction processes. Given that much of our experiences in everyday life involve complex events and are often deeply encoded, this linear function may be playing a far greater role in our memory than has been previously supposed.

Appendix A

Objects and Locations used in Radvansky et al. (2017)

Objects
broken window
revolving door
wall clock
welcome mat
oak counter
potted palm
ceiling fan
waste basket
fire extinguisher
cola machine
pay phone
bulletin board

Locations
airport
city hall
bar
movie theater
hotel
ice cream parlor
barber shop
car dealership
office building
public library
high school
laundromat

Appendix B

Materials for Experiment 2

Person
The student
The mailman
The baker
The teacher
The lawyer
The doctor
The dentist
The priest
The zookeeper
The landlord
The salesman
The clerk
The maid
The janitor
The babysitter
The receptionist
The lumberjack
The engineer

Activity
is eating.
is swimming.
is climbing.
is singing.
is dusting.
is stretching.
is running.
is sleeping.
is driving.
is painting.
is shopping.
is writing.
is bicycling.
is showering.
is yelling.
is laughing.
is hunting.
is reading.

Appendix C

Materials Used for Experiment 3

Animal	
Ape	Ant
Beaver	Bison
Chicken	Cow
Deer	Dolphin
Eagle	Elk
Frog	Fox
Goat	Goose
Horse	Hippo
Lion	Lobster
Monkey	Mouse
Owl	Otter
Pony	Panda
Rabbit	Robin
Spider	Snake
Turtle	Turkey
Wolf	Walrus
Location	
Arch	Airport
Beach	Bar
Cliff	City
Ditch	Diner
Forest	Farm
Grotto	Garden
Hill	Hotel
Island	Igloo
Meadow	Mall
Oasis	Office
Park	Plaza
Reef	Restroom
Stream	Stadium
Tunnel	Theater
Valley	Vet
Waterfall	Winery

Appendix D

Memory Retrieval Simulation Details

This simulation is also described in Fisher and Radvansky (2018). The aim of this simulation is to provide a more explicit proof of concept of our account of retention processes that can produce different patterns of observed data over different retention intervals. That is, the aim of the simulation is to capture patterns of retention and forgetting over long periods of time. We would like to be clear that the model is agnostic with regard to processes that operate during encoding at any stage, processes operating in retrieval, and the format of the representation of information in memory.

In terms of encoding, it is assumed that there is some variability in the effectiveness of learning under different circumstances. However, what is of concern here are the processes operating after the initial encoding processes. For retrieval, while various aspects of stored memory representations can influence the ease with which a memory search is able to access them, and different retrieval processes can increase or decrease success, for our purposes we ignore such influences. Presumably, a more complex model could be added on processes operating at retrieval. Instead, we simply assume that the encoding and retrieval processes are largely similar for a given set of information and emphasize the pattern of retention. The only retrieval assumption that we make is that, for our simulation if a certain proportion of the memory trace is intact, then it is possible that some sort of partial matching and reconstructive processes can lead to an accurate response. We do not specify what these processes are, given that this is a simulation of retention, not retrieval. Similarly, while the simulation does not distinguish between processes operating during recall versus recognition, we do think that for a given retrieval task, similar processes would be operating at different retention intervals. It is important to note that while forgetting results in a loss of availability, the larger theory is agnostic as to whether forgetting is due to a loss of availability or accessibility. The retention effects would be the same. We simply are concerned here with whether a memory is capable of being retrieved or not. Finally, while the simulation uses memory traces that are vectors of components, from a larger theoretical perspective, the same principles could be at work if one were to assume another representational format.

Simulation Parameters

Participants (*N*): This is the number of participants simulated.

Memory Traces (*m*): This is the total number of traces to be retained for each participant. The retrievability of each trace is determined the retention of its components and the retrieval threshold (Θ).

Learning Probability *(l)*: This is the probability that any LTM component of any trace will be originally learned.

Environmental Components (*C*): These are the components of a memory trace that are encoded from the external experience.

Potential Inferential Components (*c*): These components are encoded from inferential processes using prior knowledge. This reflects the number of *potential* inferential components that can be drawn.

Inference Formation Probability (*p*): This is the probability that a potential inference component will be drawn as determined at the level of the trace.

Overall Time Points (*t*): This is the total number of time points (simulation cycles) that retention is simulated.

LTM Consolidation (*f*): This is the probability that a trace in LTM will be fixed by the process of consolidation. When this happens, all the retained LTM components of a trace become resistant to loss for the duration it is consolidated. This probability is normally distributed.

LTM-LLM Translation (*F*): This is the probability that a LTM component is transcribed to LLM at a given time point. This probability is normally distributed by component. For the simulation, it is recommended that *F* be less than *f* because, intuitively, it seems reasonable to assume that the aim

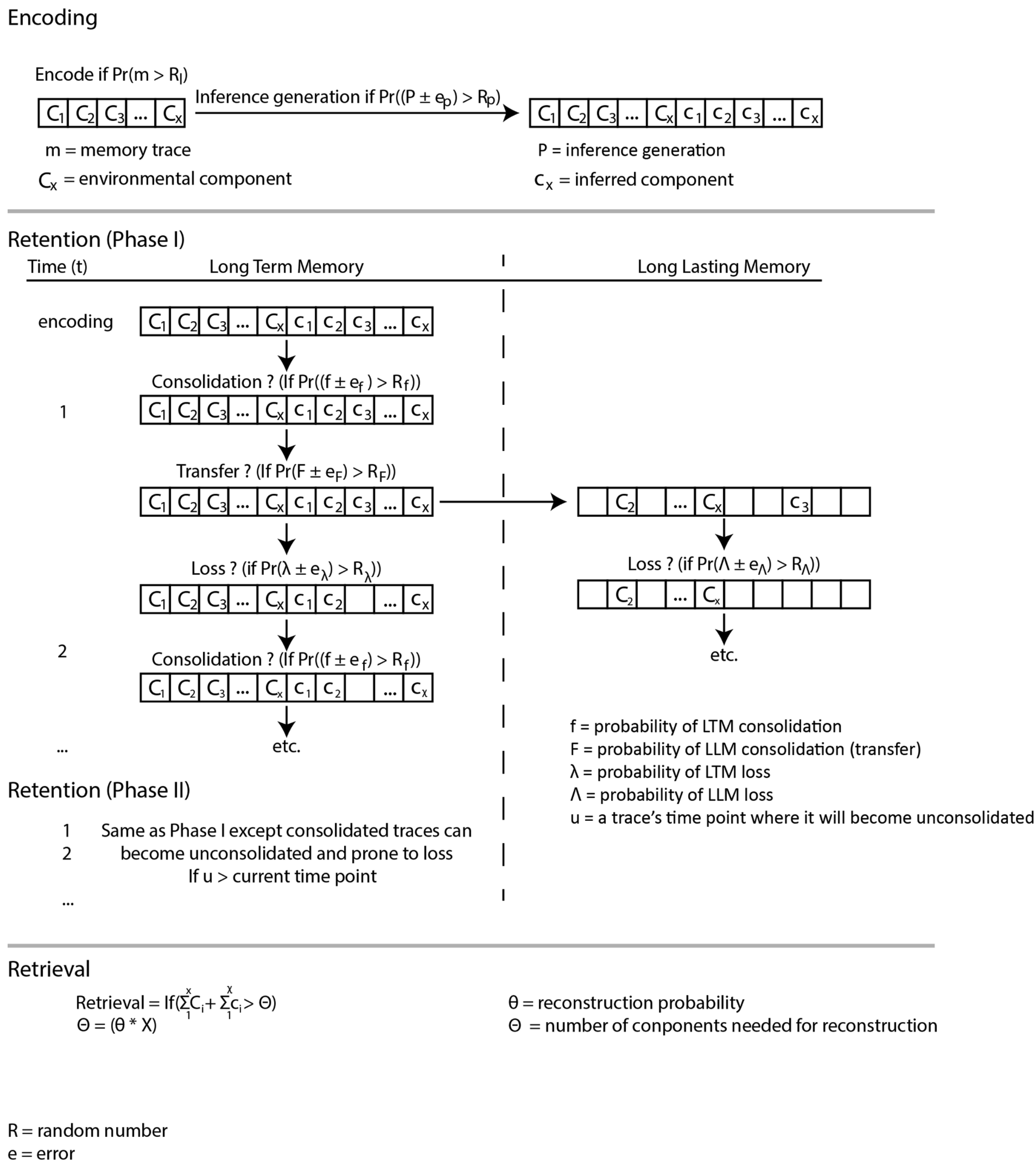

Fig. A1. A depiction of the processes operating in the memory retention simulation. Note that each instance of R is a separate random number, not the same random number applied multiple times.

of the consolidation of information in LTM is to provide it time to be transcribed to LLM.

LTM Loss (λ): This is the probability that a LTM component will be lost at a given time point. This probability is normally distributed by component.

LLM Loss (Λ): This is the probability that a LLM component will be lost at a given time point. This probability is normally distributed by component.

Consolidation Duration (*u*): This is the average duration of the consolidation of a trace in LTM. That is, how long until it becomes unconsolidated.

Retrieval Threshold (Θ): This is the proportion of components needed for a trace to be retrievable (whether by reconstruction or pattern matching). This is determined by the Θ parameter multiplied by the total number of environmental components (C). Any retained component in LTM or LLM contributes to the retrieval of the trace (note that the same component retained in both LTM and LLM only counts once).

Inferences Are Useful: If inferences are useful at retrieval, they will contribute to the retrievability of a trace (i.e., the trace having enough components to cross the threshold set by Θ). If they are not useful only the environmental components will contribute to retrieval.

Description

An outline of the simulation is provided in Fig. A1. In all cases, a random number would be between 0 and 1. The simulation begins with encoding, in which information may be stored in long term memory (LTM) as a retained trace. This trace is composed of environmental components (C) and potential inference components drawn from prior knowledge (c). The original encoding is determined by the probability parameter, l, such that $\Pr((l \pm e_l) > R_l)$, where R_l is a random number. The probability of an inference component being drawn is determined by the probability parameter, p, such $\Pr((p \pm e_p) > R_p)$, where R_p is a random number.

As retention begins, the processes of consolidation, transcription, LTM forgetting, and LLM forgetting occur iteratively at each time point. Consolidation for each trace is governed by the probability parameter f, such that $\Pr((f \pm e_f) > R_f)$, where R_f is a random number. Once consolidated, any components in the memory will not be subject to loss in LTM. Transcription is the process by which retained components of LTM traces can be copied to LLM. This is governed by the F parameter, such that $\Pr((F \pm e_F) > R_F)$, where R_F is a random number. LTM loss refers to the forgetting of components within a LTM trace, and is exponential over time. This is governed by the λ parameter, such that $\Pr((\lambda \pm e_\lambda) > R_\lambda)$, where R_λ is a random number. Similarly, LLM loss refers to the forgetting of components within an LLM trace, and is exponential over time. This loss is governed by the Λ parameter, such that $\Pr((\Lambda \pm e_\Lambda) > R_\Lambda)$, where R_Λ is a random number. For both LTM and LLM, a component that is lost remains inaccessible from that memory store for all subsequent time points. While memories in LTM may be consolidated, this consolidation is only temporary. Once a time point has passed, then the memory becomes unconsolidated, and is again open to loss. This unconsolidation is governed by the u parameter $\pm e_u$.

A trace is retrievable at any time if a sufficient number of its components are accessible. This level is determined by the Θ parameter (the Θ parameter multiplied by X, the maximum number of environmental components for each item trace). For example, if a trace had 10 environmental components and retrieval involved a Θ value of .7, the Θ threshold would require 7+ accessible components for the trace to be retrieved with the idea that partial matching or reconstruction of some nature would allow for an accurate response. These components could be either environmental or inference components,[9] and they could come from memories in either the LTM store or the LLM store. Importantly, for the simulation, if a particular component is accessible in *both* LTM and LLM, it is only counted once toward the threshold. Therefore, only a sufficient number of unique components for a trace will contribute to its retrieval.

References

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. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*(1), 1.
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.
Burtt, H. E., & Dobell, E. M. (1925). The curve of forgetting for advertising material. *Journal of Applied Psychology, 9*(1), 5.
Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36*(2), 438–448.
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.
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.
Craig, C. S., Sternthal, B., & Olshan, K. (1972). The effect of overlearning on retention. *Journal of General Psychology, 87*, 85.
De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12.
Ebbinghaus, H. (1885/1913). Memory: A contribution to experimental psychology (H. A. Ruger & C. E. Bussenius, Trans.). New York: Columbia University, Teacher's College. (Reprinted 1964, New York: Dover). ())
Fisher, J. S., & Radvansky, G. A. (2018). Patterns of forgetting. *Journal of Memory and Language, 102*, 130–141.
Goetschalckx, L., Moors, P., & Wagemans, J. (2018). Image memorability across longer time intervals. *Memory, 26*(5), 581–588.
Hellyer, S. (1962). Supplementary report: Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental Psychology, 64*(6), 650.
Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*(2), 96–101.
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.
Kristo, G., Janssen, S. M., & Murre, J. M. (2009). Retention of autobiographical memories: An Internet-based diary study. *Memory, 17*(8), 816–829.
Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12*(1), 71.
McGaugh, J. L. (2000). Memory – a century of consolidation. *Science, 287*(5451), 248–251.
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.
Murre, J. M., & Chessa, A. G. (2011). Power laws from individual differences in learning and forgetting: Mathematical analyses. *Psychonomic Bulletin & Review, 18*(3), 592–597.
Myung, I. J., Kim, C., & Pitt, M. A. (2000). Toward an explanation of the power law artifact: Insights from response surface analysis. *Memory & Cognition, 28*(5), 832–840.
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.
Radvansky, G. A. (2005). Situation models, propositions, and the fan effect. *Psychonomic Bulletin & Review, 12*(3), 478–483.
Radvansky, G. A., O'Rear, A. E., & Fisher, J. S. (2017). The persistence of event models: The differential fan effect over time. *Memory and Cognition, 45*, 1028–1044.
Radvansky, G. A., Spieler, D. H., & Zacks, R. T. (1993). Mental model organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(1), 95.
Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(5), 940–953.
Reder, L. M., & Anderson, J. R. (1980). A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology, 12*(4), 447–472.
Rubin, D. C. (1982). On the retention function for autobiographical memory. *Journal of Verbal Learning and Verbal Behavior, 21*(1), 21–38.
Rubin, D. C., Hinton, S., & Wenzel, A. (1999). The precise time course of retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(5), 1161.
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.
Thompson, C. P., Skowronski, J. J., Larsen, S. F., & Betz, A. L. (1996). *Autobiographical memory: Remembering what and remembering when.* Lawrence: Erlbaum Associates Inc.
Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology, 18*(2), 225–252.
Weaver, C. A., III, & Krug, K. S. (2004). Consolidation-like effects in flashbulb memories: Evidence from September 11, 2001. *The American Journal of Psychology*, 517–530.
Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology, 9*(4), 418–455.
Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science, 18*(2), 133–134.
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.

[9] If the "Inferences are useful" switch is on.