Journal of Memory and Language 102 (2018) 130–141

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Patterns of forgetting

Jerry S. Fisher*, Gabriel A. Radvansky

University of Notre Dame, United States

ARTICLE INFO

Keywords:
Retention
Forgetting
Event models
Word lists
Narratives

ABSTRACT

In memory research, forgetting is largely assumed to occur following a relatively consistent forgetting curve. However, recent work in our lab suggests that there is a shift in the pattern of forgetting after a retention interval of about seven days. Moreover, work on narrative comprehension has shown that information at different levels of representation show different patterns of forgetting. Much of the existing work on patterns of forgetting (a) does not allow one to assess changes in forgetting because much of the data is collected either prior to or after seven days, but not sufficiently bridging this period of time, and (b) does not consider patterns of forgetting for different levels of memory representation. In this study, memory for a list of words and narrative texts was assessed up to 12 weeks after initial learning. We observed that memory for the word list showed some forgetting early on, followed by an abrupt loss after about seven days. Moreover, for the narrative text, surface form memories were forgotten to around chance level after about an hour, whereas textbase level memories were retained until about seven days when memory suddenly dropped to around chance levels, much like the word list memories. In contrast to this, memory at the event model level remained high throughout, although there was some forgetting over time. To account for this pattern of retention and forgetting, a simulation was developed as a proof of concept to illustrate our theoretical interpretation.

Introduction

One of the most enduring findings in the study of human memory is Ebbinghaus's (1885) *retention curve* (often called the *forgetting curve*). This curve is a negatively accelerating function in which the majority of forgetting occurs soon after learning, with less information being forgotten as time progresses. The basic assumption of most memory researchers is that this is a continuous function, which progresses in a relatively constant pattern in long-term memory. The aim of the current study is to more finely assess the forgetting of information over long periods of time. This was done for both (a) a standard set of memoranda in memory research, namely word lists, as well as for (b) more complex sets of information, namely memory for narrative texts.

We first review some recent evidence from our lab suggesting a shift in memory retention and forgetting after about seven days. After this, we discuss memory for different levels of representation of narrative texts, as well as two studies that have looked at memory for text at these levels at longer retention intervals. Next, we present the current study. After this we, present a simulation as a proof of concept for the patterns of retention that we observe.

Forgetting curve shifts

There has been a fair amount work on the retention curve over the years. The bulk of this work has been focused on determining the nature of this function, such as whether it is an exponential or a power function, with the consensus being that it is somewhat better described by a power function (Wixted & Ebbesen, 1991). Some have explored the idea that a power function description is a result of an averaging of memory performance across many trials or observations (Anderson & Tweney, 1997; Averell & Heathcote, 2011; but see Wixted & Ebbesen, 1997). If the decay of individual memory traces were exponential, averaging across multiple exponential functions is best fit by a power function (Murre & Chessa, 2011), namely $M = at^b$, where M is memory, t is time, a is a constant, and b is the exponent, conveying the rate of forgetting over log time. The most important component for us here is the exponent, b. For this paper, we treat a retention curve as described by a power function as the default pattern and thus our null hypothesis for retention. Any deviation from this would need to be taken into account.

Although a power function is the default assumption, a recent analysis of existing retention data by Pettijohn and Radvansky (2017) and Csik and Radvansky (2018) suggests that there may be changes in the rate of forgetting over time, as defined by the exponent of the power function. For a power function, the exponent fit to a forgetting curve

* Corresponding author at: Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, United States.
E-mail address: jerry.s.fisher.111@nd.edu (J.S. Fisher).

https://doi.org/10.1016/j.jml.2018.05.008
Received 26 July 2017; Received in revised form 22 May 2018

Journal of Memory and Language 102 (2018) 130–141

should remain constant throughout the process of retention and forgetting (Wixted, 2004). Thus, if one fits a power function to early time points, the expectation is that future time points would be predicted to follow that same rate. Pettijohn and Radvansky tested this prediction by using data from 23 published studies, including 45 experiments, from Ebbinghaus (1885) to the present, which had five or more retention intervals. Power functions were fit to the data from the first four retention intervals to predict performance for later retention intervals. They found that while retention intervals that were seven days or shorter showed better memory than predicted, those intervals that were greater than seven days showed the opposite pattern, with faster than predicted forgetting. This suggests that the rate of forgetting, as captured by the power function exponent, may increase after seven days.

Moreover, the study by Csik and Radvansky (2018) explicitly addressed the size of the exponent of the power function, which can provide an index of the rate of forgetting. Using data from 44 published studies, including 135 experiments, which had three or more retention intervals, the data were fit to a power function, and the exponent from that function was recorded. These exponents were then analyzed as a function of the longest retention interval for a given experiment. What was found was that from about one minute to one day, there is a decrease in the rate of forgetting, with exponents approaching zero. This parallels the idea that LTP in the hippocampus takes several hours to complete and may be aided by sleep. From about one day to about nine days, the rate of forgetting remains largely stable, with exponents being similarly closer to zero. Then, after this time, for exponents of functions lasting up to years later, the rate of forgetting increases with the values largely moving further away from zero.

Although this idea that there is a transition somewhere around seven days is seen when analyzing across retention studies, there are few studies that have sufficient time points both prior to and after this transition period of seven days within their retention range. Thus, this current study directly bridges this gap for different material types.

Levels of representation

Another factor that can affect memory retention is the type of representation that is involved. Here we make use of narrative memories because they can be readily divided into three levels of representation. Specifically, these are the surface form, the textbase, and the event model levels (Van Dijk, Kintsch, & Van Dijk, 1983). The surface form is memory for the verbatim words and syntax that were used. This type of memory is typically very short lasting, often just a few minutes (Sachs, 1967). The textbase is memory for the propositional idea units that are conveyed in a text independent of the wording. Thus, a verbatim sentence from a text and a paraphrase that conveys the same meaning would map onto the same textbase representation. Finally, while the surface form and textbase representations are memories for the text itself, the event model is a representation of the situation described by the text (Glenberg, Meyer, & Lindem, 1987). This referential representation contains information that was conveyed by the text, as well as any inferences a reader may draw based on their prior knowledge.

Schmalhofer and Glavanov (1986) developed a method of assessing each of these levels of representation in memory using a signal detection analysis. For each critical sentence from a text, there are four types of recognition probe sentences: (a) the *verbatim* sentence that actually appeared in the text, (b) a *paraphrase* of a verbatim sentence, which did not actually appear in the text, although that idea was conveyed, (c) an *inference* sentence that conveys an idea that was likely generated by readers using their world knowledge, and (d) a *wrong* sentence that, while generally thematically consistent with the text, is inconsistent with the events described by the text.

The measure of the surface form compares memory for verbatim probes (hits) with memory for paraphrases (false alarms). Both probe types capture idea units that were presented in the text, but only the verbatim convey the actual words and syntax used. The measure of the textbase compares memory for the paraphrases (hits) with memory for the inferences (false alarms). Both of these probe types were never actually mentioned in the text, and are consistent with the situation described by the text, but only the paraphrase conveys idea units that were actually present in the text. Finally, the measure of the event model compares memory for the inference probes (hits) with memory for the wrong probes (false alarms). Neither of these probes types were mentioned in the text, nor did they convey idea units actually present in the text; they were both thematically consistent with the text, however, but only the inferences were actually consistent with the situation described by the text. This method of assessing different levels of text representation has been used widely over the years (Bohay, Blakely, Tamplin, & Radvansky, 2011; Fletcher & Chrysler, 1990; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Narvaez, Radvansky, Lynchard, & Copeland, 2011; Radvansky, Copeland, & von Hippel, 2010; Radvansky, Copeland, & Zwaan, 2003; Radvansky, Gibson, & McNerney, 2014; Radvansky, Zwaan, Curiel, & Copeland, 2001; Zwaan, 1994).

Of particular interest here are two studies that have reported changes in the different levels of representation over different periods of time. The first is a study by Kintsch et al. (1990). Of the two experiments reported by Kintsch et al., we focus on the first. This experiment assessed memory for the surface form, textbase, and event model levels for people of four different retention groups: immediate testing after reading, and testing after 40 min, two days, or four days. What was found, as can be seen in Fig. 1, was that there was forgetting for the surface form and textbase measures, but there was no clear evidence of forgetting for the event model level. Given this surprising finding, it is important to (a) assess whether there is any forgetting at the event model level, (b) understand the pattern of forgetting, if there is any, and (c) provided an account for what gives rise to such a long-lasting representation of memory.

The second study of interest is by Radvansky et al. (2001). Of the two experiments they reported, we focus on the second, which assessed memory after a delay. This study compared younger and older adults on narrative memory at these three levels either immediately or after

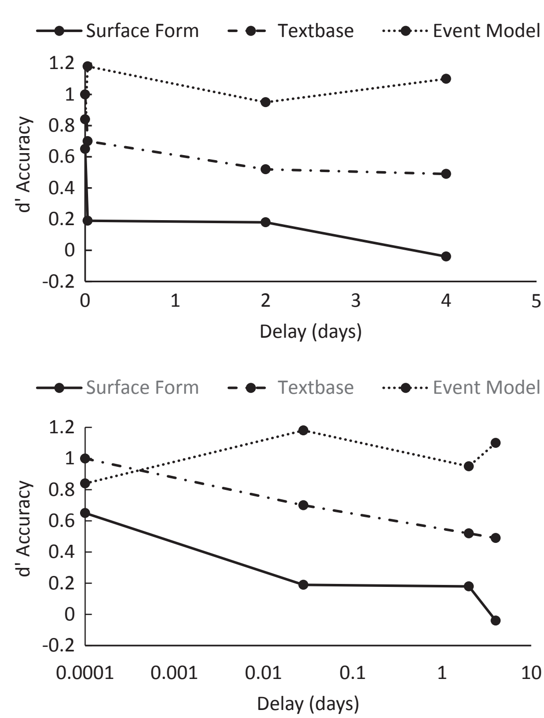

Fig. 1. Accuracy data from Kintsch et al. (1990) plotted in terms of a linear time scale (top) and a logarithmic time scale (bottom).

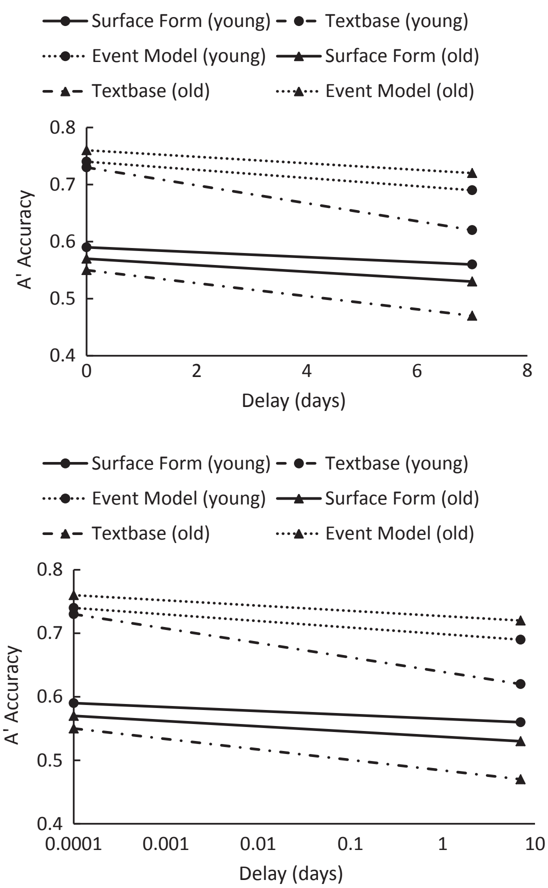

Fig. 2. Accuracy data from Radvansky et al. (2001) plotted in terms of a linear time scale (top) and a logarithmic time scale (bottom).

seven days. Unlike the Kintsch et al. (1990) study, as can be seen in Fig. 2, there was evidence of some forgetting at all three levels. The amount of forgetting at the surface form level was relatively small, however, mostly because both younger and older adults had relatively poor memory at this level to begin with, suggesting that they were at near floor level performance almost immediately. The rate of forgetting, for both age groups, was greater for the textbase level than for the event model level. Thus, putting together this study with the findings of Kintsch et al. it is unclear whether there is forgetting at the event model level, and, if so, what the pattern of this forgetting would be, and why it is less dramatic than for other levels of memory.

While both the Kintsch et al. (1990) and the Radvansky et al. (2001) studies are informative about the differential retention of levels of representation, there are shortcomings with each of them with regard to our goals here. First, for both studies, the retention intervals do not extend beyond seven days, thus, any shift in memory retention beyond the first phase cannot be observed. Second, the data in the Kintsch et al. study resulted in patterns of retention that are not very stable, particularly for the event model level. Third, for the Radvansky et al. study, there were only two time points assessed. As such, it is not possible to fit retention curves to that data. These shortcomings, along with a need to directly address the questions of interest here, led to the motivation for the current experiment.

Experiment

The current experiment directly explored whether there are differences in retention prior to and after seven days in which the early phase retention would be ceasing to have an influence. Specifically, we measured retention patterns over six time points with a range of 12 weeks. This was done for both a traditional memory study memoranda, namely a list of words, as well as memory for a narrative text. The inclusion of narrative memory also allowed us to assess the pattern of forgetting at different levels of representation for the same materials. Considering this design, there are a few predicted outcomes. First, consistent with the analysis of Pettijohn and Radvansky (2018), we predict a noticeable drop in retention at around seven days for at least some trace types. We formalize this retention pattern below as consisting of two phases. Second, also consistent with previous work, we predict that representation at the event model level will be retained to a higher degree over the 12-week retention period.

Method

Participants. Two hundred and eighty-eight (201 female; age 18–22, $M = 19.4$, $SE = 1.19$) native-English speaking students were recruited from the University of Notre Dame participant pool in exchange for partial course credit. They were assigned to one of six retention groups (Immediate, 1 h, 1 day, 1 week, 4 weeks, and 12 weeks), with each group made up of 48 participants. This sample size was selected based on prior work using the Schmalhofer-Glavanov analysis (e.g., Bohay et al., 2011).

Materials. The materials consisted of a list of twenty nouns and four narrative texts. For the word list, the nouns were 4–7 ($M = 5.6$, $SE = 0.18$) letters long, had 1 to 2 syllables ($M = 1.7$, $SE = 0.11$), and ranged in Thorndike-Lorge written frequency ratings from 98 to 2511 ($M = 720$, $SE = 137$). The narrative texts were taken from Radvansky et al. (2001). They ranged from 516 to 703 words long ($M = 621$, $SE = 79$) and were entitled: Personal Identification, The Farmer's Rebellion, New York in the Future, and the Beanie Baby Craze. One text with sample probes is given in Appendix A. In addition, a 452-word text on bullfighting was given as a practice text.

The recognition test for the four target texts involved 64 probe sentences per text (256 total probes). This test was written in JavaScript using the jsPsych library (De Leeuw, 2015) and was able to be accessed on-line. As described in the introduction, these probes were derived using the procedure developed by Schmalhofer and Glavanov (1986). There were four versions of each critical text sentence. These were (a) the *verbatim* sentence that actually appeared in the text, (b) a *paraphrase* of that sentence that conveyed the same idea, but differed in the wording, (c) an *inference* sentence that conveyed an idea that was not mentioned in the text, but which would likely have been inferred by most readers as they read, and (d) a *wrong* sentence that was thematically consistent with the topic of a text, but inconsistent with the events that were actually described. Using these four probe types, we were derived estimates of the surface form, textbase, and event model using the A′ signal detection measure (Snodgrass & Corwin, 1988).

Procedure. The first phase of the experiment took place at individual computers in the lab. People were first given the task to memorize the list of words. The list of words was presented, one at a time, for 10 s each, in a different random order for each person. After presentation, the people were told that they would read five texts. They were first given a practice text, followed by the four target texts in random order. At the beginning of each text was a title. The texts were then presented clause by clause in a self-paced manner. Participants advanced by clicking a 'continue' button at the bottom of the screen.

After the first phase, participants in the Immediate group were tested in the lab. In comparison, people in the other five retention groups were told to expect an email with a weblink for testing. They did memory testing on-line using their own computers, tablets, or smartphones. They were asked to take the test in a quiet place and with adequate time to finish. The retention intervals were as follows: 1-h group (*Median deviation* = 3 min), 1-day group (*Median deviation* = 53 min), 1-week group (*Median deviation* = 35 min), 4-week group (*Median deviation* = 48 min), and 12-week group (*Median deviation* = 183 min).

For the second phase, people were first told to recall as many words as they could from the word list. This was done by typing their responses into a textbox on the screen. When they were done, they clicked a button to end the task. There was no time limit.

Following the word list test, people were given a recognition test for the narrative texts. For the narrative recognition test, both the order of the texts and the probes within each text were randomized. At the beginning of each section, people were given the title of the narrative to indicate to them of which one was being tested. The people made their recognition decisions by clicking buttons on the screen labelled 'yes' or 'no' to indicate whether a probe sentence had been studied before. This task was not timed.

Results

In presenting the results of this study, we first provide the retention patterns for the word list data followed by the narrative text data. In addition, because the emphasis of this study is on the patterns of forgetting, and not on differences between conditions, per se, as it typically done, we do not present ANOVA analyses of our results. Instead, the data can best be understood by assessing the plots of the data in our figures, which contain error bars. For convenience, the power exponents and r^2s for each fit type are included in Table 1. The question of whether the data are better described by a two-phase power fit compared to an overall power fit is assessed using regression models and Akaike information criterion (AIC) estimates.

Word List Data. The word list recall accuracy data are shown in Fig. 3. Word recall accuracy was measured as the proportion of correctly recalled nouns. In scoring, obvious misspellings were counted as correct, but synonyms were not. As can be seen, overall retention followed a common Ebbinghaus pattern of rapid forgetting. A power function fit to the entire data pattern, such that $M = .19\,t^{-.31}$, with a good fit, $r^2 = .93$.

That said, it is also clear, especially when the data are plotted on log scale, that there is a drop at seven days compared to what be expected based on the pattern of performance prior to that time. Moreover, overall, the data takes on the appearance of a combination of two forgetting curves: one that lasts up to at least a day (phase I) and one that follows the abrupt drop that occurs at about seven days (phase II). As such, we fit two separate power functions to this data. The phase I power function included the Immediate, 1-h, and 1-day groups. This function was $M = .33\,t^{-.13}$, with a moderate fit, $r^2 = .90$[1]. The phase II power function included the 1-day, 1-week, 4-week, and 12-week groups. This function was $M = .33\,t^{-.5}$, with very good fit, $r^2 = 1.00$.

We compared the fit of an overall power function to two power function fitting by using two regression models that were derived through a logarithmic transformation to the above power expressions. Model 1 describes an overall power fit as: log(M) = .19 − .31(log x). Model 2 describes the two power function fit as: log(M) = .33 − .13(log x) for phase I and log(M) = .33 − .5(log x) for phase II. Model 2, r^2 (adj.) = .99 (Residual *SE* = .11), *AIC* = −8.19, better fit the data than Model 1, r^2 (adj.) = .91 (Residual *SE* = .34), *AIC* = −2.71. Thus, rather than the data being best fit by a single function, the two halves of the data were fit by two functions with very different exponents. The first half had a much smaller exponent than the second, consistent with the idea that forgetting is more dramatic after seven days.

Narrative text data

Text recognition at the different levels of representation was measured by calculating A′ recognition scores as describe earlier. These data are shown in Fig. 4. We consider the pattern of performance for each of the levels separately. For the surface form, memory was quite low even at Immediate testing (*M* = .58), and quickly fell to near chance (.5) for the 1-h group and onward. Because there appears to be only a single function operating here, a power function was fit to the entire data set. This function was $M = .54\,t^{-.01}$, with a poor fit, $r^2 = .30$, most likely because the data was so close to chance performance for most of the retention intervals. Overall, this pattern replicates prior work suggesting that surface form information is forgotten quickly (e.g., Sachs, 1967). What is of some interest here is that performance at this level of representation does remain largely above chance even 12 weeks later, albeit to a very small degree. The two-tailed t-tests assessing whether performance differed from chance were: Immediate: t(47) = 6.50, p < .001, 1-h: t(47) = 2.28, p = .03, 1-day: t(47) = 2.69, p = .01, 1-week: t(47) = 2.03, p = .05, 4-weeks: t(47) = 4.53, p < .001, 12-weeks: t(47) = 2.34, p = .02.

For the textbase level, a different pattern of results emerged. First, it is clear, at least early on, that performance was much better than for the surface form measure. A power function fit to the entire data pattern was $M = .61\,t^{-.04}$, with a good fit, $r^2 = .92$. However, unlike surface form memory, and like memory on the word lists, there appear to be two phases of retention at the textbase level: one that occurred for the data up to 1-day, with a rapid drop-off at the 1-week interval, followed by a second function. The power function for Immediate, 1-h, and 1-day groups was $M = .65\,t^{-.02}$, with a good fit, $r^2 = .89$. The second function, including the 1-day, 1-week, 4-week, and 12-week groups was $M = .65\,t^{-.06}$, with a very good fit, $r^2 = .95$. Also, note that, unlike the surface form, once textbase memory dropped to near chance at the 4-week and 12-week retention intervals, performance was actually not

Table 1
Exponents and fits across the different memory measures.

	Word list		Narratives					
			Surface form		Textbase		Event model	
	Exponent	r^2	Exponent	r^2	Exponent	r^2	Exponent	r^2
Overall	−.32	.88	−.01	.26	−.04	.90	−.01	.88
Phase I	−.12	.57	N.A.		−.01	.73	−.01	.69
Phase II	−.58	.99	N.A.		−.06	.94	−.02	.93

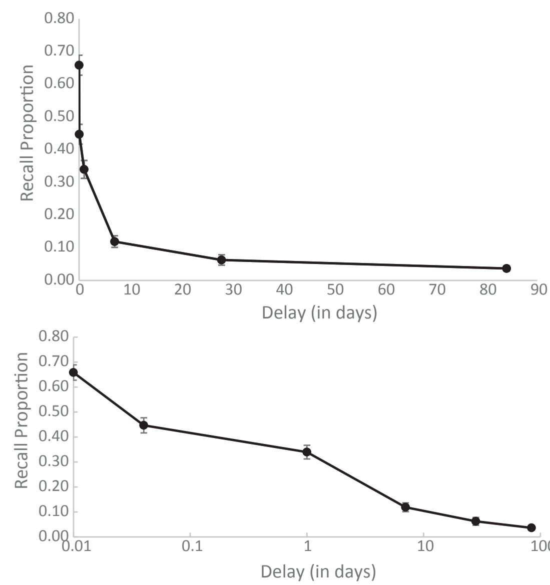

Fig. 3. Word recall proportion as a function of time on a linear (above) and logarithmic (below) scale.

[1] Note that for these split fits, the r^2 value for phase I is worse than the overall. This is likely due to the fact the power function is fit to only three data points for phase I.

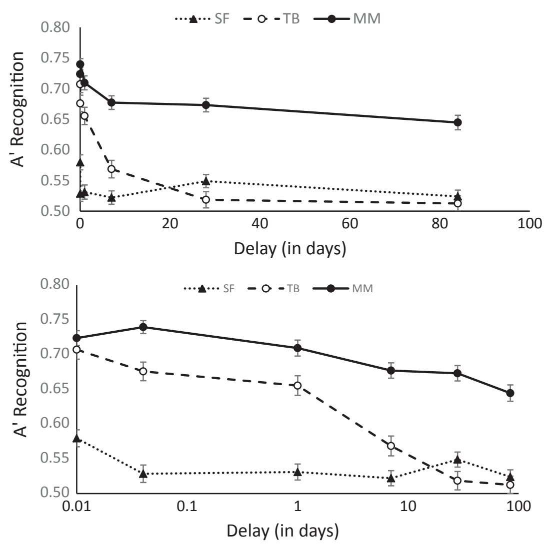

Fig. 4. Narrative text A′ retention patterns for surface form, textbase, and event model levels as a function of time on a linear (above) and logarithmic (below) scale.

different from chance. Assessments of whether performance differed from chance were: Immediate: $t(47) = 15.1$, $p < .001$, 1-h: $t(47) = 13.1$, $p < .001$, 1-day: $t(47) = 10.99$, $p < .001$, 1-week: $t(47) = 4.81$, $p < .001$, 4-weeks: $t(7) = 1.39$ $p = .17$, 12-weeks: $t < 1$. Using regression models, we found that Model 2 (two power functions), r^2 (adj.) = .97 (Residual *SE* = .02), *AIC* = −17.11, better fit the textbase data than Model 1 (a single power function), r^2 (adj.) = .90 (Residual *SE* = .05), *AIC* = −13.20.

For the event model level, yet another pattern of retention emerged. First, it is clear from Fig. 4 that memory is better than both the surface form and textbase levels, both immediately and at every retention interval. A power function fit to the entire data pattern was $M = .70\,t^{-.01}$, with a strong fit, $r^2 = .88$. Unlike the surface form and textbase level measures, memory never approached chance performance. Thus, like prior work, the event model level information is better encoded and consolidated into long-term memory. This is the level of information representation that memory is better attuned to process. While there is not the steep drop-off at the 1-week interval, if the data are analyzed separately in terms of the early and latter portions of the retention curve as was done for the word list and textbase measures, there are differences in the pattern of performance. The power function for the early data was $M = .71\,t^{-.01}$, with a poor fit, $r^2 = .44$. The second function, including the 1-day, 1-week, 4-week, and 12-week groups was $M = .71\,t^{-.02}$, with a strong fit, $r^2 = .93$. Using regression models, we found that Model 2, r^2 (adj.) = .88 (Residual *SE* = .02), *AIC* = −19.27, better fit the data than Model 1, r^2 (adj.) = .81 (Residual *SE* = .02), *AIC* = −16.68. Thus, there is a small change in the pattern of forgetting in the two phases of retention, with forgetting being greater after seven days compared to prior to this time.

General discussion

The presented study explored two primary issues. First, whether there is a shift in the pattern of forgetting prior to and after about seven days, and second, whether there are differences in the pattern of forgetting for information at different levels of representation.

Shift in forgetting after about seven days

With regard to the first issue, by using retention intervals both prior to and after seven days, consistent with recent work by Pettijohn and Radvansky (2018), we found evidence for a change in the retention curves with increased forgetting after about seven days compared to prior to this. This is most clearly seen with the word list memory as well as the textbase level representations for narrative memory. There was also some evidence with event model level memory, although that effect is much subtler. This effect was absent with the surface form memory, however those data were very close to baseline throughout.

Thus, overall there may be two phases of long-term memory retention. Phase I occurs prior to seven days, and is marked by less rapid forgetting. Phase II occurs by about seven days and is marked by more rapid forgetting. This idea of two phases of retention and forgetting would be more compelling if there were supportive evidence in the existing memory literature apart from our own study. For our search of the literature, we established a number of criteria that needed to be satisfied to make this assessment. First, any study needed to have a sufficient number of retention intervals both before and after any transition period to fit a retention curve. For our purposes, this would need to be at least three points prior and three points after seven days. Thus, studies that had sufficient number of observations prior to seven days, but not after, were excluded (e.g., Nunoi & Yoshikawa, 2016; Radvansky, O'Rear, & Fisher, in press; Runquist, 1983), as well as studies that had sufficient numbers of observations after seven days, but not before (e.g., Bahrick, Hall, & Da Costa, 2008; Conway, Cohen, & Stanhope, 1991). Second, the response rates needed to be sufficiently above chance to allow for forgetting over time to be observed. Thus, studies in which retention had essentially dropped to chance by the seven day mark, were excluded (e.g., Krueger, 1929; Luh, 1922; Wickelgren, 1975). Third, the study should involve traditional assessments of clearly identifiable content items from memory. Thus, studies that assessed other information, such as letter sequences (but not identities) were excluded (e.g., Bean, 1912).

To our knowledge, outside of the current study, there is only one other study that satisfies our criteria. This study, done by Carpenter, Pashler, Wixted, and Vul (2008), involved three experiments. In Experiments 1 and 2, people were given list of trivia facts (e.g., greyhounds have the best eyesight of any dog) to read, whereas in Experiment 3, people learned English-Swahili word pairs (e.g., somo–friend). In Experiment 1, the facts were studied once, whereas in Experiments 2 and 3, the facts and pairs were studied three times each. Moreover, in each experiment, half of the items were simply studied, and for the other half, people were given a test for retrieval practice. The retention intervals used in this study were 5-min, 1-day, 2-days, 7-days, 14-days, and 42-days.

We fit power functions to the data from each condition of each experiment, to derive the exponent, and a measure of fit (r^2), both overall as well as for the first three retention intervals (phase I) and the second three retention intervals (phase II). Table 2 presents the results of this assessment. As can be seen, in all experiments, and in all conditions, dividing the data into two phases resulted in better fits overall. Moreover, and more importantly, the exponents for phase I were consistently smaller than the exponents for phase II, consistent with the idea that there is increased forgetting in phase II relative to phase I. As further support, we subjected the Carpenter et al. (2008) data to a regression analysis of an overall model versus a two function model. In all cases, the two function model provided a much better fit to the data than an overall model. These results are presented in Table 3.

Thus, we can be more confident that there is a shift in the pattern of retention and forgetting at or around seven days. Why might this occur? At this point, we cannot say for sure, however we can suggest a possibility that could be tested in the future. Work in neuroscience suggests that there are at least two phases of retention (consolidation). One is the fast, synaptic consolidation that occurs largely in the hippocampus, perhaps through the process of long-term potentiation (LTP) (e.g., Kandel, Dudai, & Mayford, 2014). Moreover, from what we know about LTP, this is a time-limited form of consolidation that persists somewhere in the neighborhood of about seven days (Wixted & Cai, 2013), and so, is not a form of permanent memory storage. The other is slow,

Table 2
Exponents and fits across the different experiments and conditions for Carpenter et al. (2008). Model 1 refers to a single function fit to the data. Model 2 refers to a two function fit that is separated by seven days.

		Model 1		Model 2 – Phase 1		Model 2 – Phase 2	
		Exponent	r^2	Exponent	r^2	Exponent	r^2
Experiment 1	Study-Test	−.09	.56	−.01	.96	−.36	.98
	Study only	−.10	.61	−.02	.95	−.38	.98
Experiment 2	Study-Test	−.07	.44	−.01	.76	−.44	.98
	Study only	−.08	.56	−.01	.94	−.37	1.00
Experiment 3	Study-Test	−.11	.82	−.06	.99	−.25	1.00
	Study only	−.12	.89	−.08	.94	−.17	.95

Table 3
Comparison of regression models for Carpenter et al. (2008).

		Model 1			Model 2		
		r^2 (adj)	Residual SE	AIC	r^2 (adj)	Residual SE	AIC
Experiment 1	Study-Test	.46	.3	−3.33	.99	.05	−10.48
	Study only	.52	.3	−3.33	.98	.05	−10.48
Experiment 2	Study-Test	.29	.29	−3.51	.98	.05	−10.48
	Study only	.45	.27	−3.88	1	.03	−13.14
Experiment 3	Study-Test	.78	.19	−5.71	1	.01	−18.86
	Study only	.86	.16	−6.61	.97	.07	−8.72

systems consolidation that occurs across larger cortical brain systems (Abraham, 2006). This type of consolidation results in memories becoming more independent of the hippocampus.

One study by Takashima et al. (2006) used fMRI recordings collected while participants remembered pictures of natural landscapes 1, 2, 30, and 90 days after learning. Neural activity in the hippocampus decreased over time, whereas activity in the cortex, particularly in areas of the frontal lobe, increased over time. Thus, it seems likely that there are two phases of retention, with one phase including LTP-retained memories in the hippocampus and a second phase without it. Thus, the shift in the pattern of retention and forgetting at or around seven days that we are observing may be due to a shift in consolidation from an early phase involving fast, synaptic, hippocampal retention to a second phase lacking this after phase I retention has run its course. What is lacking at this point is neuroimaging evidence that directly speaks to this issue.

Our findings bring into question some basic, common ideas about memory. One has to do with the long-term consequences of long-term memory consolidation. One line of thought is that the longer a longer memory has had to consolidate, the more resistant it will be to forgetting. In other words, the more time that has passed, the more likely memories will have been consolidated to the point that they are taken out of the pool of memories that can be forgotten. Likewise, consolidation may be part of the explanation for Ribot's gradient, which is the finding that retrograde amnesia is more likely to affect recent, less consolidated memories, and less likely to affect older, more consolidated memories (see Wixted, 2004). However, the current work, the study by Pettijohn and Radvansky (2018), and Carpenter et al. (2008) show that, if anything forgetting is more rapid than would be expected after about seven days. Thus, the passage of time is not accompanied by increased consolidation and retention of declarative memories. Instead, the passage of time is accompanied by continued forgetting, which, if anything, becomes more pronounced.

Note, that this does not preclude the possibility that some memories may be consolidated and stored to the point where forgetting is minimal or absent. We acknowledge that this is a possibility. However, this does not appear to be the fate of the majority of our memories. What the cognitive preconditions and mechanisms are that allow this permanent storage to occur is unknown at this time.

Levels of representation

The second primary issue address in this study is the differential patterns of forgetting for different memory representations. We addressed this by assessing memory at three levels of representation for narrative texts. Consistent with prior research, we found that memory for precise wording, the surface form, was not remembered well, but was quickly forgotten (Sachs, 1967). Apart from this, what is also interesting is that although memory at the surface level was very close to chance, it was consistently above it. This is consistent with results showing that sometimes people can remember more or less the precise wording of something, even months later (e.g., Kintsch & Bates, 1977).

At the textbase level, forgetting was not as rapid as that for the surface form. This is consistent with prior research showing that people retain gist information longer (e.g., Frost, 2000). However, unlike previous research, we observed that there was a big drop in this kind of memory around seven days. Overall, people can remember the basic idea that they are presented with over a period of a few days, but then they have trouble discriminating what they heard or read from what they inferred.

At the event model level, forgetting was much less than that for the other two levels. This is less consistent with the Kintsch et al. (1990) study, which found no evidence of forgetting at this level, and more consistent with the Radvansky et al. (2001) study in which some evidence of forgetting was evident. Overall, while forgetting does occur at this level of representation, it is substantially reduced relative to the surface form and textbase levels. Over long periods of time, what is remembered is not the description of an event, but our memory of the event being described.

Simulation

The findings of the current study are clear, and they are consistent with the above-mentioned principles of memory retention,

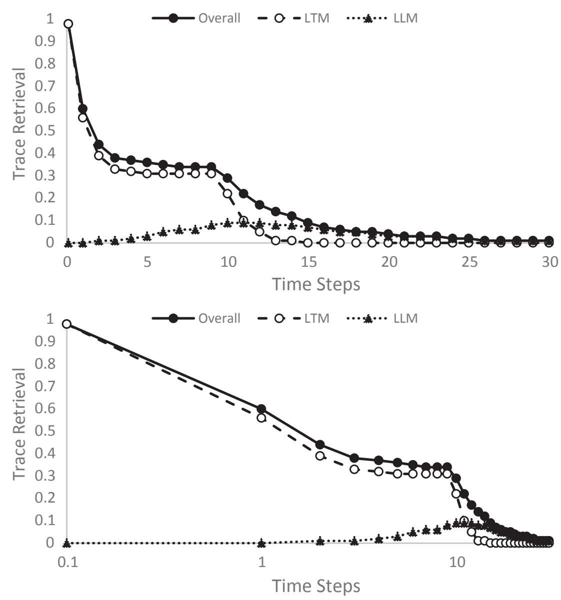

Fig. 5. Word list memory simulation as a function of time on a linear (above) and logarithmic (below) scale. LTM refers to the trace retrieval only from the hippocampal storage of the trace components. LLM refers to trace retrieval only from the cortical storage of the trace components. For the overall trace performance, both LTM and LLM contribute components to the retrieval of a trace.

consolidation, and representation. Outlined below is a simulation that serves as a proof of concept for how these concepts may come together[2]. Specifically, we sought to show that (a) retention, in some cases, shows a drop in performance around seven days, (b) not all memory traces show a clear drop, and (c) different types of memory show different patterns of forgetting.

Outline. A brief summary of the simulation is provided here, with more details provided in Appendix B. This simulation assumes that each memory trace is stored as a vector of components (e.g., Hintzman, 1984), which may be either encoded from the environment or inferences drawn from general world knowledge. This simulation also assumes two general memory stores. Following McGaugh (2000), we refer to these as long-term memory (LTM) and long-lasting memory (LLM). The LTM store, roughly analogous to hippocampal storage, allows for the temporary consolidation of knowledge for a period of time. In comparison, the LLM store, roughly analogous to cortical storage, allows for the enduring storage of knowledge. Learning involves the encoding of memory traces into LTM. Once there, over time, there is a probability that the components of a LTM trace will be consolidated (and therefore be immune to forgetting). Over time, the components of memory traces in LTM may be transcribed to LLM, or forgotten. Components that have been transcribed to LLM also have a probability of being forgotten. Importantly, the forgetting of memory trace components in this simulation is exponential over time. This is consistent with the idea that individual memories may be forgotten according to an exponential function, while the average of these traces reflects a power function (e.g., Murre & Chessa, 2011). Another important feature of the simulation is that successful retrieval is a function of the proportion of components available in a trace, not as a function of the trace as a whole. The assumption is that if a sufficient proportion of the components of trace are available, some process of reconstruction or partial matching can allow for accurate retrieval.

Word List Simulation. We first simulated the pattern of forgetting that we observed with the word list data. For this, we simulated 48 participants, each retaining 20 memory items (i.e., words). We specified a .98 probability that each word would be encoded into LTM with the idea that these words were presented long enough to be well-encoded, although some small errors may have occurred. Moreover, we assumed a set of three environmental components for each item, along with a .50 ± .05 probability of an additional inference component being drawn. The small number of components was chosen based on the idea that individual words are simpler items than sentences, and are unlikely to evoke much in the way of inferences. The .50 value for making an inference was arbitrarily set to convey our thought than such inference was likely to occur at some chance level.

During retention, we had 30 overall time steps. The probability that a LTM trace would be consolidated was .2 ± .05 to capture the idea that this is a process that takes time to unfold. The probability for a LTM component to be transcribed into LLM was set at .1 ± .05 per time step with the idea that the need for consolidation of information in LTM is to allow for a slower encoding process in LLM to occur. 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 was chosen with the idea that information in LTM can be lost relatively quickly if it is not consolidated. Similarly, the loss probability for a component in LLM was set at .05 ± .03 per unit of time with the idea that information in LLM is more durable over long periods of 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 .75 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 selected based on the idea that if part of a word or its meaning were accessible, then the rest of the word could be reconstructed. The results of this first simulation are shown in Fig. 5. As can be seen, there is a drop in memory over time during phase I, and then after around 10 time steps, at the onset of phase II, there is an abrupt drop in performance. The data are well fit by a power function for both phase I ($r^2 = .97$) and phase II ($r^2 = .98$), and the exponent of the function in phase I ($-.25$) is less than that in phase II (-3.28).

Note that Fig. 5, in addition to showing overall accuracy performance (the solid line with closed circles), also illustrates memory for the underlying components. The long-term memory traces (the dashed line and open circles) shows how these memories show a stabilization of traces after an initial drop followed by a second drop after a given time point. The long-lasting memory traces (the dotted line with the closed triangles) shows how these traces have begun to be established soon after the information is encountered and continues to build while the memories are in long-term memory. After the long-term memory traces are forgotten, only the influence the long-lasting traces is observed.

Narrative Simulation. For narrative recognition memory, although the narrative memory literature traditionally refers to three levels of representation, the surface form, textbase, and event model, the simulation assumed that there is a single memory representation. The three levels of representation are captured by how much of the memory trace is available to allow a positive response to be made.

For this simulation, the same number of participants, memory items, and parameters were used as the word list simulation with only a few exceptions so that the resulting simulation would not be a result of changing many parameters for each case. First, we assumed that each narrative item was represented by ten environmental components, and we allowed a .50 probability that up to five additional inference components may be drawn. This was done because narrative sentence information is more complex than single word, and the range of inferences that are likely to be drawn would be more extensive.

At retrieval, for recognition, for the surface form we assumed that .9 of original environmental components needed to be present in the trace to allow for an accurate response at this level. Furthermore, for the

[2] This simulation can be found at http://ec2-52–204-56–150.compute-1.amazonaws.com/pilot/RetentionSimulationAWS.html.2.

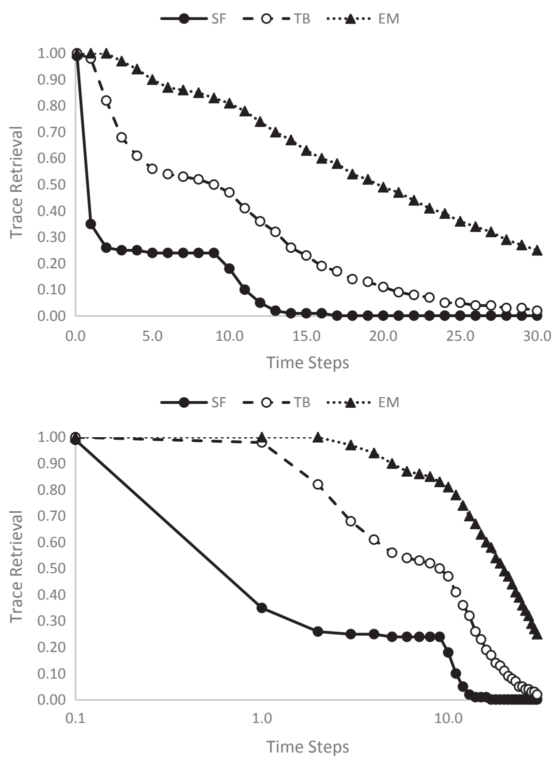

Fig. 6. Narrative memory simulation as a function of time on a linear (above) and logarithmic (below) scale. This graph only shows the overall trace performance and not the retrieval from LTM or LLM stores only. However, these values can be ascertained in the simulation.

surface form, any inferences made at encoding were not useful for retrieval. These choices were made because accurate surface form memory requires precise memory of both the word used and the syntax of the sentence structure. Deviations along these lines would line up more with the surface structure, and inferences could result in inappropriate retrieval because an inference was not actually present in the text.

In comparison, for the textbase level, we assumed a threshold where at least .5 of the environmental components were needed for an accurate response. This choice was made because accurate textbase memory is not dependent on as precise a memory as the surface form, and is open to more partial matching and reconstruction. Additionally, like surface form, no inferences at encoding were useful for retrieval. Inferences could result in inappropriate retrieval because an inference was not actually present in the text.

Finally, for the event model level, we assumed that at least .3 components needed to be present for a response to be made. At this level, these could be either environmental or inference components. This choice was made because accurate event memory is not dependent on as precise a memory as the surface form or textbase, and is open to even more partial matching and reconstruction. Inferences are appropriate here as the issue is whether a memory probe is consistent with the described event, not necessarily the text itself.

As can be seen in Fig. 6, performance for the surface form items was very low soon after original encoding, and dropped to chance relatively quickly. In comparison, for the textbase, as with word list memory, performance declined during phase I, and then after 10 time steps, during phase II, there was an abrupt drop in performance. The data are well fit by a power function for both phase I (r^2 = .82) and phase II (r^2 = .98), although the exponent of the function in phase I −.18) is less than that in phase II (−2.94). Finally, for the event model level, forgetting was more consistent throughout, without an appreciable drop in performance in the shift from phase I to phase II retention.

Note that the break down of performance for long-term memory and long-lasting memory that was done for the word list simulation is not done here. This is because in Fig. 5 we wanted to highlight the differences in memory performance at different levels of representation, similar to what was done in the Kintsch et al. (1990) study. Needless to say, the basic pattern of any such break down would be show a similar pattern.

Alternative accounts

While the finding that forgetting is reduced at the event model level, an alternative account is that this may not be a result of the superior retention for these types of representations, or even that is a different kind of memory for event model knowledge. Specifically, it may be that forgetting for the described event is just as pronounced as it is for other information, such as textbase knowledge. What makes it appear that forgetting is reduced are processes such as partial matching or reconstruction. That is, knowledge is being lost. However, people may be able to make responses because whatever information is left in the memory trace is consistent enough with memory in the probe. That is, there is a partial match, and it will appear as though forgetting is reduced at the event model level. Alternatively, because people have extensive schema and script knowledge for a wide variety of event types, this knowledge can be used to fill in the gaps brought about by forgetting. As a result, it appears as though forgetting is reduced at the event model level, when in fact it is not.

While the second of these two alternative accounts is intuitively appealing, it is unlikely that it accounts for the patterns of forgetting that we are observing. First, it conflates the idea of event models with schemas and scripts. When talking with other cognitive scientists, it is not unusual for them to think that they are one and the same. They are not. An event model is an episodic representation of a specific event, whereas schemas and scripts are more semantic knowledge about large classes of similar events. For example, your memory a specific time you crashed your car is an event model, whereas your schema of what happens in car crashes in general is more schematic knowledge.

Second, the idea that the knowledge we are tapping without event model memory is simply reconstructed knowledge does not hold up to scrutiny. Keep in mind that our event model measure assessed the difference between accepting inferences that are consistent with a particular described event (inference probes) and statements that are thematically consistent with the story, but not actually part of the described event (wrong probes). If people were solely using reconstructed knowledge from schemas and script to support responding to these items, then performance should be the same for both of these probe types because they are both similarly consistent with prior world knowledge, and our event model measure would be at chance. This clearly did not happen. Thus, people must be using memories that are consistent with specific events, namely event models, and not general knowledge that applies to a wide range of events.

Conclusion

This study contributes to a growing set of behavioral evidence against the assumption that forgetting follows a continuous function. Rather, forgetting seems to be better explained by two phases of retention, shifting from one to the next around seven days. Additionally, we build upon existing paradigms that separate retention into three levels of representation: surface form, textbase, and event model. Similar to previous findings, we found little forgetting at the event model level for periods of up to 12 weeks. This result is not a function of people reconstructing knowledge from semantic memory, but is a result of a memory of the events themselves. Future work is aimed at exploring the patterns of forgetting for different kinds of information, as well as assessing how different aspects of event memory are better or worse retained over time.

There are a number of implications of this research. One example is in terms of education. What the results show is that the basic propositional content of knowledge that is learned is retained for about seven days before it begins to be lost. However, a broader understanding of material may persist for long periods of time, although with some forgetting. This could explain why people are able to readily comprehend much of what is going on in a novel, television, or movie series, even after a long delay between exposures to the content. There may be some detailed information that is lost over time, with more details lost with longer periods of time, but the essential understanding and memory of the basics of the story are intact.

Appendix A

A.1. Sample article: personal identity

One of the great espionage problems is the search for a reliable way of determining a person's identity. This is particularly important for agents working abroad. These agents were often forced to make contacts using only sketchy information. Several disasters abroad were caused by poor identification. For example, four years ago, several undercover agents died when they thought that the people they were meeting were their contacts, when in fact they were agents working for the other side. Lead agent, Linda Gill, was shot first. She died within minutes, exposing the mission. Before the rest of the group could react to the obvious danger, two more agents, Max Eagle and James Romney, took a bullet and went down like stones. This prompted the CIA in Washington to create a Board of Identification. This was at the end of Nicolas Elder's term as Agency Head. The Board was empowered to award twenty-thousand dollars to the first person who developed a method of determining identity accurately ninety-nine percent of the time for a wide variety of people. There had been a number of attempts to solve this problem. One early idea was to have fingerprints taken at predetermined meeting sites. These sites would be strategically located across the world. A match could be made between a fingerprint and a stored file. The similarity between the two could be used to determine identity. Later, some engineers approached the identity challenge. They considered a retinal scan method. One year, Les Busby discovered that each retina had a different pattern that varied from person to person. Busby reasoned that this could be used as an identification method. This idea was based on the variations in peoples' retinas. These patterns would be distinctive no matter where a person was from. Busby even devised a special retinal scan helmet for people to wear. This method of determining identification captured the imagination of many the agency's administrators. Among those administrators were Cassell, Haynes, Hartley, and Nelson. A final idea was to use the DNA-based computer imaging system. A DNA imaging system is a device of great accuracy that can be used in most everyday situations. Early chemical- and spectral-based DNA identification methods were too cumbersome to be used abroad due to environmental changes. John Harrison was a self-taught computer game programmer. Early last year, Harrison invented and constructed four practical DNA identification systems. He completed his first system in April and submitted it to the Identification Board, but was turned down. The initial test of one of Harrison's systems was made in June. This was done abroad at a diplomatic conference. This first test of a DNA-based system was a grand success. He then built three more instruments, each smaller and more accurate than its predecessor. In August, Harrison's fourth system was tested on a trip to Egypt. It was found to be in error for only one person in a thousand. Although his systems all met the standards set up by the Board of Identification, he was not awarded any money until November, when he received five thousand dollars. A prominent member of the Board was Phil Marks. He was more impressed by the engineers. Marks thought that the programmer's device was less reliable than the work of the 'real' scientists. After several months, Harrison was taken under the wing of Senator Morris. Harrison ultimately claimed his reward money the following year. The newer DNA image identification systems are, broadly speaking, small, light-weight devices. A DNA sampling tube is hidden in a purse, briefcase, or clothing. As such, it remains available wherever the agent travels. The recent identification systems may be accurate to within one in ten thousand people.

A.2. Sample probes: personal identity

Inference: Several disasters were caused by poor identification.
Paraphrase: Identity could be determined by getting the similarity between the two.
Inference: Harrison was a highly skilled programmer and brilliant inventor.
Wrong: Marks himself was an amateur engineer.

Appendix B

B.1. Memory retrieval simulation details

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.

B.2. 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 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.

B.3. Description

An outline of the processing operating in 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 given time point if a sufficient number of its components are still accessible. This level of sufficiency 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 item trace to be retrieved with the idea that a partial matching process or reconstruction of some nature would allow for an accurate response. These accessible components could be either environmental or inference components[3], 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.

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

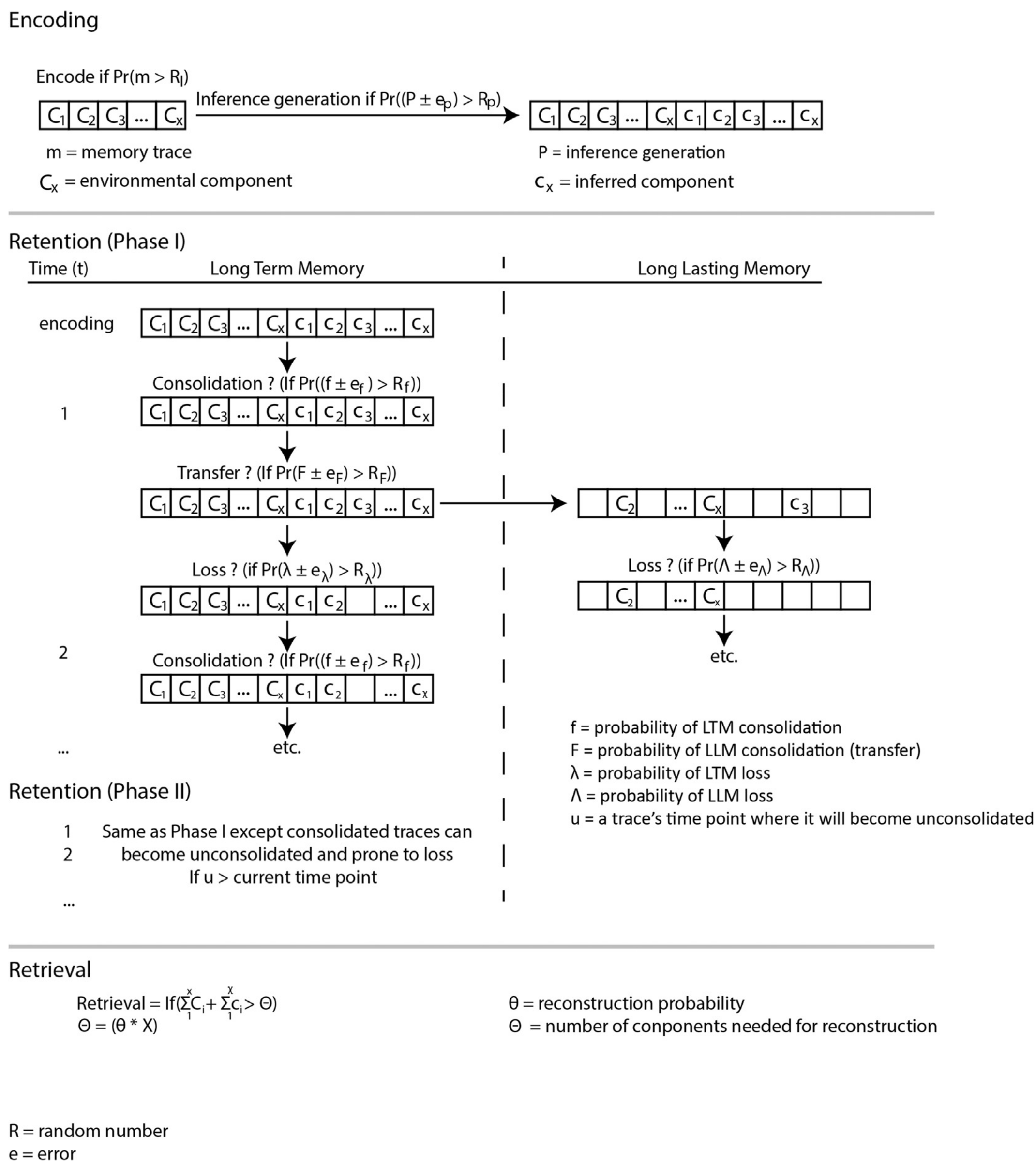

Fig. A1. A depiction of the processes operating in the memory retention simulation. Note that each instance of R is a separate random number, not the same random number applied multiple times.

References

Abraham, W. C. (2006). Memory maintenance the changing nature of neural mechanisms. *Current Directions in Psychological Science, 15*(1), 5–8.

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., Hall, L. K., & Da Costa, L. A. (2008). Fifty years of memory of college grades: Accuracy and distortions. *Emotion, 8*(1), 13.

Bean, C. H. (1912). *The curve of forgetting*. Press of the New era printing Company.

Bohay, M., Blakely, D. P., Tamplin, A. K., & Radvansky, G. A. (2011). Note-taking, review, memory and comprehension. *American Journal of Psychology, 124*, 63–73.

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.

Conway, M. A., Cohen, G., & Stanhope, N. (1991). On the very long-term retention of knowledge acquired through formal education: Twelve years of cognitive psychology. *Journal of Experimental Psychology: General, 120*(4), 395.

Csik, A. M., & Radvansky, G. A. (2018). Variations in forgetting rates (submitted for publication).

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). ())

Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: Recognition memory for three types of textual information. *Discourse Processes, 13*(2), 175–190.

Frost, P. (2000). The quality of false memory over time: Is memory for misinformation "remembered" or "known"? *Psychonomic Bulletin & Review, 7*(3), 531–536.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26*(1), 69–83.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*(2), 96–101.

Kandel, E. R., Dudai, Y., & Mayford, M. R. (2014). The molecular and systems biology of memory. *Cell, 157*(1), 163–186.

Kintsch, W., & Bates, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory, 3*(2), 150–159.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29*(2), 133–159.

Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12*(1), 71.

Luh, C. W. (1922). The conditions of retention. *Psychological Monographs*, 31(3), i.

McGaugh, J. L. (2000). Memory – A century of consolidation. *Science, 287*(5451), 248–251.

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.

Narvaez, D., Radvansky, G. A., Lynchard, N. A., & Copeland, D. E. (2011). Are older adults more attuned to morally-charged information? *Experimental Aging Research, 37*, 398–434.

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

Pettijohn, K. A., & Radvansky, G. A. (2018). *Shifting forgetting rates* (submitted for publication).

Radvansky, G. A., Copeland, D. E., & von Hippel, W. (2010). Stereotype activation, inhibition, and aging. *Journal of Experimental Social Psychology, 46*, 51–60.

Radvansky, G. A., Copeland, D. E., & Zwaan, R. A. (2003). Aging and functional spatial relations in comprehension and memory. *Psychology and Aging, 18*, 161–165.

Radvansky, G. A., Andrea, E. O., & Fisher, J. S. (2017). Event models and the fan effect. *Memory & Cognition, 45*(6), 1028–1044.

Radvansky, G. A., Gibson, B. S., & McNerney, M. W. (2014). Working memory, situation models, and synesthesia. *American Journal of Psychology, 127*, 325–342.

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging, 16*, 145–160.

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

Sachs, J. S. (1967). Recopition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2*(9), 437–442.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language, 25*(3), 279–294.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34–50.

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., ... Fernandez, G. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences, 103*(3), 756–761.

Van Dijk, T. A., Kintsch, W., & Van Dijk, T. A. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Wickelgren, W. A. (1975). Age and storage dynamics in continuous recognition memory. *Developmental Psychology, 11*(2), 165–169.

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

Wixted, J. T., & Cai, D. J. (2013). Memory consolidation. *The Oxford handbook of cognitive neuroscience, Volume 1: Core topics*, 1, 436.

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.

Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 920–933.