Research Article

Psychological Science
23(6) 625–634

DOI: 10.1177/0956797611430961
http://pss.sagepub.com

A Power-Law Model of Psychological Memory Strength in Short- and Long-Term Recognition

Chris Donkin[1] and Robert M. Nosofsky[2]
[1]University of New South Wales and [2]Indiana University

Abstract

A classic law of cognition is that forgetting curves are closely approximated by power functions. This law describes relations between different empirical dependent variables and the retention interval, and the precise form of the functional relation depends on the scale used to measure each variable. In the research reported here, we conducted a recognition task involving both short- and long-term probes. We discovered that formal memory-strength parameters from an exemplar-recognition model closely followed a power function of the lag between studied items and a test probe. The model accounted for rich sets of response time (RT) data at both individual-subject and individual-lag levels. Because memory strengths were derived from model fits to choices and RTs from individual trials, the psychological power law was independent of the scale used to summarize the forgetting functions. Alternative models that assumed different functional relations or posited a separate fixed-strength working memory store fared considerably worse than the power-law model did in predicting the data.

Keywords

forgetting, memory, short-term memory, power laws, power functions, forgetting curves, response times

Received 8/19/11; Revision accepted 11/1/11

One of the classic quantitative laws of cognitive psychology is that forgetting curves are well described by power functions (e.g., Anderson & Schooler, 1991; Wickelgren, 1974; Wixted & Ebbesen, 1991). For example, Wixted and Ebbesen (1991) and Wixted and Carpenter (2007) showed that diverse measures of forgetting, including the proportion of correct responses in free recall of word lists, recognition judgments of faces, and savings in relearning lists of nonsense syllables, were well described as power functions of the retention interval. However, these results involved relations between the retention interval and a variety of empirical dependent variables (Wixted & Ebbesen, 1991, p. 414). Each such dependent variable is associated with some particular scale of measurement, and the precise quantitative law that describes relations among the variables must depend on the scale of measurement that is chosen. For example, the precise form of the function that relates forgetting to the retention interval will vary depending on whether performance is measured by the proportion of correctly recognized items or by d'.

In the present work, we tested participants in a speeded old/new recognition paradigm involving both short- and long-term probes. The key independent variable was the lag between a test probe and a corresponding item on the study list. *Lag* is measured by how many items back into the memory set the probed item is. The most recently presented study-list item has a lag of 1, the next most recent item has a lag of 2, and so forth. We applied a modern exemplar-based recognition model (Nosofsky, Little, Donkin, & Fific, 2011) to account for individual-subject choice probabilities and response times (RTs) observed in the task. Application of the formal model allows one to estimate a set of parameters that describe the strength with which each exemplar from the study list is stored in memory. In applying the model, we made a remarkable discovery: Except for a small residual primacy effect, the estimated memory strengths were almost a perfect power function of the lag with which the original exemplars were presented on the study lists.

Whereas the previous evidence for a power law relied on measured relations between a variety of empirical dependent variables and the retention interval, the present regularity involved a latent psychological variable (memory-strength parameters). Thus, the obtained evidence for a power law of memory was found at a deeper, theoretical level of analysis.

Corresponding Author:
Chris Donkin, School of Psychology, University of New South Wales, NSW, 2052, Australia
E-mail: christopher.donkin@gmail.com

Furthermore, because the estimated memory strengths were based on model fits to choices and RTs from individual trials, the derived power law was independent of the scale used to summarize the empirical forgetting functions.

This work makes two other key contributions to the literature. First, although previous research has demonstrated remarkable generality and precision for a power law of forgetting, the function has been used to describe limited amounts of data. In the examples reported by Wixted and Ebbesen (1991), for instance, the individual forgetting functions were composed of four or five data points. By comparison, when we combined the exemplar-based recognition model with the power law of memory strength, we accounted for enormously rich sets of RT-distribution data at both the individual-subject and individual-lag levels. Extremely accurate accounts of mean correct RTs and error RTs as well as hit rates, false-alarm rates, and d' as a function of lag emerged naturally from the formal model as well.

Second, our research also broached the question of whether the assumption of distinct working memory and long-term memory stores is needed to account for the present forms of data. A long-standing debate is whether working memory involves a time- or capacity-limited system that is distinct from the system that underlies long-term memory (e.g., Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Cowan, 2001; Oberauer, 2002). An alternative view is that the same principles and processes govern both short- and long-term memory (e.g., G. D. A. Brown, Neath, & Chater, 2007; Crowder, 1993; McElree, 2006; Nairne, 2002). The present recognition paradigm involved both short- and long-term probes. All of the recognition data were captured by a model based on the assumption that psychological memory strength is a simple and continuous function of lag. This brings into question whether distinct working memory and long-term memory systems are operating. Nevertheless, we compared one such separate-systems model to the power-law model proposed here.

Experiment

Following a recent paradigm reported by Öztekin, Davachi, and McElree (2010), we presented observers with 12-item memory sets consisting of letters or words. Each set was followed by a test probe, and observers were required to make speeded recognition judgments of whether the probe was old or new. To establish a direct correspondence between psychological and experimenter-defined lag, we provided explicit instructions to control rehearsal strategies. Individual subjects were tested for extended periods, which allowed us to collect detailed RT-distribution and choice-probability data at the level of individual subjects and individual lags.

Participants

Four participants, 2 in the letters condition and 2 in the words condition, completed ten 1-hr sessions on separate days. Participant 4 was the first author. The first 3 participants were reimbursed $12 per session (including a bonus for accuracy above 75%).

Stimuli

In the letters condition, items were 24 English letters (all letters in the alphabet, excluding *A* and *E*). In the words condition, a list of 1,535 one-syllable, three- to six-letter words was generated using the MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm).

Procedure

At the start of each trial, a study list was created by randomly sampling 12 items from the stimulus set. Items were presented in the center of a computer monitor at a visual angle of approximately 3° to 4°. Each trial began with a 500-ms fixation cross. Twelve items were then presented for 500 ms each, with a 100-ms break between items. Participants were given clear and repeated instructions regarding rehearsal. They were asked to say each item (silently, to themselves) once as it was presented and to avoid rehearsing previously presented stimuli or developing any special strategies for remembering items.

After the study list, an asterisk appeared for 500 ms to signal the probe. The probe then appeared and remained on screen until a response was made. Participants indicated whether the probe was old ("F" key) or new ("J" key). Feedback indicating whether their response was correct or incorrect was presented for 1,000 ms, followed by a 1,500-ms intertrial interval.

The probe was a member of the study list (a target) on half of the trials and a randomly sampled item not on the study list (a lure) on the other half of the trials. On target trials, each serial position was probed equally often, with the order of all trials randomized within a block. Each session comprised five blocks of 48 trials. For each participant, 10 sessions yielded 1,200 trials on which the probe was a lure and 100 trials of each serial position when the probe was a target.

Results

Responses faster than 200 ms or slower than 5,000 ms were excluded from analysis. For each participant, and separately for correct and incorrect responses in each of the 13 trial types (12 serial positions and 1 lure), we removed responses slower than 3 standard deviations above the mean RT in that condition. We removed 1.08% of responses in total.

Accuracy (as measured by d') and mean RT for correct responses are plotted as a function of lag for each participant in Figure 1. Although participants were more accurate in recognizing words than letters, the influence of study-probe lag was the same in both conditions: At early lags, performance declined quickly (i.e., RT slowed and d' decreased); at later lags, performance continued to decline, but at a slower rate. There was also a small primacy effect, in which performance

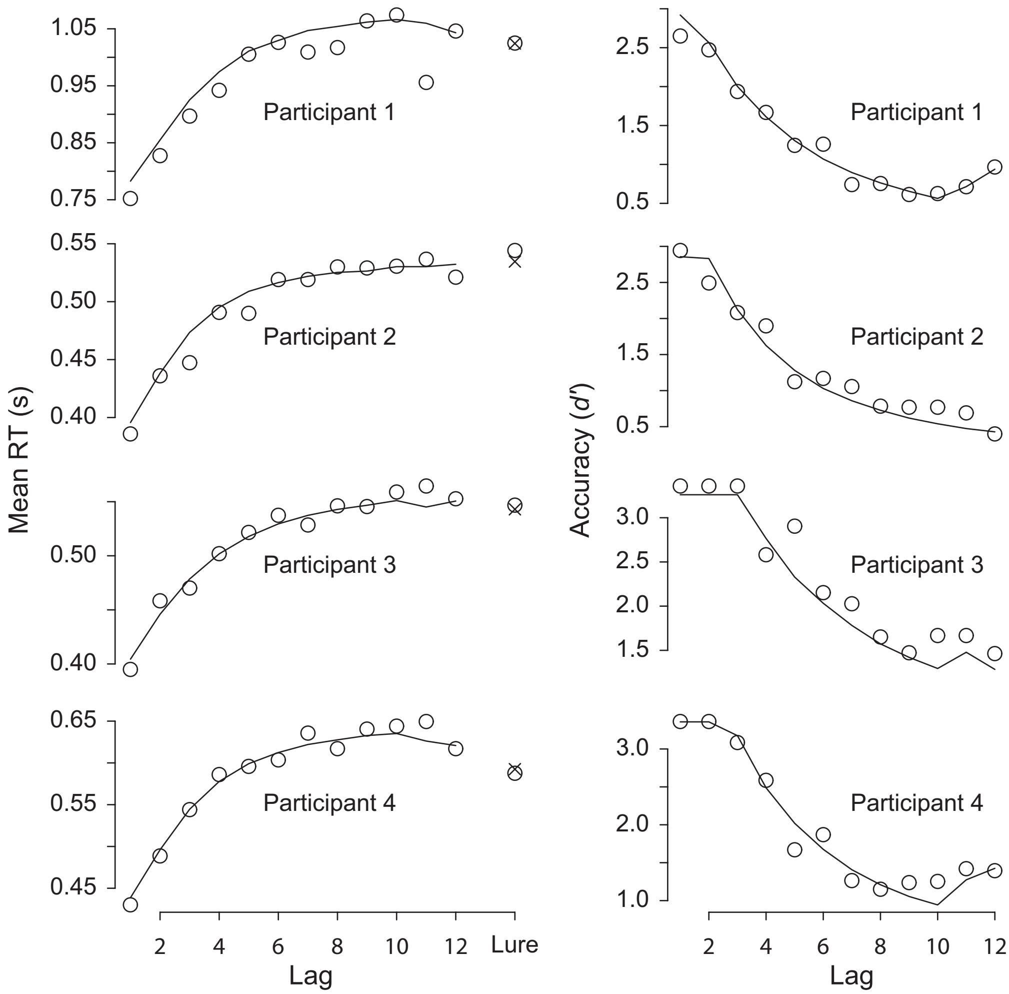

Fig. 1. Mean response time (RT; left column) on trials answered correctly and accuracy (*d*′; right column) as a function of target-probe lag. For RTs, the graphs also show the results for lures. Results are shown separately for each of the 4 participants. Circles represent observed data. Predictions from the exemplar-recognition power-law model are shown using black lines for targets and *X*s for lures. (In computing observed and predicted values of *d*′, maximum hit rates were set at .99.)

for items at the greatest lags (first and second serial positions of the study list) saw a slight boost. Öztekin et al. (2010, Fig. 2) observed a similar pattern of results, although their data were averaged across participants and pairs of adjacent serial positions.

Figure 2 shows detailed RT distributions for each of the study-probe lags for Participant 3. (The RT distributions for the other participants were essentially identical to those of Participant 3; see Figs. S1–S3 in the Supplemental Material available online.) The cumulative-distribution-function plots shown in the figure provide an efficient means of simultaneously illustrating accuracy and the form of the correct (hit) and incorrect (miss) RT distributions in each lag condition. Each plot is made up of quantile estimates from correct and incorrect RT distributions. The quantile estimates (diamond symbols) show the RT below which .10, .30, .50, .70, and .90 of the responses in that distribution fall. The positions of the quantiles on the *x*-axis reflect the speed at which responses are made, so that slower distributions stretch further to the right. The heights of the quantiles indicate, separately for correct and incorrect trials, the absolute cumulative proportion of responses with RTs below the quantile cutoff. Therefore, note that the relative heights of the correct and incorrect distributions reflect the proportion of correct versus incorrect responses at each cumulative RT. The curves reach asymptotes at the overall correct and incorrect response proportions at each lag.

Figure 2 shows that there was a systematic effect of lag on the RT distributions. As lag increases, the median of each distribution (third diamond within each plot) shifts to the right and the distribution becomes more positively skewed (the points spread out to the right). This tendency is pronounced at the early lags and slows down for the larger ones. Figure 3 shows the RT distributions (for all participants) for trials on

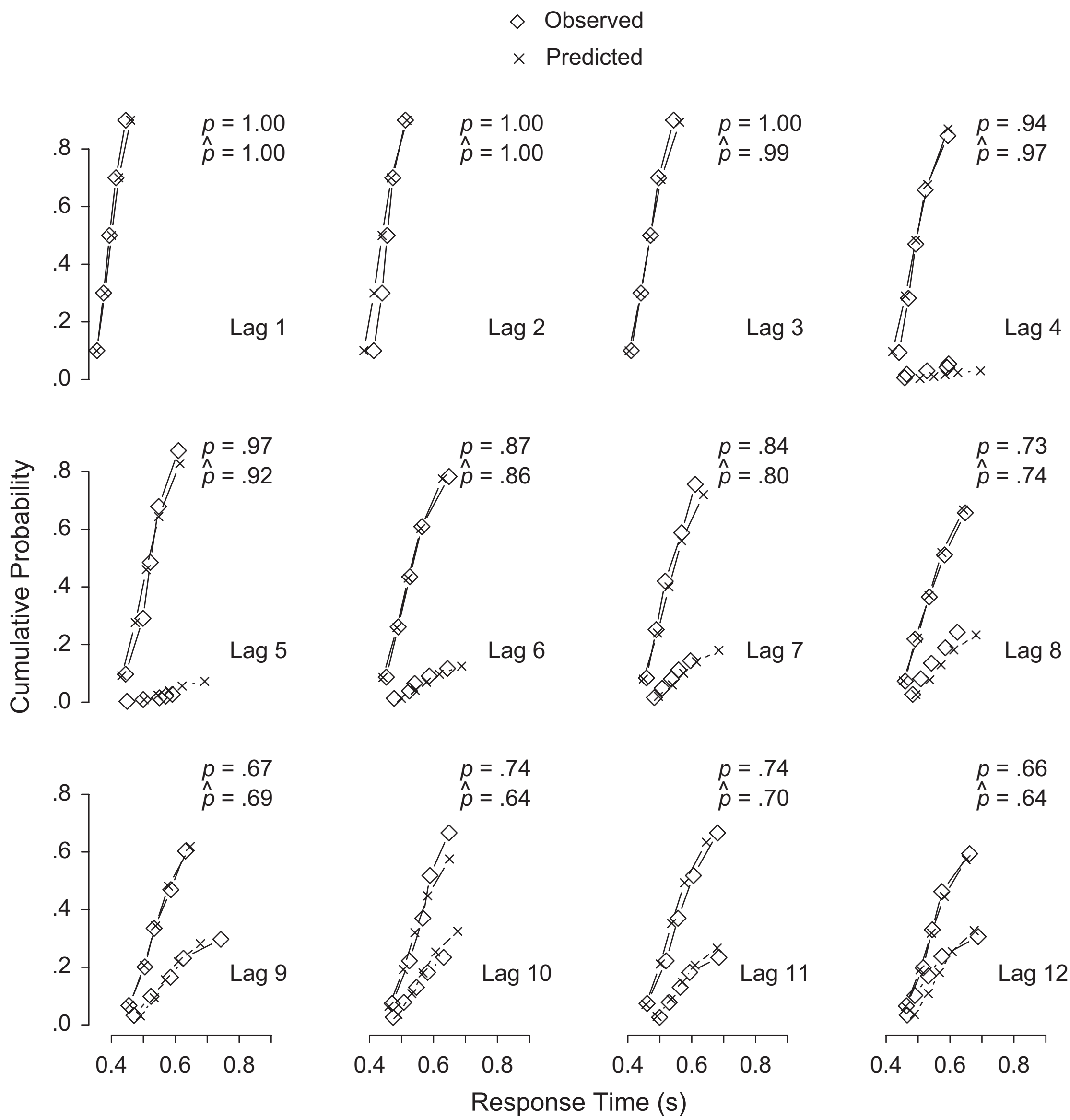

Fig. 2. Cumulative distribution functions for Participant 3 at each of the 12 lags. For Lags 4 through 12, the upper function presents results for trials answered correctly, and the lower function presents results for trials answered incorrectly. For Lags 1 through 3, there were no incorrect responses. Both observed data and predictions from the exemplar-recognition power-law model are shown. For each lag, the observed and predicted proportion of correct responses are shown as p and $\hat{p}$, respectively.

which the probe was a lure. The general form of the lure RT distributions is the same as for the targets.

Next, we demonstrate that all of these results were captured in precise quantitative fashion by a simple version of an exemplar-recognition model that assumed that memory strength was a power function of lag.

The modeling framework

Because we have presented the exemplar-recognition model in previous articles (Nosofsky et al., 2011; Nosofsky & Palmeri, 1997), here we provide only a capsule summary. According to the model, the items from the study list are stored as individual exemplars in memory. Presentation of a test probe causes the exemplars to be activated and retrieved. The retrieved exemplars drive an evidence-accumulation process that leads the observer to decide whether the probed item had been presented in the study list ("old" item) or had not ("new" item).

Formally, according to the model, presentation of probe i leads to overall "activation" (A_i) of the items in the memory set, given by

$$A_i = \sum_j m_j \times s_{ij}, \qquad (1)$$

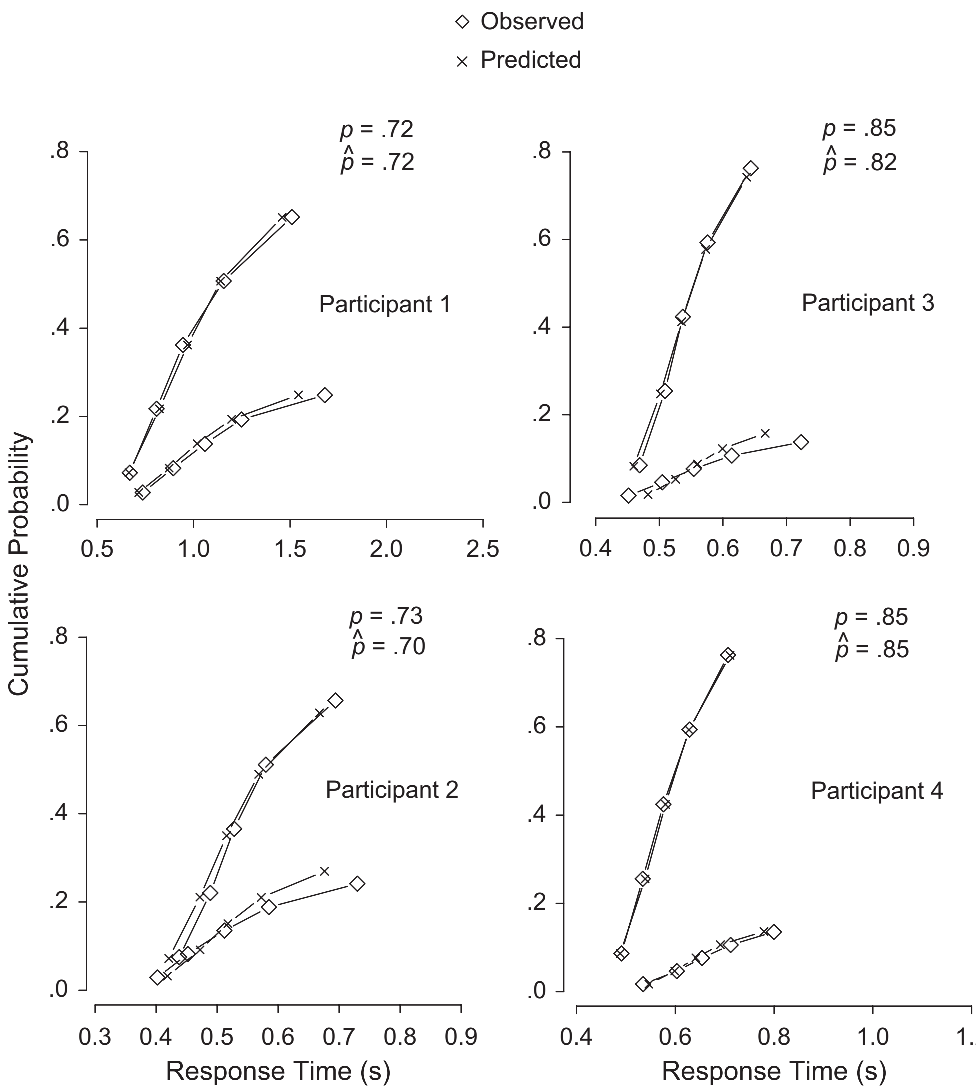

Fig. 3. Cumulative distribution functions for the 4 participants on target-absent (lure) trials. The upper function presents results for trials answered correctly, and the lower function presents results for trials answered incorrectly. Both observed data and predictions from the exemplar-recognition power-law model are shown. For each lag, the observed and predicted proportion of correct responses are shown as p and $\hat{p}$, respectively.

where m_j is the memory strength associated with the item at lag j, and s_{ij} is the similarity of the probe to the item at lag j. For the present paradigm, an exceedingly simple model of similarity was applied: If the probe matched the item at lag j, then s_{ij} was set equal to 1; if the probe did not match the item at lag j, then s_{ij} was set equal to a free parameter s ($0 < s < 1$). In the general version of the model, the individual memory strengths m_j were freely estimated parameters; past evidence from paradigms with small memory-set sizes indicates that the magnitude of the memory strengths decreases with increasing lag (Nosofsky et al., 2011; see also Kahana & Sekuler, 2002).

On the basis of results from applying the general model (see Model Fits), we also applied a special case, in which it was assumed that memory strength was a decreasing power function of lag j:

$$m_j = \alpha \times j^{-\beta}, \tag{2}$$

where β measures the rate at which memory strength declines with lag. (The parameter α is a scaling parameter; without loss of generality, its value can be set equal to 1 in the present model.) Because of small primacy effects in the data, the memory strengths for items in Serial Positions 1 and 2 were scaled by free parameters γ_1 and γ_2.

According to the exemplar model, the mean rate of evidence accumulation supporting an "old" response is given by

$$p_i = A_i/(A_i + k), \tag{3}$$

where k is a drift-rate-criterion parameter. The mean rate of evidence accumulation supporting a "new" response is simply

$q_i = 1 - p_i$. So, for example, if a probe leads to high activation of the memory-set items (e.g., because it matches a recently presented member of the memory set), then there will be a high rate of evidence accumulation supporting an "old" response.

If the accumulating evidence first reaches an "old" response threshold (R_{old}), then the observer responds "old"; if the evidence first reaches a "new" response threshold (R_{new}), then the observer responds "new." The decision time is determined by the time that it takes the accumulating evidence to reach either threshold. In past applications of the model, the evidence-accumulation process was modeled as a *random walk*, in which a single counter moved probabilistically on each step toward either the "old" or "new" threshold. As explained in the appendix, in the present experiment, we applied a slight variant in which the evidence-accumulation process was modeled in terms of linear-ballistic accumulation (S. D. Brown & Heathcote, 2008).

Finally, as is commonly assumed in numerous RT models, the exemplar model included parameters for between-trial variability in the start point of the evidence-accumulation process (a), between-trial variability in the evidence-accumulation rate (σ), and a mean residual time (μ) for processes not associated with decision making (e.g., encoding and response execution). Details are provided in the appendix.

Fitting procedure

Models were fit to the choice and RT from each individual trial using analytic expressions of likelihood (see S. D. Brown & Heathcote, 2008, for details). The general and power-law versions of the exemplar model share seven parameters (see previous discussion, the appendix, and Table 1): s, k, R_{old}, R_{new}, a, σ, and μ. The general (i.e., free-memory-strength) model estimated 12 further m_j parameters, one memory strength for each

Table 1. Parameter Estimates and Bayesian Information Criterion (BIC) Values for Each Participant for Each Version of the Exemplar Model

	Participant 1			Participant 2			Participant 3			Participant 4		
Parameter	General	Power law	Dual system	General	Power law	Dual system	General	Power law	Dual system	General	Power law	Dual system
a	0.50	0.49	0.55	0.11	0.11	0.13	0.07	0.07	0.07	0.00	0.02	0.04
R_{old}	0.77	0.77	0.79	0.25	0.25	0.25	0.25	0.25	0.25	0.30	0.32	0.32
R_{new}	0.78	0.78	0.78	0.29	0.29	0.28	0.27	0.28	0.28	0.28	0.30	0.29
μ	0.20	0.20	0.24	0.17	0.17	0.19	0.16	0.14	0.15	0.10	0.08	0.10
s	0.06	0.03	0.003	0.04	0.04	0.02	0.03	0.05	0.04	0.08	0.05	0.03
k	0.05	0.08	0.02	0.09	0.14	0.15	0.13	0.22	0.35	0.06	0.15	0.18
σ	0.19	0.19	0.17	0.17	0.17	0.17	0.11	0.10	0.11	0.11	0.11	0.11
γ_1	—	2.51	0.97	—	1.00	0.26	—	1.22	0.56	—	2.12	0.79
γ_2	—	1.55	0.57	—	1.00	0.42	—	1.30	0.69	—	1.64	0.71
β	—	1.60	—	—	1.52	—	—	1.11	—	—	1.38	—
m_{wm}	—	—	1.000	—	—	1.000	—	—	1.000	—	—	1.000
m_{ltm}	—	—	0.01	—	—	0.06	—	—	0.20	—	—	0.09
m_1	1.000	—	—	1.000	—	—	1.000	—	—	1.000	—	—
m_2	0.413	—	—	0.210	—	—	0.230	—	—	0.334	—	—
m_3	0.190	—	—	0.142	—	—	0.200	—	—	0.166	—	—
m_4	0.143	—	—	0.088	—	—	0.113	—	—	0.111	—	—
m_5	0.087	—	—	0.050	—	—	0.103	—	—	0.069	—	—
m_6	0.084	—	—	0.043	—	—	0.077	—	—	0.071	—	—
m_7	0.044	—	—	0.043	—	—	0.074	—	—	0.046	—	—
m_8	0.040	—	—	0.029	—	—	0.056	—	—	0.043	—	—
m_9	0.035	—	—	0.026	—	—	0.050	—	—	0.043	—	—
m_{10}	0.033	—	—	0.024	—	—	0.052	—	—	0.044	—	—
m_{11}	0.040	—	—	0.022	—	—	0.053	—	—	0.046	—	—
m_{12}	0.057	—	—	0.013	—	—	0.045	—	—	0.053	—	—
BIC	3,677	**3,611**	3,683	−4,039	**−4,087**	−3,766	−2,902	**−2,954**	−2,639	−1,810	**−1,871**	−1,721

Note: The parameters of the models are defined as follows (see the text and the appendix for further details): a = start-point variability; R_{old} = threshold for "old" responses; R_{new} = threshold for "new" responses; μ = mean residual time; s = similarity; k = drift-rate criterion; σ = drift-rate variability; γ_i = primacy scaling factor for Serial Position i; β = power-function decay rate; m_{wm} = memory strength for items in working memory; m_{ltm} = memory strength for items in long-term memory; m_j = memory strength for lag j. Without loss of generality, m_1 was set to 1 in the general version of the model, m_{wm} was set to 1 in the dual-store version of the model, and α was set to 1 in the power-law version of the model. The parameters a, R_{old}, R_{new}, and μ are measured in seconds. For each participant, the smallest BIC value is in boldface.

lag. The special-case power-law model instead estimated the decay-rate β and primacy parameters γ_1 and γ_2.

Model comparisons were made using the Bayesian information criterion (BIC), which combines measures of absolute fit and a penalty for the number of free parameters to determine which model provides the most parsimonious explanation of the data. Smaller BIC values indicate the preferred model. Details are provided in the appendix.

Model fits

Table 1 reports the BIC values and best-fitting parameter estimates for the general and power-law versions of the exemplar model for each participant. In all cases, the BIC yielded by the power-law model was superior to the one yielded by the more general version. This result indicates that the additional parameters utilized by the more general model did not provide enough of a benefit in quality of fit to offset its increased complexity.

To provide further evidence for the utility of the power-law model, we fitted a variety of competing functions (the same ones considered previously by Wixted & Ebbesen, 1991) to the Lag 1 to Lag 10 memory strengths estimated from the general version of the exemplar model. We found the two free parameters of each function that minimized the sum of squared deviations between the function's predictions and the estimated memory strengths. The results of this analysis are shown in Table 2. Impressively, for all 4 participants, the power function provided the best description by far of the estimated memory strengths. Indeed, as shown in Figure 4, when one plots the estimated memory-strength parameters from the general model against lag, the plots closely follow a power function for all participants.

Most important, the version of the exemplar model that assumes a lag-based power law provides an outstanding account of all of the detailed RT-distribution data. To demonstrate this point, we plotted its predictions of the full RT distributions for correct and incorrect responses in both target-present and target-absent conditions (Figs. 2 and 3). (Recall that Figure 2 shows the target-present results for Participant 3 only; the results for the other 3 participants are equally impressive.) The agreement between the model and the data was excellent, especially given that only 10 parameters were used to simultaneously fit each participant's 26 RT distributions (2,400 data points per participant). Because the model captured the detailed form of the RT distributions for trials answered correctly and incorrectly at each lag, it naturally also accounted extremely well for the overall accuracy and mean RT data (see Fig. 1).

Dual-store model

As another source of comparison, we fitted a representative from the class of dual-store models to the RT-distribution data. In the dual-store model, we assumed that, as each individual item was presented, it entered into a working memory buffer of Size 4 (e.g., Cowan, 2001; Oberauer, 2002). Once the buffer size was exceeded, one of the items already existing in the buffer was displaced into a separate long-term store. The displaced item was assumed to be chosen at random. On the basis of the all-or-none properties deemed to characterize working memory in recent work (e.g., Rouder et al., 2008; Zhang & Luck, 2009), we assumed that all items present in the working memory buffer had equal memory strength m_{wm}. By contrast, items displaced into long-term memory had (lower) average strength m_{ltm}. We also granted the dual-store model the same flexibility as the power-law model by estimating separate primacy-strength parameters for the items in Serial Positions 1 and 2. The manner of computing overall activation and evidence accumulation for this dual-store model was the same as for the other versions of the exemplar model.

Note that the dual-store model is a mixture model. On some proportion of trials, the overall activation yielded by a positive probe will be high because the probed item resides in the working memory buffer. On the remaining proportion of trials, the overall activation will be low, because the item has been displaced into long-term memory. Items with shorter lags will yield a higher proportion of trials with high activations, because there is a higher probability that they still exist in the working memory buffer.

The fits of the dual-store model are shown along with those of the other versions of the exemplar model in Table 1. In all cases, the dual-store model provided a substantially worse account of the data than did the power-law model. Except for

Table 2. Sum of Squared Deviations (SSDs) for Each of the Six Candidate Forgetting Functions (Wixted & Ebbesen, 1991) Fit to Estimated Memory-Strength Parameters

Function	Participant 1	Participant 2	Participant 3	Participant 4
Power: $m = \alpha \times t^{-\beta}$	**0.0027**	**0.0044**	**0.0138**	**0.0008**
Exponential: $m = \alpha \times e^{-\beta t}$	0.0148	0.0205	0.0508	0.0190
Hyperbolic: $m = \alpha/(1 + \beta \times t)$	0.0649	0.1443	0.0955	0.0087
Linear: $m = \alpha - \beta \times t$	0.3537	0.4542	0.4096	0.4007
Logarithmic: $m = \alpha - b \times \log(t)$	0.1349	0.2348	0.2087	0.1790
Exponential power: $m = \alpha \times e^{-2 \times \beta \times \sqrt{t}}$	0.0037	0.0165	0.0329	0.0079

Note: Smaller SSD values indicate better-fitting functions. The smallest SSD for each participant is in boldface. t = lag; m = memory strength.

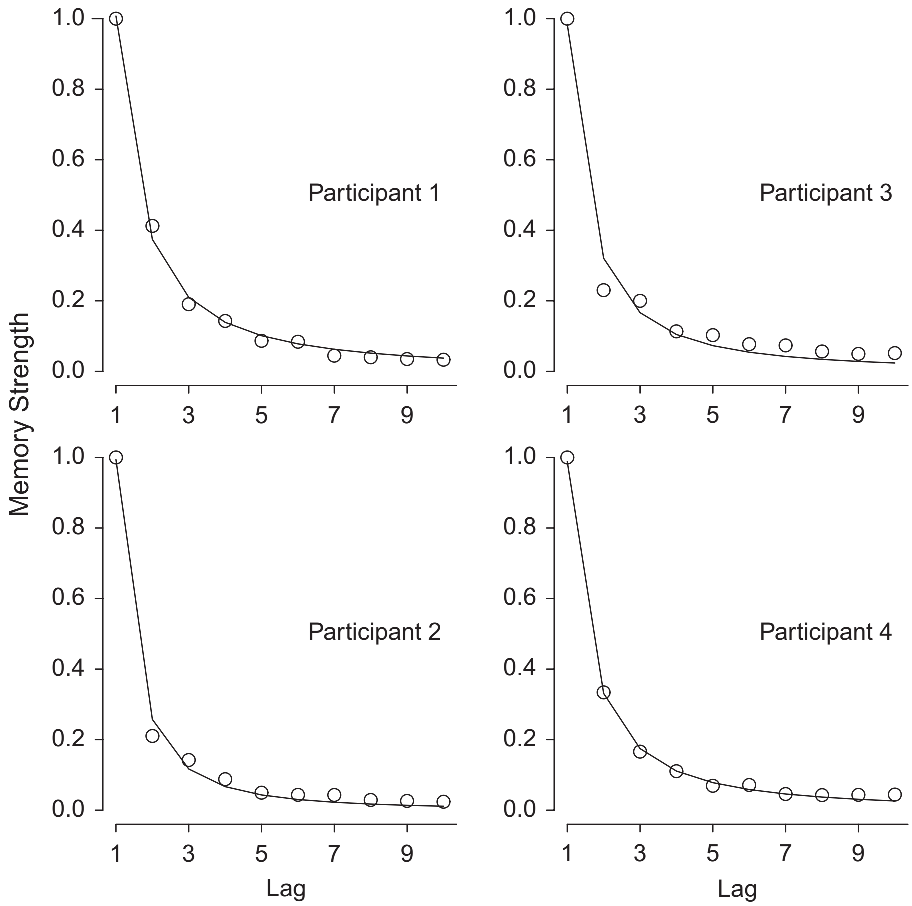

Fig. 4. Memory strength in the general model (circles) and the predicted memory strength from the best-fitting power function (solid lines) as a function of lag. Results are shown separately for each of the 4 participants.

Participant 1, it also fared much worse than did the free-memory-strength version. In our view, the version of the dual-store model that we tested is a reasonable representative, but we acknowledge that a wide variety of alternative dual-store models could be formulated. The precision of fit achieved by the power-law model provides a challenging yardstick to gauge such alternative formulations.

Discussion

In the experiment reported here, we found strong evidence for an exemplar-recognition model that assumed a power-function relation between psychological memory strength and lag. The model, despite its parametric simplicity, produced a remarkably good fit to data from a probe-recognition task, in which lists of 12 items (made of either letters or words) were presented. The model accounted simultaneously for full RT distributions for correct and incorrect responses across all serial positions of target-present as well as target-absent trials. A wide variety of alternative functions failed to capture the relation between estimated memory strength and lag. An alternative dual-store model that posited an all-or-none, limited-capacity working memory buffer also fared considerably worse than did the power-law model.

Because our analysis was aimed at fine-grained RT distributions observed for individual subjects at individual lags, the support for the power law is not due to artifacts involving averaging across subjects (e.g., Anderson & Tweney, 1997; Myung, Kim, & Pitt, 2000; Wixted & Ebbesen, 1997). Furthermore, because the memory-strength parameters were required to account for enormously rich sets of data, there is a great deal of precision in their estimates. Therefore, the support for the power law is unlikely to be due to flexibility in fitting noisy estimates (Lee, 2004).

Analogous to findings in other work involving the discovery of laws at a psychological level of analysis (e.g., Shepard, 1987), the present regularity was uncovered within the framework of a highly successful formal model. It is important to acknowledge, therefore, that the discovered regularity involving the memory strengths is model dependent. However, because the model combines well-established principles of exemplar-based recognition and evidence accumulation for which there is a good deal of consensus, we believe that the discovered regularity is intriguing and important.

Natural questions arise regarding the generality of our findings and the conditions under which the power function holds. From one perspective, past research already suggests considerable generality for a power law of forgetting, because it has been found to hold widely for empirical forgetting functions. Wixted (2004) considered a variety of possible reasons for the emergence of the empirical power law. He suggested that the best explanation was that "the underlying memory traces themselves individually exhibit an ever-decreasing rate of decay" (p. 875), which is a property consistent with power-law forgetting. Our model-based evidence for such power-law decreases in psychological memory strength lends support to Wixted's suggestion. Nevertheless, future research needs to test more directly for the generality with which the psychological power law holds. An interesting possibility is that the power-function relation might hold generally, but the parameters of the function might be systematically influenced by experimental factors, such as the similarity of the items within the study lists and the time interval between each item presentation.

Finally, the present theoretical message has been to provide evidence for a power law of memory strength at a psychological level of analysis. The work does not provide any explanation as to why the power law is favored over various alternatives. In addition, in its present preliminary form, the power law is stated with respect to lag. Future research needs to unpack the detailed psychological and neurological mechanisms, such as decay, interference, and consolidation, which contribute to the form of the lag-based forgetting functions. Such research will lead to the development of more fully specified models and laws of psychological memory strength.

Appendix: Model Details and Fitting Procedure

In past applications of the exemplar model (e.g., Nosofsky et al., 2011), the evidence-accumulation rates defined in Equation 3 were used to drive a discrete random-walk process. In the experiment reported here, we applied a continuous linear-ballistic-accumulation (LBA) process (S. D. Brown & Heathcote, 2008). Past work suggests that the LBA approach yields predictions that are essentially the same as random-walk and diffusion models (Donkin, Brown, Heathcote, & Wagenmakers, 2011). An advantage of the LBA approach relative to the random-walk model, however, is that analytic expressions have been developed to allow maximum-likelihood fits to RT-distribution data without requiring simulation. In addition, the continuous nature of the LBA model provides for more graceful parameter-search methods.

For the current exemplar-based LBA model, two accumulators were established, one for "old" responses and one for "new" responses. On each trial, the evidence-accumulation rate on each accumulator was an independent and randomly sampled value from normal distributions with means given by the values p_i and q_i computed from Equation 3, and common standard deviation σ. The start point of each accumulator is an independent and randomly sampled value from uniform distribution $[0, a]$. As assumed in all LBA applications, evidence accumulates at a linear and ballistic rate (i.e., without moment-to-moment noise) until a response threshold is reached in either accumulator. The response whose threshold was reached is made, and the time taken for evidence to reach that threshold is the decision time for that trial. The predicted RT is the sum of the decision time and nondecision time, μ.

The fits of the models were evaluated using the BIC:

$$\text{BIC} = -2 \times \ln L + P \times \ln N, \quad \text{(A1)}$$

where L is the (maximum) likelihood of the data given the model parameters, P is the number of free parameters, and N is the total number of data points. Smaller BIC values indicate the preferred model. An identical pattern of model-fitting results and conclusions was obtained when we used the alternative Akaike information criterion as a criterion of fit.

Acknowledgments

The authors would like to thank Jerome Busemeyer, Rich Shiffrin, and Jim Townsend for helpful discussions, and two anonymous reviewers for their comments on an earlier version of this manuscript.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by Grant FA9550-08-1-0486 from the Air Force Office of Scientific Research to Robert M. Nosofsky.

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

References

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

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Advances in the psychology of learning and motivation: Vol. 2. Research and theory* (pp. 89–195). New York, NY: Academic Press.

Baddeley, A. D., & Hitch, G. J. L. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

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

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition*, *21*, 142–145.

Donkin, C., Brown, S. D., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, *18*, 61–69.

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, *42*, 2177–2192.

Lee, M. D. (2004). A Bayesian analysis of retention functions. *Journal of Mathematical Psychology*, *48*, 310–321.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp. 155–200). San Diego, CA: Academic Press.

Myung, J., Kim, C., & Pitt, M. A. (2000). Toward an explanation of the power law artifact: Insights from response surface analysis. *Memory & Cognition*, *28*, 832–840.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81.

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*, 280–315.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 411–421.

Öztekin, I., Davachi, L., & McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? Neural evidence in support of a single store. *Psychological Science*, *21*, 1123–1133.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences, USA*, *105*, 5975–5979.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

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

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

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

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, *2*, 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*, 731–739.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*, 423–428.