Fundamental Law of Memory Recall

Michelangelo Naim[1+], Mikhail Katkov[1+], Misha Tsodyks[1*]

1 Department of Neurobiology, Weizmann Institute of Science, Rehovot 76000, Israel
+ these authors contributed equally to this work

*** misha@weizmann.ac.il**

Abstract

Free recall of random lists of words is a standard way to probe human memory. We proposed the associative search process that can be mathematically solved, providing an analytical prediction for the average number of words recalled from a list of an arbitrary length. Previously reported free recall performance depended on experimental details. Since recall could be affected by variability in words acquisition, we designed a protocol where participants performed both recall and recognition trials, using the latter to estimate the number of acquired words. The results closely match theoretical prediction. We conclude that memory recall operates according to a stereotyped search process common to all people.

Keywords: Model, Neural Network, Free recall, Working Memory, Theory

Significance Statement

The main contribution of this work is that an analytical expression for a performance of human participants in a high-level cognitive task (memory recall of random lists of words) that was derived mathematically from a set of simple assumptions, was confirmed experimentally to a remarkable precision. This level of precision of an analytical model is common for physical theories, but is believed to be impossible for biological systems. The results show that some aspects of our cognition are universal for all people and can be predicted theoretically from first principles.

INTRODUCTION

Humans exhibit remarkable proficiency in reciting poems, participating in performances and giving long talks. However, recalling a collection of unrelated events is challenging. To understand human memory one needs to understand both the ability to acquire vast amounts of information and at the same time the limited ability to recall random material. The standard experimental paradigm to address the later question is free recall (e.g. see Kahana 2012). Typical experiments involve recalling randomly assembled lists of words in an arbitrary order after a brief exposure. It was observed over the years that when the presented list becomes longer, the average number of recalled words grows but in a sublinear way, such that the fraction of words recalled steadily decreases (Binet and Henri 1894, Standing 1973, Murray et al. 1976). The exact mathematical form of this relation is controversial and was found to depend on the details of experimental procedures, such as presentation rate (Waugh 1967). In some studies, recall performance was argued to exhibit a power-law relation to the number of presented words (Murray et al. 1976), but parameters of this relation were not determined precisely.

Several influential computational models of recall were developed in cognitive literature that incorporate interactive probabilistic search processes (see e.g. Raaijmakers and Shiffrin 1980, Gillund and Shiffrin 1984, Howard and Kahana 2002, Laming 2009, Polyn et al. 2009, Lehman and Malmberg 2013). In particular, the closest to our approach is the famous 'Search of Associative Memory' model (SAM) that considers sequential sampling and retrieval steps triggered by associations between words that are built up during the acquisition of the presented list (Raaijmakers and Shiffrin 1980). This and other cognitive models have multiple free parameters that can be tuned to reproduce the experimental results on recall quite precisely, including not only the number of words recalled but also the temporal regularities of recall, such as primacy, recency and temporal contiguity effects (Murdock Jr 1962, Murdock and Okada 1970, Howard and Kahana 1999). However, most of the free parameters lack clear biological meaning and cannot be constrained before the data is collected, hence the models cannot be used to predict the recall performance but only explain it a posteriori.

In our recent publications (Romani et al. 2013, Katkov et al. 2017) we proposed a *deterministic* step-by-step associative algorithm based on two basic principles:

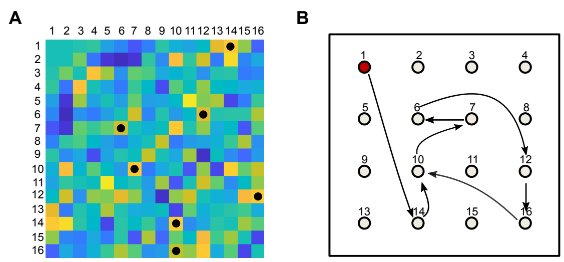

Figure 1. Associative search model of free recall.
(**A**) Matrix of overlaps for a list of 16 items (schematic). For each recalled item, the maximal element in the corresponding row is marked with a black spot.
(**B**) A graph with 16 nodes illustrates the words in the list. Recall trajectory begins with the first node, and converges to a cycle after the 10^{th} node is visited for the second time.

- Memory items are represented in the brain by sparse neuronal ensembles in dedicated memory networks;
- Next item to be recalled is the one that has a representation with a largest overlap to the current one, unless this item is the one that was recalled on the previous step.

The relation of this model to SAM is discussed later. We showed that transition rule proposed above can be implemented in attractor neural networks via modulation of feedback inhibition (Recanatesi et al. 2015, 2017). It is illustrated in Fig. 1 (more details in Methods), where the matrix of overlaps between 16 memory representations is shown in the left panel. When the first item is recalled (say the 1st one in the list), the corresponding row of the matrix, which includes the overlaps of this item with all the others, is searched for the maximal element (14th element in this case), and hence the 14th item is recalled next. This process continues according to the above rule, unless it points to an item that was just recalled in the previous step, in which case the next largest overlap is searched. After a certain number of transitions, this process begins to cycle over already visited items, such that no new items can longer be recalled (Fig. 1b). As shown in (Romani et al. 2013), when memory representations are very sparse, the overlaps between the items can be approximated by a *random* symmetric matrix and the resulting model does not have a single free parameter. Moreover, one can derive a universal expression for the average number of recalled words from a list of length L, that we call Recall Capacity (RC):

$$\begin{aligned} RC &= k \cdot \sqrt{L} \\ k &\approx 2.1 \end{aligned} \tag{1}$$

We emphasize that Eq. (1) does not have any free parameters that could be tuned to fit the experimental results, rather both the exponent and coefficient of this power law expression are a result of the assumed recall mechanism and hence cannot be adjusted; in other words this equation constitutes a true prediction regarding the recall performance as opposed to earlier theoretical studies. Here we present the results of our experiments designed to test this prediction.

RESULTS

The universality of the above analytical expression for RC seems to be at odds with previous experiments that show that performance in free recall task strongly depends on the experimental protocol, for example presentation speed during the acquisition stage (see e.g. Murdock Jr 1960, 1962, Roberts 1972, Howard and Kahana 1999, Kahana et al. 2002, Zaromb et al. 2006, Ward et al. 2010, Miller et al. 2012, Grenfell-Essam et al. 2017) and the extent of practice (Klein et al. 2005, Romani et al. 2016). Since most of the published studies only considered a limited range of list lengths, we performed free recall experiments on Amazon Mechanical Turk® platform for list lengths of $8, 16, 32, 64, 128, 256$ and 512 words, and two presentation speeds: 1 and 1.5 seconds per word. To avoid practice effects, each participant performed a single free recall trial with a randomly assembled list of words of a given length. The results confirm previous observations that recall performance improves as the time allotted for acquisition of each word increases, approaching the theoretical prediction of Eq. (1) from below (see Fig. 2a).

We reasoned that some or all of the deviation of the observed RC from the theoretically predicted one could be due to the non-perfect acquisition of words during the presentation phase of the experiment, such that some of the presented words are not encoded in memory well enough to be candidates for recall. It seems reasonable that acquisition depends on various factors, such as attention, age of participants, acquisition speed, etc. We therefore conjectured that differences in acquisition could be the main cause of variability in published studies, while subsequent recall proceeds according to the universal search process proposed in (Romani et al. 2013). One should then correct Eq. (1) for RC, replacing the number of presented words L with the number of effectively acquired words M:

$$\begin{aligned} RC &= k \cdot \sqrt{M} \\ k &\approx 2.1 \end{aligned} \tag{2}$$

To test this conjecture, we employed additional *recognition* experiments in order to independently evaluate the number of words effectively acquired by participants. Following Standing [1973], we presented participants with pairs of words, one from the list just presented and one a randomly chosen lure, requesting them to report which word was from the presented list. The number of words acquired was then estimated from the number of correct recognitions (see Methods). Importantly, in order to control for the inter-subject variability, same subjects performed recall and recognition experiments for each list length. Moreover, each participant performed a single recognition test, to avoid the well known effect of 'output interference' between subsequent recognition tests (see e.g. Criss et al. 2011).

Fig. 2b shows the estimated average number of acquired words M as a function of list length L, compared to the results of (Standing 1973) who used presentation rate of 5.6 seconds per word (see Methods for details of analysis). Results confirm that acquisition improves with time allotted to presentation of each word. Standard error of the mean for the number of acquired words across participants, for each list length and each presentation speed, was estimated with a bootstrap procedure by randomly sampling a list of participants with replacement (Efron and Tibshirani 1994, see Methods).

In Fig. 2c experimentally obtained RC (yellow and green lines) is compared with the theoretical prediction of Eq. (2) (black line), where M is the average number of encoded words, estimated in the recognition experiment. Remarkably, agreement between the data and theoretical prediction is very good for both presentation speeds, even though the number of acquired and recalled words is very different in these two conditions for each value of list length. We also performed multiple simulations of our recall algorithm (Romani et al. 2013, Katkov et al. 2017) and found that it captures the statistics of the recall performances as accessed with bootstrap analysis of the results (see Fig. S1 in Supplementary materials).

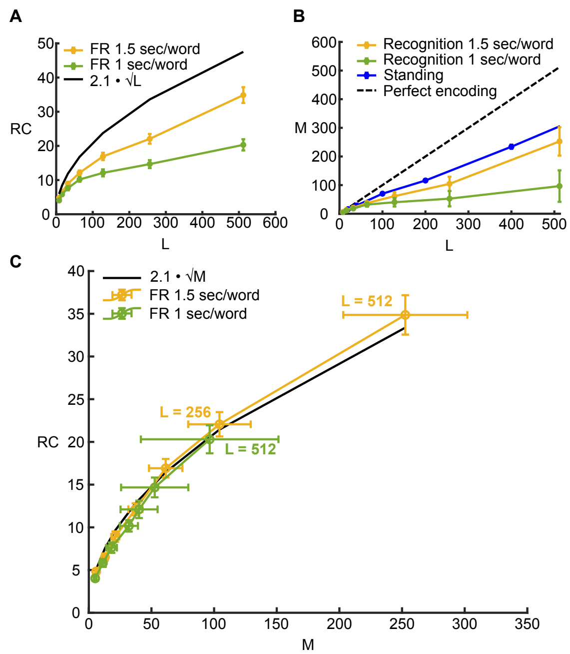

Figure 2. Recall and recognition performance. (**A**) Average number of words recalled as a function of the number of words presented. Black line: Eq. (1). Yellow line: experimental results for presentation rate 1.5 sec/word. Green line: experimental results for presentation rate 1 sec/word. The error in RC is a standard error of the mean. (**B**) Estimated average number of acquired words for lists of different lengths. Black dashed line corresponds to perfect encoding, yellow line corresponds to presentation rate 1.5 sec/word and green line to presentation rate 1 sec/word. The error in M is computed with bootstrap procedure (Efron and Tibshirani 1994). Blue line corresponds to the results of (Standing 1973). (**C**) Average number of words recalled as a function of the average number of acquired words. Black line: theoretical prediction, Eq. (2). Yellow line: experimental results for presentation rate 1.5 sec/word. Green line: experimental results for presentation rate 1 sec/word. The error in RC is a standard error of the mean, while the error in M is computed with bootstrap procedure (see Methods for details

DISCUSSION

The results presented in this study show that average performance in free recall experiments can be predicted from the number of words effectively acquired during presentation with remarkable precision by the analytical, parameter-free expression Eq. (2), derived from a deterministic associative search model of recall. The relation between these two independently measured quantities holds even though both of them strongly depend on the presentation speed of the words. Hence it appears that memory recall is a much more universal process than memory acquisition, at least when random material is involved. Since our theory is not specific to the nature of the material being acquired, we conjecture that recall of different types of information, such as e.g. randomly assembled lists of sentences or pictures, should result in similar recall performance.

Our recall model can be viewed as a radically simplified version of the classical 'Search of Associative Memory' model (SAM), see Raaijmakers and Shiffrin 1980. While the detailed comparison between the models goes beyond the scope of this paper, we point out some of the crucial similarities and differences here. In both models, recall is triggered by a matrix of associations between the items, which in SAM is built up during presentation according to a rather complex set of processes and updated during retrieval of each new item, while in our model is simply assumed to be a fixed, structure-less symmetric matrix (see Fig. 1 above). Since inter-item associations in SAM depend on the presentation order, the model reproduces some of the order-specific aspects of recall, such as primacy and recency effects, while we assume that associations are dominated by long-term representations, which accounts for diversity in the intrinsic recall probability of different words (Katkov et al. 2015) but fails to account for order-specificity of recall (but see Katkov et al. 2015 for the extended version of our model that does account for temporal effects). Subsequent recall in SAM proceeds as a series of attempted probabilistic samplings and retrievals of memory items, until a certain limiting number of failed attempts is reached after which recall terminates. In our model, this is replaced by a deterministic transition rule that selects the next item with the strongest association to the currently recalled one. As a result, recall of new items terminates automatically when a certain transition repeats and the algorithm begins to cycle over already recalled items, without a need to any arbitrary stopping rule. Finally, SAM assumes that all the presented words are stored into long-term memory to different degrees, i.e. could in principle be recalled, while in the current study we assume that only a certain fraction of words are effectively acquired to become candidates for recall, the process that we don't model explicitly but rather access with recognition experiments. We consider it little short of a mystery that with these radical simplifications, the model predicts the recall performance with such a remarkable precision and without a need to tune a single parameter, which suggests that despite all the simplifications, it faithfully captures a key first-order effect in the data. Future theoretical and experimental studies should be pursued to access which aspects of the models are valid and which are crucial for the obtained results.

ACKNOWLEDGMENTS

This work is supported by the EU-H2020-FET 1564 and EU-M-GATE 765549 and Foundation Adelis. We thank Drs. Eli Nelken and Larry Abbott for helpful comments on the manuscript.

Methods

Participants, Stimuli and Procedure

In total 723 participants, were recruited to perform memory experiments on Amazon Mechanical Turk® (https://www.mturk.com). Ethics approval was obtained by the IRB (Institutional Review Board) of the Weizmann Institute of Science. Each participant accepted an informed consent form before participation and was paid from 50 to 85 cents for approximately $5-25$ min, depending on the task. Presented lists were composed of non-repeating words randomly selected from a pool of 751 words produced by selecting English words (Healey et al. 2014) and then maintaining only the words with a frequency per million greater than 10 (Medler and Binder 2005). The stimuli were presented on the standard Amazon Mechanical Turk® web page for Human Intelligent Task. Each trial was initiated by the participant by pressing "Start Experiment" button on computer screen. List presentation followed 300 ms of white frame. Depending on the experiment, each word was shown within a frame with black font for 500, 1000 ms followed by empty frame for 500 ms. After the last word in the list, there was a 1000 ms delay before participant performed the task. The set of list lengths was: 8, 16, 32, 64, 128, 256 and 512 words. Each participant performed experiment A (free recall) and Experiment B (recognition) with lists of the same length. In more details

- 348 participants performed the two experiments with presentation rate of 1.5 sec/word: 265 participants did both experiments for only one list length, 54 for two list lengths, 18, 9 and 2 for 3, 4 and 5 list lengths respectively.
- 375 participants performed the two experiments with presentation rate of 1 sec/word: 373 participants did both experiments for only one list length, 2 for two list lengths.

Experiment A - Free recall. Participants were instructed to attend closely to the stimuli in preparation for the recalling memory test. After presentation and after clicking a "Start Recall" button, participants were requested to type in as many words as they could in any order. After the completion of a word (following non-character input) the word was erased from the screen, such that participants were seeing only the currently typed word. Only one trial was performed by each participant. The time for recalling depended on the length of the learning set, from 1 minute and 30 seconds up to 10 minute and 30 seconds, with a 1 minute and 30 seconds increase for every length doubling. The obvious misspelling errors were corrected. Repetitions and the intrusions (words that were not in the presented list) were ignored during analysis.
Experiment B - Recognition task. In recognition trial, participants were shown 2 words, one on top of another. One word was randomly selected among just presented words (target), and another one was selected from the rest of the pool of words. The vertical placement of the target was random. After presentation and after clicking a "Start Recognition" button, participants were requested to click on the words they think was presented to them during the trial. Each list was followed with 5 recognition trials per participant, but only the first trial was considered in the analysis. Time for all trials was limited to 45 min, but in practice each response usually took less than two seconds.

Analysis of the results

The average number of recalled words (RC) for each list length and its standard error were obtained from the distribution of the number of recalled words across participants.

The average number of words acquired for each list length L was computed from the results of recognition experiments as in (Standing 1973). Suppose that M out of L

words are remembered on average after an exposure to the list, the rest are missed. The chance that one of the acquired words is presented during a recognition trial is then M/L, while the chance that a missed word is presented is $1 - M/L$. We assume that in the first case, a participant correctly points to a target word, while in the second case, she/he is guessing. The fraction of correct responses C can then be computed as

$$C = \frac{M}{L} + \frac{1}{2} \cdot \left(1 - \frac{M}{L}\right) . \tag{3}$$

Hence the average number of remembered words can be computed as

$$M = L \cdot (2C - 1) . \tag{4}$$

In order to estimate a standard error of the mean for the number of acquired words across participants, for each list length, we performed a bootstrap procedure (Efron and Tibshirani 1994). We generated multiple bootstrap samples by randomly sampling a list of N participants with replacement N times. Each bootstrap sample differs from the original list in that some participants are included several times while others are missing. For each bootstrap sample b out of total number B, with $B = 500$, we compute the estimate for the average number of acquired words, $M(b)$, according to Eq. (4). The standard error of M is then calculated as a sample standard deviation of B values of $M(b)$:

$$se_B = \sqrt{\sum_{b=1}^{B} \frac{\left(M\left(b\right) - \bar{M}\right)^2}{B-1}} , \tag{5}$$

where $\bar{M} = \sum_{b=1}^{B} \frac{M(b)}{B}$.

Recall model

Our recall model is presented in more details in Romani et al. 2013, Katkov et al. 2017. In this contribution we simulated a simplified version of the model, where we approximate the matrix of overlaps between random sparse memory representations by a random symmetric L by L matrix where L is a number of words in the list, and each element is chosen independently from a normal distribution. A new matrix is constructed for each recall trial. The sequence $\{k_1, k_2, \ldots, k_r\}$ of recalled items is defined as follows. Item k_1 is chosen randomly among all L presented items with equal probability. When n items are recalled, the next recalled item k_{n+1} is the one that has the maximal overlap with the currently recalled item k_n, excluding the item that was recalled just before the current one, k_{n-1}. After the same transition between two items is experienced for the second time, the recall is terminated since the model enters into a cycle.

References

Alfred Binet and Victor Henri. Mémoire des mots. *L'Année psychologique*, 1(1):1–23, 1894.

Amy H Criss, Kenneth J Malmberg, and Richard M Shiffrin. Output interference in recognition memory. *Journal of Memory and Language*, 64(4):316–326, 2011.

Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

Gary Gillund and Richard M Shiffrin. A retrieval model for both recognition and recall. *Psychological review*, 91(1):1, 1984.

Rachel Grenfell-Essam, Geoff Ward, and Lydia Tan. Common modality effects in immediate free recall and immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(12):1909, 2017.

M Karl Healey, Patrick Crutchley, and Michael J Kahana. Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology: General*, 143(4):1553, 2014.

Marc W Howard and Michael J Kahana. Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4):923, 1999.

Marc W Howard and Michael J Kahana. A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3):269–299, 2002.

Michael J Kahana, Marc W Howard, Franklin Zaromb, and Arthur Wingfield. Age dissociates recency and lag recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3):530, 2002.

Michael Jacob Kahana. *Foundations of human memory.* OUP USA, 2012.

M Katkov, S Romani, and M Tsodyks. Memory retrieval from first principles. *Neuron*, 94(5):1027–1032, 2017.

Mikhail Katkov, Sandro Romani, and Misha Tsodyks. Effects of long-term representations on free recall of unrelated words. *Learning & Memory*, 22(2):101–108, 2015.

Krystal A Klein, Kelly M Addis, and Michael J Kahana. A comparative analysis of serial and free recall. *Memory & Cognition*, 33(5):833–839, 2005.

Donald Laming. Failure to recall. *Psychological Review*, 116(1):157, 2009.

Melissa Lehman and Kenneth J Malmberg. A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, 120(1):155, 2013.

DA Medler and JR Binder. Mcword: An on-line orthographic database of the english language, 2005.

Jonathan F Miller, Christoph T Weidemann, and Michael J Kahana. Recall termination in free recall. *Memory & cognition*, 40(4):540–550, 2012.

Bennet B Murdock and Ronald Okada. Interresponse times in single-trial free recall. *Journal of Experimental Psychology*, 86(2):263, 1970.

Bennet B Murdock Jr. The immediate retention of unrelated words. *Journal of Experimental Psychology*, 60(4):222, 1960.

Bennet B Murdock Jr. The serial position effect of free recall. *Journal of experimental psychology*, 64(5):482, 1962.

DJ Murray, Carol Pye, and WE Hockley. Standing's power function in long-term memory. *Psychological Research*, 38(4):319–331, 1976.

Sean M Polyn, Kenneth A Norman, and Michael J Kahana. A context maintenance and retrieval model of organizational processes in free recall. *Psychological review*, 116(1): 129, 2009.

Jeroen GW Raaijmakers and Richard M Shiffrin. Sam: A theory of probabilistic search of associative memory. In *Psychology of learning and motivation*, volume 14, pages 207–262. Elsevier, 1980.

Stefano Recanatesi, Mikhail Katkov, Sandro Romani, and Misha Tsodyks. Neural network model of memory retrieval. *Frontiers in Computational Neuroscience*, 9:149, 2015.

Stefano Recanatesi, Mikhail Katkov, and Misha Tsodyks. Memory states and transitions between them in attractor neural networks. *Neural Computation*, 29(10): 2684–2711, 2017.

William A Roberts. Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology*, 92(3):365, 1972.

Sandro Romani, Itai Pinkoviezky, Alon Rubin, and Misha Tsodyks. Scaling laws of associative memory retrieval. *Neural computation*, 25(10):2523–2544, 2013.

Sandro Romani, Mikhail Katkov, and Misha Tsodyks. Practice makes perfect in memory recall. *Learning & Memory*, 23(4):169–173, 2016.

Lionel Standing. Learning 10000 pictures. *The Quarterly journal of experimental psychology*, 25(2):207–222, 1973.

Geoff Ward, Lydia Tan, and Rachel Grenfell-Essam. Examining the relationship between free recall and immediate serial recall: the effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5): 1207, 2010.

Nancy C Waugh. Presentation time and free recall. *Journal of Experimental Psychology*, 73(1):39, 1967.

Franklin M Zaromb, Marc W Howard, Emily D Dolan, Yevgeniy B Sirotin, Michele Tully, Arthur Wingfield, and Michael J Kahana. Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4):792, 2006.

Supplemental Information

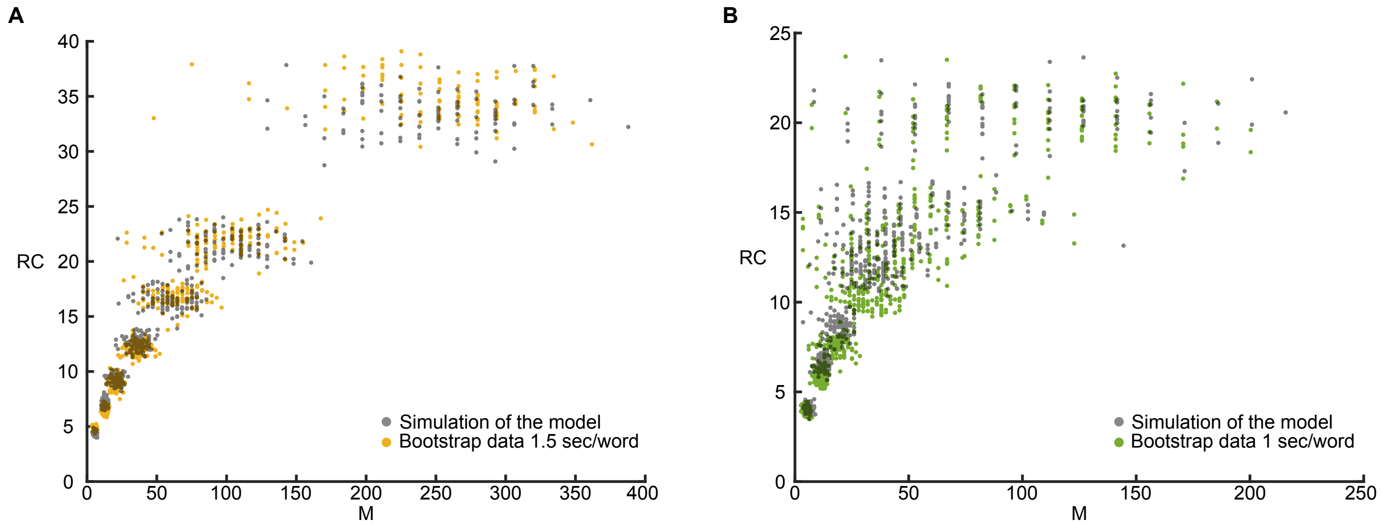

Figure S1. Bootstrap analysis and comparison to model simulations.
(**A**) 1.5 seconds per word presentation rate; (**B**) 1 seconds per word presentation rate.

100 bootstrap samples for each list length are shown with colored dots with coordinates $M(b)$ and $RC(b)$, where $RC(b)$ is an average number of recalled words computed for each bootstrap sample b. Black dots show corresponding simulation results, obtained as follows. From the results of recognition experiment, we calculate, for each list length L, the fraction of correct recognitions across the participants, c, and therefore the probability $p = (2c - 1)$ that a presented word is acquired. With these two numbers, we simulate multiple recognition and recall experiments. For recognition experiment, we draw a binomial random variable with probability c for each participant independently, simulating their recognition answers, from which we compute the number of acquired words averaged for all participants as explained in the Methods. We then drew L binomial variables with probability p for each participant, simulating the acquisition of words by this participant during the recall experiment. With the number of acquired words known for each participant, we run the recall model (see Methods) to obtain the average recall performance over participants. Every simulation described above produced 7 pairs of results (M, RC), one per list length. We repeated the whole procedure 100 times, same as the number of bootstrap samples.