

## Fundamental Law of Memory Recall

Michelangelo Naim<sup>1,†</sup>, Mikhail Katkov<sup>1,†</sup>, Sandro Romani<sup>2</sup>, and Misha Tsodyks<sup>1,3,\*</sup>

<sup>1</sup>*Department of Neurobiology, Weizmann Institute of Science, Rehovot 76000, Israel*

<sup>2</sup>*Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA*

<sup>3</sup>*The Simons Center for Systems Biology, Institute for Advanced Study, Princeton, New Jersey 08540, USA*

 (Received 8 May 2019; revised manuscript received 16 May 2019; published 10 January 2020)

Human memory appears to be fragile and unpredictable. Free recall of random lists of words is a standard paradigm used to probe episodic memory. We proposed an associative search process that can be reduced to a deterministic walk on random graphs defined by the structure of memory representations. The corresponding graph model can be solved analytically, resulting in a novel parameter-free prediction for the average number of memory items recalled ( $R$ ) out of  $M$  items in memory:  $R = \sqrt{3\pi M/2}$ . This prediction was verified with a specially designed experimental protocol combining large-scale crowd-sourced free recall and recognition experiments with randomly assembled lists of words or common facts. Our results show that human memory can be described by universal laws derived from first principles.

DOI: [10.1103/PhysRevLett.124.018101](https://doi.org/10.1103/PhysRevLett.124.018101)

Human cognition is typically considered to be too complex to be described by physics-style universal mathematical laws (see a notable exception in the form of a universal law of generalization proposed in [1]). Human memory in particular is a critically important mental capacity that includes multiple processes, most crucially acquisition, maintenance, and recall (see, e.g., [2]). While human memory capacity for information is practically infinite, all of the mentioned processes are not entirely reliable, for example recall is often a challenging task even when information being recalled is encoded in memory. An important advantage for studying recall is that it can be precisely quantified with a classical paradigm of “free recall” (see, e.g., [3]). Typical experiments involve recalling randomly assembled lists of words in an arbitrary order after a brief exposure. When the presented list becomes longer, the average number of recalled words grows but in a sublinear way ([4–6]). The exact mathematical form of this relation is controversial and was found to depend on the details of experimental procedures, such as presentation rate ([7]). In some studies, recall performance was found to exhibit a power-law relation to the number of presented words ([6]), but parameters of this relation were extremely variable across different experimental conditions.

These observations seem to rule out any possibility that memory recall can be described by a universal mathematical law that would hold for all experimental conditions and

all people. Yet in this study we demonstrate with new experiments that most of the variability in recall can be accounted for by measuring the acquisition and maintenance of information during the presentation phase of the experiment, and when that is controlled, recall itself is much more predictable. Moreover, relation between the number of items in memory and the average fraction of it that can be successfully recalled is described by a parameter-free analytical expression derived here from the deterministic model introduced in [8,9]. In other words, despite the overall unpredictability of human memory, some aspects of it obey simple universal laws.

The proposed recall process is based on two principles: (i) memory items are represented in the brain by overlapping random sparse neuronal ensembles in dedicated memory networks. (ii) The next item to be recalled is the one with a largest overlap to the current one, excluding the item that was recalled on the previous step. The first principle is a common element of most neural network models of memory (see, e.g., [10,11]), while the second one is specific to our model, expressing an associativity principle in neuronal terms. More specifically, item representations are chosen as random binary  $\{0,1\}$  vectors where each element of the vector chosen to be 1 with small probability  $f \ll 1$  independently of other elements. Overlaps are defined as scalar products between these representations. The model is illustrated in Fig. 1 (more details in Supplemental Material [12]), where the matrix of overlaps (“similarity matrix,” or SM) between 16 memory representations is shown in Fig. 1(a). Figure 1(b) is a graph that shows the transitions between memory items induced by the SM. When the first item is recalled (say the first one in the list), the corresponding row of the matrix, which includes the overlaps of this item with all the others, is

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

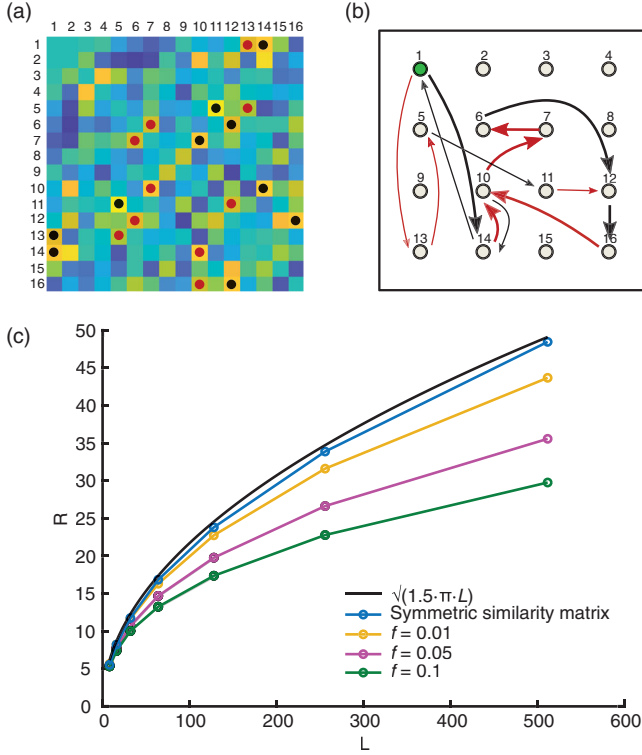


FIG. 1. Associative search model of free recall. (a) SM (similarity matrix) for a list of 16 items (schematic). For each recalled item, the maximal element in the corresponding row is marked with a black spot, while the second maximal element is marked with a red spot. (b) A graph with 16 nodes illustrates the items in the list. Recall trajectory begins with the first node, and proceeds to an item with the largest similarity to the current one (black arrow) or the second largest one (red arrow) if the item with the largest similarity is the one recalled just before the current one. When the process returns to the 10th item, a second subtrajectory is opened up (shown with thinner arrows) and converges to a cycle after reaching the 12th node for the second time. (c) Comparison between simulations with random symmetric similarity matrix (blue line) and SM defined by random sparse ensembles with sparsity  $f = 0.01$  (yellow line),  $f = 0.05$  (magenta line),  $f = 0.1$  (green line), and  $N = 100\,000$  number of neurons. Each point is the mean of 10 000 simulations. Black line corresponds to theoretical  $\sqrt{\frac{3}{2}\pi L}$ .

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 (black arrows), unless it points to an item that was just recalled in the previous step, in which case the next largest overlap is searched (red arrows). After a certain number of transitions, this process begins to cycle over already visited items. This happens either the first time a previously recalled item is reached again, or the process could make some number of transitions over previously recalled items [items 10,14,1 in Fig. 1(b)] to open up a new trajectory (items 13,5,11,12) until finally converging to a cycle. After the cycle is reached, no new items can be recalled.

In our previous publication ([8]) we showed that the average number of recalled items (recall capacity, or  $R$ ) scales as a power-law function of the number of items in the list,  $L$  with exponent that depends on sparseness parameter  $f$ . Here we focus on the sparse limit of this model,  $f \ll 1$ , when one can neglect the correlations between different elements of the SM and replace it by a random symmetric matrix (see, e.g., [16], for biological motivation for considering a very sparse encoding). We show below that while the corresponding graph model has a history-dependent transition rule and hence is more complex than the standard family of graphs resulting from random mappings (see, e.g., [17]), it can still be solved analytically in terms of the average number of items visited before converging to a cycle.

It is instructive to first consider the simpler case of a fully random asymmetric SM with independent elements. In this case, transitions between any two items are equally likely, with probability  $1/(L-1)$ . When an item is reached for the second time the process enters into a cycle. Therefore the probability that  $k$  out of  $L$  items will be retrieved is simply

$$P(k;L) = \left(1 - \frac{1}{L-1}\right) \left(1 - \frac{2}{L-1}\right) \dots \left(1 - \frac{k-2}{L-1}\right) \frac{k-1}{L-1} \\ \simeq \frac{k}{L} e^{-\sum_{i=1}^k (i/L)} \simeq \frac{k}{L} e^{-k^2/2L}, \quad (1)$$

where we considered a limit of large number of items in the list ( $L \gg 1$ ) and assumed that  $L \gg k \gg 1$ , which is confirmed *a posteriori* below. The average number of recalled words can then be calculated as

$$R = \langle k \rangle = \sum_{k=2}^L k \frac{k}{L} e^{-k^2/2L}, \\ \approx \sqrt{L} \int_0^\infty x^2 e^{-x^2/2} dx, \\ = \sqrt{\frac{\pi L}{2}}, \quad (2)$$

which is a well known result in random graphs literature ([17,18]).

When the SM is symmetric, as in our case, the statistics of transitions in the corresponding graph is more complicated (see the Supplemental Material [12] for more details about the derivation). In particular, the probability for a transition to one of the previously recalled items scales as  $1/(2L)$  rather than  $1/L$  as in the case of asymmetric SMs, and hence the average length of trajectory until the first return converges to  $\sqrt{\pi L}$ . Moreover, with probability  $1/3$  the trajectory then turns towards previous items and opens up a new route until again hitting a previously recalled item, etc. Taken together, the chance that recall trajectory enters a cycle after each step asymptotically equals to  $1/(2L) \times 2/3 = 1/(3L)$ , as opposed to  $1/L$  for the fully

random matrix, and hence the  $R$  can be obtained by replacing  $L$  by  $3L$  in Eq. (2):

$$R \approx \sqrt{\frac{3\pi}{2}} L \approx 2.17\sqrt{L}, \quad (3)$$

see Fig. 1(c) for the comparison of this analytical estimate with numerical simulations of the model. We emphasize that Eq. (3) does not have any free parameters that could be tuned to fit the experimental results. Hence, both the exponent and coefficient of this power law expression are a result of the assumed recall mechanism; in other words, this equation constitutes a true prediction regarding the asymptotic recall performance for long lists of items as opposed to earlier theoretical studies. Here we present the results of our experiments designed to test this prediction.

The universality of the above analytical expression for  $R$  seems to contradict our everyday observations that people differ in terms of their memory effectiveness depending, e.g., on their age and experience. Moreover, it is at odds with previous experimental studies showing that performance in a free recall task strongly depends on the experimental protocol, for example presentation rate during the acquisition stage (see, e.g., [19–27]) and the extent of practice [28,29]. Since most of the published studies only considered a limited range of list lengths, we performed free recall experiments on the Amazon Mechanical Turk® platform for list lengths of 8, 16, 32, 64, 128, 256, and 512 words, and two presentation rates: 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. (3) from below [see Fig. 2(a)].

We reasoned that some or all of the variability in the experimentally observed  $R$  could result from the variability in the number of words that remain in memory as candidates for recall after the list is presented. In particular, some of the words could be missed at presentation, while others could be acquired but later erased or degraded. It seems reasonable that acquisition depends on various factors, such as attention, age of participants, acquisition speed, etc. One should then correct Eq. (3) for  $R$ , replacing the number of presented words  $L$  with the number of words in memory after the whole list is presented,  $M$ :

$$R \approx \sqrt{\frac{3\pi}{2}} M. \quad (4)$$

To test this conjecture, we designed a novel experimental protocol that involved performing both recall and *recognition* experiments on the same group of participants. Each participant performed one recognition and one recall trial with lists of the same number of words (but different words

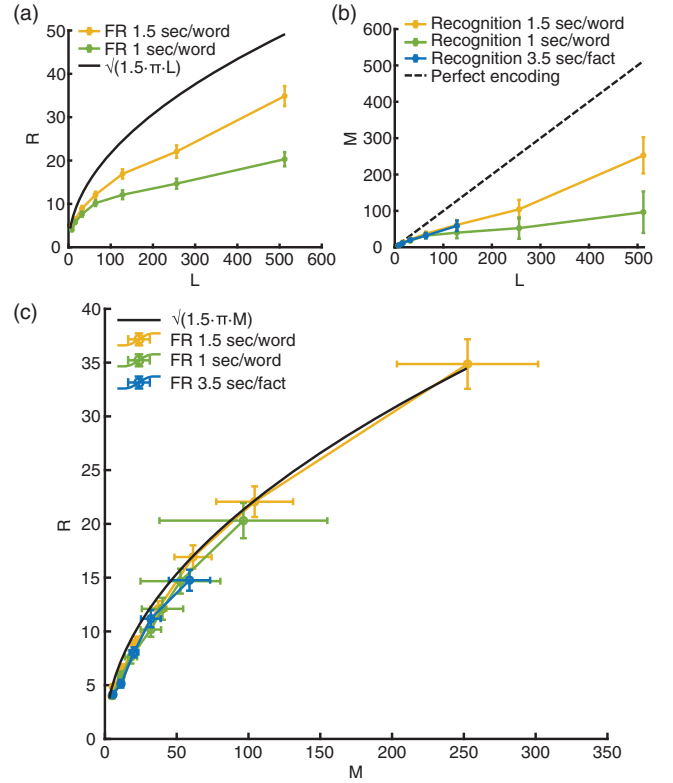


FIG. 2. Human recall and recognition performance. (a) Average number of words recalled as a function of the number of words presented. Black line: Eq. (3). Yellow line: experimental results for presentation rate 1.5 sec/word. Green line: experimental results for presentation rate 1 sec/word. The error in  $R$  is a standard error of the mean. (b) Estimated average number of encoded words or sentences for lists of different lengths. Black dashed line corresponds to perfect encoding, green line corresponds to presentation rate 1 sec/word and yellow line to presentation rate 1.5 sec/word; blue line corresponds to lists of short sentences (see text for details). The error in  $M$  is computed with bootstrap procedure [30]. (c) Average number of words (sentences) recalled as a function of the average number of encoded words (sentences). Green line: experimental results for presentation rate 1 sec/word. Yellow line: experimental results for presentation rate 1.5 sec/word. Blue line: experimental results for short sentences. The error in  $R$  is a standard error of the mean, while the error in  $M$  is computed with bootstrap procedure (see the Supplemental Material [12] for details).

between recognition and recall) and under identical presentation conditions, including presentation rate, in order to independently evaluate the average number of words in memory, and the average number of words recalled. Following [5], at the end of presentation we showed each participant a pair of words, one from the list just presented (target) and one randomly chosen lure, requesting to report which word was from the list. The average number of words remaining in memory ( $M$ ) was then estimated from the fraction of correctly recognized words ( $c$ ) by assuming that if a target word was still encoded at the end of presentation, it

will be chosen during recognition test, otherwise the participant will randomly guess which of the two words is a target:  $M = L(2c - 1)$ . Importantly, each participant performed a single recognition test, to avoid the well known effect of “output interference” between subsequent recognition tests for a single list (see, e.g., [31]).

Figure 2(b) shows the estimated average  $M$  as a function of list length  $L$  (see the Supplemental Material [12] for details of the analysis). Results confirm that  $M$  increases with time allotted to presentation of each word. Standard error of the mean for the number of encoded 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 ([30], see the Supplemental Material).

In Fig. 2(c) experimentally obtained  $R$  (yellow and green lines) is compared with the theoretical prediction of Eq. (4) (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 rates, even though the number of encoded 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 ([8,9]) and found that it captures the statistics of the recall performances as assessed with bootstrap analysis of the results (see Fig. S1 in the Supplemental Material [12]).

Experiments presented above, as well as the vast majority of previous recall experiments, were performed with lists of words. To test the generality of our model prediction, we generated a set of 325 short sentences expressing common knowledge facts, such as “Earth is round” or “Italians eat pizza,” etc. We repeated our experiments with random lists of 8,16,32,64, and 128 such sentences, each presented for 3.5 seconds (see the Supplemental Material [12] for more details of the analysis). As shown in Figs. 2(b) and 2(c), performance with lists of sentences is very close to that of words with 1.5 words per second presentation rate, albeit with some small deviations towards lower levels.

**Discussion.**—The results presented in this Letter show that the relation between the number of words in memory and the number of recalled words conforms with remarkable precision to the analytical, parameter-free expression Eq. (4), 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 number of presented words and on the presentation rate. We further confirmed the generality of Eq. (4) by repeating the experiments with lists of short sentences expressing common knowledge facts. Hence it appears that memory recall is a more universal process than memory acquisition and maintenance. The crucial aspect of the model is the similarity matrix between the items that

determines the recall transitions, but the precise nature of this matrix beyond its statistics across the presented lists and/or across participants does not have to be specified. It seems plausible that different people will have different similarity matrices, reflecting their unique language experience, which makes direct estimation of it rather challenging. However, our previous study ([32], see also [33]) showed that recall transitions are sensitive to the measure of semantic similarity called latent semantic analysis (LSA), which represents the number of times two words appear together in a representative corpora of natural text ([34]). This indicates that there is some degree of universality in interword similarities across all people.

Several influential computational models of recall were developed in cognitive psychology that incorporate interactive probabilistic search processes (see, e.g., [35–40]). These 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 [20,22,41]. However, most of the free parameters lack clear biological meaning and cannot be constrained before the data are collected; hence the models cannot be used to predict the recall performance but only explain it *a posteriori*. Our recall model can be viewed as a radically simplified version of the classical search of associative memory model (SAM), see [35]. 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, while in our model is simply assumed to be a fixed, structure-less symmetric matrix (see Fig. 1). Subsequent recall in SAM proceeds as a series of attempted probabilistic sampling 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 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 remain in memory at the end of presentation to become candidates for recall. This assumption is confirmed *a posteriori* by the collapse of  $R$  vs  $M$  curves for different presentation rates (see the Supplemental Material [12] for more detailed argumentation). We also neglected the well-documented effects of short-term memory on free recall (see, e.g., [42]), which are very small in our data (see the Supplemental Material [12]).

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 the



need to tune a single parameter. Future theoretical and experimental studies should be pursued to probe which aspects of the model are valid and which are crucial for the obtained results.

This research has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2); EU-M-GATE 765549 and Foundation Adelis. S. R. is supported by the Howard Hughes Medical Institute. We thank Drs. Mike Kahana and Eli Nelken for helpful comments.

\*misha@weizmann.ac.il

†M. N. and M. K. contributed equally to this work.

- [1] R. N. Shepard, Towards a universal law of generalization for psychological science, *Science* **237**, 1317 (1987).
- [2] Y. Dudai, The neurobiology of consolidation, or, how stable is the engram? *Annu. Rev. Psychol.* **55**, 51 (2004).
- [3] M. J. Kahana, *Foundations of Human Memory* (Oxford University Press, USA, 2012).
- [4] A. Binet and V. Henri, Mémoire des mots, *L'Année Psychol.* **1**, 1 (1894).
- [5] L. Standing, Learning 10 000 pictures, *Quart. J. Exp. Psychol.* **25**, 207 (1973).
- [6] D. J. Murray, C. Pye, and W. E. Hockley, Standing's power function in long-term memory, *Psychol. Res.* **38**, 319 (1976).
- [7] N. C. Waugh, Presentation time and free recall, *J. Exp. Psychol.* **73**, 39 (1967).
- [8] S. Romani, I. Pinkoviezky, A. Rubin, and M. Tsodyks, Scaling laws of associative memory retrieval, *Neural Comput.* **25**, 2523 (2013).
- [9] M. Katkov, S. Romani, and M. Tsodyks, Memory retrieval from first principles, *Neuron* **94**, 1027 (2017).
- [10] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554 (1982).
- [11] M. V. Tsodyks and M. V. Feigel'man, The enhanced storage capacity in neural networks with low activity level, *Europhys. Lett.* **6**, 101 (1988).
- [12] See the Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.124.018101> for details of recall model and its analytical solution; experimental details; analysis of experimental results; and effects of short-term memory on recall performance, which includes Ref. [13–15].
- [13] M. K. Healey, P. Crutchley, and M. J. Kahana, Individual differences in memory search and their relation to intelligence, *J. Exp. Psychol. Gen.* **143**, 1553 (2014).
- [14] D. A. Medler and J. R. Binder, Mcword: An on-line orthographic database of the english language, 2005, <http://www.neuro.mcw.edu/mcword/>.
- [15] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, Distributed representations of words and phrases and their compositionality, *NIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems* (2013), Vol. 2, pp. 3111–3119.
- [16] R. Q. Quiroga and G. Kreiman, Measuring sparseness in the brain: Comment on Bowers (2009), *Psychol. Rev.* **117**, 291 (2010).
- [17] B. Harris, Probability distributions related to random mappings, *Ann. Math. Stat.* **31**, 1045 (1960).
- [18] J. Katz, A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone, *Handbook of Applied Cryptography* (CRC press, Boca Raton, 1996).
- [19] B. B. Murdock, Jr., The immediate retention of unrelated words, *J. Exp. Psychol.* **60**, 222 (1960).
- [20] B. B. Murdock, Jr., The serial position effect of free recall, *J. Exp. Psychol.* **64**, 482 (1962).
- [21] W. A. Roberts, Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses, *J. Exp. Psychol.* **92**, 365 (1972).
- [22] M. W. Howard and M. J. Kahana, Contextual variability and serial position effects in free recall, *J. Exp. Psychol. Learn. Mem. Cogn.* **25**, 923 (1999).
- [23] M. J. Kahana, M. W. Howard, F. Zaromb, and A. Wingfield, Age dissociates recency and lag recency effects in free recall, *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 530 (2002).
- [24] F. M. Zaromb, M. W. Howard, E. D. Dolan, Y. B. Sirotin, M. Tully, A. Wingfield, and M. J. Kahana, Temporal associations and prior-list intrusions in free recall, *J. Exp. Psychol. Learn. Mem. Cogn.* **32**, 792 (2006).
- [25] G. Ward, L. Tan, and R. Grenfell-Essam, Examining the relationship between free recall and immediate serial recall: The effects of list length and output order, *J. Exp. Psychol. Learn. Mem. Cogn.* **36**, 1207 (2010).
- [26] J. F. Miller, C. T. Weidemann, and M. J. Kahana, Recall termination in free recall, *Mem. Cogn.* **40**, 540 (2012).
- [27] R. Grenfell-Essam, G. Ward, and L. Tan, Common modality effects in immediate free recall and immediate serial recall, *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 1909 (2017).
- [28] K. A. Klein, K. M. Addis, and M. J. Kahana, A comparative analysis of serial and free recall, *Mem. Cogn.* **33**, 833 (2005).
- [29] S. Romani, M. Katkov, and M. Tsodyks, Practice makes perfect in memory recall, *Learn. Mem.* **23**, 169 (2016).
- [30] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (CRC Press, Boca Raton, 1994).
- [31] A. H. Criss, K. J. Malmberg, and R. M. Shiffrin, Output interference in recognition memory, *J. Mem. Lang.* **64**, 316 (2011).
- [32] S. Recanatesi, M. Katkov, S. Romani, and M. Tsodyks, Neural network model of memory retrieval, *Front. Comput. Neurosci.* **9**, 149 (2015).
- [33] M. W. Howard, B. Jing, K. M. Addis, and M. J. Kahana, Semantic structure and episodic memory, *Handbook of Latent Semantic Analysis* (Lawrence Erlbaum Associates, Mahwah 2007), pp. 121–141.
- [34] T. K. Landauer and S. T. Dumais, A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge, *Psychol. Rev.* **104**, 211 (1997).
- [35] J. G. W. Raaijmakers and R. M. Shiffrin, SAM: A theory of probabilistic search of associative memory, *Psychology of Learning and Motivation* (Elsevier, New York, 1980), Vol. 14, pp. 207–262.

- [36] G. Gillund and R. M. Shiffrin, A retrieval model for both recognition and recall, *Psychol. Rev.* **91**, 1 (1984).
- [37] M. W. Howard and M. J. Kahana, A distributed representation of temporal context, *J. Math. Psychol.* **46**, 269 (2002).
- [38] D. Laming, Failure to recall, *Psychol. Rev.* **116**, 157 (2009).
- [39] S. M. Polyn, K. A. Norman, and M. J. Kahana, A context maintenance and retrieval model of organizational processes in free recall, *Psychol. Rev.* **116**, 129 (2009).
- [40] M. Lehman and K. J. Malmberg, A buffer model of memory encoding and temporal correlations in retrieval, *Psychol. Rev.* **120**, 155 (2013).
- [41] B. B. Murdock and R. Okada, Interresponse times in single-trial free recall, *J. Exp. Psychol.* **86**, 263 (1970).
- [42] M. Glanzner and A. R. Cunitz, Two storage mechanisms in free recall, *J. Verb. Learn. Verb. Behav.* **5**, 351 (1966).