Statistics of free memory recall

Mikhail Katkov and Misha Tsodyks*
The Simons Center for Systems Biology, Institute for Advanced Study, Princeton, New Jersey 08540, USA and Department of Neurobiology, Weizmann Institute of Science, Rehovot 76000, Israel

(Received 27 December 2021; accepted 18 May 2022; published 1 August 2022)

This article is part of the Physical Review Research *collection titled* *Physics of Neuroscience.*

Numerous studies analyzed the performance of participants in free recall of randomly assembled lists of words with the focus on the average number of words recalled for different experimental parameters such as list length, presentation speed, etc. The distribution of performance around the mean was not systematically studied, even though it is well-known that recall is an unpredictable process resulting in highly variable results over different trials. We recently introduced the mathematical model of free recall that reproduced well the average performance of human participants in experiments with randomly assembled lists of words or short sentences. The model assumes that during recall, each memory item currently recalled triggers a recall of a next item based on the random symmetric matrix of similarity measures between items in the list. When applying the model to experimental data, a crucial assumption was made that upon presentation, a certain fraction of presented items remain in memory that are candidates for recall, and that the number of such items can be estimated with recognition experiments performed by the same group of participants under identical conditions of item presentation as in the recall experiments. It is not clear whether this assumption is valid under different experimental paradigms and with different groups of participants. Here we propose that calculating the variance of recall performance allows one to formulate interesting predictions that can be tested without performing recognition experiments. Comparison of model predictions with experimental data on young and old participants indicates that the same recall algorithm is involved in both groups, even though old participants may have fewer candidate memory items for recall after presentation.

DOI: 10.1103/PhysRevResearch.4.033090

I. INTRODUCTION

Memory recall is notoriously unreliable both in real-life situations and in laboratory settings. Very often people cannot recall events or facts that they clearly still remember, which is attested for by later spontaneous or cued retrieval. Therefore, there is a general belief that memory phenomena cannot be described by the laws derived from simple principles, neither on a behavioral level nor physiological. On the behavioral side, cognitive psychologists have developed several classical paradigms to study human memory, where participants are presented with lists of randomly assembled items of various lengths. In particular, two paradigms that are relevant for the current study are recognition and free recall. In the first paradigm, participants are asked to determine whether a particular item was earlier presented to them. This paradigm can be used, under some assumptions elaborated later, to estimate the number of items that remain in memory by the time of the test. In the second paradigm, participants are simply requested to recall as many items as they can, in an arbitrary order. Comparing the results in these two paradigms, the gap was observed between the number of items that can be remembered and those that can be recalled [1], indicating that some remembered items cannot be readily accessed, for unknown reasons. Many detailed mathematical models of recognition and recall have been proposed in psychological literature over the years [2], however, the origin of this gap has not been specifically addressed. On the physiological side, the dominating view is that memory is a collective effect in dedicated neuronal networks. It is generally assumed that sparse random groups of neurons encode each specific memory item via an attractor mechanism and become active when memory is recalled or perceived [3,4]. Experimental confirmation for this picture is provided by the vast literature on stimulus-specific persistent activity observed during memory tasks in monkeys (see, e.g., Ref. [5]). More recently, electrical recordings in human participants also found specific neuronal activations just prior to stimulus recall [6,7].

We recently introduced a phenomenological model where free recall is implemented as a trajectory on a highly diluted graph [8]. The size of the graph is supposed to be the number of items that remain in memory after the presentation. This idea emerged from our simulations of sparse associative memory network of [4] to which we added a neuronal adaptation and temporarily modulated inhibition to cause transitions between attractors corresponding to different items in the remembered list [8,9]. This model requires some mechanism of isolating the remembered items from all other items in the

*mtsodyks@gmail.com

long-term memory of the participant, which could come about either via restoring items in a separate network dedicated to short-term memory (e.g., hippocampus) or binding them together via a hierarchical scheme involving a higher-order encoding of a context of the experiment [10,11]. When simulating the network, we observed that transitions tend to occur in such a way that the next attractor activated by the network has the largest overlap with the currently active one [9,12]. We therefore considered the matrix of overlaps between remembered items as a symmetric similarity matrix (SM) that defines the *deterministic* search process responsible for recall. The gap between the number of remembered items and the number of recalled ones is naturally explained by the fact that the recall process enters a cycle after some number of transitions, after which no new items can be recalled. In the sparse coding limit, the SM matrix can be approximated as a random symmetric matrix, resulting in a parameter-free mathematical model of free recall. This model can be solved to produce an analytical expression for the average number of items that can be recalled (R) given the number of remembered items (M) [13]:

$$< R >= \sqrt{\frac{3\pi}{2}M}. \quad (1)$$

Since this equation has no free parameters to tune, it appears to be incompatible with numerous observations that recall performance depends on various aspects of experimental protocol, such as, e.g., presentation speed of the material or the age of participants [14–23]. These observations could, however, be reconciled with the universality of Eq. (1) if one assumes that all the above observations are explained by changes in M. To test for this possibility, we recently performed recall and recognition experiments on the same sets of participants under equivalent experimental conditions and found that in all situations that we tested (seven different list lengths from eight to 512 words and two different presentation speeds), the experimental results lie surprisingly close to the predicted relation given by the above equation [13]. Moreover, we showed that the same relation also holds when lists of short sentences expressing well-known facts are used instead of single words, demonstrating some degree of universality of the obtained results.

In this paper, we present a further mathematical analysis of our recall model by considering the *distribution* of the number of recalled items for a given experimental condition. We begin by calculating the second moment of the number of items that can be recalled, $< R^2 >$. The motivation for this analysis, beyond mathematical interest, are severalfold. First, we want to see how well the model predicts the distribution of R obtained experimentally. Second, as mentioned above, estimating the number of items in memory from recognition experiments is not direct and is based on several assumptions. The basis for this estimation is the relation between the number of items in memory and the probability for the correct answer to a 2AFC recognition trial (C), where a randomly chosen item from the presented list of length L is paired against a lure:

$$C = \frac{1}{2}\left(1 + \frac{M}{L}\right). \quad (2)$$

This relation is derived by assuming that if the presented item chosen for recognition is in memory, a participant points to it securely; if this item is not in memory, a participant chooses randomly between the presented item and a lure. To justify Eq. (1), it has to be further assumed that all of the items that are in memory become candidates for recall on equal footing. These simplifying assumptions could be violated in several different ways. For example, the dichotomy between an item being in memory or not (forgotten) may be false, and one could recognize an item based on very partial information, without it being firmly remembered. One could also imagine that remembered items do not constitute a homogeneous group that are equal candidates for recall, rather they could be broken into two or more groups that cannot be reached from each other. The close agreement between the model and experimental results achieved in Ref. [13] indicates that assumptions described above apparently hold for those experiments but it is not clear how general this conclusion is when other experimental conditions or other groups of participants are involved. For example, it is possible that when delay is introduced between presentation and test, or when older participants are involved, not all memory items that can be successfully recognized as familiar can be recalled. As we will see below, considering the second moment of the number of recalled items could be a way to address these issues by deriving the mathematical predictions that can be tested without knowing the number of words in memory (M), i.e., without the need for recognition experiments.

II. RESULTS

A. Model

Our model of free recall was introduced in previous publications [8,13,24]. Briefly, for a given set of items being recalled (i.e., items that are in memory after the presentation), we consider a random symmetric matrix of inter-item similarities. This matrix could, e.g., represent the overlaps between long-term neuronal encodings of the presented items, but its precise nature does not have to be agreed on for the purpose of this paper. We hypothesize that free recall proceeds according to the following deterministic transition rule: The first item to be recalled is chosen randomly. After that, each time an item is recalled, the next item that has the largest similarity to the current one is chosen, excluding the one that was recalled just before the current one. For a given SM, the process just described is an entirely deterministic one, hence it invariably converges into a cycle after which no new items can be recalled, and the length of the recall trajectory corresponds to the number of items recalled on a particular trial (R). Moreover, as follows by the transition rule, recall is entirely determined by the position of the largest and second-largest elements in each row of SM and hence distribution of recall trajectories does not depend on the distribution of matrix elements of the SM as long as they are independent from each other, subject to symmetry constraint. The model can be illustrated by a graph with each node corresponding to a memory item and two arrows from each node point to nodes with either largest or second-largest elements of the corresponding row of the SM (see Fig. 1 for an example of a SM and the corresponding recall trajectory).

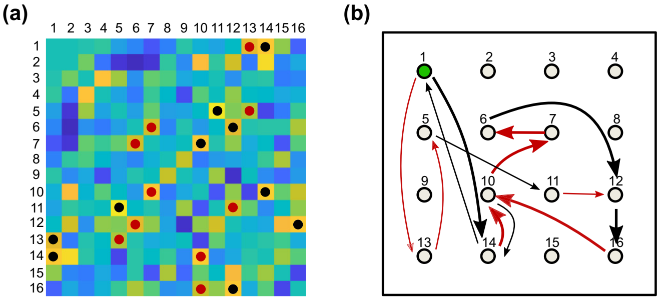

FIG. 1. Associative search model of free recall. (a) SM (similarity matrix) for a list of 16 items (schematic). For each recalled item, the maximal element in the corresponding row is marked with a black spot, while the second maximal element is marked with a red spot. (b) A graph with 16 nodes illustrates 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 tenth 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. Adapted from Ref. [13].

To calculate the statistics of R over different realizations of SMs, in Ref. [13] we first considered a simpler model where the SM is chosen to be a matrix of random and independent elements (i.e., asymmetric), which can be solved analytically. Indeed, since in this model transitions between all pairs of items have the same probability $\frac{1}{M-1}$ (where M as before is the number of items in memory), the probability to recall R items before colliding with a previously recalled one and converging to a cycle is given by

$$P(R;M) = \left(1-\frac{1}{M-1}\right)\left(1-\frac{2}{M-1}\right) \dots \left(1-\frac{R-2}{M-1}\right)\frac{R-1}{M-1}. \tag{3}$$

This expression can be further simplified in the asymptotic limit of large number of items ($M \gg 1$) as

$$P(R;M) \simeq \frac{R}{M} e^{-\sum_{i=1}^{R}\frac{R}{M}} \simeq \frac{R}{M} e^{-\frac{R^2}{2M}}, \tag{4}$$

from which all the moments of R can easily be computed.

We now turn to our original recall model where SM is a random symmetric matrix. This model is mathematically considerably more challenging, because (i) due to the SM symmetry, the probability for the transition between a given pair of items, say $i \to j$, depends on previous transitions involving these items; and (ii) the transition rule has memory, hence collision with the previously recalled item does not necessarily imply that the process enters into a cycle, rather a cycle only begins when the transition to the same item repeats for the second time. As we showed in Ref. [13], asymptotically for large M the probability for the recall process to collide with each of the previously recalled nodes is $\frac{1}{2M}$, i.e., one-half of the corresponding probability in the model with random SM. Moreover, after the collision, the probability for the process to enter into a cycle is 2/3, otherwise the process proceeds along the previous trajectory in the opposite direction until splitting into a new set of items (see Fig. 1 for an example). The probability to collide with each of the previously recalled item and enter into a cycle is therefore given by $\frac{1}{2M}\frac{2}{3} = \frac{1}{3M}$, as opposed to $\frac{1}{M}$ in the model with fully random SM; hence the overall asymptotic probability that the process enters into a cycle after recalling R items is given by Eq. (4) with substitution $M \to 3M$,

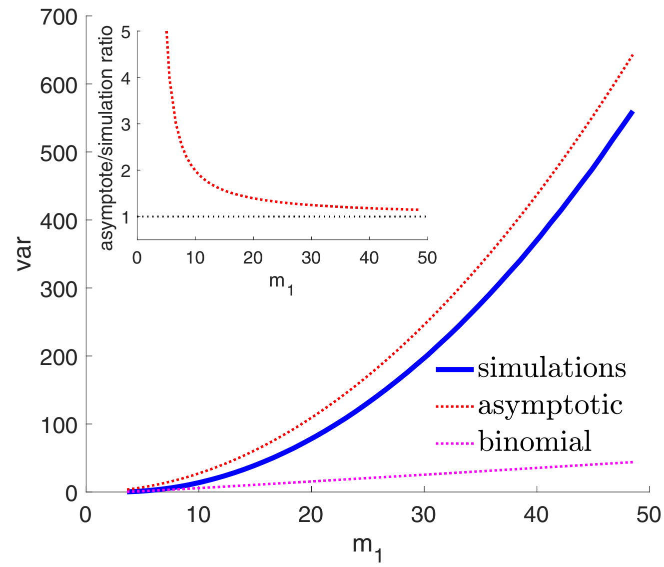

FIG. 2. Simulations of recall model. Relation between first moment and variance obtained in simulations (blue), using Eq. (7) (red) and assuming binomial model (magenta). The ratio between variance obtained using asymptotic formula and the ones obtained in simulations is shown in the inset and is slowly approaching 1.

$$P(R;M) \simeq \frac{R}{3M} e^{-\frac{R^2}{6M}}, \tag{5}$$

resulting in the following asymptotic expressions for the first two moments of R, $m_1 = \langle R \rangle$ and $m_2 = \langle R^2 \rangle$:

$$m_1(M) \simeq \sqrt{\frac{3\pi}{2}M},$$
$$m_2(M) \simeq 6M. \tag{6}$$

Eliminating M from these equations results in the following relation between the variance and the mean for the number of recalled items:

$$\sigma^2 \simeq \left(\frac{4}{\pi}-1\right)m_1^2, \tag{7}$$

where $\sigma^2 = m_2 - m_1^2$, i.e., variance in this model scales quadratcaly with the mean. Since this equation does not explicitly depend on M, it can, in principle, be tested in recall experiments alone without performing recognition experiments. One should keep in mind, however, that the above asymptotic formulas represent the leading order of the corresponding moments for large M and ignore lower order terms, in particular, the linear correction in the relation between the variance and the mean of Eq. (7). Currently, we don't have an analytical expression for the corrections, hence we resorted to numerical simulations of the model to access the precision of the asymptotic expressions (Fig. 2, see Appendix B for details). Indeed, one can see the convergence of the asymptotic value of variance to the one obtained in simulations (Fig. 2, inset) but this convergence happens at rather large values of m_1. In the following analysis, we therefore used the numerical dependence of variance on m_1 rather than the asymptotic one. As a simple alternative to our model, we considered a scheme where each item is recalled independently from the others with probability $p = \frac{m_1}{M}$, which results in the binomial distribution for the number of recalled items with mean m_1 and variance,

$$\sigma_B^2 = m_1\left(1 - \frac{m_1}{M}\right), \tag{8}$$

which asymptotes to m_1 for large M. In summary, this naive model results in a much more narrow distribution of recalls compared to the model introduced above.

The above results shown in Fig. 2 allow us to test predictions of the model without knowing the number of words in memory, i.e., without performing recognition experiments. However, before applying this analysis to experimental data, one has to take into account that they were obtained assuming that the number of items in memory (M) is fixed for all trials. This is, however, not the case when different lists are presented to the diverse group of participants (remember in our experiments each participant performed a single recall trial). Hence the above results provide the *lower bound* for the variance for a given first moment rather than the true estimate, and therefore should not be directly compared to experimental data. To get the possible range of variance values for a given mean, we derive its *upper bound* value. As shown in Appendix A, the upper bound for the variance is achieved when in some of the trials all words in the list are remembered ($M = L$) and in the others none are remembered ($M = 0$):

$$\overline{\sigma}_u^2 = \overline{m}_1 \frac{m_1^2(L) + \sigma^2(L)}{m_1(L)} - \overline{m}_1^2,$$
$$0 \leqslant \overline{m}_1 \leqslant m_1(L), \tag{9}$$

where $\overline{m_1}$ and $\overline{\sigma}^2$ stand for the mean and the variance of recalls for a distribution of M, while $m_1(L)$ and $\sigma^2(L)$ refer to the mean and the variance of R for the fixed number of words in memory, $M = L$. As opposed to the lower bound, the upper bound for the variance given the mean explicitly depends on the number of items in the list, L. For a given set of data with the particular list length, the position of the point with coordinates $(\overline{m_1}, \overline{\sigma}^2)$ depends on the spread of M values across trials (participants), being closer to lower (upper) bound for narrower (wider) distributions of M between trials, correspondingly.

B. Experiments

Before applying the above analysis to the data of Ref. [13], we consider the large data set of multiple free recall experiments performed by 270 participants over the course of seven daily sessions, each session consisting of 16 lists of 16 words, each list presented and recalled one by one. This data set was collected in the laboratory of Prof. Michael Kahana (UPenn) and initial part of it was presented in Ref. [22] where more experimental details can be found. A larger part of this set was previously analyzed by us in Ref. [25]. As we showed

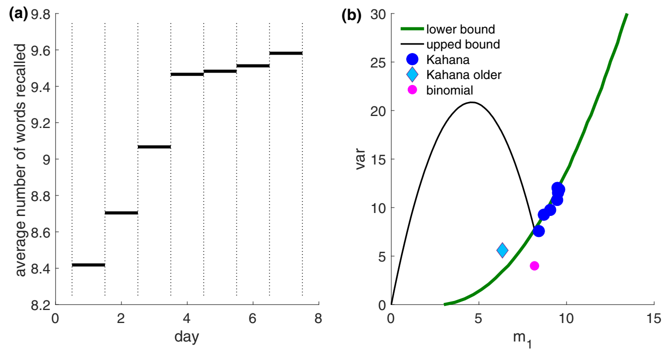

FIG. 3. Variance versus first moment in Kahana experiment. (a) Average daily performance in recall experiment. (b) First moment and variance of the recall of random lists of words by students, separately for each session (dark blue circles) and for older participants (light blue diamond). Dark green (black) line is the lower (upper) bound for the variance given the mean. Magenta circle represents the mean and the variance of the binomial model described in the text.

in that publication, recall performance gradually increases for the first four daily sessions before saturating, accompanied by developing a more structured recall, in particular, exhibiting either chaining (recalling groups of words in the original order) or chunking (recalling groups of consecutively presented words one after another before switching to another group). We therefore considered each daily session separately for the analysis of the average and variance of the number of words recalled. The average performance over the days is shown in Fig. 3(a), and the variance versus mean results for each day are shown in Fig. 3(b), overlaid on the relevant range of possible values of $\overline{m_1}$ and $\overline{\sigma}^2$ as predicted by the model. Remarkably, the results for the first session are very close to the right upper corner of the relevant range, compatible with the model if all the participants remember all 16 words at each trial ($M = L$), which appears reasonable for such short lists when experiments are performed by students in the laboratory. In other words, given that the average recall performance is the maximal possible for $L = 16$ according to the theoretical Eq. (1), the model correctly predicts both the mean and the variance of the performance across all trials on the first day without any tuning. For subsequent days, the average performance exceeds the maximal one predicted by the model, which we believe is explained by various deviations from the fully random recall assumed in the model, as analyzed in Ref. [25]. For comparison, the naive binomial model predicts the variance that is significantly smaller than the observed one.

We also considered an additional, smaller data set obtained in the same laboratory with 38 older participants (61–85 years old) performing the same experiments [26]). Unsurprisingly, older participants recall fewer words on average, consistent with the wide literature on the effects of aging on memory (see, e.g., Ref. [27]). Since recognition experiments were not performed on old participants for lists of the same length, we cannot directly estimate the average number of words they remember after presentation. Computing the variance is thus the only way to compare the data to the model. Indeed, we see that the variance of recall performance of these older participants lies within the region compatible with the model, in fact, quite close to the lower bound, indicating that recall in

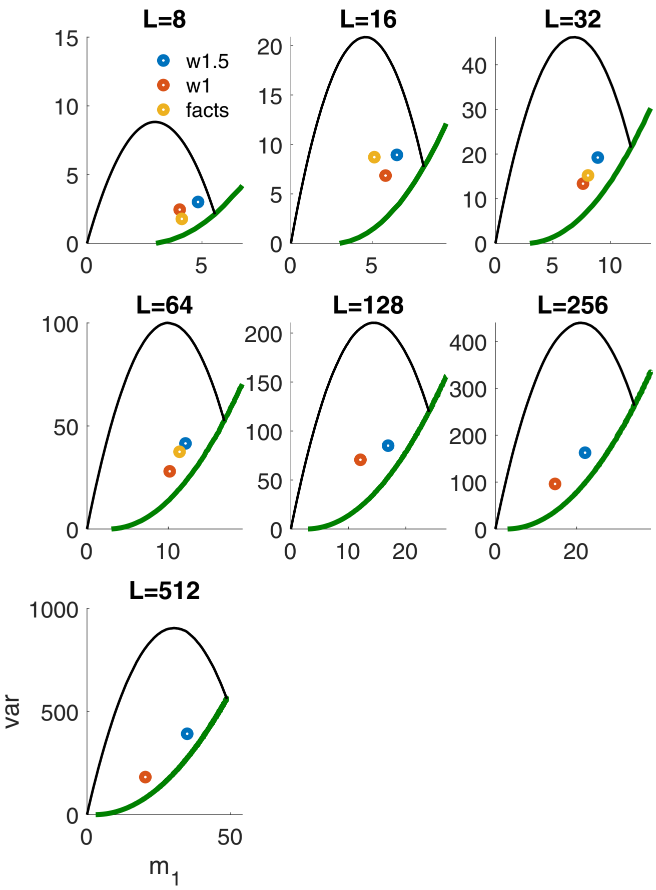

FIG. 4. Variance versus mean of recall performance. Relation between mean (m_1, x axis) and variance (var, y axis) for the number of recalled items obtained in Naim *et al.* [13] are shown. Experimental results for each presentation length ($\boldsymbol{L}$) are shown in separate panel. Green curve corresponds to theoretical lower bound described by Eq. (7), and black line corresponds to L-dependent upper bound described by Eqs. (9). Blue circles correspond to lists of words with the presentation speed 1.5 s per word; orange circles correspond to random lists of words with the presentation speed 1 s per word; and orange circles correspond lists of short sentences representing well-known facts.

old participants is governed by the same search algorithm as in younger participants.

We then reanalyzed the results presented in our previous publication [13]. In that study, groups of participants performed recall and recognition experiments under identical presentation conditions chosen from the set of 18 conditions, namely, seven list lengths of nouns ($L = 8, 16, 32, 64, 128, 256$) presented at two speeds 1 s per word and 1.5 s per word and four lists of general facts with four lengths ($L = 8, 16, 32, 64$) presented at 3 s per fact. Each participant performed one recall trial of a particular list and one recognition trial for another list of the same type (words or facts, length and presentation speed). In the recognition trial, one item from the list was paired with a novel item and the participant was requested to report which one of them was taken from the presented list (two-alternative forced choice protocol). Only one recognition trial per list was employed to prevent a well-known effect of response interference (see Ref. [28]). For each experimental condition, the average (over the corresponding group of participants) number of items that were encoded in memory after the presentation (M) was then estimated from the number of correct recognition answers using Eq. (2), but the procedure did not allow estimation of M for each individual participant. Here we reanalyzed this data and calculated the mean and the variance of the number of recalled items across participants for each experimental condition. In Fig. 4, we present these results separately for each list length. As we can see, all data points lie in the appropriate regions between lower and upper limits of variance calculated above, i.e., the data is compatible with the model. We then used both measures to independently estimate, for each condition, (i) the average number of items in memory after the presentation ($\langle M\rangle$) and (ii) the distribution of the number of recalled items, $P(R)$. The difficulty of this estimation lies in the fact that we have no way to know how M is distributed across trials (participants). We therefore assume that distribution of M for each condition is given by the Gaussian distribution truncated between 0 and L:

$$P(M) \sim e^{\alpha(M-\mu)^2}$$
$$0 \leqslant M \leqslant L. \quad (10)$$

The rationale for this choice is that the truncated Gaussian distribution has a maximal entropy over all distributions bounded in a region, for fixed mean and variance [29]. Moreover, the truncated Gaussian distribution, by varying α and μ in the range of $-\infty$ to $+\infty$, gives the whole range of distributions from a single delta function at the intermediate value of M to the sum of two delta functions at $M = 0$ and $M = L$, corresponding to the lower and the upper bound of the variance given the mean. We therefore find the parameters of the truncated Gaussian distribution of M that results in the best fit to both moments of R measured in our experiments (see

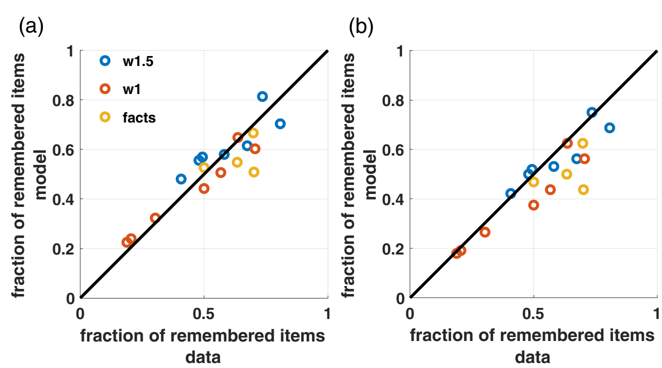

FIG. 5. Number of remembered items. Comparison between the number of remembered items M estimated in recognition experiment (x axis [13]) and using only recall data for the same experimental condition (y axis). (a) M along y axis is computed as the average of the inferred distribution of M from the mean and the variance of recall (see details in Appendix C). (b) Same as (a) but M is assumed to be the same for all trials for a given condition. Both: As in Fig. 4, blue circles correspond to random lists of words with presentation speed 1.5 s per word; orange circles correspond to random lists of words with presentation speed 1 s per word and orange circles correspond to lists of short sentences representing well-known facts.

TABLE I. Kolmogorov-Smirnov test for estimated recall distributions. *P* values for two sample Kolmogorov-Smirnov tests are presented for each condition in Naim *et al.* [13] experiments.

Based on two-moment estimation				Based on first moment estimation			
L	1.5 s per noun	1 s per noun	Short sentences	L	1.5 s per noun	1 s per noun	Short sentences
8	0.927	0.635	0.715	8	0.007	0.007	0.017
16	0.997	1.000	0.389	16	0.196	0.364	0.001
32	0.343	0.942	0.996	32	0.005	0.796	0.436
64	0.997	0.046	0.938	64	0.341	0.408	0.691
128	0.999	0.768		128	0.528	0.020	
256	0.981	1.000		256	0.379	0.044	
512	0.466	0.647		512	0.407	0.038	

Appendix C). We emphasize that this procedure only uses the results of the recall experiment. We then calculated the average value of M from this distribution and compared it to the same value estimated independently with recognition experiments (see Ref. [13]). As we see from Fig. 5(a), two estimates are quite close for all experimental conditions. Moreover, we found that the distributions of recalls are similar to the experimentally obtained ones (see Table I for Kolmogorov-Smirnov statistics and Fig. 6 for examples; other comparisons are shown in Appendix C). It is important to note that in these estimates, recall variance plays an important role; if neglected, simply assuming that all participants remembered the same number of items M (such that $m_1 \approx \sqrt{\frac{3\pi}{2}M}$), theoretical predictions for $\langle M \rangle$ (=M in this case) and $P(R)$ are less accurate (compare Figs. 5(a) and 5(b) and left and right parts of Table I). In summary, we demonstrate with the above analysis that the model accounts well for the whole distribution of recall performance for all experimental conditions of Ref. [13] and not only for the average performance.

The inferred distributions of M shown in Fig. 6 are quite wide, indicating that the number of items in memory after presentation of the list is very different for different people (trials). Since each participant performs only one recognition task (with one presentation of a correct item versus lure), we cannot know how much of this variability is due to different acquisition abilities of different people and how much is due to different lists and other factors like attention, etc. Moreover, we don't believe that our estimation of distributions of M is very reliable because the effect of $P(M)$ on mean and variance of recall performance is not very strong. To illustrate this issue, we simulated recall performance for $L = 256$ for four different distributions of M of constant mean, from a single delta function, corresponding to the lowest bound for the variance of recall, to a sum of two delta functions at $M = 0$ and $M = L$, corresponding to the upper bound of the variance. Results are shown in Fig. 7. One can see that the recall distribution corresponding to the flat distribution of M is not very different from the one corresponding to a single delta-function distribution of M; nevertheless, as we showed above, inferred distributions of M agree better with recognition experiments.

III. DISCUSSION

Given that human memory is a complex multistage process, it is remarkable that some aspects of it can be described by universal analytical expression like Eq. (1). In the current contribution, we showed that statistics of free recall of randomly assembled lists of words is faithfully described by the model introduced in Ref. [13], giving further support to the hypothesized search process based on long-term representations of memory items. Moreover, going beyond the average recall performance and analyzing the variance allowed us to

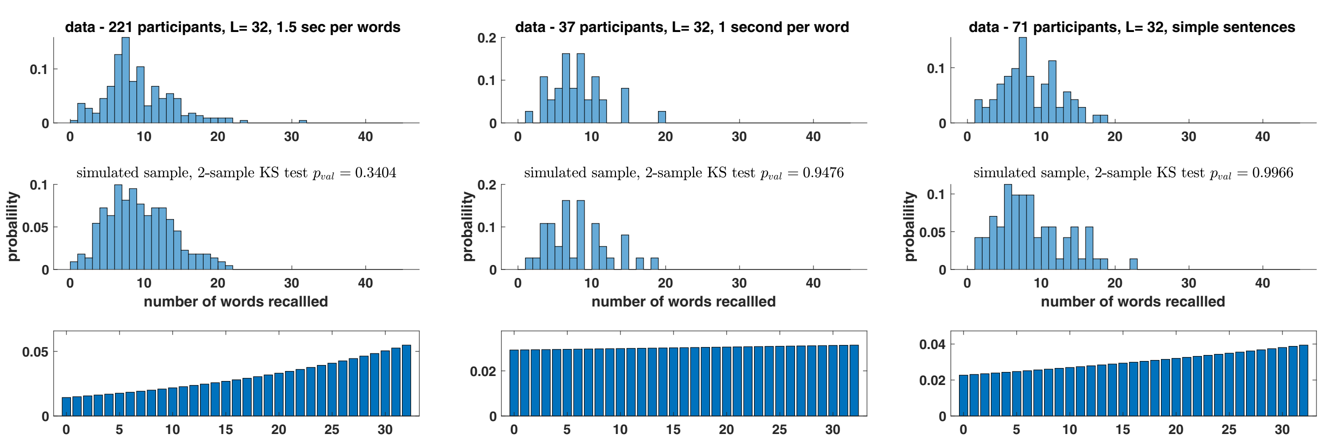

FIG. 6. Top two rows: Comparison between distributions of the number of words recalled for the data and simulated by the model for $L = 32$. Bottom row: Estimated distribution for the number of remembered words.

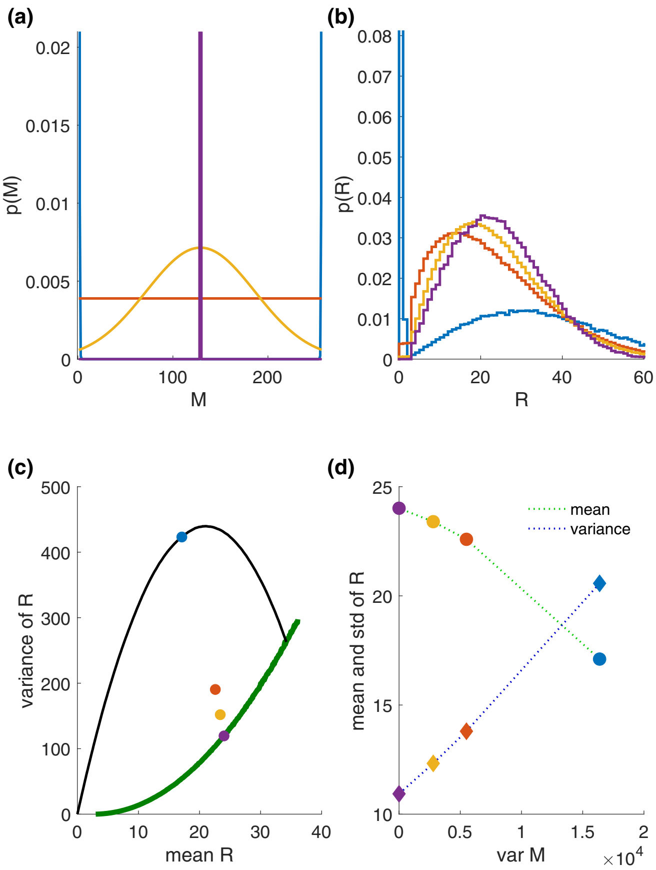

FIG. 7. Sensitivity of the distribution of R to changes of parameters of truncated Gaussian distribution of M. (a) Four distributions of M are shown with mean of 128 and increasing variance, from a single delta-function (magenta) to two delta functions at $M = 0$ and $M = L = 256$ (blue). (b) Distributions of R corresponding to distributions of M shown in (a), obtained in simulations. (c) Mean and variance of R for distributions of M from (a), overlaid on the lower and upper bounds of recall variance. (d) Mean and standard deviation of R as a function of variance in M [numerically computed using distributions shown in (a)]. All four distributions of M were computed using truncated Gaussian ansatz of Eq. (10) with parameters $\mu = 128$ and $\alpha = -0.305176, 0, 0.00305176, 30.5176$ for magenta, red, yellow, and blue distributions, correspondingly.

estimate the number of memory items that were stored in memory immediately prior to recall that agreed well with independent estimates from recognition experiments. This result supports the assumption made in Ref. [13] that recognition experiments can indeed be used to estimate the number of memory items that are candidates for recall following the acquisition. Other models assume that each item presented during acquisition remains in memory to a certain extent and recognition is based on the strength of a familiarity signal for a given item (see, e.g., Ref. [30]). Whether these models can account for the relation between the results of recognition and recall experiments presented in Ref. [13] is an open issue for future studies.

In addition to our experimental results obtained with MTurk, we analyzed the data set from the laboratory of Kahana obtained in the laboratory conditions, with young participants (students) performing multiple recall experiments with lists of 16 words over the course of several daily sessions [22]. As we showed previously in Ref. [25], recall performance averaged across participants exhibited a steady improvement for the first four days before reaching a steady level. Interestingly, we observed that performance on the first day was well fit to the model predictions in terms of both moments, if participants kept all 16 words in memory immediately after presentation. Unfortunately, this prediction of the analysis cannot be directly verified because participants of this study did not perform recognition experiments on individual lists of 16 words. We also considered recall statistics for the small group of 38 older participants collected by Kahana laboratory and presented in Ref. [26]. Unsurprisingly, old participants recalled fewer words on average than young ones (6.34 vs 8.7). The variance of the number of recalled words for these participants was well within the region compatible with our model. It therefore appears that recall in old participants follows the same process as in younger ones but with fewer words that remain in memory after presentation. This tentative conclusion contradicts previous observations that recognition performance of older people is not different from younger ones [31]. In the more recent experiments in Kahana laboratory [32], 38 old participants, whose recall performance we analyzed above, on average exhibited poorer recognition performance compared to their younger peers when tested at the end of the recall session with all the words presented on the same day (16 lists with 256 words in total), in particular, they more often indicated new words as old (false alarms), with no observable difference in correct identifications. The authors applied their previously introduced Temporal Context Model 2 (TCM2) to these data to search for differences between old and young participants. TCM2 is a rather complex model characterized by multiple processes with a correspondingly large number of parameters, and the best account for the results requires changing six of them, in particular, corresponding to the interaction between representations of words and contexts. Given these controversies, we believe that more careful studies of recognition with old participants are warranted.

To conclude, we believe that considering statistics of memory performance, in particular, higher moments of the distribution of performance measures, is a fruitful approach to test various models against the data. The relation between first and second moments in our previously performed experiments is compatible with the free recall model of Ref. [13] but other experimental conditions, in particular when delays are introduced between acquisition and recall or when older participants are involved, should be more carefully studied in future experiments.

ACKNOWLEDGMENTS

We would like to thank Drs. Michael Kahana and Nicolas Lenner for helpful discussions that triggered this study. 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, Is-

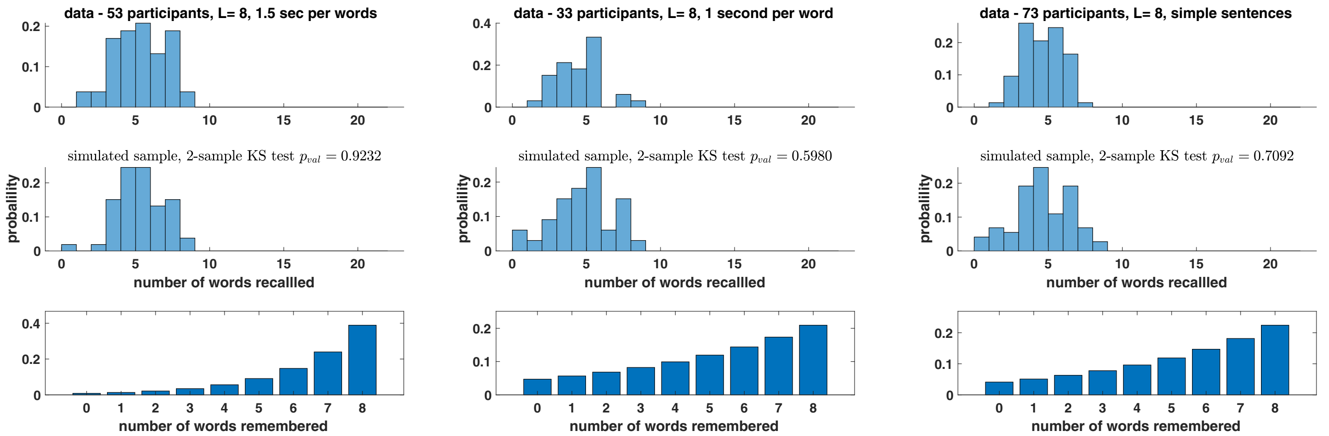

FIG. 8. Top two rows: Comparison between distributions of the number of words recalled for the data and simulated by the model for $L = 8$. Bottom row: Estimated distribution for the number of remembered words.

raeli Science Foundation Grant No. 1657/19, and Foundation Adelis.

APPENDIX A: DERIVATION OF THE UPPER BOUND OF THE VARIANCE OF THE NUMBER OF RECALLED WORDS

Here we consider the maximal value for the second moment of the number of recalled items given the first moment, for lists of length L. The corresponding expression is given by the equation

$$\overline{m_2} = \int_0^L dMP(M)m_2(M), \tag{A1}$$

where $P(M)$ is the probability density function (pdf) for the number of items in memory and $m_2(M)$ is the second moment of the number of recalled items for fixed value of M [i.e., when $P(x) = \delta(x - M)$]. We are interested in finding $P(M)$ that maximizes this expression, given the constraint on the first moment:

$$\overline{m_1} = \int_0^L dMP(M)m_1(M). \tag{A2}$$

Additional constraint should guarantee the normalization condition, $\int_0^L dMP(M) = 1$. Since the expressions for the second moment (that is being maximized) and both constraints are linear functionals of $P(M)$, it is easy to demonstrate that at the maximum, the function $P(M)$ has to be zero everywhere except for a discrete set of points, i.e., given by a sum of delta functions. We can then show that for our form of functions $m_1(M)$ and $m_2(M)$ the maximum is achieved for $P(M) = (1 - c)\delta(M) + c\delta(M - L)$, i.e., sum of delta functions at two extreme values of M. To prove that this is indeed the case, consider adding a third delta function at the intermediate value of M: $P(M) = (1 - c - c_1)\delta(M) + c\delta(M - L) + c_1\delta(M - M_1)$, where $0 < M_1 < L$. Substituting this ansatz to Eq. (A1) and taking into account the constraint for the first moment from Eq. (A2) results in the following expression for the second moment as a function of c_1:

$$\overline{m_2} = \overline{m_1}\frac{m_2(L)}{m_1(L)} + c_1 m_1(M_1)\left(\frac{m_2(M_1)}{m_1(M_1)} - \frac{m_2(L)}{m_1(L)}\right). \tag{A3}$$

Given the above form for the moments, the ratio of the second and first moments, $\frac{m_2(M)}{m_1(M)}$ is an increasing function of M, hence the last term in brackets in this equation is negative and the

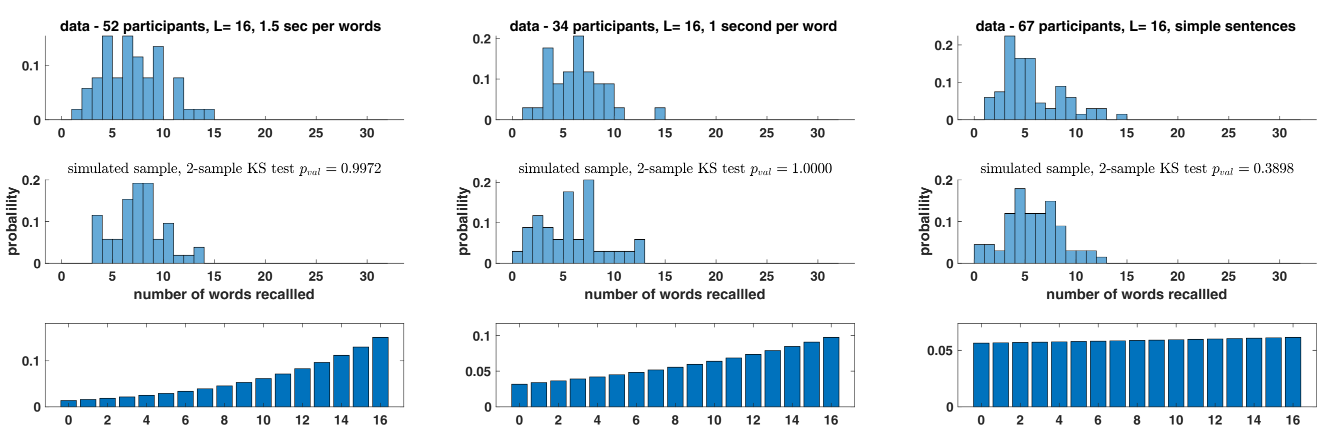

FIG. 9. Top two rows: Comparison between distributions of the number of words recalled for the data and simulated by the model for $L = 16$. Bottom row: Estimated distribution for the number of remembered words.

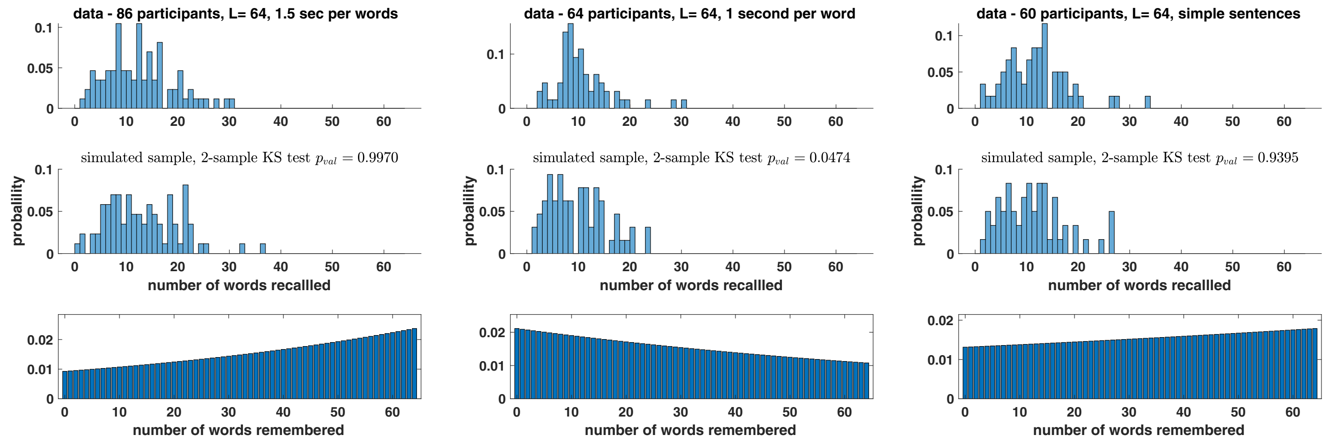

FIG. 10. Top two rows: Comparison between distributions of the number of words recalled for the data and simulated by the model for $L = 64$. Bottom row: Estimated distribution for the number of remembered words.

maximum of the second moment, and hence the variance, is obtained for $c_1 = 0$, resulting in Eqs. (9) of the main text.

APPENDIX B: SIMULATION OF MOMENTS FOR SYMMETRIC RANDOM MATRIX

To obtain first and second moments of recall performance for fixed M, $m_1(M)$, and $m_2(M)$, we simulated 10^5 trials of the free recall model [13] for each list length M from 4 to 512 using software provided in Ref. [24]. To obtain the upper bound, we used Eqs. (9).

APPENDIX C: RECALL DISTRIBUTIONS

Figure 8–13 show the comparison between distributions for the number of recalled items obtained in Ref. [13] and simulated ones. In these simulations, instead of assuming that all participants remember the same number of items, we assumed that there is a distribution of the number of items participants remember before recall. Since we have no additional information, we used truncated Gaussian distributions as a maximum entropy distribution with mean and variance constrained. In particular, this class of distributions can describe the case where all participants remember the same number of items as a limiting case.

The truncated Gaussian has two free parameters to fit, as seen from Eq. (10). We optimized the squared difference between m_1 and m_2 measured in Ref. [13] and obtained in simulations with much larger weight assigned for m_1 than m_2 differences (see below). More specifically, we proceeded in iterations, where around 25 iterations were sufficient to find the optimal values of μ and α, with errors defined by finite sampling. Each iteration consisted of drawing 20 random samples of μ and α, normally distributed, with each sample having different standard deviations. More specifically, for samples $k = 1..20$, standard deviations were $10^{-k/4}$. For each set of μ and σ, we simulated 10^6 recall trials. For each trial, we drew the number of items available for recall M_i, $i = 1..10^6$ from the truncated Gaussian with the current parameters. Then we simulated recall for this M_i using our model from Refs. [8,13,24] to obtain the number of recalled items R_i. Once all R_i were collected, we computed m_1 and m_2. The score was

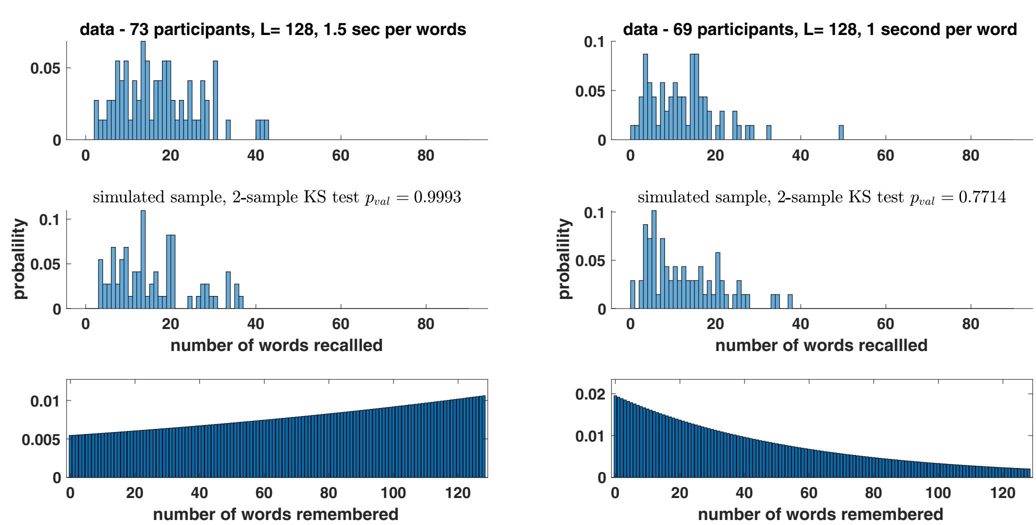

FIG. 11. Top two rows: Comparison between distributions of the number of words recalled for the data and simulated by the model for $L = 128$. Bottom row: Estimated distribution for the number of remembered words.

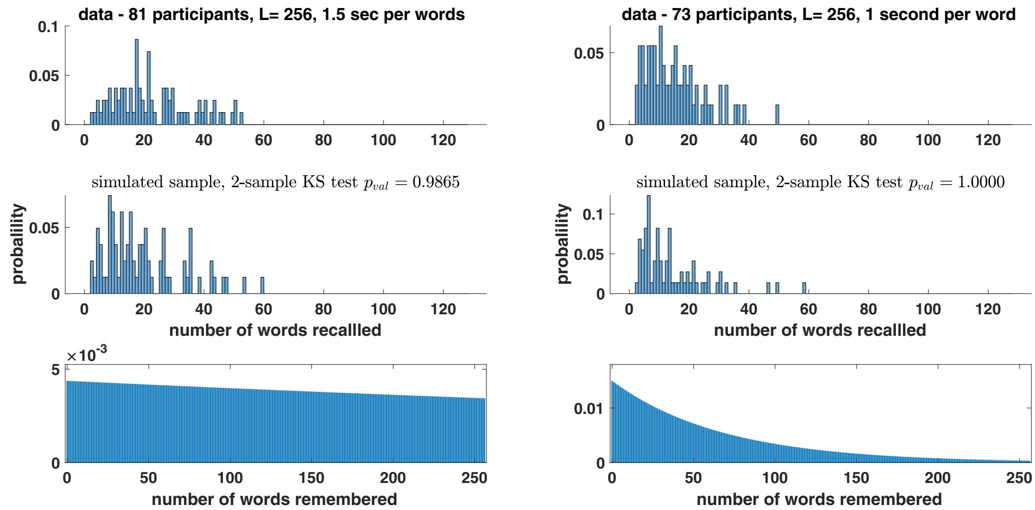

FIG. 12. Top two rows: Comparison between distributions of the number of words recalled for the data and simulated by the model for $L = 256$. Bottom row: Estimated distribution for the number of remembered words.

computed as $10^9(m_{1,d} - m_{1,s})^2 + (m_{2,d} - m_{2,s})^2$, where $m_{\circ,d}$ are moments obtained in experiment, and $m_{\circ,s}$ are moments obtained form simulations. Then the set of parameters providing the smallest score was chosen for the next iteration. Having large weight on m_1 difference practically fixes m_1 to its experimental value. Despite the fact that we only optimized for two moments, the overall distributions are very similar for optimal values, such that the Kolmogorov-Smirnov test could not reject the SM hypothesis (at 5% threshold) for all 18 cases except one ($L = 64$, presentation speed $1s$ per word). In Figs. 8 to 13, in the middle rows, subsamples of the same size as data were randomly selected from the sample used to compute the Kolmogorov-Smirnov test, to visualize data, and simulation distributions.

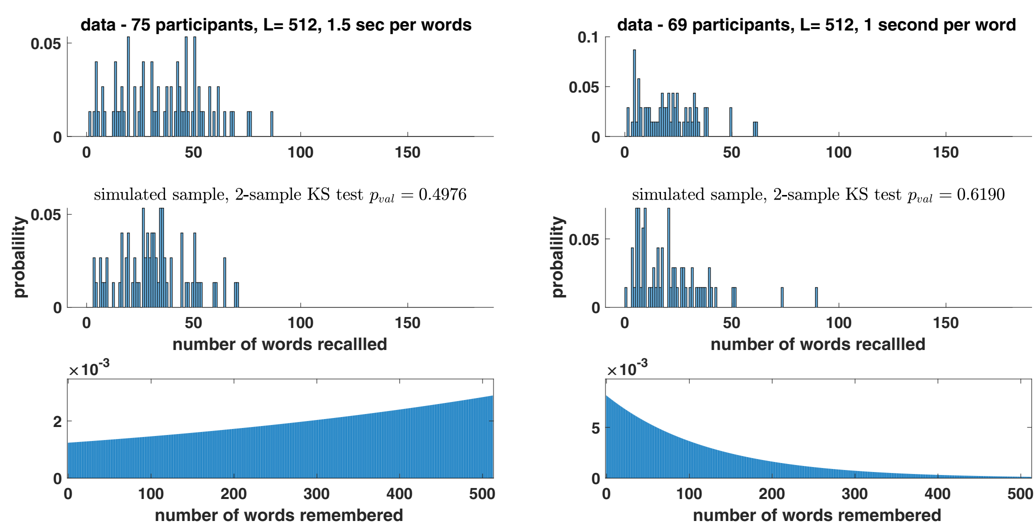

FIG. 13. Top two rows: Comparison between distributions of the number of words recalled for the data and simulated by the model for $L = 512$. Bottom row: Estimated distribution for the number of remembered words.

[1] L. Standing, Learning 10000 pictures, Q. J. Exp. Psychol. **25**, 207 (1973).
[2] M. J. Kahana, Computational models of memory search, Annu. Rev. Psychol. **71**, 107 (2020).
[3] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
[4] M. V. Tsodyks and M. V. Feigel'man, The enhanced storage capacity in neural networks with low activity level, Europhys. Lett. **6**, 101 (1988).
[5] Y. Miyashita, Neuronal correlate of visual associative long-term memory in the primate temporal cortex, Nature (London) **335**, 817 (1988).
[6] H. Gelbard-Sagiv, R. Mukamel, M. Harel, R. Malach, and I.

Fried, Internally generated reactivation of single neurons in human hippocampus during free recall, Science **322**, 96 (2008).

[7] S. A. Josselyn and S. Tonegawa, Memory engrams: Recalling the past and imagining the future, Science **367**, eaaw4325 (2020).

[8] S. Romani, I. Pinkoviezky, A. Rubin, and M. Tsodyks, Scaling laws of associative memory retrieval, Neural Comput. **25**, 2523 (2013).

[9] S. Recanatesi, M. Katkov, S. Romani, and M. Tsodyks, Neural network model of memory retrieval, Front. Comput. Neurosci. **9**, 149 (2015).

[10] M. J. Kahana, Associative retrieval processes in free recall, Mem. Cognit. **24**, 103 (1996).

[11] M. Naim, M. Katkov, S. Recanatesi, and M. Tsodyks, Emergence of hierarchical organization in memory for random material, Sci. Rep. **9**, 10448 (2019).

[12] S. Recanatesi, M. Katkov, and M. Tsodyks, Memory states and transitions between them in attractor neural networks, Neural Comput. **29**, 2684 (2017).

[13] M. Naim, M. Katkov, S. Romani, and M. Tsodyks, Fundamental law of memory recall, Phys. Rev. Lett. **124**, 018101 (2020).

[14] B. B. Murdock Jr., The immediate retention of unrelated words, J. Exp. Psychol. **60**, 222 (1960).

[15] B. B. Murdock Jr., The serial position effect of free recall, J. Exp. Psychol. **64**, 482 (1962).

[16] 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).

[17] M. W. Howard and M. J. Kahana, Contextual variability and serial position effects in free recall, J. Exp. Psychol.: Learning, Memory, Cognition **25**, 923 (1999).

[18] 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. Cognit. **28**, 530 (2002).

[19] K. A. Klein, K. M. Addis, and M. J. Kahana, A comparative analysis of serial and free recall, Mem. Cognit. **33**, 833 (2005).

[20] 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. Cognit. **32**, 792 (2006).

[21] 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. Cognit. **36**, 1207 (2010).

[22] J. F. Miller, C. T. Weidemann, and M. J. Kahana, Recall termination in free recall, Mem. Cognit. **40**, 540 (2012).

[23] R. Grenfell-Essam, G. Ward, and L. Tan, Common modality effects in immediate free recall and immediate serial recall, J. Exp. Psychol.: Learn. Mem. Cognit. **43**, 1909 (2017).

[24] M. Katkov, S. Romani, and M. Tsodyks, Memory retrieval from first principles, Neuron **94**, 1027 (2017).

[25] S. Romani, M. Katkov, and M. Tsodyks, Practice makes perfect in memory recall, Learn. Mem. **23**, 169 (2016).

[26] M. K. Healey and M. J. Kahana, A four-component model of age-related memory change, Psychol. Rev. **123**, 23 (2016).

[27] P. Verhaeghen, A. Marcoen, and L. Goossens, Facts and fiction about memory aging: A quantitative integration of research findings, J. Gerontol. **48**, P157 (1993).

[28] A. H. Criss, K. J. Malmberg, and R. M. Shiffrin, Output interference in recognition memory, J. Mem. Lang. **64**, 316 (2011).

[29] T. F. Irvine Myron Tribus and J. P. Hartnett, *Rational Descriptions, Decisions and Designs* (Pergamon Press, Elmsford, NY, 1969).

[30] R. M. Shiffrin and M. Steyvers, A model for recognition memory: Rem-retrieving effectively from memory, Psychon. Bull. Rev. **4**, 145 (1997).

[31] D. Schonfield and B.-A. Robertson, Memory storage and aging, Can. J. Psychol. **20**, 228 (1966).

[32] M. K. Healey, P. Crutchley, and M. J. Kahana, Individual differences in memory search and their relation to intelligence, J. Exp. Psychol.: General **143**, 1553 (2014).