Practice makes perfect in memory recall

Sandro Romani,[1,4] Mikhail Katkov,[2,4] and Misha Tsodyks[2,3]

[1]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA; [2]Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel; [3]Department of Neuroscience, Columbia University, New York, New York 10032, USA

A large variability in performance is observed when participants recall briefly presented lists of words. The sources of such variability are not known. Our analysis of a large data set of free recall revealed a small fraction of participants that reached an extremely high performance, including many trials with the recall of complete lists. Moreover, some of them developed a number of consistent input-position-dependent recall strategies, in particular recalling words consecutively ("chaining") or in groups of consecutively presented words ("chunking"). The time course of acquisition and particular choice of positional grouping were variable among participants. Our results show that acquiring positional strategies plays a crucial role in improvement of recall performance.

[Supplemental material is available for this article.]

In the classical free recall paradigm, participants are asked to repeat a list of randomly assembled words in an arbitrary order. It was found that the recall is not always perfect even for the shortest lists, and for longer lists the fraction of recalled words is steadily decreasing (Roberts 1972; Standing 1973; Murray et al. 1976, Ward et al. 2010). The variability of recall performance across participants is very large (Healey et al. 2014), but its origins, and the potential contribution of practice, are not clear (Dallet 1963). Most studies of recall capacity present the data averaged across many participants. Uncovering the origins of recall variability is of great interest for general understanding of human memory. In particular, variability could reflect inherent differences in generic memory abilities, an ability to improve with practice ("learning to learn" Keppel et al. 1968), and/or application of different strategies of recall (Healey et al. 2014). In (Romani et al. 2013) we proposed that recall could be limited by random overlaps between long-term distributed neuronal representations of words. In Katkov et al. (2015), we showed that this model is broadly compatible with the large data set of immediate free recall collected in the laboratory of M. Kahana (UPenn, Miller et al. 2012). A closer inspection of the data revealed however that a small fraction of participants exhibited a performance well above what could be expected from the model with random encodings. We thus set out to consider individual differences in recall, with a particular focus on participants exhibiting very good performance. One hundred seventy participants were instructed to recall multiple lists of 16 words randomly assembled for each trial that were not repeated over the course of the experiment. Each participant performed 112 recall trials, divided into seven daily sessions.

As reported in previous studies, the variability in performance across participants is very large; we also found that this variability is steadily increasing from session to session (Supplemental Fig. 1). Surprisingly, some of the participants recall all presented words ("perfect trials", PT) in a relatively large number of trials. Overall, the fraction of PTs was 3%, but the distribution across participants was highly nonuniform, ranging from 0 (for 80 participants) to 42 PTs out of 112 (Fig. 1A). Comparing this distribution to the one obtained by randomly shuffling the number of words recalled in each trial between the participants, we selected the group of 19 participants who each achieved more than 10 PTs, i.e., 0.08% of shuffled control ("perfect participants"). These participants are the main focus of our study.

In Figure 1 (B–D, upper panels) we present recall trials of three selected participants from this group. All recalled words are shown from bottom to top in the order of their recall (output position), in color that corresponds to their serial position in the presented list (input position, from blue to yellow). The first participant (Fig. 1B) consistently exhibited a "chunking" strategy, characterized by a tendency to recall the words in four sets of four consequently presented words. Indeed, as seen in the middle panel of Figure 1B (color code here indicates the chunk number from 1 to 4), this participant almost invariably completed the recall of a chunk before recalling words from a different chunk. Note that the order of chunks, as well as the order of words within a chunk exhibits a high degree of variability; however, the recall always begins with the last chunk (yellow stripes at Fig. 1B, middle panel). The second participant (Fig. 1C) initially exhibited a random order of recalled words except for a pronounced recency effect (first words recalled were the ones presented at the end of the list; see yellow stripes near the first output position). This participant then gradually developed a new recall strategy characterized by a tendency to divide the words in five sets according to their position in the presented list (33334 chunking; chunk 1: words 1–3, chunk 2: words 4–6, chunk 3: words 7–9, chunk 4: words 10–12, chunk 5: words 13–16). The recall performance of this participant increased as he adopted this strategy (average number of words recalled increased from 10.6 on trials 1–58 to 13.3 on trials 59–112; $P < 10^{-9}$, two-sample t-test). Finally, the third participant (Fig. 1D) exhibited a similar initial behavior as the second one, but at the 41st trial he abruptly switched to "forward chaining", i.e., tendency to recall words consecutively in the order in which they are presented (blue to yellow: extreme version of temporal contiguity, see Howard and Kahana 1999). The
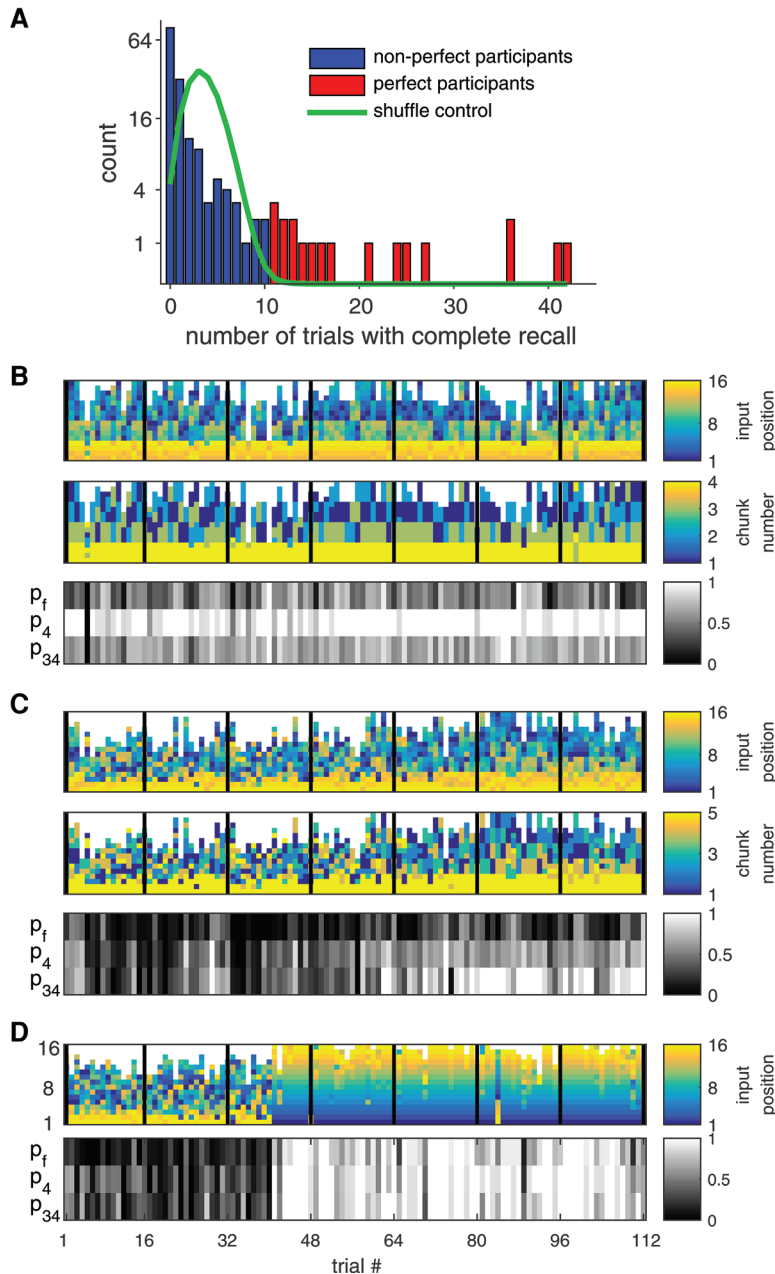
Figure 1. Variability of performance and positional recall strategies (*A*) Number of perfect trials across participants. Red, perfect participants; blue, nonperfect participants; green, bootstrap distribution. (*B*) All 112 trials of a selected participant. (*Upper* panel) recalled words are shown from *bottom* to *top* in the order of their recall, and their color corresponds to their input order in the presented list, from blue to yellow. The number of words recalled at each trial is given by the number of colored squares in the corresponding vertical column. Black vertical lines denote the end of each daily session. (*Middle* panel) same data but colors represent the chunk number of a word according to its position on the presented list, for 4444 grouping (chunks 1–4: words 1–4, 5–8, 9–12, 13–16, correspondingly). (*Lower* panel) probability measures for three strategies, computed for each trial of this participant using maximum likelihood. The rows denoted by p_f, p_4, p_{34} correspond to forward, 4444, and 33334 chunking strategy, respectively (see text and Materials and Methods). (*C*) Same as *B* for another participant, (chunks 1–5: words 1–3, 4–6, 7–9, 10–12, 13–16, correspondingly). (*D*) Another participant, who abruptly develops a forward chaining strategy after the 41st trial.

were not repeating across trials, this remarkable change in recall pattern reflects a genuine acquisition of a new strategy rather than learning of a particular list.

Figure 1 illustrates the role of positional grouping and learning in recall performance. We thus set out to characterize the ordering of recalled words exhibited by different participants. Since the recall performance of some participants was highly nonstationary (e.g., Fig. 1C,D), we aimed to specify ordering strategies *on a trial-by-trial basis*, as opposed to averaging over all trials of a particular participant (cf. Sederberg et al. 2010; Healey et al. 2014). The number of possible groupings of a list of 16 words is very large. To simplify the analysis, we first selected three strategies illustrated in Figure 1, as we found that they play a dominant role in the recall of perfect participants (see Supplemental Information section Variety of Strategies). To quantify the extent of strategy applications for a given trial, we developed a single-trial measure that specifies the probability that a certain strategy is applied, based on the maximum-likelihood principle (p_f, p_4, p_{34} for forward chaining, 4444, and 33334 chunking, correspondingly; see Supplemental Methods). The largest of these probabilities, p_{max}, reflects the extent of positional grouping for a given trial. To illustrate this analysis, we show the probabilities for the three strategies for the participants whose recall is reported in Figure 1B–D (see lower panels). For the first two participants, only one of the probabilities was high when they used the corresponding chunking strategies. When the third participant switched to forward chaining, all three probabilities increased. This is because trials where the words are recalled in the correct order can also be divided in sets of consequently presented words.

We used the probability measures to quantitatively analyze how strongly participants applied different recall strategies.

Across all participants, the correlation between the average p_{max} value exhibited by a given participant and the number of perfect trials that he achieved is significant (correlation coefficient 0.69). We compared the extent of positional strategy applied by perfect participants versus the rest, by computing the average p_{max} values for all the trials of the corresponding groups characterized by the specific number of words recalled (WRs) (Since p_{max} tends to increase for trials with large number of words recalled, we present the results separately for all the trials with a given WR. We exclude trials with WR < 6 because for these trials recency effect artificially boosts the values of p_{max}). Perfect participants exhibited stronger

change in recall order structure is accompanied by a sudden increase in performance (height of each column; average number of words recalled increased from 11.8 on trials 1–40 to 13.3 on trials 41–112; $P < 10^{-21}$, two-sample *t*-test). Since the presented lists

positional grouping than the rest of the participants for all WRs (Fig. 2A). Focusing specifically on PTs, the histogram of p_{max} is characterized by a large peak at $p_{max} = 1$ (44% of the trials), i.e., in almost half of the perfect trials of these participants the order of recall was fully compatible with one of the three positional strategies (Fig. 2B).

The above analysis of positional recall strategies was entirely based on the order in which presented words were recalled. To independently confirm different recall strategies, we considered the time it takes to recall a new word for progressive output positions. In previous studies, it was shown that chunking in serial recall is reflected in the timing of recall (Kahana and Jacobs 2000; Maybery et al. 2002). When only selecting the trials with strong chunking ($p_{max} = 1$), we observed pronounced peaks in the time it takes to recall a word following a transition from one chunk to another (Fig. 2D, middle and right columns; upper row: average over all corresponding trials; lower row: individual trials for perfect participants). For trials assigned to the 4444 strategy, peaks are observed at the output positions 5, 9, and 13, i.e., when participants begin to recall new chunks. For the 33334 strategy, the first peak is at the fifth output position, because most of the time participants begin the recall with the last chunk of size 4 (see Fig. 1C). A surprising result seen in Figure 2D is that recall progresses with approximately constant speed, except for transitions between the

chunks (this is especially true for perfect participants—see red curves in Fig. 2D). This result is in stark contrast to a pronounced overall slow-down of recall obtained when taking into account all the trials—the average recall time increases by a factor of six toward the last output position (see Fig. 2C; Murdock and Okada 1970; Katkov et al. 2015). We conclude that positional strategies strongly influence the temporal progression of recall; in particular, transitions within a chunk are much faster than between chunks and do not become slower for subsequent chunks.

As seen in Figure 1, the performance of some participants improved over the course of the experiment, both in terms of the number of words recalled and the number of perfect trials. We thus computed the recall performance in subsequent trials, averaged over all participants (Fig. 3A). Averaged over each daily session, the data show a gradual and very modest (one word) improvement in performance until saturating after the fourth session, similar to a typical pattern observed in other types of learning, such as e.g., perceptual learning (Lu et al. 2011). Within sessions, there is an interesting repeating pattern that begins with a brief increase of performance in the first trial compared with the previous day. This effect is similar to the increase in performance after the night break that was reported in perceptual learning (Karni and Sagi 1993). The secondary peak toward the end of each session corresponds to a short break that participants took after the 12th trial. Perfect participants showed a much stronger improvement with practice (reaching three words; Fig. 3B), compared with others. Moreover, perfect participants were improving their performance for the whole duration of the experiment.

The strongest practice effects were observed in the number of perfect trials. The overall fraction of PTs was steadily increasing from 1% over the first two sessions to 6% toward the end of the experiment (Fig. 3C). For the group of perfect participants, the fraction of PTs rose dramatically to 30%. Finally, we considered the evolution of strategy applications over subsequent trials. At each trial we computed the average value of p_{max} over participants that recalled a specific number of words (WR). We then average the obtained values over all WRs. The corresponding graphs show an increase of strategy application for both perfect and nonperfect participants, which is more pronounced for the former group (red vs. blue curves in Fig. 3D, upper panel). Mean p_{max} for perfect participants increased from 0.45 to 0.62 between first and last two sessions, respectively ($P < 10^{-19}$, two-sample t-test). The average difference in p_{max} between perfect participants and the rest increased from 0.06 to 0.18 for early and late trials ($P < 10^{-9}$, two-sample t-test). Analysis of individual strategies (lower panels of Fig. 3D), reveals the emergence of chunking dominating the positional strategies toward the end of the experiment (last two rows in each of the lower panels; see also the number of trials with $p_{max} = 1$ assigned to different strategies, Fig. 2D). We thus conclude that with
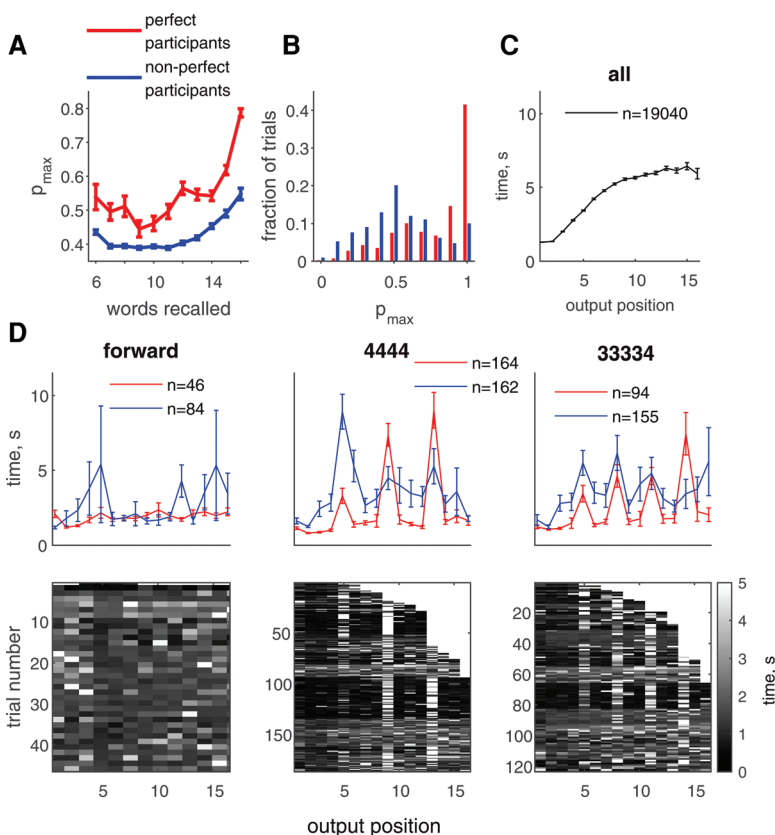

Figure 2. Perfect trials and perfect participants. (*A*) Average p_{max} values for trials with different numbers of words recalled. Red, perfect participants; blue, nonperfect participants. Error bars: SEM. (*B*) Histogram of p_{max} over the perfect trials, for perfect participants (red) and the rest (blue). (*C*) Average recall times for subsequently recalled words across all trials. (*D*) (*Top* row) Average recall times for subsequently recalled words for specific strategy, separately for perfect participants and the rest (only trials with $p_{max} = 1$ are used, strategy indicated in the title; see Materials and Methods). Red: perfect participants, blue: the rest. The numbers in the legends show the number of trials used to compute average recall times. (*Bottom* row) color-coded recall times in the individual trials for perfect participants (trials are sorted according the number of recalled words).
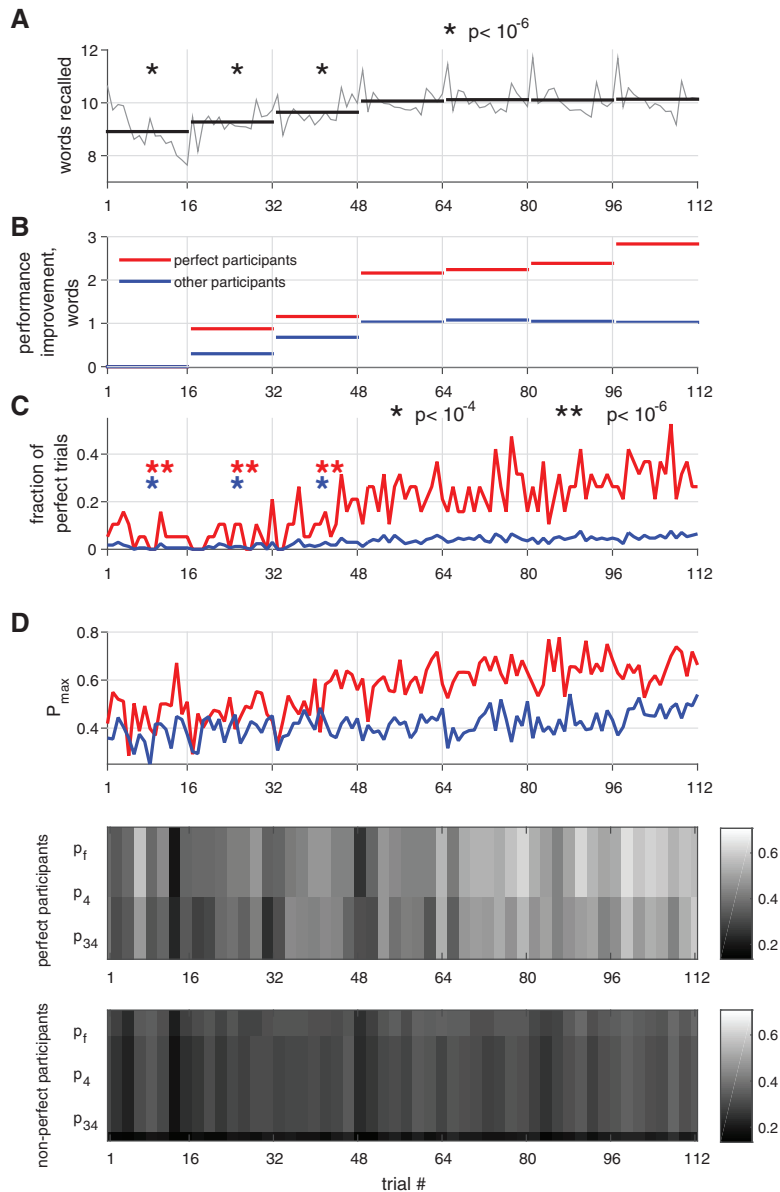
Figure 3. Practice-induced improvement in recall performance. (*A*) Number of words recalled over subsequent trials, averaged over all participants. Vertical grid lines demark daily sessions consisting of 16 trials each. Black horizontal lines: average performance in the daily sessions. (*B*) Day to day improvement in performance for perfect (red) and remaining (blue) participants. (*C*) Fraction of perfect trials for perfect (red) and remaining participants (blue). (a,c), stars: daily performances differing significantly from that of the last day—two-sample (composed of individual trials) *t*-test. (*D*) (*Upper* panel) average p_{max} values over consecutive trials (see text and Materials and Methods for details). Red: perfect participants, blue: rest of participants. (*Middle* and *lower* panels): average probability measures for different strategies over consecutive trials (see Materials and Methods) for perfect (*middle*) and the rest of participants (*lower* panel).

practice recall becomes progressively more structured in terms of one of the positional strategies.

In our analysis we only considered three types of strategies that were dominating for perfect participants (see Supplemental Information). We found however that a limited number of recall trials is more consistent with other types of chunking. These strategies were not consistently used by any of the perfect participants and adding them to our analysis did not change our conclusions (see Supplemental Information). It is possible that in a bigger data

set, participants who use one of these other strategies could also be found. Interestingly, no strategies with chunks longer than 5 words were observed in the data. It appears that chunk sizes are limited by short-term memory capacity that is involved in chunking (Graesser and Mandler 1978; Cowan 2001).

The results presented in this study reveal diverse dynamics of recall with randomly assembled lists of words. Overall, practice results in performance improvement, expressed both in the increase in the number of words recalled and in the number of perfect trials. The average degree of improvement was very modest compared with more robust "learning to learn" effects in other experimental paradigms where participants had multiple exposures to the same lists (Dallet 1963; Keppel et al. 1968). Some of the participants however achieved very high performance, reaching a complete recall of many presented list ("Perfect Participants"). Compared with the rest, perfect participants exhibited stronger positional recall strategies. In particular, forward chaining and chunking emerged as the most significant strategies associated with successful recall performance, with chunking dominating the trials where a strong degree of positional strategy was observed (Fig. 2D). Moreover, successful application of such strategies resulted in a dramatic speed-up of the recall process. A recent study identified temporal contiguity (closely related to chaining) as an important grouping strategy that is correlated with recall performance (Healey et al. 2014). We show in Supplemental Information that chunking in our data is not a byproduct of contiguity.

Chunking in memory recall was previously observed only when experimenters introduced pauses in the list (Shuell 1969; Reitman and Rueter 1980; Maybery et al. 2002; Farrel and Lewandowsky 2008) or when the same list was repeatedly presented for recall (Tulving 1962; Rosner 1971; Sternberg and Tulving 1977; Kahana and Jacobs 2000). The current study is the first one where chunking was shown to emerge for single presentations of lists of words and without any instructions. Such "spontaneous" chunking could arise from online segmentation of incoming stream of inputs into consecutive groups, or "events" (Kurby and Zacks 2008). Recall would then reflect this segmentation by associating each word with the representation of the corresponding event (Farrell 2012) or by simply creating temporal associations between different words within a chunk.

Our results indicate that practicing free recall of randomly assembled lists cannot substantially improve the performance for most people, possibly due to hard limitations arising from

overlapping distributed encoding of words in long-term memory (Romani et al. 2013; Katkov et al. 2015). Participants who acquire chunking could recall words in a chunk in rapid succession, thus increasing the performance and making the report faster. Interestingly, the recall speed observed within a chunk is similar to the recall speed of the first few words in trials without chunking (see Fig. 2C,D). Previous studies interpreted this faster retrieval as an indication that the corresponding words are held in short-term memory (see e.g., Raaijmakers and Shiffrin 1981). We hypothesize that words in a chunk are simultaneously retrieved in short-term memory. Hence, chunking suggests a way to build a hierarchical representation where words representation is replaced by chunk representation at the next level of hierarchy. We speculate that similar processes could be critical in other memory-related tasks where fast and/or composite processing is crucial, as e.g., language production (Doughty and Long 2008; Kormos 2014).

Our study raises a host of issues that are not addressed by the current models of memory recall (Kahana 2012), in particular how some of the participants progressively learn to more robustly implement a positional strategy, and why different strategies are selected. Since chunking is a crucial process involved in many different types of memory-related cognitive activity (see e.g., Kurby and Zacks 2008), a better understanding of the mechanisms by which participants acquire chunking could have wide implications beyond the free recall paradigm. Interesting further research questions would be for how long acquired recall strategies are kept by participants and whether their acquisition improves other memory-related cognitive processing, such as e.g., language comprehension.

Acknowledgments

References

Cowan N. 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* **24:** 87–114.

Dallet KM. 1963. Practice effects in free and ordered recall. *J Exp Psychol* **66:** 65–71.

Doughty CJ, Long MH. 2008. *The handbook of second language acquisition*. Wiley, Oxford, UK.

Farrell S. 2012. Temporal clustering and sequencing in short-term memory and episodic memory. *Psychol Rev* **119:** 223–271.

Farrell S, Lewandowsky S. 2008. Empirical and theoretical limits on lag recency in free recall. *Psychon Bull Rev* **15:** 1236–1250.

Graesser A II, Mandler G. 1978. Limited processing capacity constrains the storage of unrelated sets of words and retrieval from natural categories. *J Exp Psychol* **4:** 86–100.

Healey M, Crutchley P, Kahana MJ. 2014. Individual differences in memory search and their relation to intelligence. *J Exp Psychol* **143:** 1553–1569.

Howard MW, Kahana MJ. 1999. Contextual variability and serial position effects in free recall. *J Exp Psychol Learn Mem Cogn* **25:** 923–941.

Kahana MJ. 2012. *Foundations of human memory*. Oxford University Press, New York.

Kahana MJ, Jacobs J. 2000. Interresponse times in serial recall: effects of intraserial repetition. *J Exp Psychol Learn Mem Cogn* **26:** 1188–1197.

Karni A, Sagi D. 1993. The time course of learning a visual skill. *Nature* **365:** 250–252.

Katkov M, Romani S, Tsodyks M. 2015. Effects of long-term representations on free recall of unrelated words. *Learn Mem* **22:** 101–108.

Keppel G, Postman L, Zavortink B. 1968. Studies of learning to learn: VIII: the influence of massive amounts of training upon the learning and retention of paired-associate lists. *J Verbal Learn Verbal Behav* **7:** 790–796.

Kormos J. 2014. *Speech production and second language acquisition*. Routledge, London.

Kurby CA, Zacks JM. 2008. Segmentation in the perception and memory of events. *Trends Cogn Sci* **12:** 72–79.

Lu ZL, Hua T, Huang CB, Zhou Y, Dosher BA. 2011. Visual perception learning. *Neurobiol Learn Mem* **95:** 145–151.

Maybery MT, Parmentier FBR, Jones DM. 2002. Grouping of list items reflected in the timing of recall: implications for models of serial verbal memory. *J Mem Lang* **47:** 360–385.

Miller JF, Weidemann CT, Kahana MJ. 2012. Recall termination in free recall. *Mem Cognit* **40:** 540–550.

Murdock B, Okada R. 1970. Interresponse times in single trial free recall. *J Exp Psychol* **86:** 263–267.

Murray DJ, Pye C, Hockley WE. 1976. Standing's power function in long-term memory. *Psychol Res* **38:** 319–331.

Raaijmakers JGW, Shiffrin RM. 1981. Search of associative memory. *Psychol Rev* **2:** 93–134.

Reitman JS, Rueter HH. 1980. Organization revealed by recall orders and confirmed by pauses. *Cogn Psychol* **12:** 554–581.

Roberts WA. 1972. Free recall of word lists varying in length and rate of presentation: a test of total-time hypothesis. *J Exp Psychol* **92:** 365–372.

Romani S, Pinkoviezky I, Rubin A, Tsodyks M. 2013. Scaling laws of associative memory retrieval. *Neural Comput* **25:** 2523–2544.

Rosner SR. 1971. The effects of rehearsal and chunking instructions on children's multitrial free recall. *J Exp Child Psychol* **11:** 93–105.

Sederberg PB, Miller JF, Howard WH, Kahana MJ. 2010. The temporal contiguity effect predicts episodic memory performance. *Mem Cognit* **38:** 689–699.

Shuell TJ. 1969. Clustering and organization in free recall. *Psychol Bull* **72:** 353–374.

Standing L. 1973. Learning 10000 pictures. *Q J Exp Psychol* **25:** 207–222.

Sternberg RJ, Tulving E. 1977. The measurement of subjective organization in free recall. *Psychol Bull* **84:** 539–556.

Tulving E. 1962. Subjective organization in free recall of "unrelated" words. *Psychol Rev* **69:** 344–354.

Ward G, Tan L, Grenfell-Essam R. 2010. 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:** 1207–1241.