



# Negative recency effects in delayed recognition: Spacing, consolidation, and retrieval strategy processes

Rona Sheaffer<sup>1</sup> · Daniel A. Levy<sup>1</sup>

Accepted: 14 February 2022 / Published online: 11 March 2022  
© The Psychonomic Society, Inc. 2022

## Abstract

While items learned immediately before testing are generally remembered better than prior items in a study list, in delayed testing this relationship is reversed, yielding a negative recency effect. To adjudicate between the strategic rehearsal and spacing accounts of this phenomenon, we examined performance of 169 participants on a delayed recognition test following multiple sessions requiring the study and immediate free recall testing of 16 lists of 16 words. This revealed a strong effect of the amount of spacing between initial study position and initial free recall position on the degree of negative recency, supporting the spacing account. Furthermore, these spacing effects were nonmonotonic, suggesting that they are mediated by consolidation processes. Additional analyses indicate that strategies and rehearsal opportunities may also contribute to the effects of within-list encoding position on subsequent long-term memory, but for recall more than for recognition.

**Keywords** Recognition · Spacing · Recall · Consolidation · Rehearsal

Imagine that you are at a departmental function, meeting unfamiliar researchers from other labs and trying to remember their names. You even use the trick you once read about, of repeating the person's name in conversation with them: "Hi George . . . nice to make your acquaintance, George . . . I really love your work, George . . .". Now imagine your chagrin the next day, when you bump into George in the corridor, but can't for the life of you remember his name. Why didn't the name repetition trick help consolidate your memory of that name?

There's a good explanation for your forgetfulness—you ignored the spacing effect. Repeated exposure to information strengthens memory for it, but spaced repetitions are most effective (Madigan, 1969; Melton, 1970; and more recently Cepeda et al., 2006; Delaney et al., 2010). Had you repeated George's name throughout the event and not in a massed fashion, the chances of retaining that information for future recall would increase significantly.

Recently, Kahana and colleagues (Kuhn et al., 2018) have appealed to the spacing effect to explain an interesting aspect of free recall. In *immediate* free recall of word lists, the most recently studied items are generally better recalled than mid-list items, an effect often attributed to the retrieval of final items from short-term memory (Davelaar et al., 2005; Talmi et al., 2005). When cognitively demanding activity intervenes between study and test, this effect dissipates. Furthermore, when after initial immediate free recall and subsequent cognitive activity, a *final* free recall test is administered, the prominent recency effect observed in immediate free recall is actually reversed, with list-end items being remembered more poorly than earlier items ( Craik, 1970). In other words, there is a positive within-list recency effect for immediate retrieval, and a negative within-list recency effect for delayed re-retrieval. Initially, this negative recency effect was adduced to provide evidence in favor of dual-store models of word list memory. Craik argued that items from the end of a word list were held in short-term memory buffer for the least amount of time. Therefore, although they had a higher probability of being immediately retrieved, they have the least strength in longer-term memory. Single store models sought alternative

---

✉ Daniel A. Levy  
daniel.levy@idc.ac.il

<sup>1</sup> Baruch Ivcher School of Psychology, Interdisciplinary Center Herzliya, 8 HaUniversita St, 46150 Herzliya, Israel

explanations of the negative recency effect. Contrary to accounts that attribute it to strategic rehearsal processes (e.g., Tan & Ward, 2000), Kuhn et al. (2018) provide evidence<sup>1</sup> that negative recency is a function of spacing (i.e., the amount of time or the number of cognitive steps intervening between initial study and immediate free recall). This challenges the strategic rehearsal processes account, which ascribes importance only to an item's list position in the presentation phase and not to its output position in the immediate recall phase. Kuhn and colleagues propose that since items presented late in an encoding list are likely to be immediately recalled early on, those items have the least amount of spacing between activations, and therefore the subsequent memory for them is weaker than for items earlier in the encoding sequence. However, the crucial factor is not study list position alone, but the degree of spacing resulting from the study list position and the output position in the immediate free recall test. This account can accommodate the negative recency phenomenon within a single-store model of list memory retrieval patterns.

If the spacing of prior exposures is responsible for the long-term negative-recency effect, we should expect to see that effect not only on subsequent (final) free recall, but on other expressions of memory traces, such as recognition tests. This is because the second exposure via recall is said to serve as a further encoding event ( Craik, 1970; Kuhn et al., 2018). That becomes apparent upon considering that retrieval in initial free recall is a paradigmatic case of retrieval practice (recently reviewed by McDermott, 2021). Although retrieval practice affects recollection more strongly than familiarity, in a multi-list paradigm like the current one, it was found to improve overall recognition (Chan & McDermott, 2007, Exp. 3). Similarly, spacing effects are reported for recognition memory, especially for its recollective aspect (Glenberg, 1976; Hintzman, 1969; Parkin & Russo, 1993; Zhao et al., 2015). Therefore, to the extent that the mechanisms responsible for the spacing effect yield negative recency for recall, they should yield negative recency for recognition. If the effect of study-recall spacing duration is found for subsequent (final) recognition testing as well, it would provide additional support for the explanation put forward by Kuhn and colleagues; absent such a finding, the evidence in favor of the spacing account and against the strategic retrieval account would be equivocal.

List serial position effects on delayed (subsequent/final) recognition tests have been explored in several classic studies. Craik et al. (1970) reported a negative recency effect in delayed recognition following immediate free recall tests, parallel to the negative recency effect found for final free

recall (Craik, 1970). In contrast, Cohen (1970) reported positive recency effects in delayed recognition following immediate free recall tests. Engle (1974) noted that in the study of Craik et al. (1970), inclusion in recognition success analysis was conditional on being successfully recalled in the immediate test, while in Cohen (1970) inclusion in the recognition analysis was not conditional on prior production in immediate free recall. Engle (1974) therefore directly compared conditional and unconditional analyses, as well as manipulating presentation rate. He reported that words recalled in initial free recall (IFR) exhibited negative recency at all presentation rates, but that was not the case for non-recalled words. Engle (1974) also reported that subsequent recognition confidence ratings for words that had been recalled in IFR increased linearly with IFR output position, reflecting greater subsequent memory strength. Similarly, McCabe and Madigan (1971) reported that the final (recency) item in a five-element sequence of word pairs was identified with the lowest confidence level in delayed recognition after immediate free recall. In contrast, Engle and Durban (1977), examining both auditory and visual presentations, found a small positive recency effect, and Darley and Murdock (1971), who employed a three-alternative forced choice recognition test, found no recency effects.

Notably, delayed recognition accuracy (or confidence) as a function of specific study-output lag for each IFR item was not reported in any of the abovementioned studies. Lacking that data, even in those studies that do report a negative recency effect, it is difficult to adjudicate between Kuhn et al.'s (2018) claim that spacing is responsible for long-term negative recency in delayed recognition and an alternative explanation—that list-end items strategically receive more immediate rote rehearsal (McCabe & Madigan, 1971; Reitman, 1970). In that alternative account, those list-end items are also said to be produced in initial output positions based on registration in short-term memory, while primacy items, when immediately remembered, are retrieved from long-term memory formed during initial encoding, and so are also more successfully identified in delayed recognition tests.

Furthermore, there is reason to suspect that in a final recognition test a different pattern of effects might be observed than those reported by Kuhn et al. (2018) for final free recall. Numerous studies by Zacks and colleagues (e.g., Swallow et al., 2009; reviewed in Radvansky & Zacks, 2017) have documented the strong impact of proximity to event boundaries on subsequent memory. In recognition memory, the most potent retrieval cue is the memorandum itself, which is used as the mnemonic probe—a copy cue (Tulving, 1983). In a test of recognition, the strategic retrieval processes employed in free recall may be overshadowed by copy cue strength and proximity to boundaries—i.e., study list end and/or to an early output position in the initial free recall

<sup>1</sup> Based on the analysis of a large database of free recall performance (the Penn Electrophysiology of Encoding and Retrieval Study [PEERS]); see references to numerous studies of these data provided by Kuhn et al. (2018).

stage—which lead to strong encoding might yield the best final recognition performance.

Therefore, examination of negative recency in delayed recognition, and identification of its possible mechanism if occurring, require detailed information about the lags between study presentation and production in IFR. Fortunately, the PEERS data set includes a number of experimental sessions in which delayed recognition tests were administered at the conclusion of learning and immediate free recall testing of all lists. Furthermore, unlike earlier reports, this data enable examination of the exact serial order of recall in immediate free recall, enabling more precise quantification of the study-to-free-recall lag. We were therefore able to assess whether the negative recency effect in delayed recognition is indeed a function of spacing between initial encoding and immediate free recall in the lists presented during the earlier phases of the experiment.

In addition to investigating whether final recognition performance would support the spacing hypothesis of Kuhn et al. (2018) by performing the same analyses they executed, but on the recognition data, we took advantage of the scope of the PEERS database to conduct additional analyses tracking recognition probabilities as a function of list recency and within-list recency. The most important of these analyses examines the degree of monotonicity of the effects of spacing. As detailed below, this revealed some unexpected findings.

## Method

The current investigation is based recognition data collected as part of Experiment 1 of the Penn Electrophysiology of Encoding and Retrieval Study (PEERS).<sup>2</sup> One hundred seventy-one participants participated in Experiment 1 (consisting of seven experimental sessions).<sup>3</sup> Two of these participants were excluded from the present analyses due to data corruption, leaving  $N = 169$ . Each of the seven sessions consisted of 16 lists of 16 words presented one at a time on a computer screen. Each study list was followed by an immediate free recall test, and each session ended with a comprehensive recognition test. Half of the sessions were randomly chosen to include a final free recall test, which took place before the recognition test.

Earlier PEERS publications report details of the method (for a more complete description, see Healey et al., 2014; Lohnas & Kahana, 2014). In brief, each item was on the screen for 3,000 ms, followed by jittered 800–1,200-ms interstimulus interval.<sup>4</sup> After the last item in the list, there was a 1,200–1,400-ms jittered delay, after which a tone sounded, a row of asterisks appeared, and the participant was given 75 seconds to vocally recall the just-presented items in any order. If a session was randomly selected for final free recall (FFR) test, participants performed the FFR test following the 16th immediate free recall (IFR) test. An instruction screen informed participants that they had 5 minutes to recall all the items from the preceding lists in any order. An old/new recognition test was administered after either FFR or after the last list's IFR test. In total, 320 words were presented serially on the computer screen, with target/lure ratio varying with session, and targets comprising 80%, 75%, 62.5%, or 50% of the total trial items. Participants were instructed to indicate for each word whether the test word had been presented previously. Recognition trials were self-paced. Feedback on accuracy and reaction time was provided after each trial.

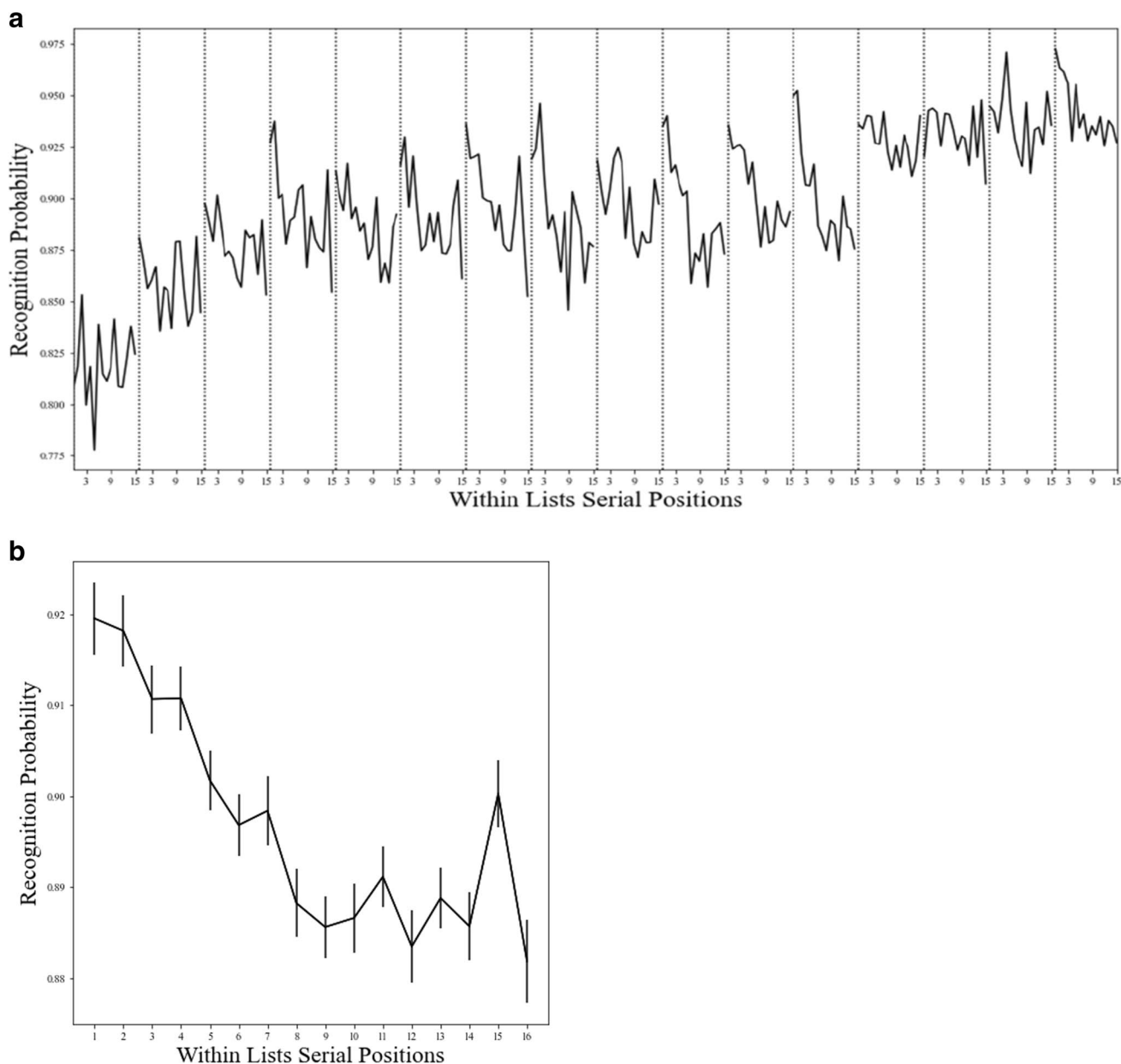
## Results

Before turning to the analyses relevant to the question under consideration, it is instructive to get a sense of the overall patterns of performance in delayed recognition. One interesting pattern is the probability of final recognition as a function of the item's serial position during encoding, considering both the list that each item came from and the item's serial position within that list, for a total of 256 ( $16 \times 16$ ) possible positions. We therefore conducted a repeated-measures analysis of variance (ANOVA), with list number (Lists 1–16) and within-list serial position (Slots 1–16) as within-subject factors. Mauchly's test of sphericity indicated that degrees of freedom in both main effects required correction; Greenhouse–Geisser epsilon was applied accordingly. We found a main effect of between-list position,  $F(11.34, 1905.58) = 47.92, p < .001$ , and a main effect of within-list position,  $F(12.72, 2136.08) = 10.02, p < .001$ , but no interaction between those factors,  $F(83.19, 13976.51) = 1.09, p = .270$ . As can be seen in Fig. 1a (top), there is a positive long-term

<sup>2</sup> Raw data from the PEERS studies was obtained from Michael Kahana's web page (<http://memory.psych.upenn.edu>).

<sup>3</sup> Kuhn et al. (2018) include the data of PEERS Experiments 2 and 3 in their analyses. Since the former involved continual-distractor free recall, and the latter included some participants performing externalized free recall (reporting intrusions), we did not include those data in our analyses.

<sup>4</sup> Words were either presented concurrently with a task cue indicating one of two judgments (size or animacy) to be made for that word, or with no encoding task. Kuhn et al. (2018) report that the encoding task had minimal effects on recall performance, and, importantly, nearly identical negative-recency effects across encoding conditions. We have therefore followed their method and aggregated data across these conditions.

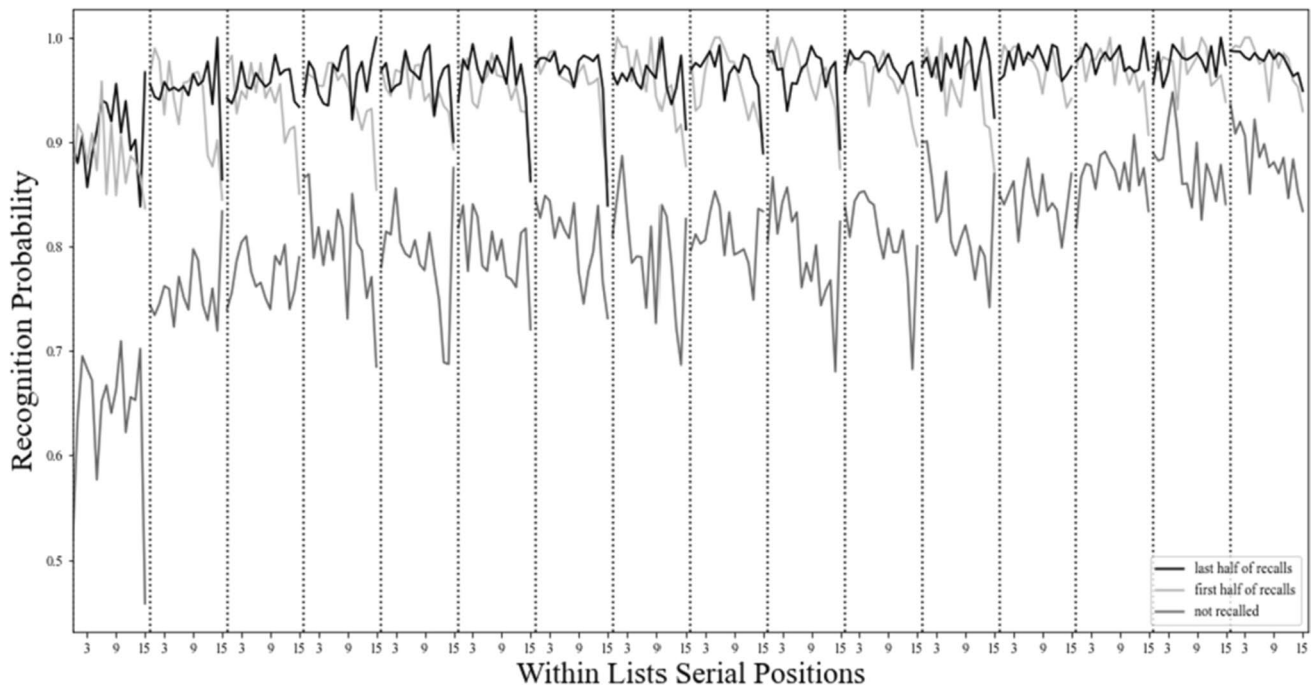


**Fig. 1** Final recognition as a function of serial position during encoding. All 16 lists of Fig. 1a (up) are averaged together in Fig. 1b (down), to create a within-list serial position curve. Error bars indicate *SEM*

recency effect *across* lists, such that, as might be expected, participants recognized more items from recent than from remote lists; this is reflected in the strong positive-slope linear trend in the across list effect,  $F = 347.1$ ,  $p < .001$ , and the absence of a significant quadratic trend,  $F = 3.61$ ,  $p > .05$ . This detrimental effect of study–test delay, with the concomitant number of intervening lists (possibly leading to retroactive interference on subsequent recognition), is not surprising. Moreover, since for the investigation of the spacing effect proposal the key factor is the within-list position at study (and at initial free recall), and

the across-list position is not relevant, nor did it interact with the within-list effects, we will not further comment on that finding.

To provide a clearer picture of the shape of the *within*-list serial position curve, we aggregated the data across all 16 lists; in this analysis, we included all items, whether or not they were successfully recalled in the initial free recall (IFR) stage. The shape of the average serial position curve presented in Fig. 1b (bottom) illustrates that there is also a primacy effect within each list. In fact, recognition probability appears to decrease with each



**Fig. 2** Final recognition as a function of recency of encoding, for three classes of items: those recalled in early output positions in IFR (first half of recalls), those recalled in late output positions in IFR (last half of recalls), and items that were not recalled in IFR

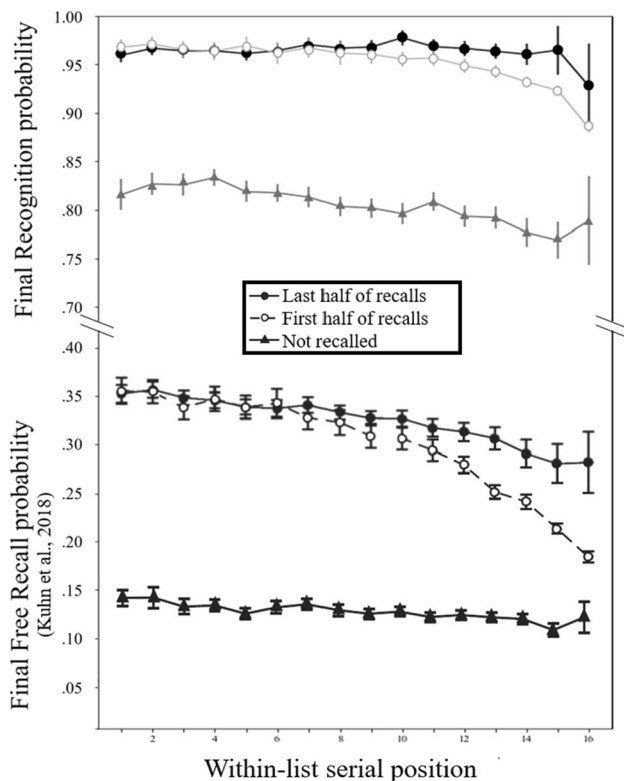
additional item on the list, until reaching its lowest value at the middle of the list (Item 9), followed by the asymptotic mid-list plateau observed in classic serial position curves for immediate free recall. At the penultimate point, recognition probability rises again (Item 15). This is reflected in the presence of both linear ( $F = 49.32$ ,  $p < .001$ ) and quadratic ( $F = 26.46$ ,  $p < .001$ ) components in the within-list effect. However, the comparison with Fig. 3 and the free recall data provided by Lohnas and Kahana (2014); Fig. 3a) indicates that this spike is not an indication of advantage in principle, but rather a function of the greater relative proportion for the words in this position achieving successful initial free recall (~81% vs. an average of 59% for items in Positions 9–14), for which subsequent recognition is better. Notably, and once again unlike classic serial position curves for immediate free recall, there is a sharp decline in final recognition success for the final 16th item, despite items in that position having been initially recalled at a rate of ~94% (Lohnas & Kahana, 2014; Fig. 3a). The implications of these performance trends will be explored in the Discussion.

To examine the proposal of Kuhn et al. (2018) regarding spacing effects on negative recency, we proceeded to analyze recency effects, examining memory for later list items compared to earlier items. We assume that the mere retrieval of an item produces learning (e.g., Karpicke & Roediger, 2008). Therefore, to avoid confounds of the effects of final free recall before long-term recognition, all the analyses

presented here (including Fig. 1, above) consider final recognition data only for PEERS Experiment 1 sessions that did not include the FFR test.<sup>5</sup> Moreover, the analyses consider recognition data separately for items that participants did or did not initially recall during their IFR trials, except for the detailed spacing analysis, which considers the number of items between encoding and subsequent IFR, and is therefore based solely on the former items.

To test the hypothesis that spacing between encoding and initial retrieval influences final recognition, we classified the initially recalled items into two categories according to their output positions, defining early output positions as the first half of outputs, and late output positions as the second half (as in Kuhn et al., 2018). This partitioning resulted in three classes of items: not initially recalled, recalled early, and recalled late. Figure 2 shows the probability of final recognition for these three item types as a function of recency of encoding, considering both the item's list number and the item's serial position within that list. It illustrates three major effects in final recognition. First, participants correctly recognized almost all the items that they recalled during IFR, as was reported for FFR by Kuhn et al. (2018). Second, participants recognized more items

<sup>5</sup> In practice, we found very similar negative-recency effects for words in both FFR and no-FFR sessions, but still believe it preferable to avoid that potential confound.



**Fig. 3** *Upper part of panel:* Final recognition probabilities as a function of within-list position, averaged across all 16 lists of Fig. 2. Error bars depict 1 SEM (calculated using the methods described in Cousineau, 2005, and Morey, 2008). The large SEM seen at Positions 15 and 16 for items recalled late and unrecalled items results from a large drop in the number of items recalled at these positions (see Appendix, Supplementary Table 1, for the exact number of items recognized in each position). *Lower part of panel:* Final free recall probabilities, as reported by Kuhn et al., 2018. For both recall and recognition, the negative recency effect is strongest (i.e., retrieval probability is relatively weaker) for items recalled in early output positions but attenuated (i.e., retrieval probability is relatively stronger) for items recalled in later output positions

from recent than from remote lists. However, this effect was seen primarily for items that participants failed to recall in IFR. Third, we found a pronounced within-list negative-recency effect, which was most dominant for items that were recalled in early output positions during IFR. To provide a clearer picture of the relationship between this negative recency and item type, in Fig. 3 we aggregated the data across all 16 lists.<sup>6</sup> Like the results of Kuhn et al. (2018); Fig. 2a) regarding free recall, Fig. 3 demonstrates that the negative recency effect is

<sup>6</sup> Figure 3 portrays mean performance for each serial position across all participants rather than means of individual participants' performance, because of the different numbers of participants contributing data to the later positions, which required the mixed-models analysis conducted on this data, as detailed.

**Table 1** Results of a mixed model analysis comparing recognition probability between the first and last half of retrievals, separately for each of the serial positions (retrieval position half was used as fixed factor and participant as a random factor)

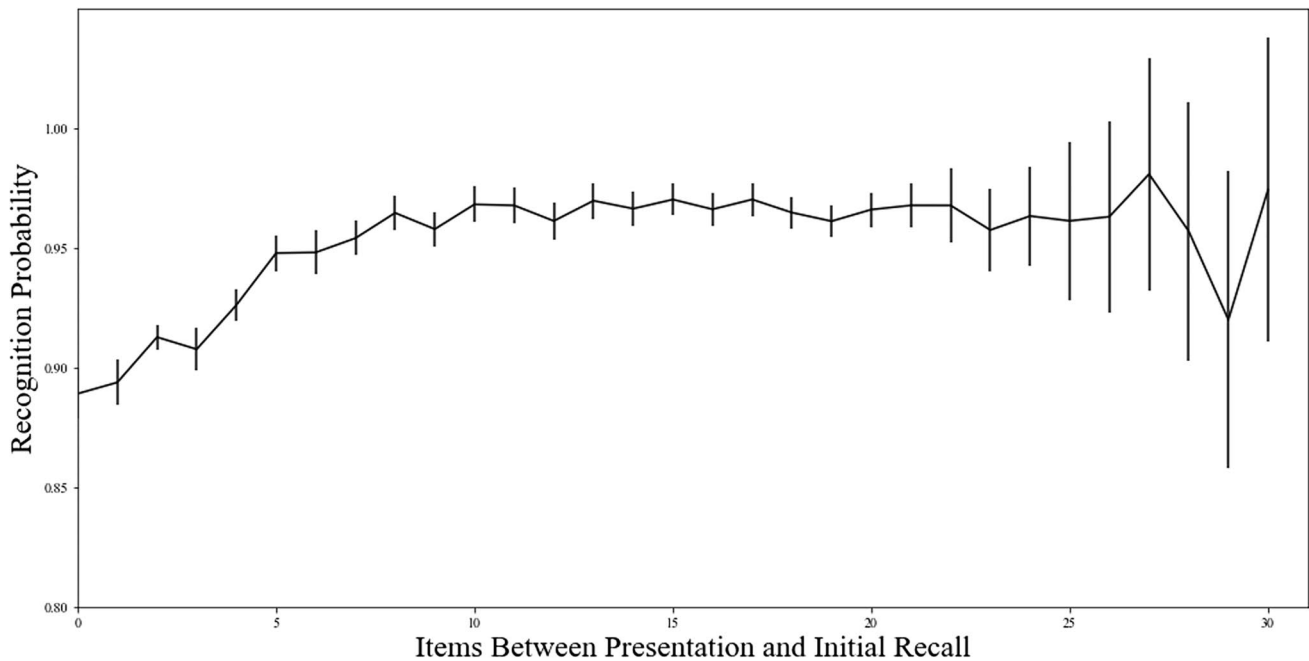
Serial Position	<i>N</i> (first half)	<i>N</i> (second half)	<i>df</i>	<i>F</i>	<i>p</i> value
1	166 (3)	166 (3)	(1,330)	.142	.706
2	165 (4)	165 (4)	(1,327)	1.033	.310
3	166 (3)	166 (3)	(1,330)	.227	.634
4	163 (5)	167 (2)	(1,328)	.077	.782
5	164 (4)	165 (4)	(1,326)	3.162	.076
6	161 (6)	165 (4)	(1,323)	1.125	.290
7	163 (4)	165 (4)	(1,326)	.583	.446
8	164 (3)	166 (3)	(1,327)	.127	.722
9	164 (4)	167 (2)	(1,329)	.566	.452
10	160 (9)	166 (3)	(1,323)	2.965	.086
11	166 (3)	166 (3)	(1,329)	2.666	.103
12	167 (2)	165 (4)	(1,329)	14.718	.000
13	164 (5)	167 (2)	(1,329)	5.158	.024
14	169 (0)	167 (0)	(1,333)	7.145	.008
15	166 (3)	144 (5)	(1,307)	54.677	.000
16	168 (1)	101 (6)	(1,266)	34.224	.000

*Note.* Columns 2 and 3 present the number of participants included in the analysis for each position (participants who recalled one or more items), not including excluded outliers. The numbers in parentheses represent the number of participants that were excluded

attenuated when the analysis is restricted to either not-recalled items or items recalled late in the recall period. Thus, this finding provides additional support for the spacing-based account of negative recency ( Craik, 1970; Kuhn et al., 2018).

This inference is further supported by an analysis focusing on the initially recalled items (the two top curves in Fig. 3), which allowed us to specifically compare items recalled in the first half of recalls to those recalled in the last half, as a function of the encoding serial position (ranging from 1-16). The number of items in each of these serial positions varied greatly. Only 107 participants (out of 169) recalled one or more items from the 16<sup>th</sup> encoding position in late output positions, whereas all 169 did so when considering early output positions. This was simply because items from encoding position 16 were much more likely to be produced early in the recall sequence. Therefore, we subjected these data to mixed model analyses, which is most appropriate for unbalanced data (Baayen et al., 2008; Tibon & Levy, 2015). For each of the serial positions, outlier participants whose recognition performance was 3 SD below group mean were excluded<sup>7</sup>. Item

<sup>7</sup> On average, three participants were excluded from each serial position. See Columns 2 and 3 on Table 1 for the exact number of participants excluded from each position.



**Fig. 4** Probability of final recognition as a function of the spacing between initial presentation and initial recall. A positive correlation was found, demonstrating that the probability of recalling an item during final recognition increases as the number of items between the initial presentation and IFR increases. Error bars depict 1 *SEM* (cal-

culated using the methods described in Cousineau, 2005, and Morey, 2008). Note that the large *SEM* seen at the right end of the graph derives from the scarcity of items recalled after such a large spacing (see Appendix, Supplementary Table 2, for the exact number of items recognized for each spacing)

serial position during initial encoding, and IFR output position (first vs. second half) were fixed factors, with participant as a random factor (West, 2009). This analysis yielded a main effect of item serial position,  $F(15, 5171) = 16.38, p < .001$ , and a main effect of IFR output position,  $F(1, 5171) = 49.96, p < .001$ . Importantly, and consistent with the spacing-based account of negative recency, there was a significant interaction between serial position and IFR output position,  $F(15, 5171) = 9.66, p < .001$ . As shown in Fig. 3, the proportion of items recognized in Encoding Positions 1–11 was not affected by IFR output position, whereas the proportion of recognized items from Encoding Positions 12–16 was significantly lower if produced in the first half of IFR output positions than if produced in the last half. This is confirmed by the mixed model analyses that compared recognition probability between the first and last half, separately for each of the serial positions (see Table 1 for the complete statistical report, and Supplementary Table 1 for the detailed descriptive statistics). This indicates the specificity of the negative recency effect on final recognition to items previously produced in the first half of IFR output positions, as might be expected given the shorter spacing between the initial encoding and retrieval of these items.

To gauge the spacing account more directly, we examined the probability of correct final recognition as a function of the exact spacing between the initial presentation of the items and their recall during the IFR (measured as the

number of intervening items; possible values range from 0 to 30). As explained in the Introduction, the spacing account of negative recency posits that memory for items should improve consistently as the spacing between the two learning episodes (initial encoding and IFR) increases. Therefore, end-of-list items, which have shorter spacings than early list items on average, should be remembered less well, and this effect should be strongest for end-of-list items that are recalled early (shortest spacings of all). Consistent with the spacing account, Fig. 4 portrays the positive correlation between spacing and recognition probability, such that final recognition performance rises with the spacing between an item's position during study and its position in IFR. Following the methods used by Kuhn et al. (2018), we computed several variations of the correlation between spacing amount and recognition probability separately for each participant.<sup>8</sup> Across all possible spacings (0–30), the distribution of correlation coefficients was significantly positive (mean correlation = 0.27),  $t(167) = 15.41, p < .001$ . However, Fig. 4 indicates that such correlations might not be equivalent across

<sup>8</sup> Several participants were excluded from the correlational analyses due to perfect performance (constant recognition probability), which does not/did not allow correlation calculation. The exact number of participants excluded from each analysis can be deduced from the degrees of freedom reported.

all spacing positions. We therefore dismantled the general correlation into three separate examinations of spacing-recognition correlations for spacing amounts 0–10, 11–20, and 21–30. Indeed, these subset correlations demonstrate that the overall positive correlation derives completely from the leftmost part of the graph (spacings 0–10), mean  $r = .34$ ,  $t(166) = 15.54$ ,  $p < .001$ , as clearly seen in Fig. 4, whereas the correlation in the other two subcategories is either null (mean  $r = -.02$ ,  $t(160) = 0.67$ ,  $p = .502$ ), or negative (mean  $r = -.73$ ),  $t(168) = 29.42$ ,  $p < .001$ .<sup>9</sup> This seems to indicate that spacing between prior exposures indeed improves subsequent recognition performance, but that once sufficient spacing is provided, further spacing does not provide additional memory strength.

## Discussion

We assessed the spacing account of the delayed recency effect in retrieval of studied word lists (Kuhn et al., 2018), by examining whether its predictions would hold not only for final free recall but also for final recognition testing. We found that this was indeed the case: in comprehensive delayed recognition testing conducted on a large corpus of studied words, the probability of a word being correctly recognized was significantly influenced by the spacing between its initial presentation and its initial immediate free recall. As in the analyses reported by Kuhn et al. (2018), this was found both in comparison of halves of recall sessions and for individual spacings. However, a more nuanced examination of this relationship demonstrated that unlike the case in final free recall (Kuhn et al., 2018, Fig. 2b), this relationship was not monotonic. Increases from immediate repetition up to ~10 intervening items led to greater likelihood of correct subsequent endorsement, but additional spacing did not further improve performance.

Considering both the consonance and the dissonance between the recognition and recall spacing effects may be instructive regarding the possible mechanistic bases of the spacing effect in general (Greene, 1989). Although this issue is not discussed by Kuhn et al. (2018), the spacing account accords well with the notion that immediate post-encoding consolidation processes, including changes on the neural level, are required for the stability of newly formed memories (Ben-Yakov & Dudai, 2011; Dudai et al., 2015). Alternatively, the advantage of spaced repetitions might be in the changing temporal context (Howard & Kahana, 1999,

2002) of the activation of the word representations; association with multiple contexts might increase the likelihood of the subsequent recall and recognition of a studied probe (Smith, 1982; Smith & Handy, 2016). A third possibility, suggested by many earlier studies of within-session spacing effects (reviewed by Delaney et al., 2010), is that participants use different encoding strategies for each encounter with a stimulus in massed versus spaced presentations. This claim is supported by participant reports (Delaney et al., 2010).

Resolution of this question—whether consolidation factors, temporal context changes, or encoding processing differences are responsible for negative recency in delayed recognition—might be provided by the non-monotonic effects in the current data (Fig. 4). This finding accords with an early report of non-monotonic spacing effects on retrieval. This may be found in Glenberg (1976, Exp. 3), in which memory for trigrams was tested at lags of 8, 32 or 64 items after a second presentation, which lagged 0, 1, 8, 20, or 40 items after the initial presentation. Glenberg reported a nonmonotonic spacing advantage at the two longer retention intervals, which approached asymptote following a lag of 8 items between study presentation – comparable to our finding of an asymptotic trend beginning at a lag of ~10 items. Beyond that report, however, the variety of paradigms employed in earlier research makes it difficult to draw direct comparisons. As noted by Benjamin and Tullis (2010), few studies of spacing effects have made use of recognition testing. Additionally, nonmonotonicity in the effects of spacing on the order of days and even months between presentations (as summarized, e.g., in the meta-analysis of Cepeda et al., 2006) are likely to involve very different processes than lags of a maximum of 75 seconds as in the present study. A more fundamental difference between the present paradigm, and earlier research is that in a preponderance of studies, the second (and sometimes third, etc.) encounter with to-be-remembered information is in an additional study trial identical to the initial study trial. Under such circumstances, differences in strategic encoding processes between spaced and massed presentations (Delaney et al., 2010) might very well play a role. In contrast, in the current paradigm, the second exposure to the information is when it is produced in the IFR. In this case, the participants are not engaging in intentional encoding, but rather in intentional retrieval. It therefore seems that the two most robust candidates for explaining the current results are consolidation factors and temporal context changes.

We propose that for recognition memory, which enable retrieval judgments based on familiarity, the benefits to repeated encoding might significantly depend on metabolic re-potentialization of the neuronal networks representing the words to be remembered (Feng et al., 2019; Xue et al., 2011). Once those neurons have had enough time to recoup their ability to conduct the processes required for

<sup>9</sup> The negative correlation in the third subcategory (spacing 21–30) should be viewed with caution due to the scarcity of items recalled in this subcategory and the large standard deviation; see Supplementary Table 2.



Hebbian plasticity, little further benefit might be derived from additional delay (Smolen et al., 2016). While temporal context factors may also impact on recognition, in the final recognition test paradigm, those influences might be attenuated. In this paradigm, the multiple temporally structured study–test stages of the preceding part of the experiment are disrupted by the presentation of a comprehensive set of recognition probes in random order, so those factors might be overshadowed by the strength of the recognition test copy cue, as noted in the Introduction. Therefore, in the current data as well in the comparable retention intervals in Glenberg (1976), spacing benefits become asymptotic after ~10 items. In contrast, for recall, the greater the difference between the temporal context of the initial (study) and subsequent (initial free recall) exposures to a given target word, the greater the variety of associative cues available, with the greater chance that it will be recollected in the final free recall test. As opposed to the recognition test in which the probes are presented randomly, disrupting the possibility of using temporal context retrieval of one target to cue additional targets, in free recall participants may use temporal context to perform further retrievals of items that were temporally proximal to just-retrieved words. Thus, while both temporal context and consolidation process factors affect both free recall and recognition, the two tasks may be sensitive to these influences, to different degrees. Consolidation factors might be more relevant for final familiarity-based recognition (yielding an asymptotic limit to spacing benefit), while temporal context factors might play a greater role in recollection-based final free recall (in which Kuhn et al., 2018, observed no asymptote in spacing benefits).

One aspect of the current data may require explanation through an additional process. Examination of the overall recognition probabilities for each serial position (Fig. 1b; that graph also includes the words not retrieved in initial free recall) reveals that there is a sharp drop in recognition probability for the last list item; indeed, it is the item least likely to be later recognized, despite the very strong recency effect reported for initial free recall of items in that position (~94%; Lohnas & Kahana, 2014; Fig. 3a). Although the differences are numerically small, given the power inherent in the dataset, this anomaly should not be ignored. We suggest that it results from the fact that initial free recall began immediately after list presentation. The last list item, enjoying the standard recency advantage in immediate free recall, may be reported from short-term memory, and therefore subject to minimal retrieval processing and remembered less well in the long run (Crowder, 1976; McCabe & Madigan, 1971; Reitman, 1970; Tan & Ward, 2000). In contrast, primacy and mid-list words that are not held in a short-term store require IFR retrieval by strategic recall processes. Those retrieval processes might serve as the type of deep

encoding that underlies the testing effect (Karpicke & Roediger, 2008), leading to stronger subsequent recognition in a delayed test. This distinction cannot account for the asymptotic profile of spacing effects observed for all the other input position items in the recognition test but may be relevant to understanding poor delayed recognition of the final input position items. We noted above that the negative recency effect was initially adduced to provide evidence for dual-store models of word list memory ( Craik, 1970); the finding of Kuhn et al. (2018) that the crucial factor is not study list position but rather study-to-initial recall spacing provides an alternative explanation that accommodates single-store models. Seemingly, this anomalously poor delayed recognition of the final list items might be seen as providing support for a limited version of dual-store models, applying only to the final list item. It seems to us, though, that resolution of the larger issue is better served by paradigms that combine anatomical and physiological assays and interventions with behavioral patterns (e.g., Innocenti et al., 2013; Kloth et al., 2020; Talmi et al., 2005) in providing evidence for or against process dissociation.

In conclusion, performance on delayed comprehensive recognition tasks generally supports the spacing account of the negative-recency effect observed in retrieval of word lists (Kuhn et al., 2018). Moreover, the pattern of nonmonotonicity in negative recency effects hints to a link between these behavioral findings and cellular-molecular processes implicated in the early consolidation of new learning. The rich pattern of results revealed by the large dataset including multiple tests of a large number of participants indicates that rehearsal and retrieval strategy factors may also play a role in affecting what we will subsequently remember and what we will subsequently forget.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.3758/s13421-022-01293-3>.

**Acknowledgements** The authors thank Michael Kahana and Lynn Lohnas for helpful guidance, and Noam Isachar for assistance with programming.

**Funding** This work was supported by the Israel Science Foundation grant 2497/18 to D.A.L.

## References

- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective?. *Cognitive Psychology*, 61(3), 228–247. <https://doi.org/10.1016/j.cogpsych.2010.05.004>

- Ben-Yakov, A., & Dudai, Y. (2011). Constructing realistic engrams: Poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *Journal of Neuroscience*, *31*, 9032–9042. <https://doi.org/10.1523/jneurosci.0702-11.2011>
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354. <https://doi.org/10.1037/0033-2909.132.3.354>
- Chan, J. C., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 431–437. <https://doi.org/10.1037/0278-7393.33.2.431>
- Cohen, R. L. (1970). Recency effects in long-term recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *9*, 672–678. [https://doi.org/10.1016/s0022-5371\(70\)80031-7](https://doi.org/10.1016/s0022-5371(70)80031-7)
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45. <https://doi.org/10.20982/tqmp.08.3.p182>
- Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, *9*, 658–664. [https://doi.org/10.1016/s0022-5371\(70\)80042-1](https://doi.org/10.1016/s0022-5371(70)80042-1)
- Craik, F. I. M., Gardiner, J. M., & Watkins, M. J. (1970). Further evidence for a negative recency effect in free recall. *Journal of Verbal Learning and Verbal Behavior*, *9*, 554–560. [https://doi.org/10.1016/s0022-5371\(70\)80101-3](https://doi.org/10.1016/s0022-5371(70)80101-3)
- Crowder, R. G. (1976). *Principles of learning and memory*. Erlbaum.
- Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, *91*(1), 66–73. <https://doi.org/10.1037/h0031836>
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, *112*, 3–42. <https://doi.org/10.1037/0033-295x.112.1.3>
- Delaney, P. F., Verkoeijen, P. P., & Spigel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation*, *53*, 63–147. [https://doi.org/10.1016/S0079-7421\(10\)53003-2](https://doi.org/10.1016/S0079-7421(10)53003-2)
- Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron*, *88*, 20–32. <https://doi.org/10.1016/j.neuron.2015.09.004>
- Engle, R. W. (1974). Negative recency in delayed recognition. *Journal of Memory and Language*, *13*(2), 209–216. [https://doi.org/10.1016/S0022-5371\(74\)80045-9](https://doi.org/10.1016/S0022-5371(74)80045-9)
- Engle, R. W., & Durban, E. D. (1977). Effects of modality of presentation on delayed recognition. *Perceptual and Motor Skills*, *45*, 1203–1210. <https://doi.org/10.2466/pms.1977.45.3f.1203>
- Feng, K., Zhao, X., Liu, J., Cai, Y., Ye, Z., Chen, C., & Xue, G. (2019). Spaced learning enhances episodic memory by increasing neural pattern similarity across repetitions. *The Journal of Neuroscience*, *39*, 5351–5360. <https://doi.org/10.1523/JNEUROSCI.2741-18.2019>
- Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 1–16. [https://doi.org/10.1016/S0022-5371\(76\)90002-5](https://doi.org/10.1016/S0022-5371(76)90002-5)
- Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 371–377. <https://doi.org/10.1037/0278-7393.15.3.371>
- Healey, M. K., Crutchley, P., & Kahana, M. J. (2014). Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology: General*, *143*, 1553–1569. <https://doi.org/10.1037/a0036306>
- Hintzman, D. L. (1969). Recognition time: Effects of recency, frequency, and the spacing of repetitions. *Journal of Experimental Psychology*, *79*, 192–194. <https://doi.org/10.1037/h0026943>
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 923–941. <https://doi.org/10.1037/0278-7393.25.4.923>
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269–299. <https://doi.org/10.1006/jmps.2001.1388>
- Innocenti, I., Cappa, S. F., Feurra, M., Giovannelli, F., Santarnecchi, E., Bianco, G., Cincotta, M., & Rossi, S. (2013). TMS interference with primacy and recency mechanisms reveals bimodal episodic encoding in the human brain. *Journal of Cognitive Neuroscience*, *25*, 109–116. [https://doi.org/10.1162/jocn\\_a\\_00304](https://doi.org/10.1162/jocn_a_00304)
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319* (5865), 966–968. <https://doi.org/10.1126/science.1152408>
- Kloth, N., Lemke, J., Wiendl, H., Meuth, S. G., Duning, T., & Johnen, A. (2020). Serial position effects rapidly distinguish Alzheimer's from frontotemporal dementia. *Journal of Neurology*, *267*, 975–983. <https://doi.org/10.1007/s00415-019-09662-w>
- Kuhn, J. R., Lohnas, L. J., & Kahana, M. J. (2018). A spacing account of negative recency in final free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1180–1185. <https://doi.org/10.1037/e502412013-799>
- Lohnas, L. J., & Kahana, M. J. (2014). Compound cuing in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 12–24. <https://doi.org/10.1037/a0033698>
- Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 828–835. [https://doi.org/10.1016/s0022-5371\(69\)80050-2](https://doi.org/10.1016/s0022-5371(69)80050-2)
- McCabe, L., & Madigan, S. (1971). Negative effects of recency in recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 307–310. [https://doi.org/10.1016/S0022-5371\(71\)80059-2](https://doi.org/10.1016/S0022-5371(71)80059-2)
- McDermott, K. B. (2021). Practicing retrieval facilitates learning. *Annual Review of Psychology*, *72*, 609–633. <https://doi.org/10.1146/annurev-psych-010419-051019>
- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Memory*, *9*, 596–606. [https://doi.org/10.1016/s0022-5371\(70\)80107-4](https://doi.org/10.1016/s0022-5371(70)80107-4)
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64. <https://doi.org/10.20982/tqmp.04.2.p061>
- Parkin, A. J., & Russo, R. (1993). On the origin of functional differences in recollective experience. *Memory*, *1*, 231–237. <https://doi.org/10.1080/09658219308258235>
- Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, *17*, 133–140. <https://doi.org/10.1016/j.cobeha.2017.08.006>
- Reitman, W. (1970). What does it take to remember? In D. Norman (Ed.), *Models of human memory* (pp. 469–509). Academic Press.
- Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition*, *10*, 405–412. <https://doi.org/10.3758/bf03197642>
- Smith, S. M., & Handy, J. D. (2016). The crutch of context-dependency: Effects of contextual support and constancy on acquisition and retention. *Memory*, *24*, 1134–1141. <https://doi.org/10.1080/09658211.2015.1071852>
- Smolen, P., Zhang, Y., & Byrne, J. (2016). The right time to learn: Mechanisms and optimization of spaced learning. *Nature Reviews Neuroscience*, *17*, 77–88. <https://doi.org/10.1038/nrn.2015.18>

- Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*, 236–257. <https://doi.org/10.1037/a0022160>
- Talmi, D., Grady, C. L., Goshen-Gottstein, Y., & Moscovitch, M. (2005). Neuroimaging the serial position curve: A test of single-store versus dual-store models. *Psychological Science*, *16*, 716–723. <https://doi.org/10.1111/j.1467-9280.2005.01601.x>
- Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1589–1626. <https://doi.org/10.1037/0278-7393.26.6.1589>
- Tibon, R., & Levy, D. A. (2015). Striking a balance: analyzing unbalanced event-related potential data. *Frontiers in Psychology*, *6*, 555. <https://doi.org/10.3389/fpsyg.2015.00555>
- Tulving, E. (1983). *Elements of episodic memory*. Clarendon Press.
- West, B. T. (2009). Analyzing longitudinal data with the linear mixed model's procedure in SPSS. *Evaluation & the Health Professions*, *32*, 207–228. <https://doi.org/10.1177/0163278709338554>
- Xue, G., Mei, L., Chen, C., Lu, Z. L., Poldrack, R., & Dong, Q. (2011). Spaced learning enhances subsequent recognition memory by reducing neural repetition suppression. *Journal of Cognitive Neuroscience*, *23*, 1624–1633. <https://doi.org/10.1162/jocn.2010.21532>
- Zhao, X., Wang, C., Liu, Q., Xiao, X., Jiang, T., Chen, C., & Xue, G. (2015). Neural mechanisms of the spacing effect in episodic memory: A parallel EEG and fMRI study. *Cortex*, *69*, 76–92. <https://doi.org/10.1016/j.cortex.2015.04.002>

**Open practices statement** The data and materials for all experiments are available (<http://memory.psych.upenn.edu>), and this analysis was not preregistered.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.