Psychon Bull Rev (2015) 22:779–790
DOI 10.3758/s13423-014-0746-7

BRIEF REPORT

When items 'pop into mind': variability in temporal-context reinstatement in free-recall

Talya Sadeh · Rani Moran · Yonatan Goshen-Gottstein

Abstract It is well established that performance in free-recall is mediated by an individual's ability to reinstate the study-context during retrieval. This notion is supported by an abundance of evidence and is reflected in prominent models of memory. Introspectively, however, we often feel that a memory just 'pops into mind' and its recall is not accompanied by contextual detail. Here we ask whether this introspection is honored by the cognitive system. Namely, do items one recalls vary in the extent to which their contexts are reinstated? Previous research has provided evidence that indeed recall of some items relies on only little, if any, contextual reinstatement. This evidence pertains to one aspect of context: the concurrent, static encoding context of items, as tapped by the source-memory paradigm. However, because real-life events are strongly embedded in time, it is crucial to also investigate the dynamic, temporal aspects of context. To do so, we capitalized on one of the seminal findings linking recall with temporal-context: the temporal-contiguity effect, whereby the closer two items at study, the higher the probability that they will be retrieved one after the other during test. Using the Remember/Know paradigm, we show that in free-recall, 'Remember' retrievals, which are supposedly accompanied by contextual reinstatement, produce a larger temporal-contiguity effect as compared to 'Know' retrievals. Furthermore, 'Know' retrievals are more likely to be followed by retrieval errors (e.g., intrusions) than 'Remember' retrievals. These findings provide evidence that recalled items vary in the degree to which their temporal-context is reinstated.

Keywords Familiarity and recollection · Implicit/explicit memory · Long-term episodic memory · Memory

Introduction

Perhaps the most common test used to probe memory, both in everyday life and in laboratory settings, is free-recall. In free-recall tasks, participants study a list of items (e.g., words) and subsequently retrieve them with no external cue. Numerous empirical findings have demonstrated that recall performance is driven to a large extent by the degree to which an item's context at encoding can be reinstated at retrieval (Bjork & Whitten, 1974; Fisher & Craik, 1977; Glenberg, 1979; Greene, 1990; Howard & Kahana, 1999; Kahana, 1996; Morris, Bransford, & Franks, 1977; Nadel & Moscovitch, 1997; Tulving & Thomson, 1973). The context that is bound to an item at encoding may include any information peripheral to the item, including visual, spatial, semantic, gist, schematic, emotional, or temporal information.

One of the seminal findings pertaining to the importance of item-to-context binding in driving memory performance is the *temporal contiguity effect* (also known as the *lag-recency effect*). This effect describes the finding that the closer two items are presented during the study phase (i.e., the more contiguous they

Electronic supplementary material The online version of this article (doi:10.3758/s13423-014-0746-7) contains supplementary material, which is available to authorized users.

T. Sadeh · R. Moran · Y. Goshen-Gottstein (✉)
School of Psychological Sciences, Tel-Aviv University, Tel Aviv, Israel
e-mail: goshen@post.tau.ac.il

Present Address:
T. Sadeh (✉)
Rotman Research Institute, Baycrest Centre, 3560 Bathurst Street, Toronto, Ontario, Canada M6A 2E1
e-mail: tsadeh@research.baycrest.org

are, or the smaller the absolute temporal lag[1] between their serial positions), the higher the probability that these two items will be retrieved consecutively during the test phase (Howard & Kahana, 1999; Kahana, 1996; Sederberg, Howard, & Kahana, 2008). Systematic investigations of the temporal contiguity effect have converged on the notion that this effect is mediated by the fact that adjacent events are more likely than farther events to be bound to a similar context (Howard & Kahana, 1999). In addition, recalling an event reinstates its context, which then serves as a context-similarity based cue for the following recall. Importantly, alternate, non-contextual mechanisms for the temporal contiguity effect – such as the activation of associations between adjacent stimuli – could not account for findings of the temporal contiguity effect over a wide variety of conditions, over both small and large timescales and across different lists (e.g., Howard & Kahana, 1999; Howard, Youker, & Venkatadass, 2008; Unsworth, 2008). Thus, the evidence strongly supports the idea that the temporal contiguity effect is mediated by contextual reinstatement rather than by item-item associations.

The contiguity effect is a highly robust phenomenon, in that it is found across various paradigms (final free-recall, continuous-distractor recall, paired-associates recall, recognition; Davis, Geller, Rizzuto, & Kahana, 2008; Howard & Kahana, 1999; Howard et al., 2008; Schwartz, Howard, Jing, & Kahana, 2005). Furthermore, in a recent large-scale study, the existence of the contiguity effect was shown to be highly consistent, reliably observed in almost every participant (n = 126; Healey & Kahana, 2014). Indeed, the robustness of the effect highlights the important role of item-to-context binding in driving memory performance.

While much empirical effort has focused on temporal context influences on the dynamics of free recall, one important aspect of this influence has been relatively ignored: Can memories be recalled with little or no reliance on context? Intuitively, we often feel like a memory just ‘pops into our mind’ and its recall is accompanied by little, if any, contextual detail. If so, then variability may be found in the extent to which items are bound to their contexts at encoding and/or in the efficiency of context in driving retrieval.[2] This idea has yet to be incorporated into models of recall, let alone be demonstrated empirically (but see Brainerd & Reyna, 2010).

To illustrate, consider the Temporal Context Model (TCM; Sederberg et al., 2008), a highly influential model of free recall. According to this model, upon presentation of an item at study, its pre-experimental context is reinstated, and subsequently blended into the current encoding context. An outcome of this continuous contextual-evolution (whilst items are presented for study) is that the encoding contexts of different items overlap according to a temporally decreasing gradient – i.e., the extent of overlap decreases monotonically as a function of items' temporal lag. Additionally, each item is associated with the encoding context in which it is presented. At test, as items are recalled, their encoding contexts are retrieved and are blended into the test context. The test context in turn, activates other items in proportion to the overlap between their encoding contexts and the test context.

The upshot of such a contextual mechanism is that the probability of recalling two items successively is proportional to their temporal distance. Importantly, in its current form, TCM does not incorporate variability between items in the efficiency of either item-to-context binding at encoding or in subsequent reinstatement of an item’s context at retrieval, for a given individual. However, an item may be poorly (rather than strongly) bound to its encoding context for any of several reasons (e.g., temporal variability in the efficiency of the encoding process, which is caused either by random fluctuations in attention or by the fact that some items are easier to encode relative to others). In addition, the extent of contextual reinstatement at retrieval may also vary. Critically, such sources of variability should generate systematic differences in the degree of contextual reinstatement of individual items and, consequentially, also in patterns of activation that recalled items induce among peer items. Less efficient item-to-context binding at encoding or less efficient contextual retrieval would lead to a weaker contiguity effect with respect to recall transitions originating at that item.

Indirect evidence for variance in the contiguity effect, and hence in item-to-context binding, can be found in two studies in which individual differences were examined. In one study, the magnitude of the temporal-contiguity effect was found to be correlated with recall performance across-participants (Sederberg, Miller, Howard, & Kahana, 2010). This suggests that in addition to affecting the recall dynamics (i.e., the order of recalled item), the quality of the contextual processes (encoding and reinstatement) is beneficial to the quantity of recall. On similar lines, Golomb, Peelle, Addis, Kahana and Wingfield (2008) found a reduced temporal-contiguity effect for older, as compared to younger, adults. Both these results were interpreted as reflecting across-participant variability in item-to-context binding. While these demonstrations are informative, they address individual differences, differences that would require no change in our understanding of the retrieval

---

[1] At retrieval, lags between consecutively recalled items may be either positive or negative. When the recall sequence follows the same direction as the study sequence (e.g., an item from serial-position 5 is retrieved followed by an item from serial-position 6), the lags are positive (in this case, lag +1). When the sequence at test is reversed (e.g., an item from serial-position 11 is retrieved followed by an item from serial-position 8), the lags are negative (in this case, −3). A study-list of, for instance, 12 items, would have 22 possible lags: from −11 to 11 (excluding 0).

[2] The notion of variability in the efficiency of item-to-context binding in driving recall should not be confused with the ‘contextual-variability theory’ (Lohnas, Polyn and Kahana, 2011). The contextual-variability theory refers to the dynamic nature of context which evolves over the course of time. Versions of such a theory are the premise of the current investigation, which examines whether assuming an evolving temporal context, there is within-subject variability in the influence of such a context on the dynamics of free recall.

dynamics of recall in an individual's mind. The question still remains, therefore, whether similar differences can be found within individual participants, which would require making some adjustments to our understanding of how item-to-context binding mediates recall.

Some hints for a positive answer to this question may be found in the few studies which applied the Remember/Know (R/K) paradigm (Tulving, 1985) to free recall. These studies have shown that participants subjectively describe some of their memories as being devoid of, or comprising less, contextual reinstatement. In the first such study (Tulving, 1985), a non-trivial proportion of recalled words were found to be given K judgments, thus perhaps perceived by individuals as not being accompanied by contextual reinstatement. This finding has since been replicated (Arnold & Lindsay, 2002; McDermott, 2006; Read, 1996). Furthermore, two manipulations that are understood to affect contextual processes but not non-contextual processes, indexed by R and K judgments respectively, have been found to produce corresponding effects in free-recall (i.e., levels-of-processing, LOP; Hamilton & Rajaram, 2003; divided-attention; McCabe, Roediger, & Karpicke, 2011).

Most recently, Mickes, Seale-Carlisle and Wixted (2013) asked participants to make several judgments, including R/K and source-memory, for words retrieved in a free recall paradigm. Source-memory scores were found to be greater for R items than for K items. The authors concluded that R responses reflect retrieval of item and context information, while K responses reflect retrieval of item information alone. Importantly, contextual retrieval was indexed by asking participants to recount which of two judgments (size /animacy) they had performed while studying each individual word, a task which ostensibly should not be affected by the context of other words. Thus, only a *static* and *isolated* aspect of context was probed in this study, an aspect which was confined to the relation between individual items and their co-occurring contexts. However, in such assays, the overlaps between the encoding contexts of different items as well as the dynamic role that such overlaps plays in driving following recalls, is not addressed.

This dynamic aspect of context, which is the focus of the current study, may be best understood by focusing on its temporal properties, as described, for example, by the TCM. To reiterate, in this model, context is described as a dynamic entity which evolves with time during the study and test phases, producing contextual overlaps that are operative in affecting recall-dynamics. This conceptualization highlights the temporally vivacious, interconnected nature of context, which extends beyond the static and isolated nature of context that co-occurred with a studied event (as indexed by, for example, source memory).

It has yet to be empirically demonstrated that recalled items show differential degrees of reliance on these aspects of temporal context. In this study, therefore, we sought to provide evidence that the effects of item-to-context binding during free recall, as captured by the temporal contiguity effect, may occur to different degrees of efficiency. We used the R/K paradigm to examine whether the contiguity effect would distinguish between R-retrievals and K-retrievals, and asked whether R-retrievals would show a typical contiguity effect, whereas K-retrievals would show a much-reduced effect or no effect at all. As a complementary marker of contextual retrieval, we examined the proportion of errors following R and K responses to successful retrieval. We reasoned that if K-retrievals are accompanied by reinstatement of encoding context to a lesser extent than R-retrievals, then K-retrievals should be less likely to trigger retrieval of another word from the encoding episode. As such, K retrievals would more likely be followed by retrieval errors than R retrievals.

Method

Participants

Participants were 88 native Hebrew speakers (61 women) aged 19–31 years (mean 24.4), who were paid or given course credit in return for their participation.

Materials

Stimuli consisted of 348 Hebrew nouns, 3–6 letters long (mean word length = 4.1 letters). For each participant, 29 lists of 12 words each were sampled without replacement from this pool of words. Of the 29 lists, four served as practice lists (see Procedure below) and 25 as test lists.

Procedure

The experiment consisted of 25 study-test blocks with an arithmetic distractor task between each study and test phase. In the test blocks, R/K judgments were given for each of the retrieved words.

Prior to the experiment, participants were given standard R/K instructions (Gardiner & Java, 1990) that were slightly modified so as to apply to the free-recall test. An additional modification in the standard instructions was an emphasis that the R/K judgments should be an honest reflection of each participant's subjective feelings, and that it was perfectly reasonable to use only one of the responses (R or K) for most or even all the words they retrieved. This modification in the standard instructions was implemented so that participants would not feel obliged to classify some of the words as K, even if they were actually judged as R (and vice versa).

To focus on retrieval from long-term memory, we used a delayed free-recall paradigm (Postman & Phillips, 1965) in which recency effects – arguably the product of a non-contextual short-term store (Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005) – are largely reduced. Our analysis focused, therefore, on retrieval from all serial positions.

At encoding, each word was presented visually for 1350 ms followed by an asterisk for 400 ms. Participants were required to indicate whether the word was abstract or concrete. Their judgment could be made during the entire duration of the 1750 ms encoding trial.

Following each 12-word list, an arithmetic distractor task was given for 30 s. The arithmetic task comprised a presentation of a random 3-digit number for 2 s, followed by a fixation cross which appeared on the screen for the remaining 28 s. Participants were instructed to overtly count backwards in steps of three starting with the presented number until they heard a sound that signaled that the arithmetic task was over. Because the purpose of the distractor task was only to eliminate effects of short-term memory, performance in the task was not recorded for analysis. Still, to ensure that participants were performing the arithmetic task, and not memorizing the study items, they were told that their responses were recorded and that their performance in this task was scored (see below regarding monetary prizes for performance in the experiment).

At the offset of the arithmetic task, five question marks appeared on the screen for 90 seconds. The question marks signaled participants to start recalling as many words as possible from the last list presented, in any order, until a notice preparing them for the next list appeared on the screen. Participants were instructed to type the words they recalled using the computer keyboard, and to press the ENTER button after each word they typed. Following the press on the ENTER button, the words "Remember" and "Know" appeared at the center of the screen, signaling participants to make an R/K judgment regarding the last word they had typed. R/K judgments were made by pressing on one of two designated buttons. In total, 29 lists were presented. The first four served as practice lists and were disregarded from the analysis.

As an incentive to enhance performance, participants were told that the individuals with the three best results would be awarded monetary prizes (the first prize comparable to $200, the second to $40, and the third to $25). Participants were given detailed instructions regarding the scoring method. For the memory task, they were awarded ten points for correct responses and penalized two points for incorrect responses (participants were not penalized or rewarded for repeating a word that was already retrieved within the same list). For the arithmetic task, participants were told that they would be awarded ten points for each correct response and penalized ten points for each incorrect response (however, as mentioned above, only results of the free-recall test were taken into account in the final scoring). No points were differentially awarded for R as compared to K responses.

Results

Participants recalled a mean of 4.56 (38 %) words per list (SEM = 0.14), with a mean of 24 % (SEM =2.23 %) given a K response and the remaining 76 % given an R response. A one-sample t-test revealed that the rate of R responses significantly exceeded chance (50 %; $t_{87} = 11.5$; $p < 0.001$).

The proportion of K responses represents a mean of only one item per list per participant. This result is in line with the low proportion of K responses reported in previous free-recall studies (Arnold & Lindsay, 2002; Hamilton & Rajaram, 2003; McCabe et al., 2011; McDermott, 2006; Tulving, 1985). Indeed, to ensure sufficient data for our analyses, our experiment was designed to include many participants, each recalling a large number of lists, resulting in a total of approximately 2,200 K data-points (88 participants × 25 lists). Critically, the low number of K responses[3] reduces the reliability of performance measures for K, as compared to R. This, in turn, should make it harder to reject the null hypothesis when comparing the effects of R and K, thus compromising the statistical power of the design to reveal significant differences.[4] Therefore, to the extent that the small number of K responses is problematic, it is only problematic in inflating type II errors but not when a significant effect is found. Nevertheless, when relevant, we used a surrogate data technique to further guarantee that the different proportions of baseline R and K responses did not mediate the different patterns that we observed between these two responses (the rationale of this analysis is similar to that underlying bootstrap analyses; Stine, 1989; see Supplemental Materials).

We now turn to examine whether R-retrievals show a more robust contiguity effect compared with K-retrievals. Temporal contiguity effects were examined by calculating conditional-response probabilities (CRPs)[5] using scripts provided from http://memory.psych.upenn.edu/Software. CRPs provide the probability of making transitions at a certain lag conditional on this lag being available (Howard & Kahana, 1999). CRPs were calculated as the sum of all actual transitions a participant made across a certain lag divided by the sum of all possible transitions that could be made across that lag. This calculation was done separately for each potential lag.

---

[3] A few participants had too few K responses to calculate their temporal contiguity effects for K and were, therefore, excluded from the analysis, as is reflected by the reduced degrees of freedom.

[4] These analyses would use an estimate of variability which would include K variability, thereby producing a large estimated common error term.

[5] The probabilities are conditional on the event that a transition of a certain lag would yield a studied item that has not already been retrieved. CRPs are typically plotted on a graph where the X-axis represents the lag and the Y-axis represents the CRP of each lag. The lag-recency is demonstrated on such graphs by higher CRPs for the small lags, as compared to the larger lags.

Figure 1 presents the CRP curve collapsed across R and K transitions. The figure exhibits the typical contiguity effect, with larger CRPs for smaller lags. In addition, a typical asymmetry effect is demonstrated whereby the curve is steeper for positive than negative lags. These findings provide a replication to a previous examination of the delayed free-recall paradigm (Howard & Kahana, 1999).

Figure 2 presents the R-CRP and the K-CRP curves, which were calculated separately based on transitions originating from R or K correct recalls, respectively. Consistent with our prediction, examination of Fig. 2 revealed a reduced contiguity effect for K as compared to R responses.

Next, we examined whether the different pattern of the contiguity effects for R and K was statistically significant – namely, whether the R-CRP curve was steeper than the K-CRP curve. It has been shown that the CRP curve decreases monotonically for all lags excluding the last ones (Howard, Sederberg, & Kahana, 2009), with the CRPs of the last few lags being larger than those of earlier lags. This pattern was likewise observed in our data. Still, to be consistent with earlier studies, we only examined the monotonic portion of the CRP curve and focused on the first six lags (Howard & Kahana, 1999; Spillers & Unsworth, 2011).

To compare the slopes of the R and K curves, we followed previous research (Howard, Jing, Addis, & Kahana, 2007) by fitting a power function to the curves. This was done separately for the positive lags (1 to 6) and negative lags (−1 to −6) and for the R and K responses. The power function model is defined according to:

$$CRP(l) = A\left|l\right|^{b}, \tag{1}$$

Where l is the lag, A is the power coefficient and the steepness of the CRP curve is gauged by the exponent$b$: The more negative$b$, the steeper is the CRP curve. Taking logarithms of both the CRP values and the absolute lags in Equation 1 yields a linear model with intercept $\log(A)$ and slope b :

$$\log(CRP(l)) = logA + b{*}log\left|l\right|, \tag{2}$$

We thus calculated, for each participant, separately for R and K responses, the b coefficient using standard linear regression analysis. Paired t-tests were then conducted to compare the b coefficients of R and K. Participants for whom less than two CRP lags were non-positive were excluded from the analysis.

For positive lags, the mean of the b coefficients for R was almost seven times steeper than that of K (across-participant means for R = −0.62; K = −0.09). The difference between the slopes of the two conditions was significant (t_{66} =5.3; $p < .001$) .[6] In a follow-up analysis, the mean b coefficients of both R and K were compared to zero, and a significant effect was found for R (t_{85} =12.43; $p < .001$) but not for K (t_{68} = 1.35; p = .18). For negative lags, we also found significant differences between the b coefficients of R and of K (mean R = -0.31, mean K = 0.0004; (t_{60} = 3.18; p = .002). Like for positive lags, in a follow-up analysis the mean b coefficients of both R and K were compared to zero, and a significant effect was found for R (t_{86} = 7.37; $p < .001$) but not for K (t_{61} = 0.125; p = .9). In addition, paired t-tests revealed that the CRP effects for R were significantly stronger in the positive lags than in the negative lags (t_{84} = 4.78; $p < .001$).

An additional measure of the contiguity effect is the temporal-factor score (Polyn, Norman, & Kahana, 2009). This measure is particularly stable in cases of a limited number of observations (such as is the case with regard to K responses) and unlike the CRP curve, is insensitive to transitions into errors, that is, to recalls that are followed by errors (the relevance of this insensitivity to errors is described below). The temporal-factor score is a measure of the tendency of a participant to successively retrieve items in short lags—namely, from nearby study serial positions.[7] The larger the temporal-factor score, the higher the tendency of a participant to successively retrieve items with short lags. A score of 0.5 indicates no effect of temporal contiguity. The temporal-factor score for R items was 0.6 and was significantly greater than that of K items (0.55; t_{87} = 2.9; p = .005). For both R and K the mean score was significantly above 0.5 (for R: t_{87} = 13.2; $p < .001$; for K: t_{87} = 2.87; p = .005).

A potential confound of our results pertains to the effects of output position. We found that R-retrievals were output earlier (mean output position = 2.97) than K-retrievals (mean output position = 3.31; t_{86} = 3.85; $p < .001$). In addition, the contiguity effect tends to be larger for the earlier output positions (Howard & Kahana, 1999; although using a delayed free recall task mitigates this effect). Therefore, the difference we found in the slopes of R and K responses may have been driven by output position rather than by any other differences between R and K.

---

[6] A significant result was also obtained when fitting a linear function to the CRP (means R = −0.023; K = −0.0098; t_{80} = -2.74; p = .008).

[7] The temporal-factor score is calculated as follows. Each transition between two successively-retrieved items i and j with 'absolute transition lag' of |i-j| , is ranked in a decreasing order of absolute transition lag among all available transitions from item i (i.e., ‘available’ transitions refer to transitions into item j' that have not already been recalled). For example, if the available absolute transition lags are [1 2 2 4 6], then the rank of an actual transition to 6 would be r = 1, an actual transition to 4 would be r = 2, an actual transition to 2 would be r = 3.5 (because there are the two available transitions with absolute lag 2, correspond to lags 2 and −2), and an actual transition to 1 would be ranked r = 5. The lag is then scored according to (r−1)/(n−1), where n is the number of the available lags (n = 5, in the example above). The mean across all transitions is then computed for each participant.

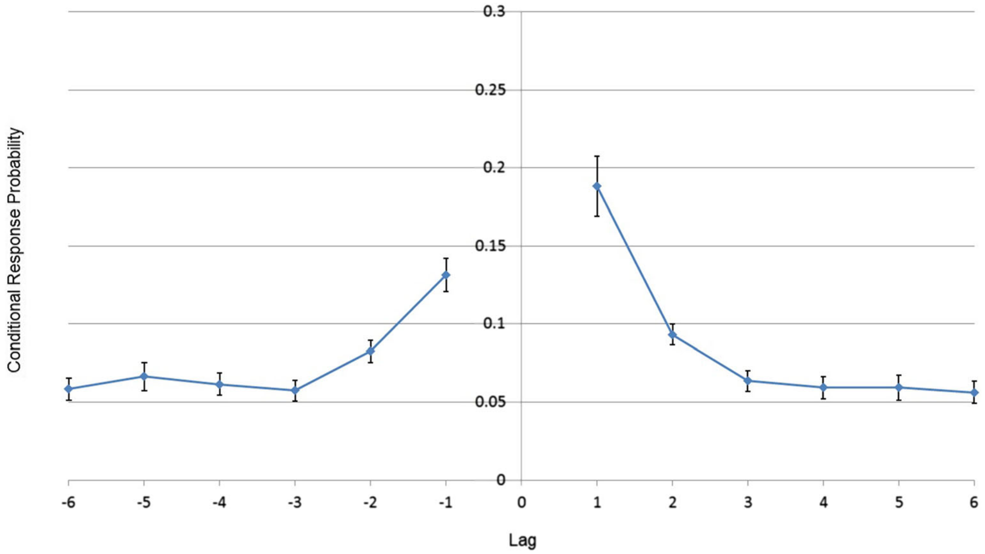

Fig. 1 The CRP curve for all retrievals (collapsed across R and K judgments). CRP measures the probability that a transition between two successively recalled items would be made across a certain lag. Lag refers to the distance, at study, between the serial-positions of two successively recalled words. Error bars reflect 95 % confidence intervals for within-subject designs (Loftus & Masson, 1994)

To rule out this potential confound, we refit the power functions to the CRP curves, this time separately for the early (1–3) and later (4–6) output positions. Unfortunately, when the data was split between the two output-position conditions, the power functions could not be fitted to many participants as their CRP curves did not include at least two non-zero coefficients. This was especially pronounced for the K responses. For example, for later outputs, we estimated the R exponent for 73 participants, the K exponent for 29 participants, and both R and K exponents for 25 participants (at least one of the R or K exponents were estimated for 77 participants). Thus, if we conducted paired t-tests, to compare R and K exponents, which were based solely on participants for whom both exponents were estimated, we would have excluded a large portion of the data. To include all participants (for which at least one of the R or K exponents was estimated) in our analysis, we conducted permutation tests (Fisher, 1935) separately for early and later outputs, rather than paired t-tests, as follows.[8]

A single 'permutation step' was conducted separately for the 'paired participants' i.e., participants who had paired R-K data, and for the 'non-paired participants' i.e., participants who had a single, either R or K, observation. First, for the paired participants, we shuffled randomly the R and K labels of the power exponents for each of the participants. This means that with probability 0.5 the exponents for a given participant either 'switch identity', so that the R and K exponent are considered as K and R exponents respectively, or maintain their 'true' identities. Notably, this reshuffling scheme maintains the paired structure of the data invariant. Second, for the non-paired participants, we shuffled randomly the R and K labels of the power exponents across participants. Notably, this reshuffling scheme maintains the number of non-paired K and R observations invariant. Following these shuffles for the paired and non-paired participants, we calculated the difference between the mean R exponent and the mean K exponent. These means were computed based on all available (shuffled) observations. We thus obtained a single 'permutation R-K contrast'. By repeating the permutation step 100,000 times, we obtained 100,000 permutation R-K contrasts. Critically, shuffling the data in this manner obliterates any systematic differences between the R and K exponents, while maintaining the paired-data structure whenever possible and controlling for the larger number of missing data points for the K relative to R labels. Consequently, the distribution of the permutation contrasts corresponds to a contrast distribution under the null hypothesis that the means of the R and K power exponents are equal. Having generated this null distribution, we computed the empirical contrast as the difference between the mean R and mean K exponents from the non-shuffled empirical data. The p-value for rejecting the null hypothesis was estimated as the proportion of permutations (out of 100,000) which yielded an effect which was larger or equal to the empirical contrast.

The same, 'R steeper than K', pattern was found for lags of both output positions groups. For outputs 1–3, the across-participant means of the positive lags were -0.590 for R and -0.028 for K $(\widehat{p} < .00001)$. The across-participant means of the negative lags were -0.257 for R and 0.046 for K $(\widehat{p} = .039)$. For outputs 4–6, the across-participant means

---

[8] An analysis that was based on non-paired, independent t-tests, which included all participants with at least one exponent, yielded the same conclusions.

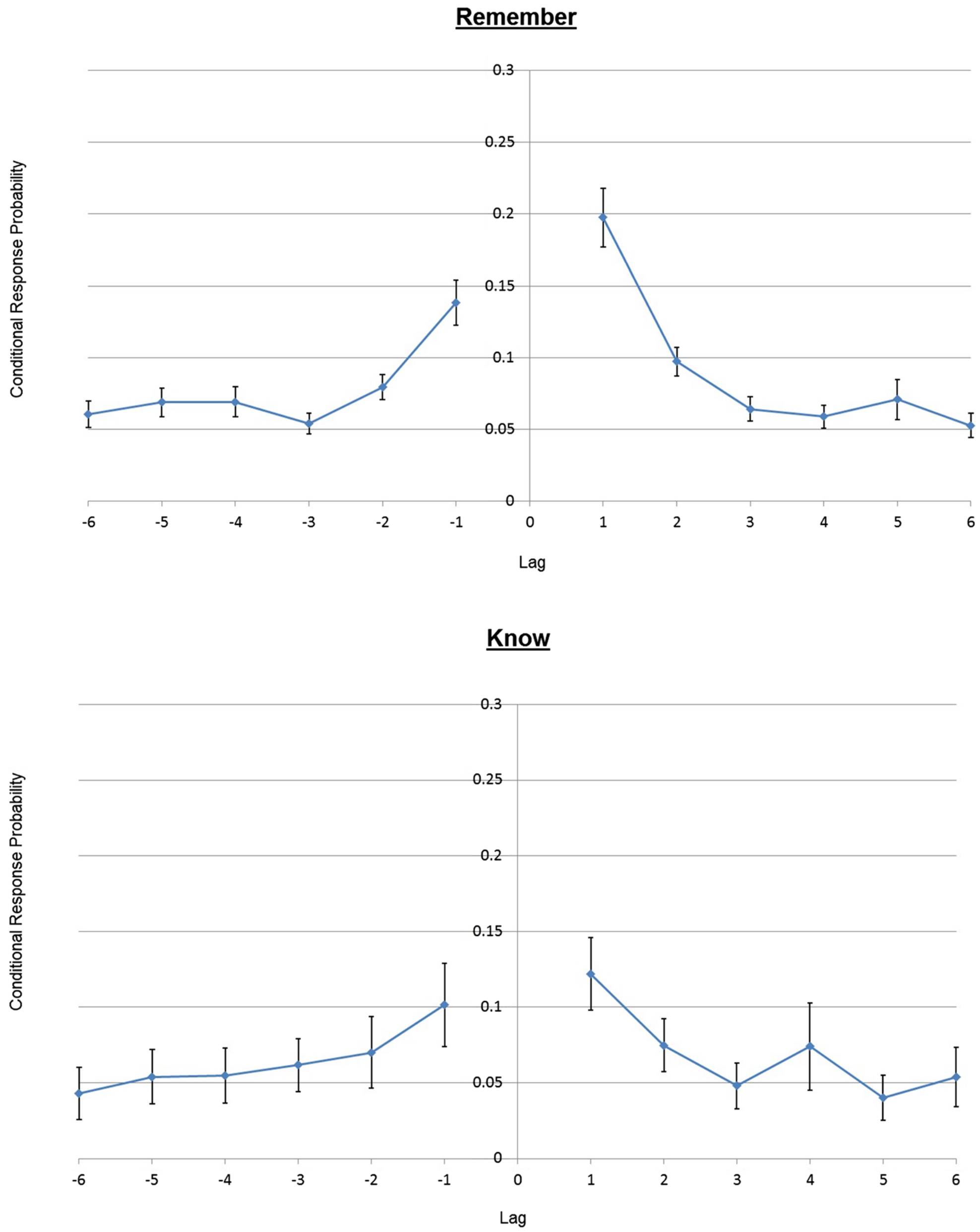

Fig. 2 CRP curves for R and K judgments. For each possible lag, we measured the probability of that lag following recall of words given R judgments, and likewise for K judgments. Error bars reflect 95 % confidence intervals for within-subject designs (Loftus & Masson, 1994)

of positive lags were −0.190 for R and 0.194 for K $(\widehat{p} < .001)$. The across-participant means of the negative lags in the later output positions were −0.092 for R and 0.005 for K. The difference, however, did not reach significance $(\widehat{p} = .240)$. The fact that the 'R steeper than K' pattern of findings was replicated when breaking down the data to early and late output positions speaks against the possibility that the differences between the contiguity effects of R and K were driven merely by differences in output positions.

For proponents of dual-store models of memory (e.g., Davelaar et al., 2005; Usher, Davelaar, Haarmann, & Goshen-Gottstein, 2008), an additional concern may be raised. The steeper CRP curve for R than for K may be, to some extent, the result of a higher residual reliance on a short-term buffer for R than for K (though the delay between the study and the test phases should presumably empty this buffer). The short term buffer consists of recency items that are usually output consecutively during the initiation of recall, and as such increase the probability of transitions in small lags (Moran & Goshen-Gottstein, 2014). Thus, if such recalls tend to elicit R (rather than K) judgments, then a steeper R-CRP curve is expected. This concern is highlighted by the fact that R-retrievals were output earlier than K-retrievals, and indeed buffer items tend to be output early. An increased R reliance on a short-term buffer would manifest in an enhanced recency effect for R (compared with K). We thus examined whether the serial position curves of R- and K-retrievals, and particularly the recency portion of these curves, differed.

Figure 3 depicts the serial-position curves of R and K, and Fig. 4 depicts the probability of first recall of R and K. As illustrated in these figures, no significant differences were found between the serial-position effects of R and K, as reflected by the overlap of the confidence intervals of the two conditions. Furthermore, as the previous CRP analysis showed, the difference between CRP steepness for R and K emerged even when only the late output positions were considered. It is unlikely that such late outputs are retrieved from a residual short term buffer.

An additional concern is that our findings regarding differential CRP slopes for R and K may be mediated by the elevated rates of error that followed K-retrievals (this finding is reported in the concluding part of the Results). In particular, the higher proportion of errors following K responses may have led to the reduced-steepness of their CRPs. This is because in calculating the conditional response probabilities – the ratio of actual transitions to possible transitions – errors do not contribute to the numerator of the CRP measure (that is, transitions to errors are not considered to be legitimate 'actual transitions'). At the same time, because legitimate transitions instead of an error were possible, the denominator is increased. Thus, when more errors are made, lower coefficients are necessarily obtained. This may translate to a lower steepness of the CRP curve. While such an 'error effect' may also result from the putative higher temporal-context involvement in R judgments (indeed, this is our argument below), it is nevertheless informative to probe whether it accounts for the entire contiguity effect.

We thus repeated our contiguity analyses– namely, that comparing the CRP curves of R and K – this time excluding transitions to errors from the data. Results of this analysis revealed the same pattern as the results when data included errors. We found that the mean b coefficients for positive CRPs for R (= -0.62) was significantly more negative than that of K (= −0.097; $t_{66} = 5.25$; $p < .001$). The differences between the negative lags of R (= −0.3) and K (=0.002) were likewise significant ($t_{60} = 3.17$; $p = .002$). This analysis, as well as the analysis of the temporal-factor scores (recall that this score is insensitive to errors), alleviates the concern that the results regarding the different effects of contiguity for R and K responses were a mere byproduct of the different proportions of errors following R and K judgments.

We now turn to our second prediction regarding errors following R and K retrievals. To reiterate, we reasoned that if correct K-retrievals rely less heavily on context, then such retrievals – as compared to R retrievals – would more likely to be followed by retrieval errors (i.e., intrusions from other lists or words that did not appear at all during the experiment). Results supported this reasoning, with 10.95 % of R-retrievals followed by errors and 27 % more errors, for a total of 13.96 %, of K-retrievals followed by errors ($t_{86} = 2.2$; $p = .03$). Here too, to examine whether this effect was entirely driven by differences in output positions between R and K (indeed, errors tend to be output later in the recall sequence; Zaromb et al., 2006), we reanalyzed the data separately for the early (1–3) and later (4–6) output positions. For outputs 1–3, the proportion of R-retrievals and K-retrievals followed by errors was virtually identical (means = 7.5 %; $t_{83} = 0.01$, $p = .99$). For outputs 4–6, the proportion of errors following K-retrievals (mean = 21 %) was, however, significantly greater than that following R-retrievals (mean = 14 %; $t_{82} = 2.76$; $p = .007$). In addition, a repeated-measures ANOVA with output (1–3/4–6) and condition (R/K) as within-subject factors revealed a significant interaction between output and condition ($F_{1,82} = 4.69$, MSE = 2.2; $p = .03$). Thus, the difference between the proportions of errors following R- and K-retrievals was evident only for later output positions. We return to this result in the Discussion.

Discussion

In our study, participants recalled lists of words and, following the recall of each word, made Remember-Know judgments. We found systematic differences between retrieval which is judged by participants to reflect remembering of study items along with associated details regarding the study episode – i.e., R judgments – as compared to retrieval which is not – i.e., K judgments. These differences pertain to reliance on temporal-context in driving retrieval, as captured by the temporal-contiguity effect. Our findings provide the first empirical demonstration that the items one recalls vary in the extent to which their temporal-contexts are reinstated.

There has been growing interest in the role of automatic, or non-contextual, processes in free recall, and particularly in elucidating the processes the R/K paradigm captures in this context (e.g., Hamilton & Rajaram, 2003; McCabe et al., 2011; Mickes et al., 2013). This paradigm is extremely popular in the study of recognition, typically in the guise of dual-process theories (Yonelinas, 2002; for a recent review of the empirical literature, see Sadeh, Ozubko, Winocur, & Moscovitch, 2014). In contrast, our findings are not dependent on a specific interpretation the R/K paradigm. Further, we do not claim to support a qualitative distinction between two processes underlying free recall (e.g., a contextual and non-contextual process), though such a claim has been made by others (Brainerd & Reyna, 2010). Rather, the importance of our study is in demonstrating that the R/K paradigm taps onto variation in reliance on temporal context by providing evidence for quantitative (not qualitative) differences between R- and K-retrievals.

The variance in reliance on context may reflect two distinct underlying retrieval processes, perhaps similar to

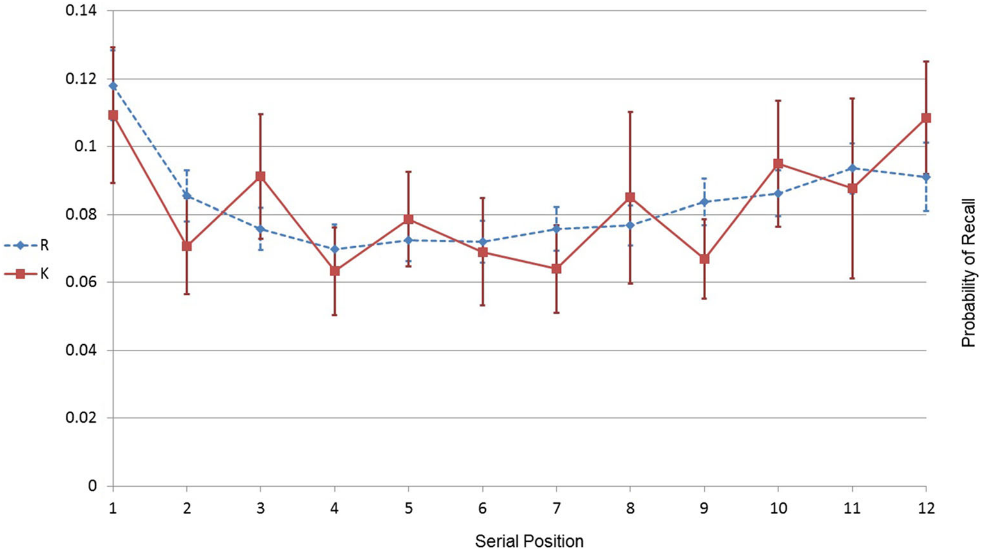

Fig. 3 Serial position curves for R and K judgments. To examine possible differences in serial-position effects for R and K, we calculated the probabilities of recall for each serial position (Howard & Kahana, 1999). Because across all serial positions the probability of retrieval for K is significantly lower than that of R, we standardized the probabilities of recall so that R and K serial-position effects would be comparable. This was achieved by dividing – for each participant and for R and K responses separately – the probability of recall for each serial-position by the total probability of recall (across all serial-positions). Critically, the comparison of the standardized scores uncovered no significant differences between the serial-position effects of R and K. Error bars reflect 95 % confidence intervals for within-subject designs (Loftus & Masson, 1994)

those proposed by dual-process models of recognition memory (Brainerd, Gomes, & Moran, 2014; Yonelinas, 2002). Alternatively, participants' performance may have been mediated by a unidimensional, composite mnemonic-evidence strength continuum, with R and K responses mapping onto different levels of mnemonic evidence (e.g., Dunn, 2004). Indeed, it is likely that similar results to those obtained in the current study would be obtained if confidence ratings were used (instead of R/K judgments). Such a pattern has been previously observed in recognition (Schwartz et al., 2005), where a temporal-contiguity effect was found only for items given high-confidence 'old' responses, and not for items given lower-confidence 'old' responses.

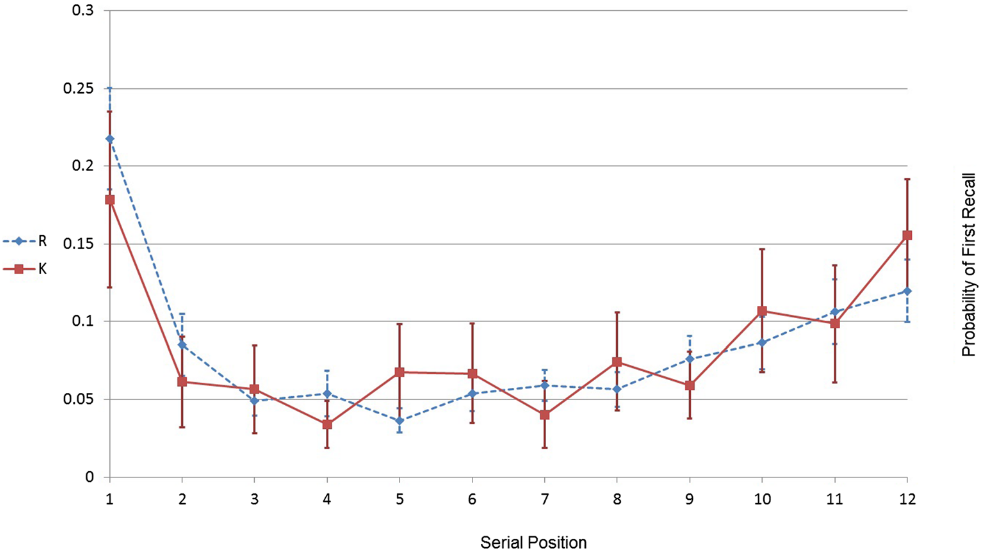

Fig. 4 Curves of probability of first recall for R and K judgments. To examine possible differences in serial-position effects for R and K, we also calculated the probabilities of first recall for each serial position (Howard & Kahana, 1999). Because across all serial positions the probability of retrieval for K is significantly lower than that of R, we standardized the probabilities of first recall so that R and K serial-position effects would be comparable. This was achieved by dividing – for each participant and for R and K responses separately – the probability of first recall for each serial-position by the total probability of recall (across all serial-positions). Critically, the comparison of the standardized scores uncovered no significant differences between the serial-position effects of R and K. Error bars reflect 95 % confidence intervals for within-subject designs (Loftus & Masson, 1994)

Note that whereas in the recognition literature there have been different suggestions regarding the relationship between R and K responses (e.g., exclusivity), this question is of no relevance for the present endeavor, as our analysis does not directly compare the proportion of R and K responses. Rather, we compare the tendency to recall a neighboring item given that the preceding item was given an R response (and likewise for K). In other words, the calculation of the temporal contiguity score for R and for K are independent of each other.

Our CRP analyses revealed a similar pattern for negative lags, namely, for successive retrievals in backward direction, as found for the positive lags. However, the CRP effects for R were more pronounced for the positive than for the negative lags. These results are consistent with the known asymmetry in contiguity effects of positive and negative lags, whereby negative lags generally show a reduced CRP effect (Howard & Kahana, 1999). The more moderate effects for negative lags are testimony to the working of temporal context in driving the contiguity effect. Temporal context continuously evolves over time, going forward in direction, as time itself (Polyn & Kahana, 2008). This results in a more enhanced tendency to recall items in a forward direction (reflected in positive lags) than in a backward direction (reflected in negative lags). If the contiguity effect were merely driven by item-to-item associations, such asymmetry would not be expected (Kahana, 2002).[9]

Successful reinstatement of the context (presumably tapped by correct R-retrievals) is more likely to trigger successful retrieval of an additional item, as compared to cases in which the context is less successfully reinstated (presumably tapped by K-retrievals). This phenomenon appears to occur only relatively late in the course of retrieval, as suggested by the demonstration of differences between the error rates following R- and K-retrievals for later (4–6), but not earlier (1–3) output positions, as well as by the significant interaction between output and condition. Possibly, the contextual retrieval process is sufficiently fluent earlier on in the course of retrieval so as to not be interrupted by events for which the context is not fully reinstated. However, later on in the course of retrieval, as recall becomes harder and less fluent, initiation of the next recall in a sequence is more profoundly affected by whether or not the context of the current item was successfully reinstated.

A caveat to our conclusions regards the content on which the metacognitive judgments of R and K are based. It is possible that participants base their judgments on whether or not they recall neighboring items. If this is the case, then the contiguity effect (namely, the tendency to recall neighboring items) and R responses are, in fact, two measures of the same phenomenon. We addressed this concern in two ways, one based on data from this study and another based on additional data from our lab. First, we analyzed responses given in a post-experiment questionnaire (cf., “the retrieval intentionality criterion”; Schacter, Bowers, & Booker, 1989) in which we asked participants to describe the content on which their R and K judgments were based. Of the 88 participants, ten explicitly mentioned that, at times, their R responses were based on whether they could recall neighboring items. We, therefore, re-analyzed the data excluding these participants from the analysis. The results after this exclusion retained their original pattern and significance (for positive lags: mean slope for R = -0.57, K = −0.11; t_{58} = 4.37; $p < .001$; for negative lags: R = −0.29, K = 0.03; t_{54} = 3.4; $p = .001$). Thus, it seems unlikely that R responses are merely a measure of the tendency of recalling neighboring items. This conclusion was further supported by additional from our lab (Sadeh, Moran & Goshen-Gottstein, in preparation), where a different measure of the contiguity effect was used, which cannot be associated with whether or not participants explicitly remember recalling neighboring items from the same list. There, participants retrieved in a surprise final-free-recall test (e.g., Craik, Gardiner, & Watkins, 1970) items from all the studied lists, in any order, while making R/K judgments (cf., Howard et al., 2008; Unsworth, 2008). Critically, we found an enhanced contiguity effect across different lists (i.e., when the lag refers to the difference between the serial numbers of the *lists* in which two consecutively recalled items appeared) for R responses relative to that found for K responses. While our original concern was that R judgments were based on whether or not neighboring items from the *same* list could be retrieved, this result indicated that R-judgments were correlated with the tendency to recall items from neighboring, *different* lists. Therefore, R-judgments did not measure the same phenomenon as the contiguity effect. Furthermore, this result has importance beyond alleviating the above concern. That the contiguity advantage for R responses persists across lists that were presented several minutes apart from each other is testimony to the veracity of the contextual mechanism – rather than an association-based mechanism (Kahana, 1996) – that is at work in producing our differential contiguity effects.

Interestingly, we found no significant differences between the serial-position curves, and specifically, the primacy and recency effects of R and K (see Figs. 3 and 4). Primacy and recency effects are interpreted by many in contextual terms – i.e., frequent access to beginning-of-list and end-of-list contexts (Howard et al., 2009). Therefore, one would expect these effects to differ between R and K. Nevertheless, primacy and

---

[9] The asymmetry in the temporal contiguity effect is captured by the TCM. According to this model, when an item is recalled two contextual components are retrieved: the encoding context of the item and its pre-experimental context. Importantly, these two components vary in their tendency to cue items in the forward and backward directions. The former, encoding context is 'bi-directional' (or 'direction blind') in that it is equally similar to the encoding context of other studied items in the forward and backward directions controlling for lag. The latter, pre-experimental contextual component, however, is directional in that it selectively overlaps with the encoding context of items in the forward direction. In total, a retrieved context more effectively cues items in the forward than in the backward direction, and an enhanced tendency to recall such items relative to backward items ensues.

recency effects have also been interpreted in non-contextual terms – enhanced rehearsal as an account for primacy effects (Rundus, 1971; Tan & Ward, 2000), and a short-term memory buffer (Davelaar et al., 2005; Farrell, 2014) or enhanced distinctiveness (Neath, 1993) as accounts for recency effects or residual recency effects, as are often observed in delayed free recall tasks such as ours. However, being a null-finding, this should not be over-interpreted. Primacy and recency may still be linked to contextual effects by using different measures of contextual processing and/or by applying different memory tasks (e.g., recognition, cued-recall) than those applied in the current study.

To conclude, the current study provides novel evidence for within-participant variance in reliance on temporal context during recall. We demonstrate that R and K items differ with regard to their ability to trigger retrieval of additional items with whom they share episodic, contextual, history. Our results highlight that the processes of item-to-context binding is not a uniform, all-or-none one. Rather, temporal-context is encoded and/or retrieved to different degrees for different items. In this sense, the introspection that an item 'pops into mind' with little or no contextual background seems to be honored by the human cognitive system.

Acknowledgments Portions of this article were part of a doctoral dissertation submitted at Tel-Aviv University by TS under the supervision of YGG. We thank Lilach Leibovitch, Alisa Lurie, and Rafael Amram for their help in data collection.

References

Arnold, M. M., & Lindsay, D. S. (2002). Remembering remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 521–529.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6*(2), 173–189.

Brainerd, C. J., Gomes, C. F. A., & Moran, R. (2014). The Two Recollections. *Psychological Review*, in press.

Brainerd, C., & Reyna, V. (2010). Recollective and nonrecollective recall. *Journal of Memory and Language, 63*(3), 425–445.

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*(5), 554–560.

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*(1), 3–42.

Davis, O., Geller, A., Rizzuto, D., & Kahana, M. (2008). Temporal associative processes revealed by intrusions in paired-associate recall. *Psychonomic Bulletin & Review, 15*(1), 64–69.

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*(2), 524–542.

Farrell, S. (2014). Correcting the correction of conditional recency slopes. *Psychonomic Bulletin & Review, 1–6*.

Fisher, R. A. (1935). *The design of experiments*. New York: Hafner.

Fisher, R. P., & Craik, F. I. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory, 3*(6), 701.

Gardiner, J., & Java, R. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*(1), 23–30.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*(2), 95–112.

Golomb, J. D., Peelle, J. E., Addis, K. M., Kahana, M. J., & Wingfield, A. (2008). Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. *Memory & Cognition, 36*(5), 947–956.

Greene, R. L. (1990). Memory for pair frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 110–116.

Hamilton, M., & Rajaram, S. (2003). States of awareness across multiple memory tasks: Obtaining a "pure" measure of conscious recollection. *Acta Psychologica, 112*(1), 43–69.

Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies?. *Journal of Experimental Psychology: General, 143*(2),575.

Howard, M. W., Jing, B., Addis, K., & Kahana, M. J. (2007). Semantic structure and episodic memory. *Handbook of latent semantic analysis*, 121–142.

Howard, M., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 923–941.

Howard, M., Sederberg, P., & Kahana, M. (2009). Reply to Farrell and Lewandowsky: Recency—contiguity interactions predicted by the temporal context model. *Psychonomic Bulletin & Review, 16*(5), 973–984.

Howard, M., Youker, T., & Venkatadass, V. (2008). The persistence of memory: Contiguity effects across hundreds of seconds. *Psychonomic Bulletin & Review, 15*(1), 58–63.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24*(1), 103–109.

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition, 30*(6), 823–840.

Loftus, G., & Masson, M. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*(4), 476–490.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of Memory and Language, 64*(3), 249–255.

McCabe, D., Roediger, H., & Karpicke, J. (2011). Automatic processing influences free recall: Converging evidence from the process dissociation procedure and remember-know judgments. *Memory & Cognition, 39*(3), 389–402.

McDermott, K. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition, 34*(2), 261–267.

Mickes, L., Seale-Carlisle, T. M., & Wixted, J. T. (2013). Rethinking familiarity: Remember/Know judgments in free recall. *Journal of Memory and Language, 68*(4), 333–349.

Moran, R., & Goshen-Gottstein, Y. (2014). The conditional-recency dissociation is confounded with nominal recency: Should unitary models of memory still be devaluated? *Psychonomic Bulletin & Review, 21*(12),332-43.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*(5), 519–533.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology, 7*(2), 217–227.

Neath, I. (1993). Contextual and distinctive processes and the serial position function. *Journal of Memory and Language, 32*(6), 820–840.

Polyn, S. M., & Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences, 12*(1), 24.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*(1), 129.

Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *The Quarterly Journal of Experimental Psychology, 17*(2), 132–138.

Read, J. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review, 3*(1), 105–111.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*(1), 63–77.

Sadeh, T., Moran, A., & Goshen-Gottstein Y. Differential long-term contextual effects for Remember and Know judgments in free recall (in preparation)

Sadeh, T., Ozubko, J. D., Winocur, G., & Moscovitch, M. (2014). How we forget may depend on how we remember. *Trends in Cognitive Sciences, 18*(1), 26–36.

Schacter, D. L., Bowers, J., Booker, J. (1989) Intention, awareness, and implicit memory: The retrieval-intentionality criterion. In S. Lewandowsky, J. C. Dunn, K. Kirsner (Eds.), *Implicit memory: theoretical issues* (pp. 47–65). Hillsdale, NJ: Erlbaum.

Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past. *Psychological Science, 16*(11), 898–904. doi:10.1111/j.1467-9280.2005.01634.x

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review, 115*(4), 893.

Sederberg, P. B., Miller, J. F., Howard, M. W., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory & Cognition, 38*(6), 689–699.

Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal-contextual retrieval from episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6),1532.

Stine, R. (1989). An introduction to bootstrap methods examples and ideas. *Sociological Methods & Research, 18*(2–3), 243–291.

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*(6), 1589–1625.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne, 26*(1), 1–12.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373.

Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory and Language, 59*(2), 223–236.

Usher, M., Davelaar, E. J., Haarmann, H. J., & Goshen-Gottstein, Y. (2008). Short-term memory after all: Comment on Sederberg, Howard, and Kahana (2008). *Psychological Review, 115*(4), 1108–1118.

Yonelinas, A. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441–517.

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., & Kahana, M. J. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 792.