

Does Depth of Processing Affect Temporal Contiguity?

Abigail M. D. Mundorf, Mitchell G. Uitvlugt, and M. Karl Healey

Michigan State University

Author Note

Correspondence concerning this article should be addressed to Abigail Mundorf (desterab@msu.edu) or M. Karl Healey (khealey@msu.edu) at Michigan State University, Department of Psychology, 316 Physics Road, East Lansing, MI. We thank G. Carter Brown and Kaitlin Ifkovits for help in developing materials and Linh Lazarus for helpful discussions. This material is based upon work supported by the National Science Foundation under Grant No. 1848972.

Abstract

Memory tends to be better when items are processed for their meaning (deep processing) rather than their perceptual features (shallow processing). This levels of processing (LOP) effect is well-replicated and has been applied in many settings, but the mechanisms involved are still not well understood. The temporal contiguity effect (TCE), the finding that recalling one event often triggers recall of another event experienced nearby in time, also predicts memory performance. This effect has given rise to several competing theories with specific contiguity-generating mechanisms related to how items are processed. Therefore, studying how LOP and the TCE interact may shed light on the mechanisms underlying both effects. However, it is unknown how LOP and the TCE interact—various theories make differing predictions. In this preregistered study, we tested predictions of three theoretical explanations: accounts which assume temporal information is automatically encoded, accounts based on a trade-off between item and order information, and accounts which emphasize the importance of strategic control processes. Participants completed an immediate free recall task where they either engaged in deep processing, shallow processing, or no additional task while studying each word. Recall and the TCE were highest for no-task lists and greater for deep than shallow processing. Our results support theories which assume temporal associations are automatically encoded and those which emphasize strategic control processes. Both perspectives should be considered in theory development. These findings also suggest temporal information may contribute to better recall under deeper processing with implications for determining which situations benefit from deep processing.

Does Depth of Processing Affect Temporal Contiguity?

Memory tends to be better for items processed according to meaning (deep processing) rather than perceptual features (shallow processing). This levels of processing (LOP) effect has been consistently observed in both recall and recognition regardless of encoding intentionality or specific deep processing task (Craik & Tulving, 1975; Eysenck, 1979; Hyde & Jenkins, 1969; Moscovitch & Craik, 1976; but see Rose & Craik, 2012). Extensive work has investigated interactions between deep processing and other aspects of memory, such as primacy and recency (Mazuryk & Lockhart, 1974) and semantic organization (Einstein & Hunt, 1980; Hyde & Jenkins, 1969). The benefits of deep processing have inspired recommendations for teaching methods, study strategies, and textbook design (Ayçiçegi-Dinn & Caldwell-Harris, 2009; Biggs, 1978; Martin et al., 1985; Seiver et al., 2019). Yet, the mechanisms through which deep processing influences memory are still not well understood (Baddeley, 1978; Craik, 2002; Eysenck, 1979).

Another widely-studied phenomenon, the temporal contiguity effect (TCE), *has* been linked to specific mechanisms but has *not* received much attention in the LOP literature. The TCE is the finding that recalling one event often triggers recall of another event experienced nearby in time (Kahana, 1996). Although the size of the effect is modulated by various factors, a TCE has been consistently observed regardless of task instructions or stimuli characteristics (Healey et al., 2019; Healey & Uitvlugt, 2019; Mundorf et al., 2021; Sadeh et al., 2015; but see Osth & Fox, 2019). The TCE also predicts memory performance (Healey et al., 2019; Sederberg et al., 2010; Spillers & Unsworth, 2011), at least for intentional encoding of unrelated words (Healey & Uitvlugt, 2019; Mundorf et al., 2021). These findings have given rise to many models of episodic memory with TCE-generating mechanisms (e.g., Davelaar et al., 2005; Farrell, 2012; Howard et al., 2015; Lehman & Malmberg, 2013).

Both LOP and the TCE have strongly influenced memory theory development, and both point to practical ways of improving memory. Yet, little work has examined how these effects interact. Theories which make the same predictions for summary measures, like overall

recall, often make divergent predictions for the TCE, making temporal contiguity a useful tool for theory testing. Considering these two effects together allows us to develop a more unified theory of memory that can explain not only each effect independently but also how they interact. Below, we outline theoretically motivated hypotheses of how LOP might influence the TCE.

Reasons to Predict Deep Processing May Increase the TCE

Deeper LOP and a larger TCE are both associated with better recall. Thus, on purely empirical grounds, a reasonable hypothesis is that deeper processing should be associated with increased temporal contiguity.

Retrieved context models (Howard & Kahana, 2002; Lohnas et al., 2015) provide a theoretical motivation for this hypothesis. These models assume memories form when items become associated with the current state of a mental context representation which drifts through a high-dimensional representational space. When an item is studied, it activates its pre-existing representation (which contains the item's pre-existing associations), the activation of previous items' representations fade, and context drifts towards this just-studied item's representation. In this way, items studied nearby in a list become associated with similar states of context. When an item is recalled, it reinstates its associated context from encoding, providing a cue for items originally studied nearby in time. This naturally produces a TCE. The size of the TCE depends on how far context drifts with each event. If items weakly activate their pre-existing contextual representation, context will drift very little; all items will form associations with a similar state of context, and the TCE will be small. If each item strongly activates its pre-existing context, mental context will drift farther toward the just-studied item's representation. Only items studied close in time will share similar contexts, enhancing the TCE. In this light, deep processing should cause context to drift farther than shallow processing because a deep processing task involves not only activating items' perceptual features (as shallow processing does) but also deeper semantic features (as suggested by

Healey & Kahana, 2016).

However, there is another possible interpretation of how LOP influence contextual dynamics. These models make a distinction between item and context representations. If deep processing acts primarily on item representations and not context, deeper processing would not increase the TCE. Examining the TCE under deep processing will help adjudicate between these competing interpretations of retrieved context models.

Reasons to Predict Deep Processing May Decrease the TCE

Other perspectives suggest deep processing should reduce the TCE. Under the item-order framework, a prominent explanation for memory phenomena like the enactment and generation effects (Engelkamp & Zimmer, 1997; Hirshman & Bjork, 1988; Nairne et al., 1991), recall depends on processing information about individual items *and* inter-item associations like temporal order. But there is a trade-off: Any manipulation that encourages item-specific processing should improve memory for specific items at the expense of memory for order. Thus, the TCE should be reduced (Lazarus et al., in prep; McDaniel & Bugg, 2008). For example, McDaniel et al. (2011) found a smaller TCE for lists of orthographically distinct items (e.g., *khaki*, *lynx*) compared to common items (e.g., *cookie*, *ruler*) and suggested the reduction was due to distinct words requiring more item-specific processing. Similarly, deeper processing may draw more attention to item-specific information (McDaniel & Bugg, 2008). The item-order account predicts deep processing should lead to better memory for items but reduced memory for order.

Finally, LOP may change participants' encoding strategies. Absent any experimenter-imposed encoding task, participants often adopt effective order-based strategies, such as linking items together to form a story (Delaney & Knowles, 2005; Hintzman, 2016; Unsworth, 2016). By encouraging serial recall, such strategies may contribute to the TCE (Bouffard et al., 2018; Unsworth et al., 2019). For participants using order-based strategies, *any* experimenter-imposed processing task that encourages focusing on individual

items should interfere with such strategies, reducing recall and the TCE. That is, even if not all participants use order-based strategies, the average TCE should be highest with no encoding task. One study found deep processing reduced recall and the TCE relative to no-task (Long & Kahana, 2017), but more work is needed to replicate these findings and compare both deep and shallow processing to no-task. The impact of encoding tasks on recall, on the other hand, likely depends on individual differences in the effectiveness of strategies employed. A task may not impair memory if participants are using ineffective strategies. Indeed, several studies report better recall for deep processing or no effect of task (Hunt et al., 2011; Hyde & Jenkins, 1969), while others report deep processing impairs memory relative to no-task (Hagen et al., 1970; Mazuryk & Lockhart, 1974).

In sum, there are theoretically motivated reasons to suspect deep processing may increase or decrease the TCE. Existing literature lacks information on which hypothesis is accurate. Here, we propose to fill this gap.

Methods

The hypotheses, methods, and analysis plan for this study were preregistered prior to data collection (https://osf.io/4abjv/?view_only=f246b1d2f32d49f898f43e20fb045465; Healey et al., 2020).

Participants studied 30 lists of words for free recall: 10 lists with no encoding task, 10 with a shallow encoding task (judging if the letter “T” was in the word), and 10 with a deep encoding task (judging if the word referred to a living thing).

Participants

Participants were Michigan State University undergraduate students who completed the experiment for course credit. Data collection began in September 2020 when Michigan State’s classes were conducted remotely due to COVID-19. Therefore, all participants completed the study online.

Sample Size and Stopping Rule

As stated in the preregistration, we planned to collect data from at least 327 participants. This target sample size was selected to provide 95% $1 - \beta$ power to detect a small effect ($d \geq 0.2$) via a two-tailed paired-sample t-test. We originally had planned to stop data collection once the target sample size had been reached or at the end of the Fall 2020 semester, whichever came first. However, COVID-19 created a higher than normal demand within our department for online studies to allow students to meet course requirements remotely. To help meet this demand, we altered our plan and continued to collect data for the entire semester even after surpassing the original target. The data from existing participants were *not* examined prior to making this decision. We collected data from 825 participants in total.

Data Exclusion and Final Sample

Eight participants were excluded for not meeting our demographic exclusion criteria: three for reporting English was not their first language, four for failing to report their first language, and one for indicating they were over 18 at one point and under 18 at another point within the same session. For the remaining participants, data was excluded for any *list* where they recalled fewer than two list items (measuring the TCE requires at least 2 recalled items) or output more than 32 responses (i.e., twice the list length). Any *participant* who had more than 10% of their lists excluded (i.e., > 3 out of 30) was completely excluded from analysis. In total, we excluded 145 participants. This high exclusion rate reflects an overall low average performance in the sample—we return to possible explanations and implications below. Among included participants, a total of 427 lists were excluded (71 from deep lists, 278 from shallow lists, and 78 from no-task lists).

The final sample included 680 participants (82.4% of the total sample); 470 were female, and the mean age was 19.6 ($SD = 1.9$). Participants in the final sample had an average of 97.9% of their lists included ($SD = 3.1\%$, *Mode* = 100%).

Materials

Participants studied 30 lists each composed of 16 words in an immediate free recall task. Lists were composed of words randomly selected from the pool of 1,638 nouns developed for the Penn Electrophysiology of Encoding and Retrieval Study (see Healey et al., 2019). Ten of the 30 lists were randomly assigned to each of the three conditions. Lists were presented in random order with the restriction that no more than two lists from the same condition were presented successively.

Before studying the first list, participants were given instructions explaining each encoding task and the free recall test that would follow each list. Full task instructions are included on the OSF page for this project. For each word in the the deep processing lists, participants were asked “Does this word refer to a living thing?”. For the shallow processing lists, they were asked “Does this word contain the letter T?”. Participants pressed the Y key for YES or the N key for NO while the word was on the screen. For the control no-task condition, participants were assigned no encoding task, were not required to make any keypress, and were free to study the words as they chose.

The letter “T” was chosen as the target letter for the shallow processing task in an effort to roughly match the expected number of YES responses in the deep processing task. To determine how many YES responses would be expected in the deep processing task, two undergraduate research assistants (i.e., from the same student body as our participants) and one author (MGU) independently rated each of the words in the pool as either living or non-living. The three raters agreed for 1,425 out of 1,638 words. Some words were more difficult to judge than others; for example, the word “chest” might be judged as living if it is interpreted as a body part but judged as non-living if it is interpreted as a container (like a “treasure chest”). For the 213 words where they disagreed, the remaining authors each made a YES/NO judgment and the modal judgment across all raters was taken as the expected response. For the deep processing task, 36.0% of the 1,638 words had an expected YES response. “T” occurs in 36.1% of the words in the pool, closer to 36.0% than any other letter.

Procedure

Each trial began with an instruction screen informing the participant which encoding task to perform for the upcoming list. To allow participants to take short breaks as needed, the instruction screen did not advance until the participant pressed SPACE. During the study phase, words were presented individually in the center of the screen for 1 s followed by a 400-600 ms jittered inter-stimulus interval. In deep and shallow lists, the relevant question was displayed above the word until participants entered a response. Then, the prompt disappeared, leaving just the to-be-studied word for the remainder of the 1 s presentation period. Following the presentation of the final word, participants had 60 s to recall as many words from the list as possible in whatever order they came to mind. Recall instructions were displayed onscreen throughout the recall period. Responses were typed individually, and participants were instructed to press ENTER after each response to submit it and clear the screen for the next response. Once the recall period had elapsed, instructions for the next list were presented.

Analyses

A spell-checking algorithm (described in Healey, 2018) checked participants' responses for spelling errors and scored their recall accuracy.

Temporal Contiguity

We used chance-adjusted temporal factor scores as our primary measure of the TCE. This analysis considers the *lag*, or distance, in serial positions between successively recalled items. For example, if a participant just recalled the item in the 3rd serial position on the study list and then recalls the item from the 5th serial position, that would be a transition of *lag* = 5 – 3 = +2. Temporal factor scores are calculated for each list by taking the $|lag|$ of each transition made by a participant, finding its percentile within the distribution of all possible $|lags|$ for that transition (Polyn et al., 2009; Sederberg et al., 2011), and then averaging across

transitions. This analysis ignores the direction of the transition (forward or backward). Transitions outside the list boundaries or to previously recalled items are not considered possible. For example, $lag = +1$ would not be possible if the just-recalled item was the last item in the list. Higher temporal factor scores indicate near-lag transitions are more likely than far-lag transitions (i.e., greater temporal contiguity). To control for primacy, recency, and other serial position effects, which may artificially inflate the TCE, we compared the actual temporal factor score to the score expected if transitions were random with respect to lag (for details on these confounds, see Mundorf et al., 2021). We calculated this chance-level expected factor score by taking the items actually recalled by each participant and permuting the order 500 times, computing a temporal factor score for each permutation. Scores are calculated for each list by subtracting the average of this chance distribution from the actual temporal factor score and dividing by the standard deviation of the chance distribution.

We used lag-conditional response probabilities (lag-CRPs) and temporal bias scores to help visualize the TCE. Lag-CRPs give the probability of making a transition of each lag conditional on the item at that lag being available (for details on how CRP is calculated, see Healey et al., 2019). Temporal bias scores, introduced by Uitvlugt and Healey (2019), are similar to the lag-CRP. However, they remove potential confounds from serial position effects in the same way as the chance-adjusted temporal factor scores. For this reason, we primarily rely on temporal bias and chance-adjusted temporal factor scores as measures of the TCE. Temporal bias for a given lag is calculated for each participant by counting the number of times a transition of that lag was actually made (actual count) and the the number of times a transition of that lag would be expected to occur if items were recalled in random order (expected count; determined through the permutation test described above). The temporal bias score is simply $\frac{\text{actual count} - \text{expected count}}{\text{expected count}}$. Cases where both the actual and expected count were zero were treated as missing values. A score above zero for a given lag indicates it occurred more often than expected by chance, and a score below zero indicates it occurred less than expected.

Semantic Contiguity

The analyses of temporal contiguity described above were part of a preregistered analysis plan. After conducting those analyses, we conducted a set of followup analyses examining semantic contiguity, or the tendency for words that are more strongly semantically related to be recalled together, to determine if LOP also affected semantic organization. These exploratory analyses were undertaken to address a potential explanation for the small size of the LOP effect on the TCE. We measured semantic relatedness between words as the cosine of the angle between their high-dimensional vector representations in Word Association Space (WAS; Steyvers et al., 2004). Measuring word relatedness with WAS $\cos(\theta)$ allows us to measure even small differences in word relatedness, even in our lists composed of randomly selected words. To quantify semantic contiguity, we used a measure analogous to chance-adjusted temporal factor scores. Chance-adjusted *semantic* factor scores are calculated in the same way as their temporal counterparts except that semantic lags are used instead of temporal lags. For a given transition, a semantic lag of 1 means transitioning to the most semantically similar available item in the list (in terms of $\cos(\theta)$), a semantic lag of 2 means transitioning to the second most similar available item, and so on.

Results

Preregistered Analyses

Overall Recall

Probability of recall is displayed in Figure 1A. Mean recall was below 0.4 in every condition, lower than in past research with similar participants (e.g., Healey & Uitvlugt, 2019) but not unusual for intentional free recall using LOP instructions (e.g., Craik & Tulving, 1975; Hunt et al., 2011). Because our primary analyses involve relative differences among conditions, low recall should not impact interpretation of the results.

Planned pairwise tests revealed higher recall for no-task ($M = 0.368$, $SE = 0.005$) than

either deep ($M = 0.316$, $SE = 0.003$), $t(679) = 15.10$, $p < .001$, $d = 0.579$, or shallow ($M = 0.271$, $SE = 0.003$) processing, $t(679) = 28.24$, $p < .001$, $d = 1.083$. This pattern is consistent with past work where no-task participants displayed higher recall than either deep or shallow processing (Hagen et al., 1970; Long & Kahana, 2017; Mazuryk & Lockhart, 1974). We also found a LOP effect; recall was higher under deep than shallow processing, $t(679) = 24.43$, $p < .001$, $d = 0.937$.

Temporal Contiguity

Chance-adjusted temporal factor scores were above chance in all conditions (Figure 1B). Planned comparisons revealed a greater TCE for no-task ($M = 1.28$, $SE = 0.03$) than deep ($M = 1.04$, $SE = 0.02$), $t(679) = 10.93$, $p < .001$, $d = 0.419$, or shallow ($M = 0.99$, $SE = 0.02$) processing, $t(679) = 13.39$, $p < .001$, $d = 0.514$. The TCE was greater for deep than shallow processing, $t(679) = 2.87$, $p = .004$, $d = 0.110$, demonstrating a LOP effect on the TCE. This effect, though significant, was small—we return to this issue in the Discussion.

Recall Dynamics Curves

Although our main focus is overall recall and the TCE, more detailed measures of recall dynamics may provide additional insight into how LOP influence memory search. Serial position curves measure recall as a function of serial position, and probability of first recall curves measure which serial positions are recalled first (Figure 2A and B). Recency was pronounced in all conditions, albeit larger for deep and shallow processing. Primacy was pronounced only for the no-task condition. This pattern is consistent with previous work where imposed processing tasks reduced primacy (e.g., Hagen et al., 1970; Long & Kahana, 2017; Mazuryk & Lockhart, 1974).

Lag-CRPs visualize the TCE as the conditional probability of making a transition of a given *lag*. Lag-CRPs displayed higher probabilities for near than far lags for all conditions (Figure 2C). The peak of the curve was largest for no-task and smallest for deep processing (cf. temporal factor scores in Figure 1B). While the no-task and shallow conditions exhibited the

forward asymmetry typically associated with the TCE (Healey et al., 2019), this asymmetry was attenuated in the deep condition. However, we urge caution in interpreting these results. Serial position effects can introduce a spurious TCE that disguises true differences between conditions, particularly when recall or primacy/recency differ substantially among conditions (Healey et al., 2019; Mundorf et al., 2021; Polyn et al., 2011; Uitvlugt & Healey, 2019), as they do here.

We can illustrate this spurious TCE by simulating data where items are recalled with no true TCE. We simulated recalls for 100,000 participants for each condition. The probability of recalling each item was set to the recall probability of the corresponding position in that condition's serial position curve. This resulted in n items recalled for each simulated participant. To simulate data with *no* contiguity, we simply randomly shuffled the items' output order. Yet, the lag-CRPs (Figure 3B), still display a TCE with forward asymmetry. These lag-CRPs are heavily influenced by recency; $lag = +1$ is highest for shallow processing, the condition with the most recency. In contrast, temporal bias curves and chance-adjusted temporal factor scores (Figure 3C and D) accurately display a null TCE for all conditions, making them a better tool for comparing across conditions.

Returning to our data, temporal bias scores (Figure 2D) were highest for no-task, particularly at $lag = +1$. Forward asymmetry was reduced for shallow and completely eliminated for deep processing. Temporal bias scores reveal the higher TCE for deep processing (see Figure 1B) is due to the symmetrically high bias for near transitions, which results in overall greater temporal contiguity than the asymmetrical shallow condition.¹

¹ We did not conduct an ANOVA on temporal bias scores because recall probability as a function of lag inherently violates the independence assumption (i.e., making transitions of some lags more often than expected by chance entails making transitions of other lags less than expected by chance). As such we avoid making claims about interactions between lag and condition. However, we can provide a rough estimate of the overall TCE by summing scores from $lag = -5$ to $lag = +5$ for each condition and comparing these summed scores. Summed temporal bias scores were higher in deep ($M = 1.96$, $SE = .08$) than shallow processing ($M = 1.58$, $SE = .09$), $t(679) = 4.01$, $p < .001$, $d = 0.154$, consistent with chance-adjusted temporal factor scores.

Exploratory Followup Analyses

While there was a significant LOP effect on temporal contiguity, the effect was small. One possible explanation for the small effect size is that deep processing may also enhance semantic contiguity. Deep processing is inherently semantic and increases semantic organization, at least in lists with a category structure (e.g., Einstein & Hunt, 1980; Koriat & Melkman, 1987). However, items can only be recalled in one order. When items are presented in random order, organizing recalls by semantic similarity inherently reduces temporal contiguity. Thus, the LOP effect on the TCE may have been attenuated by greater semantic organization in the deep condition.

Semantic Contiguity

In all conditions, chance-adjusted semantic factor scores were small but above chance (Figure 1C). A repeated measures ANOVA revealed a significant effect of condition on semantic contiguity, $F(2, 1358) = 5.21, p = .006, \eta^2_G = .005$. Post-hoc tests with a Bonferroni adjusted² $\alpha = .006$ revealed greater semantic contiguity in the no-task ($M = 0.10, SE = 0.01$) compared to the shallow condition ($M = 0.05, SE = 0.01$), $t(679) = 3.34, p = .001, d = 0.117$. There were no differences between no-task and deep ($M = 0.06, SE = 0.01$), $t(679) = 2.55, p = .011$, or deep and shallow, $t(679) = 0.61, p = 0.540$.

Individual Differences

We examined individual differences in recall, temporal contiguity, and semantic contiguity. Reliabilities for recall and the chance adjusted factor scores are reported in Table 1. While recall and temporal factor scores were fairly reliable, semantic factor scores were quite unreliable in all conditions. Thus, we do not report correlations involving semantic contiguity.

The TCE was positively correlated with recall in no-task ($r(678) = .76, p < .001$), deep ($r(678) = .65, p < .001$), and shallow ($r(678) = .72, p < .001$) lists with a Bonferroni adjusted

² Adjusted α is $.05/9$. We conducted nine post-hoc analyses: three t-tests for semantic contiguity, three correlations between TCE and recall, and three correlations between semantic contiguity and recall.

$\alpha = .006$, consistent with previous research using unrelated items (Mundorf et al., 2021; Sederberg et al., 2010; Uitvlugt & Healey, 2019).

Discussion

We tested three hypotheses for how levels of processing (LOP; deep, shallow, no-task control) should influence the temporal contiguity effect (TCE). Our first hypothesis was if deeper processing causes context drift to drift farther, the TCE should be greater for deep than shallow processing. Our second hypothesis was if deeper processing instead increases processing of item information at the expense of order information, it should reduce the TCE. Our final hypothesis was if the TCE arises from strategic control processes, *any* encoding task should disrupt it, regardless of depth.

We found both recall and the TCE were highest with no imposed processing task, were reduced under deep processing, and were further reduced under shallow processing. These results are inconsistent with the hypothesis that deep processing improves memory for items at the expense of memory for order. Instead, they support the hypothesis that deeper processing induces more context drift and the hypothesis that any imposed encoding task disrupts strategic processing. We discuss each hypothesis below.

Item-Order Account

Our results are incompatible with the item-order account, which assumes any manipulation that draws attention to item-specific processing will reduce relational processing. If deeper processing enhances item-specific processing (Eysenck, 1979; Healey & Kahana, 2016), the TCE should be reduced. Yet, deeper processing *increased* the TCE. For the item-order account to be consistent with our results, major assumptions regarding the relationship between item and order information would have to change.

Retrieved Context Models

Under retrieved context models, items form associations with the current state of mental context during study. As each new item activates its associated features, the context representation moves, or drifts, toward those features. Because of context drift, items studied nearby in time form associations with similar states of context. When an item is recalled, it reinstates its associated context from encoding, which serves as a good cue for other items studied nearby in time. In this way, retrieved context models naturally predict a TCE.

The size of the TCE depends on the distance context travels with each item studied. Context travels farther during encoding when items strongly activate their pre-existing context. If context drifts farther with each item, only items studied nearby in time become associated with similar contexts, and the TCE is large. Context changes very little, however, if items weakly activate their associated context. If context drifts only a short distance, all items form associations with similar contexts, reducing the preference for recalling nearby items together—the TCE will be small. Our finding of a greater TCE for deep than shallow processing is consistent with the hypothesis that deep processing should activate more of items' associated contexts (Healey & Kahana, 2016), causing context to drift farther. Notably, our results are inconsistent with an alternate version of these models where deep processing acts only on the item layer.

Although we framed our hypotheses purely in terms of whether the TCE was larger or smaller in deep processing and not the size of that difference, it is worth noting the observed effect size for the difference in temporal contiguity between deep and shallow processing was much smaller ($d = .110$) than we privately expected. A small effect is still compatible with retrieved context models, where the change in context drift can be large or small, creating a larger or smaller TCE. The small effect does, however, suggest that large increases in temporal contiguity are not necessary for the beneficial effects of deep processing on memory, and temporal contiguity is only a part of the LOP puzzle.

Influence of Control Processes

Our results are also consistent with accounts that assume the TCE arises from order-based encoding strategies. In the absence of an experimenter-imposed encoding task, participants often adopt strategies involving temporal organization (Delaney & Knowles, 2005; Hintzman, 2016; Unsworth, 2016) which may directly lead to a TCE (Bouffard et al., 2018; Unsworth et al., 2019). These TCE-generating strategies are often highly effective—the TCE is correlated with recall in lists of unrelated words (Healey et al., 2019; Mundorf et al., 2021; Sederberg et al., 2010; Spillers & Unsworth, 2011). This occurred in our data as well: there was a strong correlation between recall and the TCE in all conditions, larger than in most previous work (smallest $r = .66$; Healey et al., 2019; Mundorf et al., 2021; Sederberg et al., 2010).

A strategic control processes account predicts that assigning any task during encoding will interfere with order-based strategies. Consistent with this prediction, we found recall and the TCE were greatest with no task. The strategic control processes account also predicts that, since order-based strategies encourage forward transitions, forward asymmetry should be greatest when no task interferes with strategy use. Indeed, asymmetry was greatest for the no-task condition.

Differences in strategy use may also provide an explanation for differences in asymmetry among the processing conditions. If participants have limited time to study, a more time-consuming task will leave less time for order-based strategies and result in less forward asymmetry compared to a shorter task. Thus the reduced forward asymmetry for deep compared to shallow processing could be a result of the deep task taking longer to complete. Supporting this explanation, participants responded more slowly to the deep ($M = .73$ seconds) than the shallow ($M = .68$ seconds) processing task, $t(644) = 31.435$, $p < .001$ (see Supplemental Materials). While retrieved context models may be able to explain differences in asymmetry with existing mechanisms, the strategic control processes account offers a clear explanation for differences in asymmetry. Future work should consider how these approaches could be integrated to explain how different features of deep tasks, like difficulty or specificity,

might change strategy use.

Conclusions

Recall and the TCE were higher under deep than shallow processing and highest with no encoding task. Retrieved context models and a strategic control processes account are each consistent with these results. Although theories based on context drift and those which emphasize strategy have been presented as conflicting explanations (Healey et al., 2019; Hintzman, 2016), integrating these two accounts provides the most comprehensive explanation of our results. Integrating a strategy account with retrieved context models will support development of a theory with well-defined mechanisms (even for the difficult-to-define strategic control processes) that accounts for both automatic and intentional processes of memory. Both perspectives should be considered for furthering theory development and in efforts to utilize LOP to improve memory performance.

Open Practices Statement

The hypotheses, methods, and analysis plan for this study were preregistered prior to data collection. Details on the preregistration and de-identified data along with data analysis scripts are available at https://osf.io/4abjv/?view_only=f246b1d2f32d49f898f43e20fb045465.

Declarations

Funding

This material is based upon work supported by the National Science Foundation under Grant No. 1848972.

Conflict of Interest

The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval

The study procedures and materials were approved by the Michigan State University Institutional Review Board. All elements of this study were performed in line with the principles of the Declaration of Helsinki.

Consent to Participate

Informed consent was obtained from all individual participants included in the study.

Consent for Publication

Not applicable.

Availability of Data and Materials

All de-identified data analyzed in this study are available at https://osf.io/4abjv/?view_only=f246b1d2f32d49f898f43e20fb045465.

Code Availability

All data analysis scripts used in this study are available at
https://osf.io/4abjv/?view_only=f246b1d2f32d49f898f43e20fb045465.

Authors' Contributions

M. K. Healey developed the study concept. All authors contributed to the study design. Data collection was performed by M. G. Uitvlugt. A. M. D. Mundorf and M. G. Uitvlugt performed the data analysis and interpretation under the supervision of M. K. Healey. A. M. D. Mundorf drafted the manuscript, and M. K. Healey and M. G. Uitvlugt provided critical revisions. All authors approved the final version of the manuscript for submission.

References

- Ayçiçegi-Dinn, A., & Caldwell-Harris, C. L. (2009). Emotion-memory effects in bilingual speakers: A levels-of-processing approach. *Bilingualism, 12*(3), 291.
<https://doi.org/10.1017/s1366728909990125>
- Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review, 85*(3), 139–152.
<https://doi.org/10.1037/0033-295x.85.3.139>
- Biggs, J. B. (1978). Individual and group differences in study processes. *British Journal of Educational Psychology, 48*(3), 266–279.
<https://doi.org/10.1111/j.2044-8279.1978.tb03013.x>
- Bouffard, N., Stokes, J., Kramer, H. J., & Ekstrom, A. D. (2018). Temporal encoding strategies result in boosts to final free recall performance comparable to spatial ones. *Memory & Cognition, 46*(1), 17–31. <https://doi.org/10.3758/s13421-017-0742-z>
- Craik, F. I. M. (2002). Levels of processing: Past, present... and future? *Memory, 10*(5-6), 305–318. <https://doi.org/10.1080/09658210244000135>
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104*(3), 268–294.
<https://doi.org/10.1037/0096-3445.104.3.268>
- 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.
<https://doi.org/10.1037/0033-295X.112.1.3>
- Delaney, P. F., & Knowles, M. E. (2005). Encoding strategy changes and spacing effects in the free recall of unmixed lists. *Journal of Memory and Language, 52*(1), 120–130.
<https://doi.org/10.1016/j.jml.2004.09.002>

- Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning & Memory*, 6(5), 588–598. <https://doi.org/10.1037/0278-7393.6.5.588>
- Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica*, 96(1), 43–60. [https://doi.org/10.1016/s0001-6918\(97\)00005-x](https://doi.org/10.1016/s0001-6918(97)00005-x)
- Eysenck, M. W. (1979). Levels of processing in human memory. In F. I. M. Cermak L.S; Craik (Ed.). Psychology Press.
- Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, 119(2), 223–271. <https://doi.org/10.1037/a0027371>
- Hagen, J. W., Meacham, J. A., & Mesibov, G. (1970). Verbal labeling, rehearsal, and short-term memory. *Cognitive Psychology*, 1(1), 47–58.
[https://doi.org/10.1016/0010-0285\(70\)90004-6](https://doi.org/10.1016/0010-0285(70)90004-6)
- Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, 26(3), 699–720.
<https://doi.org/10.3758/s13423-018-1537-3>
- Healey, M. K., Mundorf, A. M. D., & Uitvlugt, M. G. (2020). Does depth of processing affect temporal contiguity? <https://osf.io/gkcm7>
- Healey, M. K. (2018). Temporal contiguity in incidentally encoded memories. *Journal of Memory and Language*, 102, 28–40. <https://doi.org/10.1016/j.jml.2018.04.003>
- Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, 123(1), 23–69. <https://doi.org/10.1037/rev0000015>
- Healey, M. K., & Uitvlugt, M. G. (2019). The role of control processes in temporal and semantic contiguity. *Memory & Cognition*, 47, 719–737.
<https://doi.org/10.3758/s13421-019-00895-8>
- Hintzman, D. L. (2016). Is memory organized by temporal contiguity? *Memory & Cognition*, 44(3), 365–375. <https://doi.org/10.3758/s13421-015-0573-8>

- Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 484–494. <https://doi.org/10.1037/0278-7393.14.3.484>
- Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language, 46*(3), 85–98. <https://doi.org/10.1006/jmla.2001.2798>
- Howard, M. W., Shankar, K. H., Aue, W. R., & Criss, A. H. (2015). A distributed representation of internal time. *Psychological Review, 122*(1), 24–53. <https://doi.org/10.1037/a0037840>
- Hunt, R. R., Smith, R. E., & Dunlap, K. R. (2011). How does distinctive processing reduce false recall? *Journal of Memory and Language, 65*(4), 378–389. <https://doi.org/10.1016/j.jml.2011.06.003>
- Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology, 82*(3), 472–481. <https://doi.org/10.1037/h0028372>
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24*(1), 103–109. <https://doi.org/10.3758/BF03197276>
- Koriat, A., & Melkman, R. (1987). Depth of processing and memory organization. *Psychological Research, 49*(2-3), 173–181. <https://doi.org/10.1007/BF00308684>
- Lazarus, L. T., Mundorf, A. M. D., Uitvlugt, M. G., & Healey, M. K. (in prep). An item-order model of reduced temporal contiguity for distinct items. *Manuscript submitted for publication.*
- Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review, 120*(1), 155–189. <https://doi.org/10.1037/a0030851>
- Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Intralist and interlist effects in free recall. *Psychological Review, 122*(2), 337–363. <https://doi.org/10.1037/a0039036>

- Long, N. M., & Kahana, M. J. (2017). Modulation of task demands suggests that semantic processing interferes with the formation of episodic associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(2), 167–176.
<https://doi.org/10.1037/xlm0000300>
- Martin, A., Brouwers, P., Cox, C., & Fedio, P. (1985). On the nature of the verbal memory deficit in alzheimer's disease. *Brain and Language*, 25(2), 323–341.
[https://doi.org/10.1016/0093-934x\(85\)90088-4](https://doi.org/10.1016/0093-934x(85)90088-4)
- Mazuryk, G. F., & Lockhart, R. S. (1974). Negative recency and levels of processing in free recall. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 28(1), 114–123.
<https://doi.org/10.1037/h0081971>
- McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, 15(2), 237–255.
<https://doi.org/10.3758/pbr.15.2.237>
- McDaniel, M. A., Cahill, M., Bugg, J. M., & Meadow, N. G. (2011). Dissociative effects of orthographic distinctiveness in pure and mixed lists: An item-order account. *Memory & Cognition*, 39(7), 1162–1173. <https://doi.org/10.3758/s13421-011-0097-9>
- Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior*, 15(4), 447–458. [https://doi.org/10.1016/s0022-5371\(76\)90040-2](https://doi.org/10.1016/s0022-5371(76)90040-2)
- Mundorf, A. M. D., Lazarus, L. T. T., Uitvlugt, M. G., & Healey, M. K. (2021). A test of retrieved context theory: Dynamics of recall after incidental encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(8), 1264–1287.
<https://doi.org/10.1037/xlm0001001>
- Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(4), 702–709. <https://doi.org/10.1037/0278-7393.17.4.702>

- Osth, A., & Fox, J. (2019). Are associations formed across word pairs? A test of learning by temporal contiguity in associative recognition. *Psychonomic Bulletin & Review*, 1–7. <https://doi.org/10.3758/s13423-019-01616-7>
- Polyn, S. M., Erlichman, G., & Kahana, M. J. (2011). Semantic cuing and the scale-insensitivity of recency and contiguity. *Journal Experimental Psychology: Learning, Memory and Cognition*, 37(3), 766–775. <https://doi.org/10.1037/a0022475>
- 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), 129–156. <https://doi.org/10.1037/a0014420>
- Rose, N. S., & Craik, F. I. M. (2012). A processing approach to the working memory/long-term memory distinction: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 1019–1029. <https://doi.org/10.1037/a0026976>
- Sadeh, T., Moran, R., & Goshen-Gottstein, Y. (2015). When items ‘pop into mind’: Variability in temporal-context reinstatement in free-recall. *Psychonomic Bulletin & Review*, 22(3), 779–790. <https://doi.org/10.3758/s13423-014-0746-7>
- Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review*, 18(3), 455–468. <https://doi.org/10.3758/s13423-011-0086-9>
- 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. <https://doi.org/10.3758/MC.38.6.689>
- Seiver, J. G., Pires, M., Awan, F., & Thompson, W. (2019). Retention of word pairs as a function of level of processing, instruction to remember, and delay. *Journal of Cognitive Psychology*, 31(7), 665–682. <https://doi.org/10.1080/20445911.2019.1664555>
- 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), 1532–1539.

<https://doi.org/10.1037/a0024852>

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*. American Psychological Association. <https://doi.org/10.1037/10895-018>

Uitvlugt, M. G., & Healey, M. K. (2019). Temporal proximity links unrelated news events in memory. *Psychological Science*, 30(1), 92–104.

<https://doi.org/10.1177/0956797618808474>

Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(1), 50–61. <https://doi.org/10.1037/xlm0000148>

Unsworth, N., Miller, A. L., & Robison, M. K. (2019). Individual differences in encoding strategies and free recall dynamics. *Quarterly Journal of Experimental Psychology*, 72(10), 2495–2508. <https://doi.org/10.1177/1747021819847441>

Split-half reliability for individual difference variables

Table 1

Split-half reliability for probability of recall, chance-adjusted temporal factor (TF) scores, and chance-adjusted semantic factor (SF) scores are presented here. For each condition, split-half reliability was calculated following the methodology of Sederberg et al. (2010). For each participant, we stratified their valid lists (where at least 2 list items were recalled) by condition and then randomly divided the participant's lists into two sets. In cases where the participant had an uneven number of valid lists in a given condition due to exclusions, we randomly selected which set would contain an additional list for that participant. We calculated probability of recall and chance-adjusted factor scores for each set and correlated the scores for set 1 with scores for set 2, correcting with the Spearman-Brown prediction formula ($2\rho/[1 + \rho]$). This procedure was repeated 2,000 times, where the lists assigned to each set were randomly chosen for each participant in each iteration.

| Condition | Prob. recall | Chance-adjusted TF scores | Chance-adjusted SF scores |
|-----------|--------------|---------------------------|---------------------------|
| No-task | 0.923 | 0.759 | 0.072 |
| Deep | 0.892 | 0.628 | -0.013 |
| Shallow | 0.897 | 0.671 | 0.058 |

Effects of LOP task on overall recall, temporal contiguity, and semantic contiguity

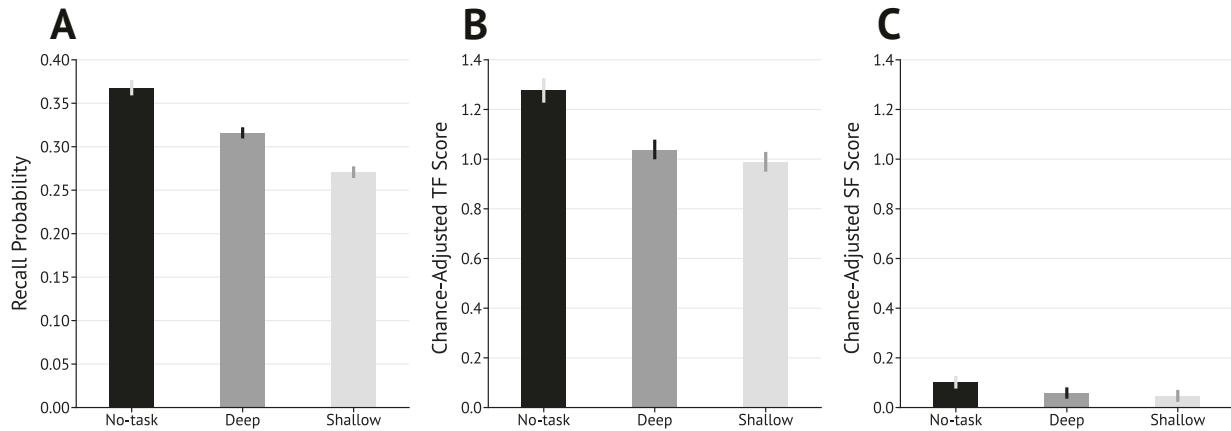


Figure 1

Measures of overall recall, temporal contiguity, and semantic contiguity for all conditions. (A) Probability of recall, (B) chance-adjusted temporal factor (TF) scores, and (C) chance-adjusted semantic factor (SF) scores for no-task, deep processing, and shallow processing lists. For temporal and semantic factor scores, chance was determined by permuting the order of recalls 500 times. Scores are calculated for each list by subtracting the average of the chance distribution from the actual temporal or semantic factor score and then dividing by the standard deviation of the chance distribution. Error bars are bootstrapped 95% confidence intervals.

Effects of LOP task on recall dynamics

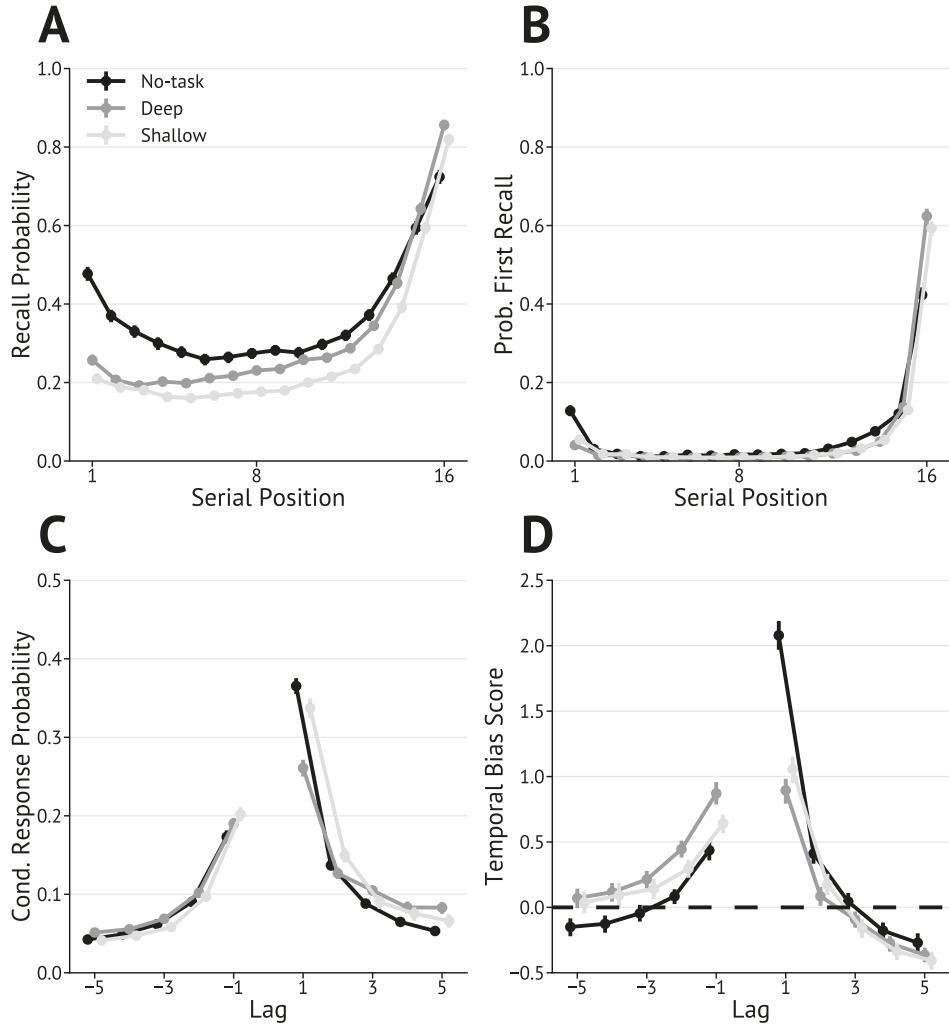


Figure 2

(A) Serial position curves, (B) probability of first recall curves, (C) lag-conditional response probabilities (lag-CRP), and (D) temporal bias scores for no-task, deep processing, and shallow processing lists. Temporal bias scores for each lag were calculated by comparing the number of times a transition of that lag was actually made to the number of times it would be expected to occur by chance. Chance was calculated by permuting the order of recalls for each list 500 times and counting on average how many times each lag occurred for each permutation. The dotted line for the temporal bias scores indicates a score of zero (no bias). Error bars are bootstrapped 95% confidence intervals.

Simulated data with no temporal contiguity

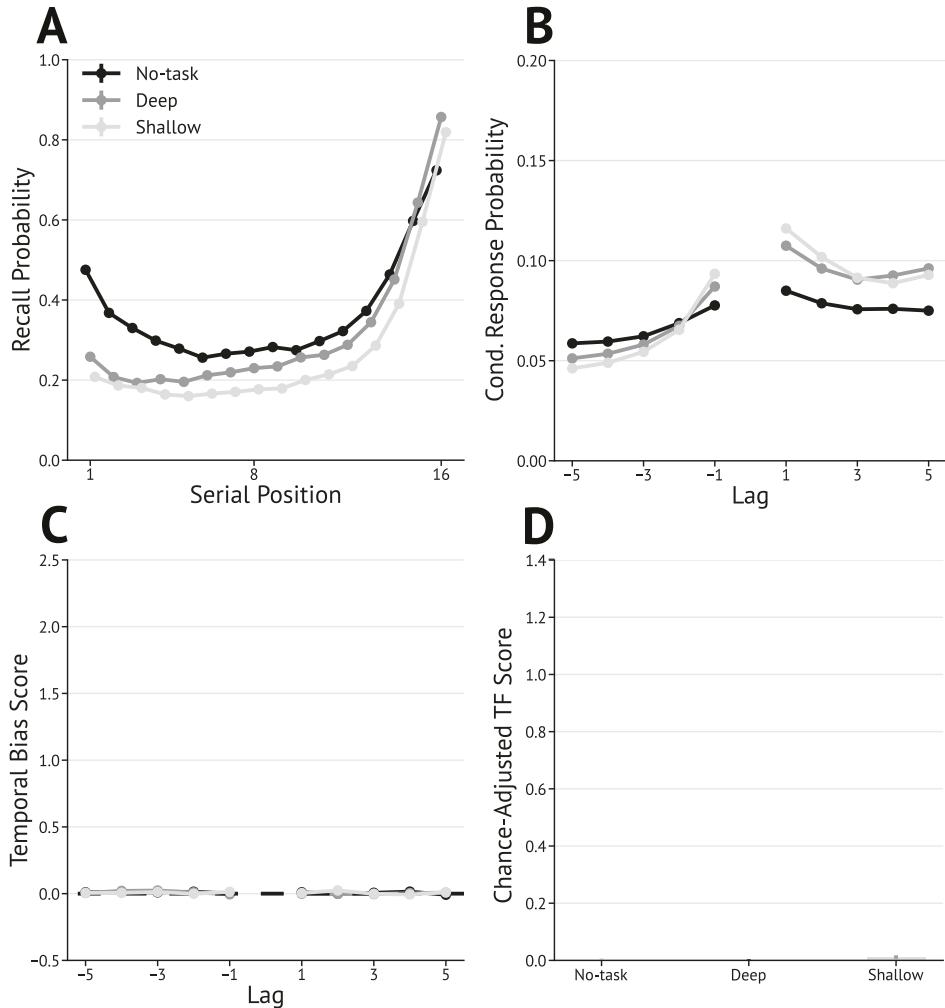


Figure 3

Simulated (A) serial position curves (SPCs), (B) lag-conditional response probability (lag-CRP) curves, (C) temporal bias curves, and (D) chance-adjusted temporal factor scores from a model where recall order was randomly selected with regard to lag to produce simulated recalls with no temporal contiguity. For this simulation we generated recalls for 100,000 simulated participants, each recalling from 1 list of 16 items. For each participant, we determined which items would be recalled using binomial distribution where the probability of the participant recalling an item from a given serial position was set to the recall probability of the corresponding serial position in that condition's serial position curve. This resulted in n recalled items. Recall order was determined by randomly shuffling the n recalled items. Despite the data being generated such that items were recalled in random order (with zero temporal contiguity), the lag-CRPs display a contiguity effect as an artifact of the recency in the simulated SPCs. We present simulated lag-CRPs on a smaller scale here in order to better display differences between conditions in this simulated data. Temporal bias curves display a null TCE, consistent with the method of data simulation. Chance-adjusted temporal factor scores are also at or near zero for all conditions (making them barely visible in this figure).