

## Temporal Contiguity in Incidentally Encoded Memories

M. Karl Healey

Michigan State University

### Author Note

I thank Mitchell Uitvlugt and Kimberly Fenn for helpful discussions. I am grateful to James S. Nairne and an anonymous reviewer for invaluable comments. Correspondence concerning this article should be addressed to M. Karl Healey (khealey@msu.edu) at Michigan State University, Department of Psychology, 316 Physics Road, East Lansing, MI.

## Abstract

Thinking of one event often triggers recall of other events experienced nearby in time. This Temporal Contiguity Effect has been extensively documented in laboratory list learning tasks, but its source is debated. Is it due to task-general automatic processes that operate whenever new memories are formed? Or is it due to task-specific encoding strategies that operate only during deliberate rote learning? I test these theories by presenting over 3500 subjects with a surprise free recall test after various incidental encoding tasks.

Experiments 1 and 2 show that temporal contiguity is dramatically reduced under incidental encoding. Experiments 3 and 4 show that although the effect is reduced, it is not eliminated—temporal information is encoded incidentally and is used to guide memory search during both free recall and serial recall. These results demonstrate that contiguity is not an artifact of strategy, but the dramatic reduction of the effect also challenges models that posit a strong link between successful memory encoding and contiguity.

*Keywords:* episodic memory; free recall; temporal contiguity

## Temporal Contiguity in Incidentally Encoded Memories

Recalling one event tends to trigger recall of other events experienced nearby in time (for a review, see Healey & Kahana, 2018). Although this Temporal Contiguity Effect (TCE) manifests in many memory tasks (Davis, Geller, Rizzuto, & Kahana, 2008; Schwartz, Howard, Jing, & Kahana, 2005), it is most readily observed in free recall where subjects study a list of words presented serially and then try to recall the words. Despite the fact that subjects are free to recall the items in any order, the order of recall tends to recapitulate the order of study (Kahana, 1996; Murdock, 1974; Postman, 1971, 1972).

The TCE can be illustrated by computing the probability of successively recalling items as a function of their distance, or lag, from each other in the study list (Kahana, 1996). For example, if after recalling the word studied in the 5<sup>th</sup> serial position, your next recall is the word from the 6<sup>th</sup> serial position, you have made a  $lag = +1$  transition. If instead you transitioned from recall of the 5<sup>th</sup> serial position to the 3<sup>rd</sup> position, you have made a  $lag = -2$  transition. For each value of lag, the conditional-response probability (CRP) is computed by dividing the number of times a transition of that lag was *actually* made by the number of times it *could* have been made (e.g., if you have just recalled the last item in the list, it is not possible to make a  $lag = +1$  transition. Transitions to already recalled items are also excluded from the counts as subjects rarely repeat items; Kahana, 1996). The lag-CRP typically is highest for  $lag = +1$  and  $-1$  (but with a forward asymmetry) and decreases sharply for larger absolute values of lag. That is, memory search tends to transition between words that were studied nearby in time.

The TCE has shaped theories of the testing effect (Karpicke, Lehman, & Aue, 2014), directed forgetting (Sahakyan, Delaney, Foster, & Abushanab, 2013), retrieval induced forgetting (Kliegl & Bäuml, 2016), childhood development (Jarrold et al., 2015), cognitive aging (Healey & Kahana, 2016; Wahlheim & Huff, 2015), event segmentation (Ezzyat & Davachi, 2014), time estimation (Sahakyan & Smith, 2014), and even perception (Turk-Browne, Simon, & Sederberg, 2012). Moreover, out of several of factors that

influence free recall (i.e., primacy, recency, and semantic similarity), the magnitude of the TCE has been found to be the most predictive of overall memory ability and general intellectual ability (Healey, Crutchley, and Kahana 2014; also see Sederberg, Miller, Howard, and Kahana 2010; Spillers and Unsworth 2011.)

Yet, we still do not know which cognitive mechanisms generate the TCE (Healey & Kahana, 2018). Here, I will consider two classes of explanation. First, that the TCE arises from task-specific mechanisms that are only engaged when we are deliberately studying a serially presented list. Second, that the TCE arises from task-general mechanisms that the memory system automatically engages whenever new memories are formed.

### **Task-Specific Mechanisms**

Control processes (Atkinson & Shiffrin, 1968; Lehman & Malmberg, 2013) allow us to strategically process information during memory encoding, maximizing recall (e.g., Delaney & Knowles, 2005; Unsworth, 2016). Some work suggests that the TCE arises from such task-specific strategies, implemented by control processes to handle the idiosyncratic demands of laboratory tasks (Hintzman, 2016). In other words, because laboratory tasks require subjects to do something they do not usually do (e.g., learn lists of largely unrelated words), they are forced to devise novel strategies to adapt to the peculiarities of the task. Such task-specific strategies, rather than task-independent memory mechanisms, could account for the contiguity effect.

As an example, the standard free recall task may encourage subjects to adopt the strategy of linking successive list items together to tell a story (Delaney & Knowles, 2005). Another example of a task-specific contiguity-generating mechanism is the method of loci, in which the list items are associated with a pre-memorized sequence of locations. Both of these strategies require subjects to pay attention to the order of presentation and recapitulate it during recall. Thus, both would produce a TCE.

But critically, subjects deploy these contiguity-generating strategies only because

they happen to be well-suited to the specifics of the task. If the specifics of the task change, subjects may adopt different strategies, and these new strategies may not generate contiguity. If the strategies are the only mechanism generating contiguity, any change in the specifics of the task that causes subjects to abandon contiguity-generating strategies should eliminate the TCE entirely. The most decisive test of this prediction is to have subjects process the list under incidental encoding conditions, which should prevent adoption of any deliberate encoding strategy, and then complete a surprise free recall test (Hintzman, 2016).

### Task-General Mechanisms

subjects obviously adopt task-specific strategies. And these strategies doubtlessly contribute to the TCE. But task-specific strategies may not be the only mechanisms that generate the TCE.<sup>1</sup> Many theories of episodic memory propose *task-general* mechanisms that automatically encode information about the temporal proximity of events when forming episodic memories, even if no specific encoding strategy is adopted. If these models are correct, a residual TCE should remain even after removing any impetus to engage encoding strategies.

As an example, some theories assume that new events form associations to a representation of time (Brown, Neath, & Chater, 2007; Howard, Shankar, Aue, & Criss, 2015). This allows recall of one event to trigger recall of temporally adjacent events via associations to adjacent temporal representations. These theories assume time is directly encoded by the memory system, but this is not the only way the memory system might automatically encode information about presentation order. Other theories assume that events experienced close together in time become associated, not with a temporal representation, but with similar states of a drifting mental context representation (Lohnas,

---

<sup>1</sup>Indeed, Hintzman (2016) suggests that in addition to engaging deliberate contiguity-generating strategies, subjects might also automatically notice similarities among temporally proximate items and therefore remember them together at recall.

Polyn, & Kahana, 2015; McGeoch, 1932; Polyn, Norman, & Kahana, 2009). This allows recall of one word to trigger recall of a word studied nearby in time via associations to a common state of mental context. Either of these mechanisms would provide the necessary ingredients to produce a TCE during free recall.

But critically, these encoding mechanisms are assumed to support memory in a range of situations and not just during laboratory list learning tasks. If the specifics of the task change, subjects may adopt different strategies, but they must still rely on the fundamental mechanisms of the memory system. Therefore, changes in specifics of the task might modulate the magnitude of the TCE by changing the contribution of context-generating strategies, but they should not eliminate the TCE entirely. Thus, these theories would predict that a residual TCE should be observed even under incidental encoding.

### **Incidental Encoding of Temporal Order**

The task-specific and task-general perspectives make competing predictions about the influence of removing the intention to encode on the size of the TCE. The literature on incidental encoding provides some data relevant to these predictions, but scholars' interpretations of these findings are mixed.

Glenberg and Bradley (1979) tested for incidental encoding of temporal associations by having subjects repeat a pair of words while trying to retain digits for a varying interval. After 81 such trials, subjects were given two surprise memory tests for the words. The first was an item recognition test (i.e., was the probe seen before); the second was either a cued recall test (given one word from a pair, recall the other) or a pair recognition test (discriminate intact from mismatched pairs). Performance was above chance on the item and pair recognition tests but was very low on the cued recall test, suggesting subjects had limited access to information about which words appeared together. A second experiment also found very low cued recall performance but above chance performance on an associative matching test. Bradley and Glenberg (1983) replicated their earlier findings

and added many control conditions, including a “sheer contiguity” condition in which the words were not presented simultaneously as in the previous experiments but merely in close temporal proximity (as is the case in free recall). In this condition, performance on the associative recognition task was not above chance. Bradley and Glenberg (1983, p. 665) concluded “that sheer temporal contiguity, that is, adjacency of processing, is not sufficient to produce the associations observed in these experiments.”

Data from Nairne (1990, 1991) suggest a different conclusion. In several studies, subjects viewed lists of serially presented words under the guise of a rating task. This incidental encoding task was followed by a surprise order reconstruction task in which subjects were shown the words and had to reconstruct their order. They could do this with considerable accuracy, even when they were shown multiple lists and required to place each word in both its correct list and its correct within-list position (Nairne, 1991). Moreover, even when subjects made a mistake on the reconstruction task, the errors were not random. Instead, order errors following incidental encoding tended to take the form of putting items in positions adjacent to the correct ones, much as they do after explicit encoding (Healy, 1974). This work suggests that subjects have relatively easy access to temporal information (for related examples of access to order information after incidental encoding, see Burns, 1996; Serra & Nairne, 1993).

But other work suggests that this knowledge of order might depend on semantic similarity. Even after explicitly studying a list for a recognition task, subjects perform poorly if they are instead given a surprise spacing judgment task, which requires them to guess the lag that separated pairs of words from the original list, *unless* the words in the pair were semantically related (Hintzman & Block, 1973; Hintzman, Summers, & Block, 1975). These results have been taken as evidence that temporal information is only encoded when subjects notice semantic similarities among items in the list (i.e., study phase retrieval, Hintzman, 2016; Hintzman & Block, 1973; Hintzman et al., 1975).

In all of these studies, the test directly asked subjects to access information about

temporal order. But being explicitly aware of the order of a list and being able to report it is not the same thing as allowing temporal information to influence memory search during free recall. It is possible that subjects could have the temporal information needed to complete an ordering task but fail to use that information to guide a free memory search. Or vice versa, subjects could have difficulty explicitly recalling order yet still implicitly access temporal information to guide memory search. Do subjects spontaneously use incidentally encoded temporal information to produce a TCE in free recall?

Among a series of studies on how the generation effect influences memory for order, Burns (1996) reported a condition in which subjects preformed an incidental encoding cover task on a list composed of 32 items (4 each from 8 semantic categories) and were then given a surprise free recall test. Burns computed a “recall order score” by counting the number of times items that were studied in adjacent serial positions were recalled in adjacent output positions (i.e., the number of  $|lag| = 1$  transitions) and dividing that count by number of items correctly recalled. The mean recall order score across 18 subjects was 0.15 ( $SD = 0.15$ ), which by a 1-sample t-test is significantly above zero. But this finding does not necessarily imply a TCE because recall order scores do not test whether  $|lag| = 1$  transitions are either more frequent than expected by chance or more frequent than longer transitions.

A more direct test for a TCE during free recall after incidental encoding comes from a series of studies reported by Nairne, Cogdill, and Lehman (2017), which were designed to investigate the causes of the survival processing effect (Nairne, Thompson, & Pandeirada, 2007). In the first experiment, subjects completed a survival processing task: subjects viewed a list of items and for each rated its relevance to a survival (or control) scenario. There was a surprise free recall test approximately two minutes after the end of the list. They measured the TCE with temporal factor scores (Polyn et al., 2009), which test for a bias toward short lags over long lags by placing the lag value of each transition made by a subject within the distribution of all lags that were possible at the time. The outcome is a



percentile score where 0.5 indicates no bias and larger scores indicate a preference for short lags. Despite a recall accuracy of approximately 45%, the temporal factor score was less than 0.51 in both conditions, which was not significantly above 0.5. A second experiment replicated this null TCE using a different processing task as a control to the survival processing task. Nairne et al. (2017) included a third experiment that used the same incidental encoding procedure as their first two experiments, but instead of a free recall test, subjects were provided with all of the words from the list and asked to reconstruct the order. This direct test of order memory did reveal a robust TCE in both the survival processing and the control condition. This finding suggests that subjects may be incidentally encoding information about temporal order but failing to spontaneously use it during free recall.

If incidental encoding really does eliminate the TCE in free recall, it would constitute strong evidence against many models of the effect (e.g., Healey & Kahana, 2018; Lohnas et al., 2015). However, further investigation is warranted before concluding that incidental encoding will always eliminate the TCE. Because the Nairne et al. (2017) study was about survival processing and not the influence of intent to encode on the TCE, it did not include an explicit encoding control condition and included some features, other than intent to encode, that are known to influence the TCE. In particular, it used a single, relatively long list (32 items). The TCE is known to be modestly reduced by both lack of task experience and long lists (Healey & Kahana, 2018). It is thus possible that aspects of the task other than incidental encoding contributed to the null TCE. Therefore, Experiment 1 directly compares the TCE in free recall after explicit versus incidental encoding.

## Experiment 1

The goal of the first experiment was to assess whether incidental encoding reduces the TCE relative to explicit encoding. All subjects viewed a list of words presented one at a time and made a simple judgment about each. subjects in the Explicit condition were

expecting a memory test; subjects in the Incidental condition were *not* expecting a memory test. After the last item, all subjects were asked to recall as many of the words as they could, in any order.

## Method

### Data Sharing

All data analyzed in this report are freely available on the author's website:  
<https://cbcc.psy.msu.edu/>.

### Subjects

Given that manipulating encoding intention might reduce, but not eliminate, the TCE, it is critical to have sufficient power to detect small effects. Sederberg et al. (2010) reported a meta-analysis of the TCE in explicit encoding studies; power calculations revealed that a sample size of 143 per condition would provide a  $1 - \beta$  power of 0.95 to detect (via a 1-tailed 1-sample t-test) an effect one-fifth the size of the average effect they reported. In order to collect enough data to meet or exceed this sample size in the Incidental conditions, subjects for all of the experiments reported here were recruited using Amazon's Mechanical Turk, a crowdsourcing website that allows for efficient collection of large volumes of high-quality data. Subjects were paid \$1.00 for participating (a rate of roughly \$10 / hr). The demographics of the Mechanical Turk community have been described elsewhere (approximately 55% female with a mean age of 32; Mason & Suri, 2012). Due to time constraints, no demographic information was collected from subjects for Experiments 1–3. Demographic information was collected in Experiment 4 (subjects in that experiment were overwhelmingly native English speakers with a high school education or a college degree; see Experiment four's Method section for details).

Subjects in the Incidental conditions were excluded from analysis if they reported in a post-experiment questionnaire that they suspected their memory would be tested while

they were performing the Incidental encoding task. The final analyzed sample was composed of 290 in the Explicit condition and 339 in the Incidental condition. Table 1 shows the total number of included and excluded subjects for each experiment.

Table 1

*Sample sizes, exclusions, and recall probability by condition.*

Exp	Condition	<i>n</i> Included	<i>n</i> Excluded (aware)	Recall Prob. (SD)
Exp1	Explicit	290	–	0.46 (0.16)
	Incidental	339	47	0.40 (0.16)
Exp2	Explicit	137	–	0.47 (0.17)
	Incidental	212	43	0.47 (0.14)
Exp3	Weight	291	31	0.42 (0.15)
	Animacy	294	31	0.37 (0.14)
	Moving Scenario	299	32	0.38 (0.14)
	Movie	329	55	0.38 (0.15)
	Relational	311	98	0.44 (0.18)
Exp4	Varying-Free	538	28	0.29 (0.12)
	Constant-Free	538	38	0.42 (0.15)
	Constant-Serial	515	34	0.37 (0.14)

## Procedure

All subjects completed two free recall lists. Each list was composed of 16 words drawn randomly from a pool of 1638 words, with the constraint that no word was used more than once for a given subject. Randomly generating lists from a large pool is critical to rule out an explanation of the TCE based on study phase retrieval. If semantically related words appear in a list, viewing one of these words may remind the subject of a related word from earlier in the list, which could allow subjects to form a new episodic link between the words that is due to semantic similarity rather than temporal proximity (Hintzman, 2016; Hintzman & Block, 1973; Hintzman et al., 1975). Drawing words from a large pool and randomly generating unique lists for each subject ensures that, across

subjects, any semantic associations are evenly distributed across lags. Therefore a study phase retrieval account of any TCE must assume that subjects are more likely to be reminded of semantic associations that occurred at near versus far lags, which is just another way of saying there is a TCE.

Words were presented one at a time on the subject's computer screen for 4 seconds. This presentation rate was deliberately chosen to be slightly faster than the 5 seconds per word used by Nairne et al. (2017) to reduce the amount of "free time" subjects have between making the judgment and the presentation of the next word. There was an inter-stimulus interval of 1 second between word presentations during which a fixation cross was displayed in the same location where the words appeared. The final word of each list was followed by a 16 second distractor period during which subjects answered math problems of the form  $A + B + C = ?$ , where  $A$ ,  $B$ , and  $C$  were positive, single-digit integers, though the answer could have been one or two digits. Subjects typed their answers to the math problems in a text box and pressed enter to submit. Upon pressing enter, a new math problem and a new blank text box appeared. Subjects were instructed to "Try to solve as many problems as you can without sacrificing accuracy. The task will automatically advance when the time is up."

Following the math distractor task, subjects in both conditions were asked to recall as many items as possible from the preceding list, in any order. Subjects typed each recalled word in a text box and pressed enter to submit the word. Upon pressing enter, the word disappeared and a new blank text box appeared such that subjects could not see their prior responses. Subjects were given 75 seconds to recall as many words as they could. To ensure subjects noticed that the recall period had begun (e.g., were not looking at the keyboard and typing their answer from the final math problem), a red screen was flashed for 500 ms before the recall instructions were displayed and the recall text box did not begin accepting input for a further 500 ms. Therefore, including the math distractor, there was a total delay of  $16 + 0.5 + 0.5 = 17$  seconds between the end of the study period and the beginning

of the recall period. A spell-checking algorithm (described in the Supplemental Materials) checked subjects' typed responses for typos and scored their recall accuracy.

**Encoding Instructions Manipulation.** Subjects were randomly assigned to either the Incidental condition or the Explicit condition. Prior to seeing the first list, subjects in both conditions were told that they would see a series of words and would make a simple judgment about each one (i.e., Would it fit in a shoebox?). The exact instructions depended on the condition. In the Explicit condition, subjects were given standard free recall instructions that described the size judgment task but emphasized memory. In the Incidental condition, subjects were given instructions only for the size judgment task and memory was never mentioned. Because the wording of the instructions are integral to the intent manipulation, they are reproduced exactly in the Supplementary Materials.

**Shoebox Task.** In both conditions, subjects were asked to make a size judgment about each word while it was present on the screen. Specifically, they were asked to judge if the word referred to an object that could fit into a shoebox. To allow for the same yes/no response across all conditions and all experiments, subjects were asked to indicate if the judgment was easy to make under the guise of norming the items for a later study. See the Supplemental Materials for the exact task instructions as well as measures taken to ensure subjects understood the task instructions.

## Quantifying the Temporal Contiguity Effect

The lag-CRP provides a visual representation of the TCE, but for several reasons it is useful to have a single number that quantifies the size of the effect. First, it is important to take into account the fact that the likelihood of successful recall is not random with respect to serial position (e.g., there are primacy and recency effects, or more generally autocorrelations in goodness of encoding), which can artificially increase the size of the TCE by disproportionately increasing the number of possible ways one can make short-lag transitions (Healey & Kahana, 2018; Hintzman, 2016). Second, the lag-CRP can also

*underestimate* the size of the TCE in some cases. Imagine a subject who successfully encodes only three items: those from serial positions 3, 8, and 16. Even if this subject’s memory search was guided by only temporal associations and they recalled the three items in perfect temporal order, they could not possibly make any transition shorter than  $|lag| = 5$ .

To address these confounds between recall accuracy and the TCE, I combined the *temporal factor score* (Polyn et al., 2009; Sederberg et al., 2010) with a shuffling procedure (Polyn, Erlikhman, & Kahana, 2011). The temporal factor score is computed by ranking the absolute value of the lag of each actual transition with respect to the absolute values of the lags of all transitions that were possible at that time, which provides a percentile score for each transition. Averaging these percentile scores across all of a subject’s transitions provides the temporal factor score. Temporal factor scores can be influenced by confounds with recall accuracy, but this influence can be estimated and removed by taking the items that a subject actually recalled for a given list, randomly shuffling (i.e., permuting) the order of recalls, and recomputing the temporal factor score. Repeating this permutation procedure many times provides a distribution of the temporal factor score expected if recall transitions are completely random with respect to lag (Healey & Kahana, 2018).

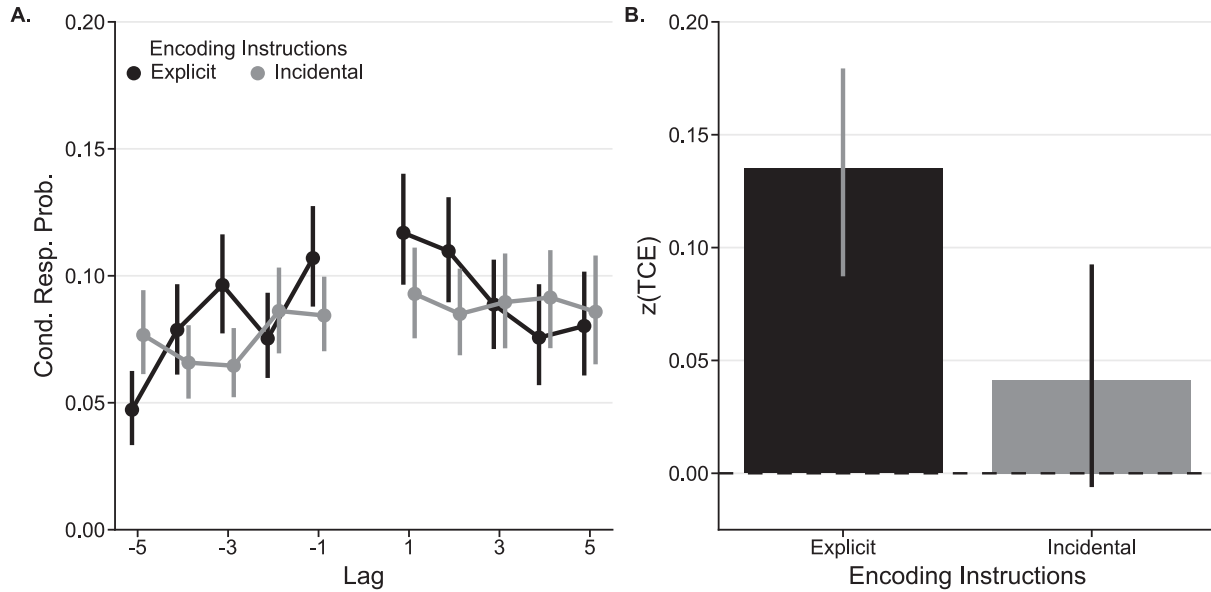
This logic was used to provide a corrected measure of the TCE for each subject. For each list, the temporal factor score was computed for the actual recall sequence and for 10,000 random permutations of the sequence. The actual temporal factor score was then converted into a z-score,  $z(\text{TCE})$ , by subtracting the permutation distribution’s mean and dividing by its standard deviation. For example, imagine that a subject recalled three words: those from serial positions 6, 5, and 8, in that order. The actual temporal factor score for that recall sequence is 0.848. We would build a distribution of the temporal factor score expected by chance by taking this recall sequence and randomly shuffling it. For example, rearranging the sequence to 5, 6, 8 gives a temporal factor score of 0.924. Rearranging the sequence to 6, 8, 5 gives a temporal factor score of 0.776. Doing this for

all six possible permutations of the sequence, we find that the distribution of possible temporal factor scores has a mean of 0.834 and a standard deviation of 0.069, which gives our observed temporal factor score a  $z(TCE) = \frac{0.848 - 0.834}{0.069} = 0.203$ . In the absence of a true TCE, the expected value of  $z(TCE)$  is zero, so we can test for a TCE by determining if the across-subject average of  $z(TCE)$  is significantly above zero.

## Results and Discussion

Figure 1 shows the lag-CRP and corrected temporal factor scores for the Explicit and Incidental conditions. The Explicit condition shows a clear TCE: the lag-CRP is highest for short lags (i.e.,  $|lag| = 1$ ) and decreases for larger lags. Moreover, the 95% confidence interval on the  $z(TCE)$  lies well above zero. By contrast, the Incidental condition shows no evidence of a TCE: the lag-CRP is nearly flat and the 95% confidence interval on the  $z(TCE)$  includes zero. These results show that removing the intent to encode can reduce the magnitude of the TCE to the point that it is statistically indistinguishable from zero.

It is notable that even in the Explicit condition, the TCE was small and more symmetric than the TCEs reported in most previous work. Across a range of variations of the free recall task, the lag-CRP typically peaks at about 0.3–0.5 (Healey & Kahana, 2018) compared to approximately 0.12 for the current Explicit condition. The critical difference is likely the amount of task experience. In a multi-list explicit encoding study, Healey and Kahana (2018) found that on a subject’s very first list, the lag-CRP peaked at approximately 0.15, but on the twelfth list, it peaked at 0.3. Figure 2 shows the TCE for the second list in the current study (note that this figure has a different y-axis scale than the figure showing the first list data). In both conditions, the  $z(TCE)$  is numerically higher than on the first list, but not significantly so (the 95% bootstrapped confidence intervals on the  $z(TCE)$  difference between list 1 and list 2 was  $-0.07 - +0.07$  for the Explicit condition and  $-0.08 - +0.07$  for the Incidental). Moreover, the TCE is still lower in the Incidental than the Explicit condition, perhaps reflecting that the Explicit condition are profiting

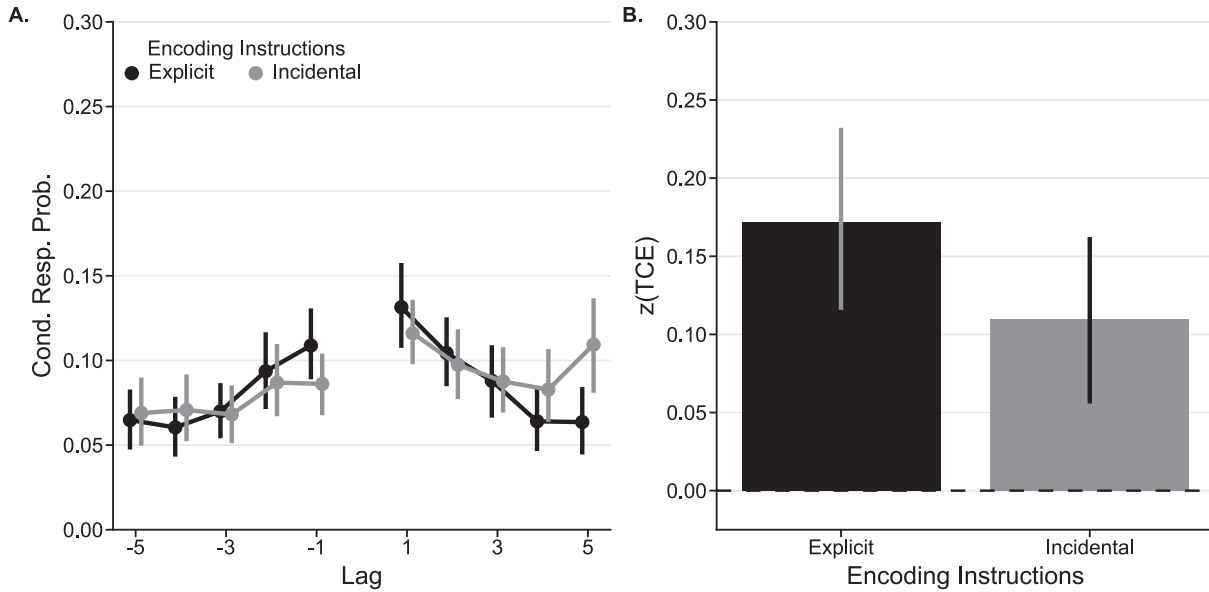


*Figure 1.* The Temporal Contiguity Effect (TCE) on the first list under Explicit versus Incidental encoding using the Shoebox Task (Experiment 1). (A) Lag-conditional response probability functions. (B) The average  $z(\text{TCE})$ .  $z(\text{TCE})$  for a given subject is computed as follows: An observed temporal factor score was computed as the average percentile ranking of the temporal lag of each actual transition in the recall sequence with respect to the lags of all transitions that were possible at that time. To determine the temporal factor score expected by chance, a permutation distribution was created by randomly shuffling the order of recalls within the sequence 10,000 times and computing a temporal factor score for each shuffling. The reported value,  $z(\text{TCE})$ , is the z-score of the observed temporal factor score within the permutation distribution. All Error bars are bootstrapped 95% confidence intervals.

from the encoding practice they gained on the first list. These findings would be difficult to account for with most existing models, because they have no mechanism to simulate practice or otherwise allow contiguity to change dramatically with task experience.

Although the TCE is our main focus, examining other aspects of recall dynamics may help shed light on why the TCE varies across conditions. Overall recall probability (Table 1) was lower in the Incidental than the Explicit condition. This is largely attributable to subjects in the Incidental condition recalling fewer items from early serial positions as can be seen in the serial position curves (SPC) in Figure 3A. Both groups show considerable recency despite the incidental encoding and a delayed test (for similar findings, see Glenberg et al., 1980; Marshall & Werder, 1972; Neath, 1993).





*Figure 2.* The Temporal Contiguity Effect (TCE) on the second list under Explicit versus Incidental encoding using the Shoebox Task (Experiment 1). (A) Lag-conditional response probability functions. (B) The average  $z(\text{TCE})$ .  $z(\text{TCE})$  for a given subject is computed as follows: An observed temporal factor score was computed as the average percentile ranking of the temporal lag of each actual transition in the recall sequence with respect to the lags of all transitions that were possible at that time. To determine the temporal factor score expected by chance, a permutation distribution was created by randomly shuffling the order of recalls within the sequence 10,000 times and computing a temporal factor score for each shuffling. The reported value,  $z(\text{TCE})$ , is the z-score of the observed temporal factor score within the permutation distribution. All Error bars are bootstrapped 95% confidence intervals.

How subjects initiate recall, as revealed by probability of first recall (PFR) curves (Figure 4B), is also informative. In delayed free recall tasks, as used here, subjects typically initiate recall by first retrieving an item from near the beginning of the list (i.e., they focus first on primacy items; Howard & Kahana, 1999). subjects in the Explicit condition showed this typical pattern, but subjects in the Incidental condition showed the opposite pattern of focusing first on recency items (Figure 3B), which is more typical of immediate recall (Hogan, 1975). On the second list, when everyone expected a memory test, these group differences in recall initiation and accuracy were largely eliminated (Figure 4).

Differences in recall dynamics between incidental and explicit encoding may be due to removing the impetus to rehearse (Glenberg et al., 1980; Marshall & Werder, 1972; Neath,

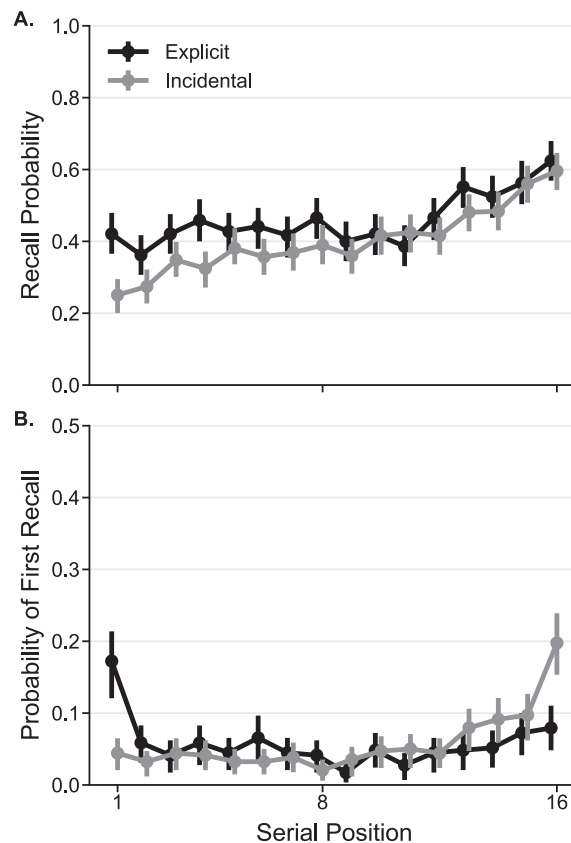


Figure 3. (A) Serial position curves and (B) probability of first recall curves on the first list under Explicit versus Incidental encoding using the Shoebox Task (Experiment 1). All error bars are bootstrapped 95% confidence intervals.

1993), which is known to increase primacy by effectively increasing the functional serial position of early list items (Brodie & Murdock, 1977; Rundus, 1971; Tan & Ward, 2000). But what about the TCE? Rehearsal could increase the TCE by causing subjects to hold adjacent items in mind at the same time (Hintzman, 2016). Although the TCE has been found under conditions designed to minimize rehearsal (Howard & Kahana, 1999), reduced rehearsal may be one factor contributing to a diminished TCE under incidental encoding.

In summary, Experiment 1 directly compared Explicit and Incidental encoding conditions and found that removing the intent to encode dramatically reduced the TCE. But because the TCE has proven to be so robust in previous studies (Healey & Kahana, 2018), I attempt to replicate the finding in Experiment 2 using a slightly different

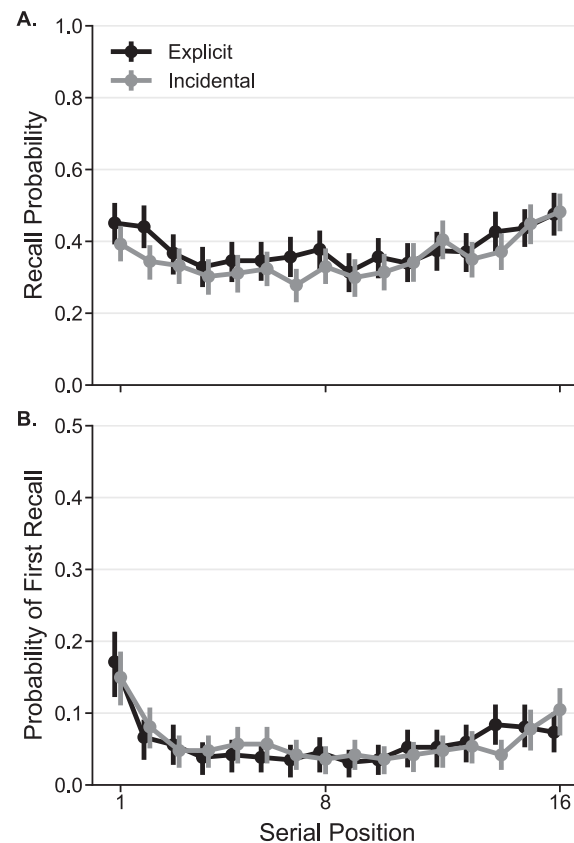


Figure 4. (A) Serial position curves and (B) probability of first recall curves on the second list under Explicit versus Incidental encoding using the Shoebox Task (Experiment 1). All error bars are bootstrapped 95% confidence intervals.

processing task.

## Experiment 2

### Method

The methods were identical to those used in Experiment 1 except for the judgment task instructions (see Table 1 for sample size information). The processing required by the Shoebox Task from Experiment 1 is quite simple. So simple that one could argue it is ineffective at forming strong memories (Eagle & Leiter, 1964), which may artificially reduce the TCE. Therefore, I wanted to retain the basic task of judging size while increasing memory performance in the Incidental condition. That is, can processing that promotes

memory do so without producing substantial contiguity? Mental imagery and self-referential processing are two effective ways to improve memory. Thus, the Front Door Judgment Task asked subjects to imagine trying to move the object referred to by each item through the front door of their house and decide whether or not it would be possible: “Specifically, for each word, you will think of the object it refers to and try to imagining yourself moving that object through the front door of your home. Ask yourself if the object would successfully fit through your front door.” Again, subjects were asked to indicate if this judgment was easy or difficult to make by pressing “Y” or “N”. See the Supplemental Materials for the exact task instructions.

## Results and Discussion

As predicted, the Front Door Task substantially improved memory accuracy in the Incidental condition. In fact, probability of recall was equal in the Explicit and Incidental conditions (Table 1), and the differences in serial position and probability first recall curves were reduced (Figure 5). Together, these findings suggest factors that influence memory accuracy, like rehearsal, played less of a role in producing differences between the Explicit and Incidental conditions in this Experiment.

Nonetheless, the Front Door Task did not produce a significant TCE on the first list under incidental encoding: Figure 6 shows that whereas the Explicit condition had a  $z(\text{TCE})$  significantly above zero, the Incidental condition showed a flattened lag-CRP and a  $z(\text{TCE})$  for which the confidence interval included zero. Analyses of the second list, which are reported in the Supplemental Materials for this and subsequent experiments, replicated the Experiment 1 finding of increased contiguity.

These results confirm that incidental encoding can dramatically reduce temporal contiguity without substantially decreasing memory performance (Nairne et al., 2017). This lack of coupling between level of recall and level of temporal contiguity has important theoretical implications, which I will consider in the Discussion.

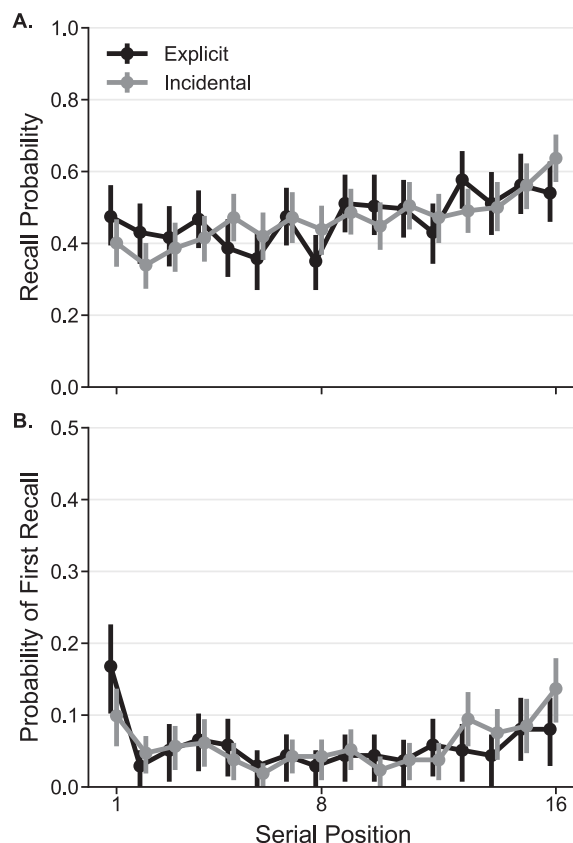
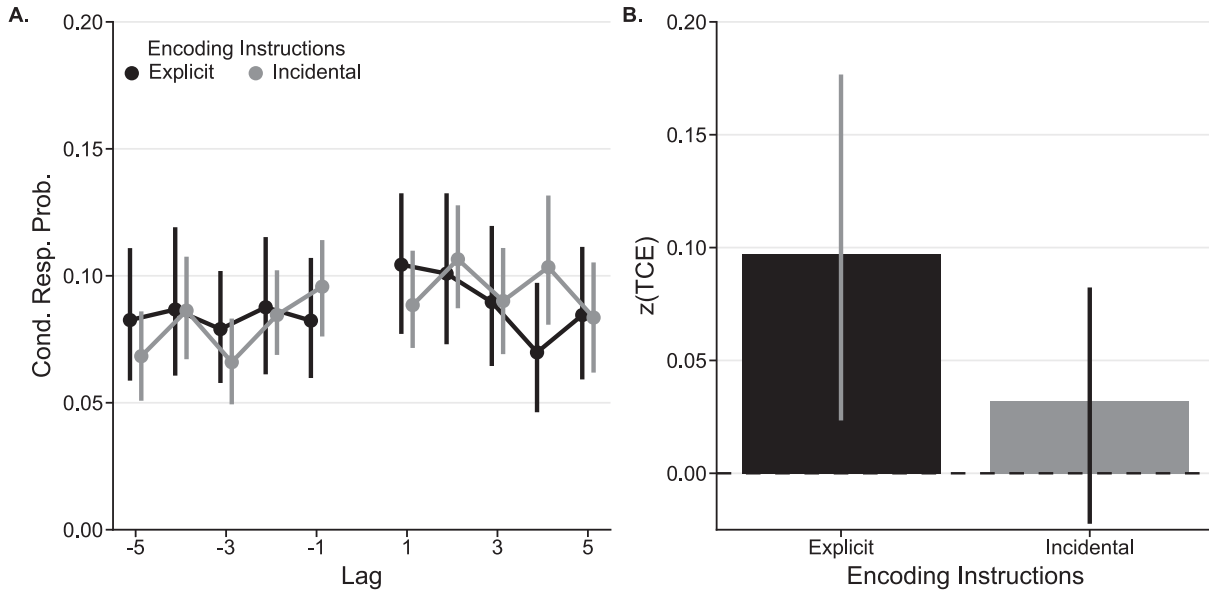


Figure 5. (A) Serial position curves and (B) probability of first recall curves on the first list under Explicit versus Incidental encoding using the Front Door Task (Experiment 2). All error bars are bootstrapped 95% confidence intervals.

Experiments 1 and 2 show that when intent to encode is absent, the TCE can be reduced to the point of being statistically indistinguishable from zero. This result is consistent with theories that ascribe the TCE to strategic control processes. Under this interpretation, the contiguity-generating processes are more or less inseparable from the intent to encode. But is intent to encode truly necessary to find a TCE? Perhaps not.

Studies showing accurate memory for serial order after incidental learning strongly suggest that subjects have access to the temporal information required to produce a TCE (Nairne, 1990, 1991). Why do they fail to use it during free recall? In the final two experiments, I consider two explanations. Experiment 3 tests the possibility that the details of the incidental encoding task determines whether or not temporal contiguity



*Figure 6.* The temporal contiguity effect (TCE) on the first list under explicit versus incidental encoding using the Front Door Task (Experiment 2). (A) Lag-conditional response probability functions. (B) The average  $z(\text{TCE})$ .  $z(\text{TCE})$  for a given subject is computed as follows: An observed temporal factor score was computed as the average percentile ranking of the temporal lag of each actual transition in the recall sequence with respect to the lags of all transitions that were possible at that time. To determine the temporal factor score expected by chance, a permutation distribution was created by randomly shuffling the order of recalls within the sequence 10,000 times and computing a temporal factor score for each shuffling. The reported value,  $z(\text{TCE})$ , is the z-score of the observed temporal factor score within the permutation distribution. All Error bars are bootstrapped 95% confidence intervals.

influences memory search. Experiment 4 tests the possibility that temporal information is encoded during incidental learning, but subjects do not automatically use it during memory search unless the memory test requires it.

### Experiment 3

Experiment 3 focuses on the possibility that temporal information is encoded automatically, but its influence is obscured by processes required by the judgment task. For example, many models produce a TCE because the representations of items studied close together are more similar to each other than they are to representations of items studied far apart. The Shoebox and Front Door Tasks encourage subjects to maintain a common

mental representation (e.g., image of a shoebox) throughout the list presentation. If this representation is incorporated into the representations of list items, it would increase the similarity of items separated by distant lags, attenuating the TCE. When effortfully memorizing, subjects likely process items in ways that are not necessary for the judgment task, perhaps decreasing the similarity of items separated by distant lags, thus increasing the TCE. That is, the judgment task might decrease the TCE in a way that is not due to the lack of intent to encode.

More generally, if intentional control processes are required to produce contiguity, it should be challenging, perhaps impossible, to observe a TCE under incidental encoding. But if control processes simply modulate the effect of automatic contiguity-generating processes (sometimes attenuating the TCE, sometimes accentuating it), it should be easy to find incidental encoding tasks that produce a TCE. Experiment 3 tests these predictions by examining five different encoding tasks.

## Method

The methods were identical to those used in Experiment 2 except for the judgment task instructions (see Table 1 for sample size information).

The question is no longer whether explicit encoding produces a larger TCE than incidental encoding, but rather whether the TCE can ever be observed under incidental encoding. Thus, in Experiment 3, all subjects were given incidental encoding instructions, but were randomly assigned to one of five different judgment tasks that varied in the type of processing required. Otherwise, the methods were identical to those used in Experiments 1 and 2 (see Table 1 for sample size information).

### Processing Task Manipulation

In all conditions, subjects were asked to make a judgment about each word as it was presented. Here, I describe the type of processing that each task was intended to discourage (or encourage). Again, to allow for the same yes/no response for each task, subjects were

asked to indicate if the judgment was easy to make under the guise of norming the items for a later study. See the Supplemental Materials for the exact task instructions.

**Weight Task.** The Weight Task was similar to the size judgment tasks used in the first two experiments except that it asked subjects to compare each item's *weight* to a common referent: a bottle of water. Specifically, they were asked to judge whether each word referred to an object that was heavier than “a standard bottle of water you’d purchase from a vending machine”. Because weight is not an easily visualizable attribute, the Weight Task might be expected to reduce the likelihood that subjects will maintain the same vivid mental image throughout the list. Thus, it may produce a larger TCE if associating each item with a common mental image tends to attenuate the TCE.

**Animacy Task.** The Animacy Task asks subjects whether each item refers to an object that is living or non-living. Like the Shoebox, Front Door, and Weight Tasks, the Animacy Task requires subjects to consider only a single attribute of each item (i.e., animacy status). But unlike the aforementioned tasks, it does not provide a reference object against which to compare each item. Thus, it further reduces the likelihood of maintaining a single vivid image throughout the list.

**Moving Scenario Task.** The Moving Scenario Task asks subjects to judge the relevance of each word to a scenario: moving to a foreign land (Nairne et al., 2017). Subjects are likely to maintain some representation of this scenario across items, but because it does not specify any pre-existing dimension, like size or weight, each item may be expected to activate many different attributes, lowering the similarity of mental representations from item to item.

**Movie Task.** The instructions for the Movie Task explain that “when you read a word, it can trigger many different thoughts” and gives the example of the word *baseball* triggering a series of thoughts: “you might have a mental image of a baseball, you might hear the crack of a bat hitting a ball, you might think of related concepts like ballpark, players, and fans...”. It then asks subjects to allow each item “to activate as many different



thoughts as possible. Then use these thoughts to generate a mental movie (like a detailed image of spending an afternoon at a baseball game or what it is like to be a player on a baseball field).” Subjects then judge whether or not it was easy to form such a mental movie. This task removes the requirement to consider each item along the same dimensions and instead encourages subjects to think deeply about the unique attributes of each item, which might be expected to cause very different mental representations to be activated with each successive item, perhaps increasing the TCE (for a different perspective on the influence of item specific processing on the TCE, see McDaniel, Cahill, Bugg, & Meadow, 2011).

**Relational Task.** The Relational Task is similar to the Movie Task, including the "baseball" example, except instead of being asked to make a new mental movie for each item, subjects are asked to “try to incorporate each new word into your existing mental movie. For example, if the next word was “owner”, you should allow it to activate many associated thoughts and then incorporate it into your existing “ballpark” movie.” This condition, which is similar to deep encoding strategies free recall subjects often adopt spontaneously (Delaney & Knowles, 2005), encourages subjects to notice semantic associations between temporally proximate items. As such, it is much like the “reminding” process Hintzman (2016) suggested contributes to the TCE (for a similar manipulation see Bower & Clark, 1969). Thus, this condition should maximize the chance of observing a TCE.

## Results and Discussion

Table 1 shows overall recall accuracy and Figure 7 shows SPCs and PFRs for the first list. The most notable difference among the groups was that, whereas the first four tasks produced SPCs and PFRs similar to those of the incidental encoding conditions of the first two experiments, the Relational Task produced a SPC and a PFR that more closely resembles those seen in multi-trial delayed recall. This suggests that the Relational Task

successfully mimicked some features of intentional encoding.

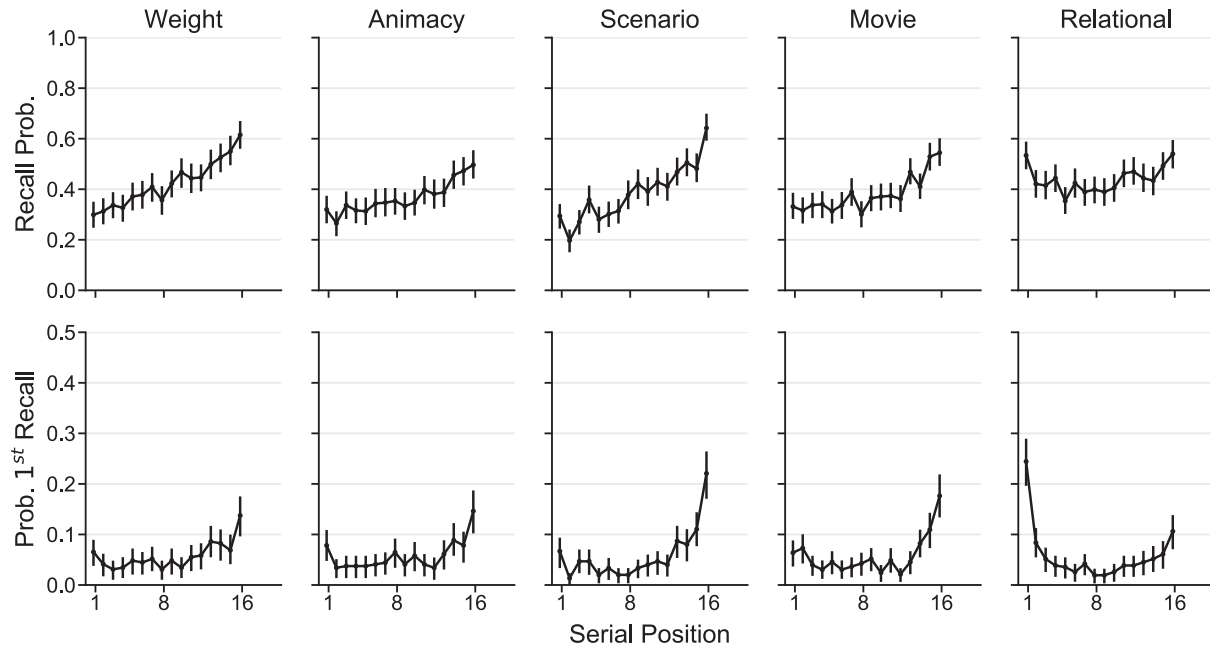


Figure 7. (Top row) Serial position curves and (Bottom row) probability of first recall curves on the first list under incidental encoding with different judgment tasks (Experiment 3). All error bars are bootstrapped 95% confidence intervals.

Critically, all of the processing tasks produced a TCE under incidental encoding conditions (Figure 8). For each task, the lag-CRP tends to decrease with increasing  $|lag|$  and the  $z(TCE)$  is significantly above zero. These results show that although the TCE can be attenuated under incidental conditions (as in Experiments 1 and 2), the lack of intent to encode, per se, does not eliminate contiguity.

Indeed, perhaps the most remarkable feature of the data is how little the size of the TCE differs among the tasks, consistent with the suggestion that the TCE is due to automatic encoding processes. The only condition for which the  $z(TCE)$  differed significantly from any other condition was the Relational Task condition, which asked subjects to integrate each item into an ongoing movie.<sup>2</sup> This suggests that encouraging subjects to notice semantic similarities among items does indeed enhance temporal

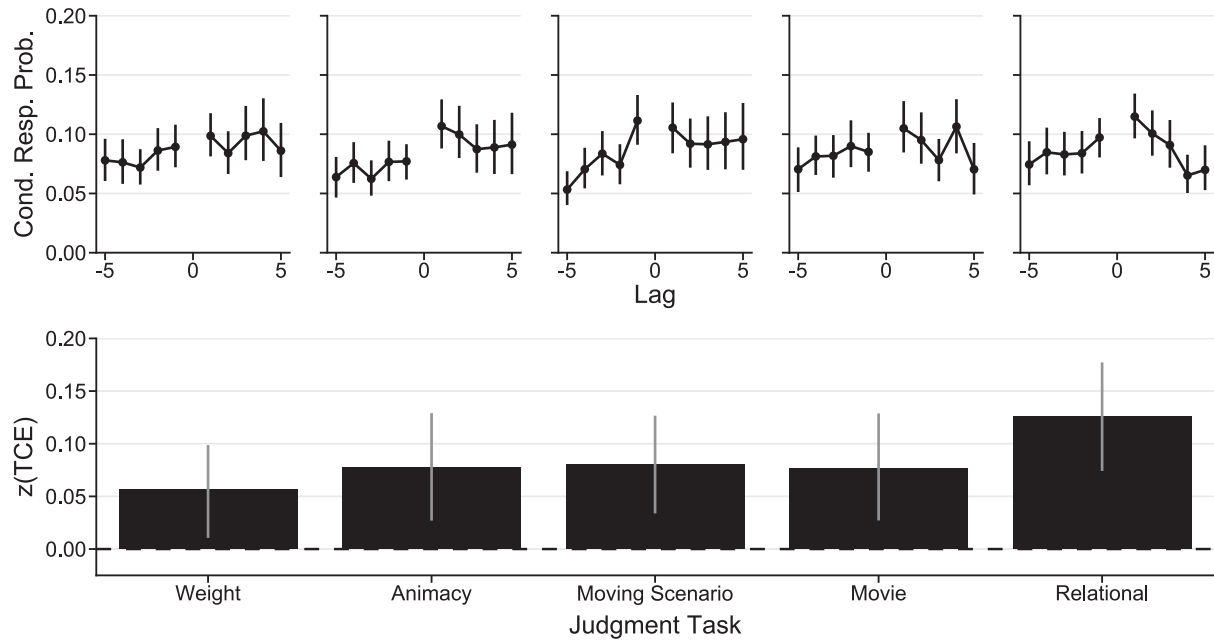
<sup>2</sup>The Relational Task also increased recall relative to most of the other tasks (see Table 1), which replicates Bower and Clark (1969)

contiguity (Hintzman, 2016), at least under incidental encoding conditions.

It should be noted that although Experiments 1 and 2 conceptually replicated Nairne et al.'s (2017) finding of no contiguity under incidental encoding using different encoding tasks, Experiment 3 failed to replicate the finding using an encoding task (Moving Scenario Task) almost identical to Nairne et al.'s. Nonetheless, the message across the present three experiments is consistent with Nairne et al.'s findings: incidental encoding *can* reduce the size of the effect relative to explicit encoding without substantially reducing recall accuracy.

The failure to exactly replicate the Nairne et al. (2017) null finding may be due to a larger sample ( $n = 299$  here versus  $N = 80$  in E1 and  $N = 80$  in E2 of Nairne et al.) providing more power to detect a small contiguity effect. But this raises the possibility that the current Experiments 1 and 2 were underpowered as well. Although sample sizes were selected to provide enough power to detect effects considerably smaller than those typically reported, as discussed above, the observed effects were even smaller than expected. Indeed comparing the significant  $z(\text{TCE})$  scores in Figure 8 with the non-significant ones in Figures 1 and 6, shows that they all of similar magnitude—of the 7 incidental encoding conditions reported across the three figures, only the Relational Task condition is significantly different from any other. Excluding the Relational Task, the average effect size of the  $z(\text{TCE})$  versus zero measured by Cohen's  $d$  was 0.16 in Experiment 3. Achieving 95% power to detect this effect would require 510 subjects, considerably more than the sample sizes in Experiments 1 and 2.

Thus, there are (at least) two mutually exclusive interpretations of the findings of Experiment 1–3. First, it could be that only encoding tasks that use a common referent tend to suppress the TCE. Second, it could be the case that a small TCE of approximately equal magnitude is present in all three experiments but they were not sufficiently powered to detect it reliably. Experiment 4 attempts to distinguish between these interpretations.



*Figure 8.* The temporal contiguity effect (TCE) on the first list under incidental encoding with different judgment tasks (Experiment 3). (Top) Lag-conditional response probability functions. (Bottom) The average  $z(\text{TCE})$ .  $z(\text{TCE})$  for a given subject is computed as follows: An observed temporal factor score was computed as the average percentile ranking the temporal lag of each actual transition in the recall sequence with respect to the lags of all transitions that were possible at that time. To determine the temporal factor score expected by chance, a permutation distribution was created by randomly shuffling the order of recalls within the sequence 10,000 times and computing a temporal factor score for each shuffling. The reported value,  $z(\text{TCE})$ , is z-score of the observed temporal factor score within the permutation distribution. All error bars are bootstrapped 95% confidence intervals.

## Experiment 4

Experiment 4 includes three conditions. The first condition is a conceptual replication of the incidental encoding task of Experiment 1 (judge each item’s size against the common referent of a shoebox), but with a larger sample size to determine if a residual TCE remains.

The remaining two conditions were designed to test two different hypotheses. The first hypothesis was that the TCE is reduced by incidental encoding tasks that require judging each list item against a common referent. This hypothesis was tested by replacing

the common referent of “shoebox” with a unique size referent for each list item (e.g., is Item 1 larger than a golf ball; is Item 2 larger than a penny; is Item 3 larger than a piano, etc.). The second hypothesis was that temporal associations generally *are* formed under incidental encoding but that subjects do not spontaneously adopt a retrieval strategy that makes use of them. This hypothesis was tested by using the same shoebox size judgment encoding task as the first condition, but replacing the surprise free recall test with a surprise serial recall test.

## Method

### subjects

For this experiment, subjects studied and recalled only a single list, which provided extra time for a short demographic questionnaire at the end of the study. To achieve 95% power, a target sample size of at least 500 subjects per condition was set. A total of 1591 individuals participated (see Table 1 for sample sizes and awareness rates by condition), of these there were 658 males, 919 females, 7 transgender individuals, and 7 individuals who preferred not to answer. There were 1547 native English speakers, 38 non-native English speakers, and 6 individuals who preferred not to answer. The mean age was 36.59 ( $SD = 11.82$ ). The highest level of education attained was less than high school for 7 individuals, high school for 398, an associates degree for 294, a bachelors degree for 632, a masters degree for 201, an advanced degree (e.g., PhD, MD, JD) for 50, and 9 preferred not to answer.

subjects were assigned to one of three conditions. All conditions had incidental encoding but varied in the judgment task and how subjects were asked to recall the words during the surprise memory test.

**Varying Size–Free Recall.** The Varying Size Task was identical to the Shoebox Task from Experiment 1 with three differences. First, instead of judging every word against a common referent (a shoebox), a different referent was randomly selected for each word

(see the Supplemental Materials for a full list of referents). Second, instead of deciding if the presented word would “fit in” a shoebox, subjects were asked if the word was “larger than” the referent to allow for cases where the referent was not a container (e.g., AFRICA: “Is it easy to judge if it is larger than a golf ball”). Third, each word was presented for 5 seconds (as opposed to the 4 second presentation rate of all other conditions) because changing the referent from item-to-item made the task more complex. Like all previous conditions, the list was followed by a 16sec math distractor and a surprise free recall test.

**Constant Size–Free Recall.** The Constant Size Task was identical to the Varying Size Task (including the “larger than” wording of the judgment task and the 5 second presentation rate) with the exception that the size referent was a shoebox for all items. The list was followed by a math distractor and a surprise free recall test.

**Constant Size–Serial Recall.** This condition was identical to the Constant Size–Free Recall except that the list was followed by a surprise serial recall task. After the math distractor, subjects were given same recall instructions used in the other conditions with one additional sentence: “Try to recall the words in the “**same order you saw them**” (including the bold emphasis).

## Results and Discussion

The Varying Size–Free Recall condition produced lower overall recall than either of the other two conditions (Figure 9A; Table 1), which is not surprising given the potential for interference between the list items and the varying size referents. subjects in the Constant Size–Serial Recall condition produced an intermediate level of recall, and not surprisingly, showed a higher probability of initiating recall from the first item in the list (Figure 9B).

Figure 10 reveals that all three conditions showed a significant TCE. Moreover, the level of the TCE did not differ significantly across conditions (The 95% bootstrapped confidence interval on difference in  $z(\text{TCE})$  between Varying–Free and Constant–Free was

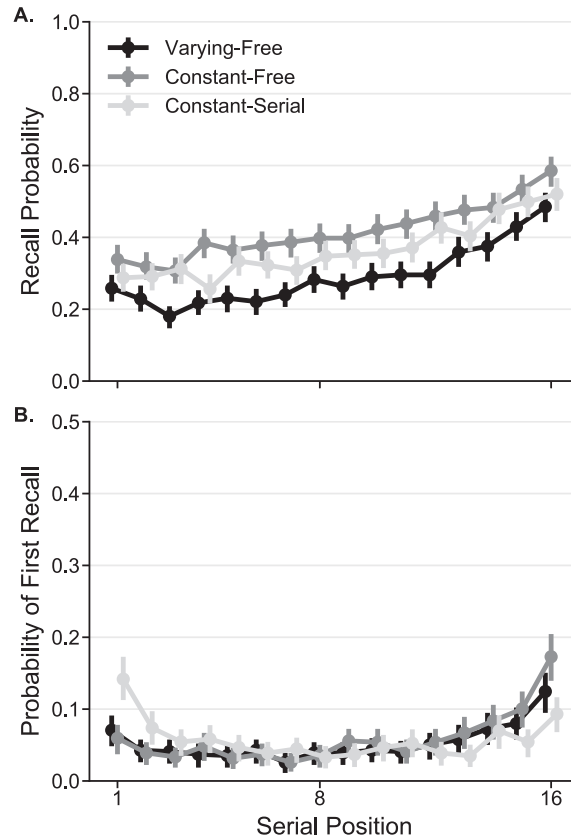


Figure 9. (A) Serial position curves and (B) probability of first recall curves on the first list as a function of encoding task variability and recall instructions (Experiment 4). All error bars are bootstrapped 95% confidence intervals.

$-0.06 - +0.07$ , between Varying-Free and Constant-Serial it was  $-0.11 - +0.02$ , and between Constant-Free and Constant Serial it was  $-0.09 - +0.01$  ). These results point to three important conclusions.

First, subjects in the Constant Size-Free Recall condition, which was almost identical to the Incidental condition in Experiment 1, encoded temporal information and spontaneously used it to guide memory search during free recall. This suggests that the failure to find a significant TCE in Experiments 1 and 2 may have been due to a lack of statistical power rather than a total absence of a TCE. To be clear, there is no doubt that incidental encoding can reduce the TCE to near zero, which is a very important finding. But, as discussed in detail below, the presence of a small residual TCE is also of

considerable theoretical consequence.

Second, changing the size referent of the judgment task from item to item had no effect on the size of the TCE. This suggests that the reduced TCE under incidental encoding is not the result of increasing the effective similarity among list items by linking each item to a common context. This conclusion should be tempered, however, by the observation that changing the referent for each item introduces the need for source monitoring (distinguishing list items from size referents during recall), which could potentially impact the TCE.

Third, explicitly asking subjects to recall the words in serial order, rather than any order they want, did not significantly increase the TCE. This suggests that the reduced TCE after incidental encoding is not entirely due to a failure to adopt a strategy of using temporal information to guide memory search. These data are consistent with the claim that incidental encoding reduces the amount of temporal information available to subjects. Although future work using a different test of order memory might reveal additional latent knowledge of temporal associations. For example, one could encourage subjects to use the order of items to aid recall without asking them to exactly reproduce the serial order.

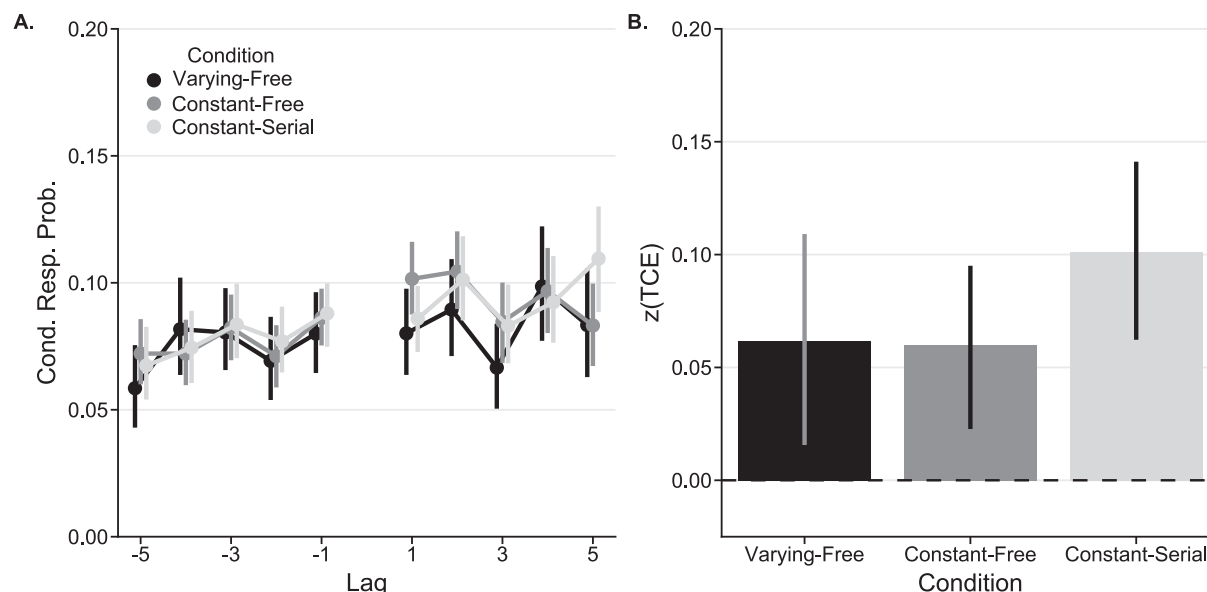
## General Discussion

The order in which events are experienced has a powerful influence on the order in which those events are recalled. The current results place an important caveat on this general observation: when the events are experienced without expectation of a memory test, the influence of order of experience on order of recall can be dramatically reduced.

### Zero or Near Zero?

The findings of these four experiments adjudicate between two theories of the TCE: Does the TCE depend on control processes that implement task-specific strategies during deliberate encoding (Hintzman, 2016)? Or does the TCE depend on automatic task-general processes that operate even when we form new memories in the absence of





*Figure 10.* The Temporal Contiguity Effect (TCE) on the first list as a function of encoding task variability and recall instructions (Experiment 4). (A) Lag-conditional response probability functions. (B) The average  $z(\text{TCE})$ .  $z(\text{TCE})$  for a given subject is computed as follows: An observed temporal factor score was computed as the average percentile ranking of the temporal lag of each actual transition in the recall sequence with respect to the lags of all transitions that were possible at that time. To determine the temporal factor score expected by chance, a permutation distribution was created by randomly shuffling the order of recalls within the sequence 10,000 times and computing a temporal factor score for each shuffling. The reported value,  $z(\text{TCE})$ , is the z-score of the observed temporal factor score within the permutation distribution. All Error bars are bootstrapped 95% confidence intervals.

deliberate study (Healey & Kahana, 2018)? The former possibility suggests the TCE should be easily eliminated by removing the impetus to engage controlled processes. The latter suggests the the TCE should be observable under most encoding circumstances. The data tell us that neither view is totally correct.

Taken as a whole, the results show that the TCE is not reliably eliminated by removing the intent to encode. A meta-analysis of the various incidental encoding conditions across the four experiments, presented in Figure 11, reveals that the average effect size is significantly above zero. To be conservative, this analysis excludes the Relational condition from Experiment 3 (because it significantly increased the TCE) and the Constant-Serial condition from Experiment 4 (because its serial recall test directly

encouraged subjects to use temporal information). Including these two conditions would result in a larger average effect size. Among the included conditions, none had an effect size that was significantly different from any other. In other words, although individual conditions may not have sufficient power to detect an effect, across conditions there is very strong evidence for a small but robust incidental TCE. This is a case where a small effect is an important effect. It provides an existence proof for temporal contiguity under incidental encoding and rules out the possibility that the TCE is an artifact of task-specific encoding strategies.

### **Implications for Theories of Episodic Memory**

Although these results strongly suggest that the TCE arises from automatic encoding processes, they just as clearly suggest that we do not yet adequately understand these processes.

For example, under the Retrieved Context family of models (e.g., Healey & Kahana, 2016; Lohnas et al., 2015; Polyn et al., 2009), which arguably provide the most comprehensive theory of TCE, the mechanisms that produce contiguity are inexorably linked to those that encode memories, making it difficult to decouple accurate recall from a large TCE. These models assume that episodic memory for an event is formed by associating list items with the current state of a context representation that changes across time. This same context representation is then used as a retrieval cue which naturally encourages memory search to focus on events that occurred near each other in time. Because a small residual TCE remained, the current data do not falsify these models, but they do set a serious challenge for modelers: How can these models account for near zero levels of contiguity in the face high recall accuracy? In other words, how can context provide an effective retrieval cue for some items, and thus produce high levels of recall, without simultaneously providing an effective cue for temporally adjacent items?

This issue is illustrated in Figure 12 which plots level of recall versus the level of the

TCE across all 12 conditions presented in this manuscript. The correlation is quite small and non-significant, suggesting that recall success may depend less on temporal contiguity than has been suggested by analyses of individual differences in free recall of explicitly encoded lists (Healey et al., 2014; Sederberg et al., 2010). Indeed, some of the conditions with the highest recall levels had the lowest levels of contiguity. Again, this finding does not falsify the models, but it does challenge them.

The precise nature of the relationship between time and context drift in these models points to one possible answer to this challenge. Although the change in context from time  $t$  to time  $t + lag$  is correlated with the passage of time, it is not *driven* by time in most Retrieved Context models (but see Howard et al., 2015, for a model in which drift is driven by time). Instead, context drift is driven by the cognitive representations activated by external and internal events. As a result, the similarity of the context representation,  $\mathbf{c}_t$ , at time  $t$  to the context representation at some other time,  $\mathbf{c}_{t+lag}$ , is partly a function of the similarity of the cognitive representations activated by the events that intervene between  $t$  and  $t + lag$ .

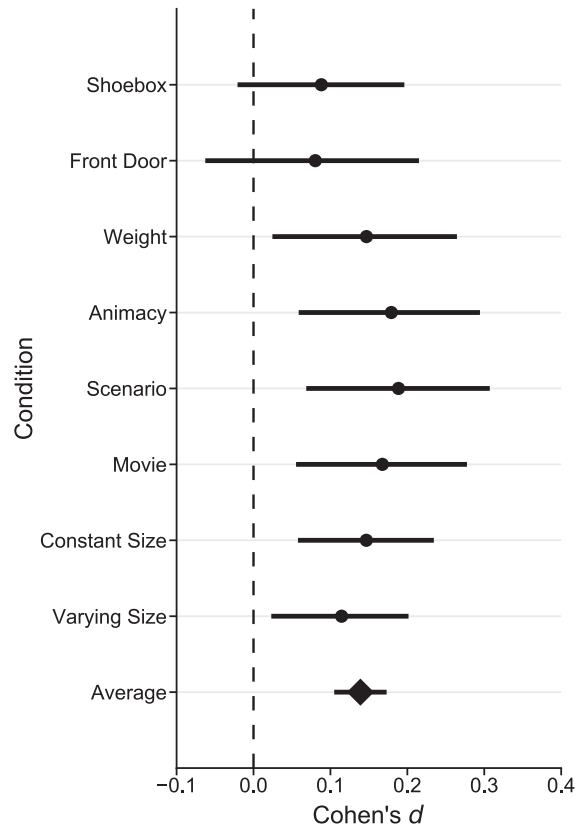
Under some incidental encoding tasks, the cognitive representations activated by successive items are likely to be similar, causing context to drift slowly and resulting in a shallow, perhaps near zero, contiguity effect. By contrast, if subjects are intending to encode items for a memory test, they are likely to engage in elaborative processing which might cause context to drift rapidly and produce a steeper contiguity effect. Experiment 4 of the current manuscript attempted to manipulate this drift rate experimentally and found no evidence that it increased the TCE, though other methods of manipulating context drift (e.g., Polyn, McCluey, & Burke, Submitted) might yield different results. A critical question for modelers then, is whether existing models can use differential drift rates to capture the large difference in TCE between Incidental and Explicit conditions while simultaneously capturing the near equal levels of recall.

But even if existing models can fit the difference between explicit and incidental

encoding by fine-tuning model parameters, we are still left with a large gap in our understanding of how encoding processes influence contiguity. How does the memory system accomplish this fine-tuning over the course of a single list without the benefit of a modeler’s fitting algorithms? Some control process must rapidly translate task instructions into an ad hoc parameterization of task-general memory mechanisms that is tailored to the task demands (Atkinson & Shiffrin, 1968; Healey et al., 2014). There have been few attempts to model how automatic memory processes interact with controlled processes to meet task demands (Lehman & Malmberg, 2013; Polyn et al., 2009). The challenge is avoiding adding a homunculus to the models that does the hard work of translating task instructions into encoding processes.

## Conclusion

In summary, these results show that control processes are not necessary to produce a TCE, but that they can powerfully influence the size of the effect and nearly eliminate it. Thus, the results point to serious limitations in existing theories of the TCE—we understand much about how memory encoding processes produce temporal contiguity, but we understand little about how these processes are controlled.



*Figure 11.* Forest plot of the effect size, Cohen's  $d$ , of the TCE on the first list in eight different conditions in which incidental encoding was followed by a free recall task along with the average effect size across the conditions. The Relational condition from Experiment 3 was excluded because it significantly increased the size of the TCE and its inclusion would inflate the estimate of the average effect size. The Constant-Serial condition from Experiment 4 was excluded because it used a serial recall test rather than a free recall test. Error bars on the individual conditions are bootstrapped 95% confidence intervals; the error bar on the average is the 95% confidence interval computed by taking the mean of the 8 condition means, calculating the standard error of that mean of means and multiplying it by  $t_{8-1} = 2.365$ .

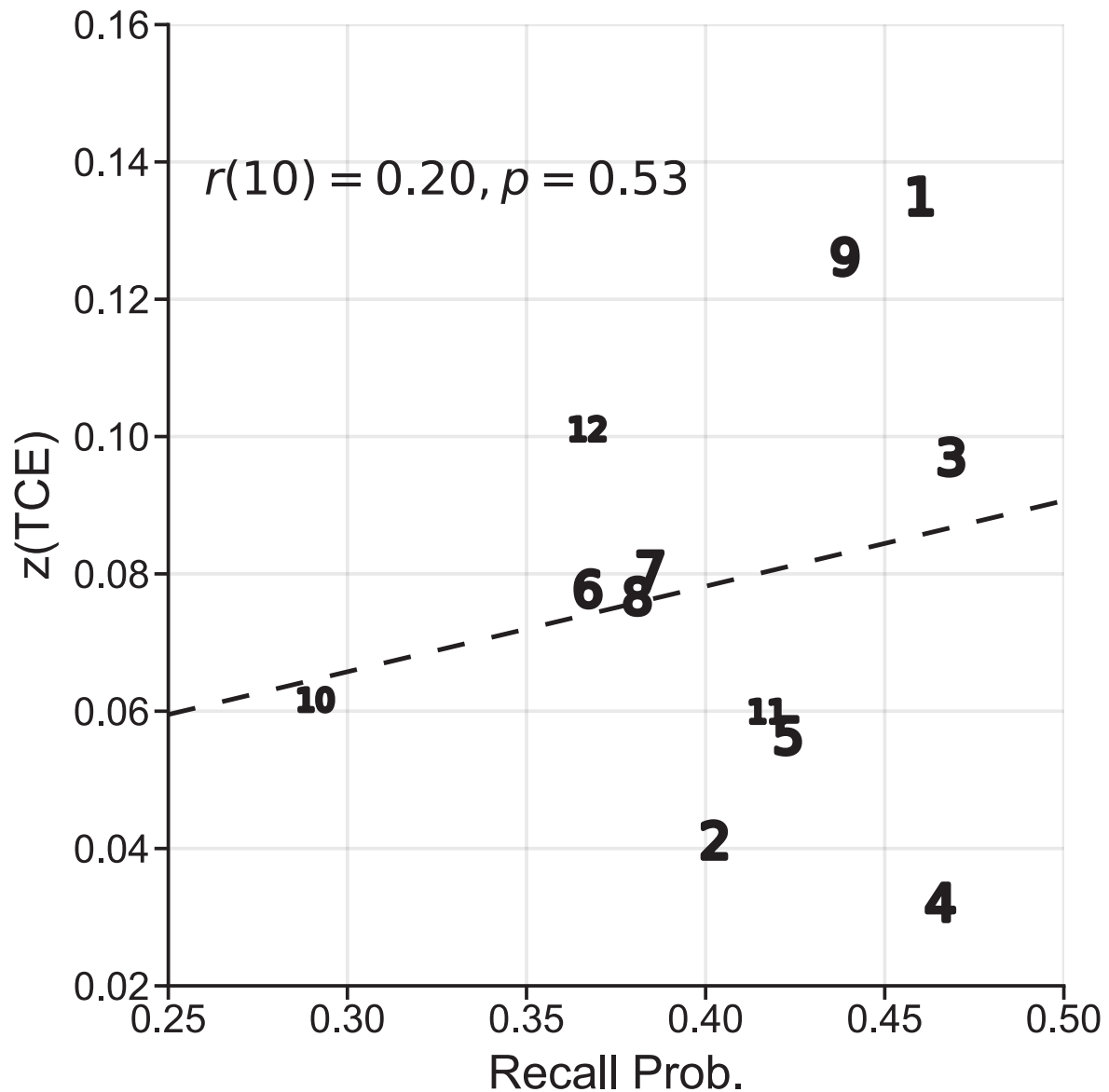


Figure 12. The Temporal Contiguity Effect on the first list versus recall accuracy on that list. Each numbered point represents the average of one of the 12 conditions reported across the four experiments: 1 = E1 Explicit; 2 = E1 Incidental; 3 = E2 Explicit; 4 = E2 Incidental; 5 = E3 Weight; 6 = E3 Animacy; 7 = E3 Moving Scenario; 8 = E3 Movie; 9 = E3 Relational; 10 = E4 Varying-Free; 11 = E4 Constant-Free; 12 = E4 Constant-Serial.

## References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, p. 89-105). New York: Academic Press.
- Bower, G. H., & Clark, M. C. (1969). Narrative stories as mediators for serial learning. *Psychonomic Science*, 14(4), 181–182.
- Bradley, M. M., & Glenberg, A. M. (1983). Strengthening associations: Duration, attention, or relations? *Journal of Verbal Learning and Verbal Behavior*, 22(6), 650–666.
- Brodie, D. A., & Murdock, B. B. (1977). Effects of presentation time on nominal and functional serial position curves in free recall. *Journal of Verbal Learning and Verbal Behavior*, 16, 185-200. doi: 10.1016/S0022-5371(77)80046-7
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3), 539–576.
- Burns, D. J. (1996). The item-order distinction and the generation effect: The importance of order information in long-term memory. *The American Journal of Psychology*, 109(4), 567.
- Davis, O. C., Geller, A. S., Rizzuto, D. S., & Kahana, M. J. (2008). Temporal associative processes revealed by intrusions in paired-associate recall. *Psychonomic Bulletin & Review*, 15(1), 64-69.
- Delaney, P. F., & Knowles, M. E. (2005, Jan). Encoding strategy changes and spacing effects in the free recall of unmixed lists. *Journal of Memory and Language*, 52(1), 120-130.
- Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology*, 68(1), 58.
- Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity.

- Neuron*, 81(5), 1179–1189.
- Glenberg, A. M., & Bradley, M. M. (1979). Mental contiguity. *Journal of Experimental Psychology: Human Learning and Memory*, 5(2), 88.
- Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., & Gretz, A. L. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 355–369.
- Healey, M. K., Crutchley, P., & Kahana, M. J. (2014). Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology: General*, 143(4), 1553–1569. doi: 10.1037/a0036306
- Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, 123(1), 23–69. doi: 10.1037/rev0000015
- Healey, M. K., & Kahana, M. J. (2018). Contiguity in episodic memory. *Submitted*.
- Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 13(6), 644 – 655. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0022537174800526> doi: [https://doi.org/10.1016/S0022-5371\(74\)80052-6](https://doi.org/10.1016/S0022-5371(74)80052-6)
- Hintzman, D. L. (2016). Is memory organized by temporal contiguity? *Memory & Cognition*, 44, 365–375.
- Hintzman, D. L., & Block, R. A. (1973). Memory for the spacing of repetitions. *Journal of Experimental Psychology*, 99, 70–74.
- Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). Spacing judgments as an index of study-phase retrieval. *Journal of Experimental Psychology: Human Learning and Memory*, 104(1), 31–40.
- Hogan, R. M. (1975). Interitem encoding and directed search in free recall. *Memory & Cognition*, 3, 197–209.
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and*



- Cognition*, 25, 923–941. doi: 10.1037/0278-7393.25.4.923
- 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. doi: 10.1037/a0037840
- Jarrold, C., Hall, D., Harvey, C. E., Tam, H., Towse, J. N., & Zarandi, A. L. (2015). What can we learn about immediate memory from the development of children’s free recall? *The Quarterly Journal of Experimental Psychology*, 68(9), 1871–1894.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24(1), 103–109. doi: 10.3758/BF03197276
- Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation*, 61, 237–284.
- Kliegl, O., & Bäuml, K.-H. T. (2016). Retrieval practice can insulate items against intralist interference: Evidence from the list-length effect, output interference, and retrieval-induced forgetting. *Journal of Experimental Psychology Learning Memory and Cognition*, 42, 202–214.
- Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, 120(1), 155–189. doi: 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.
- Marshall, P. H., & Werder, P. R. (1972). The effects of the elimination of rehearsal on primacy and recency. *Journal of Verbal Learning and Verbal Behavior*, 11, 649–653.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods*, 44(1), 1–23.
- 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.
- McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, 39, 352-70.  
doi: 10.1037/h0069819
- Murdock, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Lawrence Erlbaum and Associates.
- Nairne, J. S. (1990). Similarity and long-term memory for order. *Journal of Memory and Language*, 29, 733-746.
- Nairne, J. S. (1991). Positional uncertainty in long-term memory. *Memory & Cognition*, 19(4), 332–340.
- Nairne, J. S., Cogdill, M., & Lehman, M. (2017). Adaptive memory: Temporal, semantic, and rating-based clustering following survival processing. *Journal of Memory and Language*, 93, 304–314.
- Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. (2007). Adaptive memory: survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(2), 263.
- Neath, I. (1993). Contextual and distinctive processes and the serial position function. *Journal of Memory and Language*, 32, 820-840.
- Polyn, S. M., Erlikhman, 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.
- Polyn, S. M., McCluey, J. D., & Burke, J. F. (Submitted). Altering the flow of mental time. *Submitted*.
- 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, 129-156. doi: 10.1037/a0014420
- Postman, L. (1971). Organization and interference. *Psychological Review*, 78(4), 290.
- Postman, L. (1972). The experimental analysis of verbal learning and memory: Evolution

- and innovation. In C. P. Duncan, L. Sechrest, & A. W. Melton (Eds.), *Human memory: Festschrift in honor of Benton J. Underwood*. New York: Appleton Century Crofts.
- Rundus, D. (1971). An analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 89, 63-77. doi: 10.1037/h0031185
- Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2013). List-method directed forgetting in cognitive and clinical research: a theoretical and methodological review. *Psychology of Learning and Motivation*, 59, 131 - 189.
- Sahakyan, L., & Smith, J. R. (2014). A long time ago, in a context far, far away: Retrospective time estimates and internal context change. *Journal Experimental Psychology: Learning, Memory and Cognition*, 40(1), 86-93. doi: 10.1037/a0034250
- Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory. *Psychological Science*, 16, 898-904. doi: 10.1111/j.1467-9280.2005.01634.x
- Sederberg, P. B., Miller, J. F., Howard, W. H., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory & Cognition*, 38(6), 689-699. doi: 10.3758/MC.38.6.689
- Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition*, 21(1), 34-40.
- Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal-contextual retrieval from episodic memory. *Journal Experimental Psychology: Learning, Memory and Cognition*, 37(6), 1532-1539.
- Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1589-1626. doi: 10.1037/0278-7393.26.6.1589
- Turk-Browne, N. B., Simon, M. G., & Sederberg, P. B. (2012). Scene representations in parahippocampal cortex depend on temporal context. *Journal of Neuroscience*,

32(21), 7202 - 7207. doi: 10.1523/JNEUROSCI.0942-12.2012

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, 50–61.

Wahlheim, C. N., & Huff, M. J. (2015). Age differences in the focus of retrieval: Evidence from dual-list free recall. *Psychology and Aging*, 30(4), 768.