Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

2021, Vol. 47, No. 8, 1264–1287
https://doi.org/10.1037/xlm0001001

A Test of Retrieved Context Theory: Dynamics of Recall After Incidental Encoding

Abigail M. D. Mundorf, Linh T. T. Lazarus, Mitchell G. Uitvlugt, and M. Karl Healey
Department of Psychology, Michigan State University

The temporal contiguity effect (TCE) is the tendency for the recall of one event to cue recall of other events originally experienced nearby in time. Retrieved context theory proposes that the TCE results from fundamental properties of episodic memory: binding of events to a drifting context representation during encoding and the reinstatement of those associations during recall. If these processes are automatic, the TCE should not be dependent on any encoding strategy and should, in fact, be present regardless of encoding intentionality. Here, we ask whether this theory is compatible with recent findings that the TCE is dramatically reduced under incidental encoding, even though memory accuracy is only modestly reduced. We begin by attempting to replicate this finding in a new large-scale study with over 5,000 participants in which we manipulated encoding intentionality between participants in both delayed free recall and continual distractor free recall. A small, but reliable, TCE was observed in all conditions, although the effect was dramatically reduced in incidental encoding. In a simulation study, we demonstrated that retrieved context theory can simultaneously account for both overall recall and the strength of the TCE in incidental encoding conditions. Additional analyses revealed that the incidental TCE is not an artifact of theoretically uninteresting factors, such as recency, and is consistent with being generated by the core contextual dynamics of retrieved context theory.

Keywords: episodic memory, free recall, temporal contiguity, computational model, incidental encoding

Supplemental materials: https://doi.org/10.1037/xlm0001001.supp

Retrieved context theory proposes that all episodic memories depend on a drifting mental context representation (Howard & Kahana, 2002). The theory assumes that during encoding new memories are formed by associating experienced events with the current state of this context representation. Then, during memory search, event-to-context associations allow context to serve as a cue to recall events. Retrieved context theory forms the basis of a set of models of free recall that provide quantitatively precise accounts of dynamics of memory search (see Healey & Kahana, 2014, 2016; Howard et al., 2015; Howard & Kahana, 2002; Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008). The theory has been applied to explain many phenomena, including the testing effect (Karpicke et al., 2014), directed forgetting (Sahakyan et al., 2013), retrieval-induced forgetting (Kliegl & Bäuml, 2016), age-related change (Healey & Kahana, 2020; Wahlheim & Huff, 2015), individual differences (Healey et al., 2014; Healey & Uitvlugt, 2019), amnesia (Palombo et al., 2019; Sederberg et al., 2008), and event segmentation (Ezzyat & Davachi, 2014; Sahakyan & Smith, 2014).

In addition to providing an explanation for a broad range of memory effects, retrieved context theory also makes many clear, testable predictions. Here, we test one prediction that cuts to the very heart of the theory: Episodic memories should contain specific information about the temporal order of events, even when the memories are encoded incidentally. We begin by describing the encoding and retrieval mechanisms that lead to this prediction.

Memory for Temporal Order as an Emergent Property of Contextual Dynamics

Under retrieved context theory, memory for temporal order naturally emerges from the combination of several mechanisms that operate during encoding and retrieval. We illustrate this by describing the theory's account of the free recall task (we provide a more in-depth explanation in later sections). In standard free recall, a list of items (usually words) is presented one at a time for study. At the end of the list, the participant is asked to recall as many items as possible in whatever order they come to mind.

This article was published Online First October 28, 2021.

Abigail M. D. Mundorf https://orcid.org/0000-0002-6227-2349
Linh T. T. Lazarus https://orcid.org/0000-0003-2156-0931
Mitchell G. Uitvlugt https://orcid.org/0000-0001-9169-1000
M. Karl Healey https://orcid.org/0000-0002-2816-3536

We thank Kimberly Fenn for helpful discussions. This material is based on work supported by the National Science Foundation under Grant 1848972.

The data and materials for this experiment are available on the Open Science Framework (https://osf.io/wdhn8/).

Correspondence concerning this article should be addressed to Abigail M. D. Mundorf or M. Karl Healey, Department of Psychology, Michigan State University, 316 Physics Road, East Lansing, MI 48825, United States. Email: desterab@msu.edu or khealey@msu.edu

In retrieved context theory's account of free recall, two key mechanisms operate during the encoding phase. First, each newly presented item becomes associated with the state of context that prevailed when it was presented. As a result of these associations, context can later serve as a retrieval cue. Second, each item activates its own mental representations, which are then incorporated into the existing state of context. Incorporating previous items causes context to change, or drift. As a result of this drift, the state of context that prevails when a given item is presented is most similar to states that prevailed at nearby points in time—context drifts in an autocorrelated fashion. In this way, context carries information about the temporal distance between items. Together, these new item-context associations and the mechanism of context drift allow the context associated with one item to provide a very good cue for other items that were originally studied nearby in time.

Retrieved context theory's account of retrieval is based on two additional mechanisms. First, memory search is initiated by using the current state of the context representation as a retrieval cue. Second, upon successful recall of an item, the item's associated mental context is reactivated, or reinstated. The reinstated context then helps cue another item. Because items presented closer together in time are associated with more similar contexts, the reinstated context most strongly cues items studied immediately before or after the just-recalled item. Thus, retrieved context theory naturally predicts a temporal contiguity effect (TCE): the tendency for the recall of one item to be followed by recall of another item that was originally experienced close in time to the first. Such an effect has been widely observed in laboratory tasks (for a review, see Healey et al., 2019), including in free recall (Kahana, 1996; Murdock, 1974; Postman, 1971); recognition (Averell et al., 2016; Sadeh et al., 2015; Schwartz et al., 2005; but see Bradley & Glenberg, 1983); paired associate tasks (Campbell & Hasher, 2018; Caplan et al., 2006; Davis et al., 2008; but see Osth & Fox, 2019); and even with real-world stimuli, such as autobiographical memories and news stories (Diamond & Levine, 2019; Moreton & Ward, 2010; Uitvlugt & Healey, 2019).

Notice that the encoding and retrieval mechanisms utilized in retrieved context theory lie at the heart of both recall success and the TCE: They allow for recall of individual items and simultaneously result in a tendency for items to come to mind in temporal order. Thus, retrieved context theory predicts a TCE in almost any situation where episodic memories are formed.[1] By contrast, many competing theories of episodic memory suggest a TCE should occur under only certain circumstances (Healey et al., 2019). A theory that attributes the effect to binding in short-term or working memory, for example, would predict a TCE only when items are presented close enough in time to allow temporally adjacent items to co-occupy the short-term buffer and form interitem associations (Howard & Kahana, 1999). Similarly, theories based on control processes like strategic rehearsal predict a TCE only when items are studied intentionally (Healey, 2018; Hintzman, 2016). This divergence of predictions makes the TCE an important tool for testing retrieved context theory against other theories of episodic memory.

Here, we consider one particularly diagnostic situation: incidental encoding. Retrieved context theory makes the strong prediction that under incidental encoding, successful recall of items should be accompanied by a TCE. Moreover, a formal model of the theory should be able to precisely fit the size and shape of the effect (Healey, 2018). Conversely, many other theories predict little or no TCE under incidental encoding, such as those which posit that temporal information is only encoded through intentional use of control processes (Healey et al., 2019). In a serious challenge to retrieved context theory, some work has found that when participants are not expecting a memory test, the TCE is greatly reduced even though overall recall remains high (Healey, 2018; Nairne et al., 2017). It is unclear if this pattern of substantial recalls paired with a very small TCE in incidental encoding is consistent with the prediction of retrieved context theory that these measures should be coupled—if not, the theory would be all but falsified, and the results would instead support an account which attributes the TCE entirely to strategic control processes.

We argue that for retrieved context theory to be fully compatible with the incidental TCE, three criteria must be met: (a) a TCE must be present under incidental encoding at both short and long timescales, (b) a model of retrieved context theory must be able to simultaneously fit not only the incidental TCE, but also the number of items recalled, and (c) the nature of the incidental TCE must be consistent with being generated by a context reinstatement mechanism. We elaborate on each of these criteria subsequently.

Timescale Similarity

Retrieved context theory predicts that temporal contiguity should occur across timescales (Howard, 2004); a TCE should occur, to some degree, regardless of the time passing between item presentations. The context drift mechanism entails that, in almost all situations, the similarity of the mental contexts associated with any two items is correlated with the relative temporal distance between them. The absolute time passing between item presentations should not change this. Items studied 1 s apart will be associated with more similar contexts than items studied 10 s apart. Similarly, items studied 10 s apart will be associated with more similar contexts than those studied 100 s apart. Many other memory models would not predict this timescale similarity (Healey et al., 2019). For example, theories based on binding in a short-term buffer (e.g., Raaijmakers & Shiffrin, 1981) often assume that items presented close together in time are more likely to occupy the buffer simultaneously (Phillips et al., 1967). As such, binding in a short-term buffer would produce a TCE when items are encoded 1 s apart, but not when there is a distraction-filled delay between items because a filled delay should prevent multiple items from being active in the short-term buffer at any given moment and abolish the TCE (Howard & Kahana, 1999).

Supporting retrieved context theory's prediction that the TCE should be consistent regardless of timescale, a significant TCE has been found using continual distractor free recall (CDFR), in which a distractor task of sufficient duration to fill the buffer intervenes between each item presentation (Howard & Kahana, 1999). A

[1] Under retrieved context theory, episodic memories are inherently dependent on the formation of item-context associations; thus, it predicts a TCE in almost all situations. However, there are technically some special cases where the models based on retrieved context theory can form episodic memories without producing a TCE. These include cases where context does not drift during encoding and cases where new item-context associations are formed during study but are not reinstated at recall.

TCE has been observed at longer timescales as well, even when items are separated by hours (Cortis Mack et al., 2017).

In these previous CDFR experiments, however, participants deliberately encoded stimuli. Therefore, an explanation based on binding in a short-term buffer would still be able to account for the TCE in CDFR if deliberate control processes are not fully disrupted by the distractor task. For example, a participant might engage in rehearsal during the distractor task or may intentionally bring to mind a previous item while studying a new item. Such control processes would allow items separated by long delays to still co-occupy short-term memory and form temporal bonds. In incidental encoding, when participants are unaware their memory will be tested, there is no reason to engage such deliberate control processes, and thus a short-term buffer explanation would predict a null TCE. In contrast, retrieved context theory would still predict a TCE in CDFR even under incidental encoding. A few studies have examined recall and temporal contiguity in realistic situations, where items are presumably incidentally encoded (Moreton & Ward, 2010; Uitvlugt & Healey, 2019), but these data are not well-suited for modeling. Examining recall with incidental encoding in a well-controlled CDFR task is key: Failing to find a significant TCE in an incidental CDFR task would support accounts based on binding in a short-term buffer and pose a serious challenge to retrieved context theory.

A Link Between Recall and the TCE

Another key feature of retrieved context theory is that both memory for individual items and memory for temporal order arise from the same mechanisms. As a result, the framework predicts a tight coupling between overall recall and temporal contiguity. Stronger associations between items and the prevailing mental contexts at encoding causes the recall of a given list item to generate a stronger reinstated context cue for recalling adjacent list items. This will tend to increase both recall and the TCE. Alternatively, failing to form or reinstate item-context associations should decrease both recall and the TCE (e.g., Palombo et al., 2019). Indeed, a correlation between the size of the TCE and recall levels has been reported for tasks using intentional encoding instructions (Healey et al., 2014; Sederberg et al., 2010; Spillers & Unsworth, 2011). In incidental encoding tasks, this relationship is much weaker. Whereas overall recall is slightly lower in incidental compared with intentional encoding tasks (Glenberg et al., 1980; Marshall & Werder, 1972; Nairne et al., 2017; Neath, 1993), the TCE is greatly reduced (Healey, 2018; Nairne et al., 2017). Moreover, variations in recall levels across incidental encoding tasks do not correlate with the size of the TCE (Healey, 2018). It is unclear if models of retrieved context theory can simulate this degree of decoupling between overall recall and contiguity. Thus, to fully explain the dynamics of incidental encoding, retrieved context theory must be able to simultaneously fit a pattern of substantial recall and a small TCE.

The TCE Must Be Distinct From the Recency Effect

Under retrieved context theory, two distinct factors can contribute to the TCE. The first, as we have already discussed, is context reinstatement. During encoding, autocorrelated context drift allows context to carry information about item order. During recall, the context associated with a just-recalled item is brought to mind through context reinstatement and is incorporated into the current mental context, causing context to drift. This context then serves as a cue which will tend to activate items studied nearby in time because the context carries information about temporal order.

Although context reinstatement is a key mechanism of retrieved context theory, it is not the only factor which contributes to the TCE. Contiguity can also be generated through recency. At the beginning of the recall period, the theory assumes the current state of mental context is used as a cue to begin recalling items. All else being equal, the autocorrelated drift of context during the study phase ensures that this end-of-list context is most strongly associated with end-of-list items. Thus, these end-of-list items are likely to be recalled first (as indeed they are in immediate free recall tasks; Hogan, 1975; Howard & Kahana, 1999; Laming, 1999). Because end-of-list items were studied close together in time, the bias to recall recent items first entails successively recalling several events that were experienced close in time—the recency effect naturally enhances the TCE (Davelaar et al., 2005; Kahana, 1996). Indeed, as we will see, a strong recency effect is sufficient to produce a TCE even in the absence of any other contiguity-generating mechanisms.

Thus, under retrieved context theory, the TCE observed with intentional encoding is produced by a combination of both context reinstatement and recency mechanisms. Under any straight-forward interpretation, the theory would predict that both mechanisms also operate during incidental encoding. It is possible, however, that this straight-forward interpretation is wrong and that context reinstatement makes no contribution to the TCE in incidental encoding.[2] If so, it would indicate that although both intentional and incidental encoding produce a contiguity effect, they represent very different types of contiguity. This would severely limit the explanatory scope of retrieved context theory. Therefore, we test whether the incidental TCE can be attributed to recency alone.

The Present Study

For retrieved context theory to be fully compatible with the nature of the TCE under incidental encoding, the three criteria specified in the preceding text must be met: (a) a significant TCE must be observed for incidental encoding in CDFR; (b) a model of retrieved context theory must be able to simulate temporal contiguity and overall recall in incidental encoding; and (c) the TCE in incidental encoding must be consistent with being generated not only by recency, but also by context reinstatement. To test these criteria, we collected data from participants in both intentional and incidental encoding conditions in either a delayed free recall (DFR) or a CDFR task. We tested whether a TCE was present in all conditions. Then, we evaluated the ability of a specific model of retrieved context theory to simultaneously fit overall recall and temporal contiguity. To determine if recency alone can explain the incidental TCE, we also conducted novel analyses that controlled for the effects of recency. Finally, we examined the necessity of the context drift and reinstatement mechanisms that are central to retrieved context theory.

[2] We thank Per Sederberg for suggesting this possibility.

Method

Participants and Design

All data analyzed in this report are freely available on the Open Science Framework at https://osf.io/wdhn8/. The methods closely follow those of Healey (2018) except as noted here. Based on the average effect size calculated for a variety of incidental encoding conditions in Healey (2018), achieving 95% power to detect the TCE in incidental encoding requires a sample size of at least 510 participants per condition. However, in addition to detecting the effect, another major goal of this study was to fit a computational model. Because the quality of model fits is limited by variance in the data, we sought to test as many participants as possible. Thus, we set a goal of obtaining data from at least 1,500 participants per condition for a total of 6,000 participants.

We recruited 6,641 participants using Amazon Mechanical Turk (MTurk). No participants who contributed to MTurk data collection for Healey (2018) were permitted to participate because of the similarity of these tasks. Participants were randomly assigned to one of four conditions in a 2 (encoding type) × 2 (distractor) between-participants design.

Data were excluded for one participant due to a technical error that prevented them from completing the task. Data were also excluded from participants in the incidental encoding conditions who indicated on a postexperiment questionnaire that they were aware their memory would be tested and from participants in any condition who recalled fewer than two items or more than 24 items (i.e., twice the length of the list). As a result, 304 incidental participants were excluded for reporting awareness of a memory test, 893 additional participants were excluded for recalling fewer than two correct (i.e., nonintrusion) items in a row, and three more participants were excluded for providing more than 24 responses. Excluding these participants increased overall recall for all groups, as many participants were excluded for recalling too few items (see Table 1). However, these exclusions did not alter the general pattern of results for any of our analyses.

We also considered the possibility that some intentionally encoding participants may have been cheating, since there was no way to supervise participants during the task. Less than 2% of participants correctly recalled more than 80% of list items. Of these, only one reported items in perfect serial order, as might be expected if a participant was cheating by writing down the words as they appeared.

After all exclusions, a total of 5,441 participants (81.9% of the original sample) were included in our analyses. See Table 1 for a breakdown of sample sizes and exclusions for each condition. For the final sample, the average age of participants was 35.4 years (*SD* = 11.1), 58.9% were female, 55.3% reported an education level of bachelor's degree or above, and all reported English as their first language. Participants were paid $1.00 for taking part in the study (a rate of roughly $10.00/hr). All participants completed written informed consent, and all procedures were approved by Michigan State University's Institutional Review Board.

Table 1
Sample Sizes and Exclusions for Awareness and for Number of Recalls by Condition for Intentional and Incidental Encoding With a Delayed Free Recall (DFR) or Continual Distractor Free Recall (CDFR) Task

Condition	*n* included	*n* excluded (for awareness)	*n* excluded (for recalls)
Intentional DFR	1,487		174
Incidental DFR	1,306	163	191
Intentional CDFR	1,406		253
Incidental CDFR	1,242	141	277

Procedure

All participants viewed a single list of 12 words and were asked to make an animacy decision for each word ("Is it easy to judge if this word refers to something that is alive?"), indicating their decision (yes/no) with a key press. Participants were told that their responses would be used to guide stimuli selection for a future study. The animacy judgment task served to ensure participants were attentive during the encoding period and provide a plausible reason for participants in the incidental conditions to be viewing a list of words. In the intentional conditions, participants were informed that a free recall test would follow the study period; participants in the incidental conditions were given no such warning. The full text of the instructions given to participants is provided in the online supplemental material.

For each participant, words were drawn from a pool of 1,638 positive or neutrally valanced nouns, a list developed for the Penn Electrophysiology of Encoding and Retrieval Study (PEERS; Healey & Kahana, 2014; Lohnas & Kahana, 2014; Miller et al., 2012). Words were presented for 4 s each, with a 1-s interstimulus interval between words. No instructions were given on whether participants should read the words aloud or silently.

Manipulating the presentation of distractors allows for a test of the model's prediction of approximate timescale similarity: A TCE should be present regardless of the amount of time (provided the intervening cognitive activity causes context to drift) between items. In the DFR control conditions, a 16-s distractor period followed only the final word. In the CDFR conditions, the same 16-s distractor period followed the presentation of *each* word. Each distractor period began with a 1-s blank screen after the presentation of the previous word, followed by 13 s of a math distractor task. The distractor task required solving addition problems in the form A + B + C = ?, where A, B, and C were positive, single-digit integers. Participants were instructed to try to "solve as many problems as you can without sacrificing accuracy." The task automatically advanced after the 13 s had passed. To ensure participants noticed when the distractor task ended, a red screen was presented for 2 s after the last problem. Thus, the total duration of each distractor period was 1 s + 13 s + 2 s = 16 s.

Following the end-of-list distractor task, participants in all conditions were given 75 s to recall as many words from the list as possible, in any order. Participants typed recalled words into a text box on the computer screen. Once a word had been entered, the word disappeared, and a blank text box was available for the next recall. A spell-checking algorithm (described in Healey, 2018) checked participants' typed responses for typos and scored their recall accuracy.

After the recall period, participants completed a demographic and strategy use questionnaire. Strategy data are not analyzed in this article.

Results

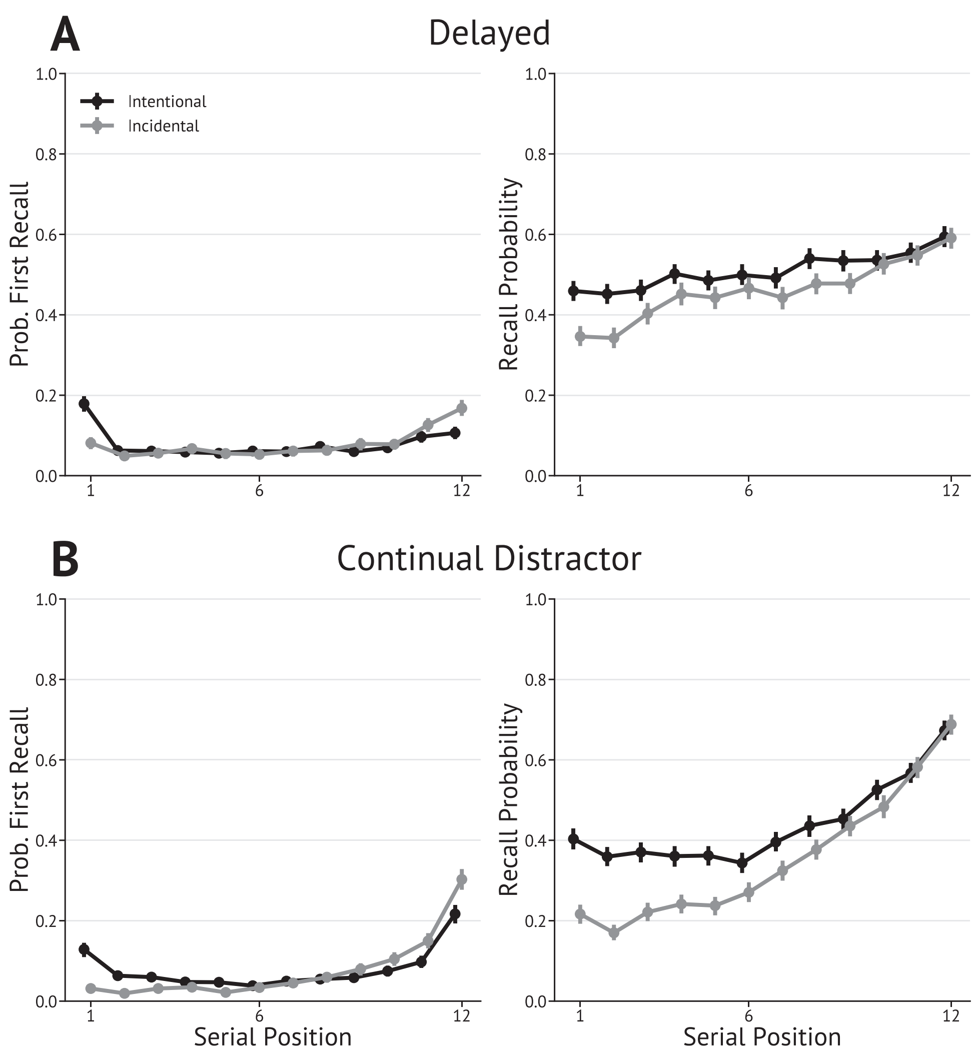
How Does Encoding Intentionality Affect Recall Initiation and Recall Accuracy?

Figure 1 displays probability of first recall (PFR) curves and serial position curves (SPCs) for intentional and incidental encoding groups for both DFR and CDFR. PFR curves display the probability that an item from a given serial position will be recalled first. The PFR curves reveal greater recency in incidental encoding and greater primacy in intentional encoding. Whereas participants in the CDFR conditions tended to initiate recall with items from the end of the list, participants in the DFR groups were more likely to initiate recall with beginning-of-list items. These patterns are similar to those seen in other free recall experiments in which participants completed a semantic orienting task for each word (Healey, 2018; Howard & Kahana, 1999).

We analyzed recall accuracy by examining SPCs, which give the probability that an item from each serial position will be recalled (see Figure 1). There is a clear recency effect for all four groups, albeit stronger in the CDFR conditions. The intentional conditions display slightly higher recall than the incidental conditions, particularly for words near the beginning of the list. In all conditions, the primacy effect is small. Although most work with DFR has found a strong primacy effect (see Bjork & Whitten, 1974; Neath, 1993; Unsworth, 2008), attenuated primacy has been reported occasionally when participants must complete a secondary task during encoding (Bhatarah et al., 2006; Healey, 2018;

Figure 1

Probability of First Recall (PFR) Curves and Serial Position Curves (SPCs) for Intentional Versus Incidental Encoding



Note. (A) Delayed free recall. (B) Continual distractor free recall. Error bars are bootstrapped 95% confidence intervals.

Howard & Kahana, 1999; Marshall & Werder, 1972; Murdock, 1965) or they are otherwise discouraged from rehearsing (Healey, 2018; Tan & Ward, 2000). Thus, in the present study, the semantic orienting task used during encoding could account for the reduced primacy in DFR.

Is There a TCE in Incidental Encoding?

To analyze the TCE, we examined lag-conditional response probabilities (lag-CRPs). The unit of analysis for the lag-CRP is a transition: recalling one list item and moving on to recall another list item. Lag-CRP gives the probability of making transitions of different temporal distances and directions within a list, measured in terms of lag between items' original positions within the study list. For example, recalling the third item in the list and transitioning to the fifth item would be $lag = +2$ transition. Recalling the third item in the list followed by the first item would be $lag = –2$ transition. For each lag, a CRP is calculated by dividing the number of times a transition of that lag was actually made by the number of times a transition of that lag was possible. All list items that had not been previously recalled by the participant are considered possible. A lag of +1, for example, would not be possible if the just-recalled item were the last item in the list. Similarly, if a participant recalls the fourth item in the list, followed by the fifth item, and then transitions to a new word, a lag of –1 would not be possible for that transition, because the item one position back (Item 4) has already been recalled. Lag-CRPs are typically highest for $|lags| = 1$ and sharply decrease for farther $|lags|$. There is also generally a bias for near forward lags, resulting in a forward asymmetry (Kahana, 1996).

In all four conditions, presented in Figure 2, the lag-CRP is highest for short lags, decreases for longer lags, and displays the forward asymmetry typically associated with a TCE. The lag-CRPs for the intentional conditions are consistent with previous findings that a nonzero TCE is present regardless of distractor task (Howard & Kahana, 1999). The lag-CRP is flatter in the incidental conditions, particularly in the forward direction, consistent with reports of a reduced TCE under incidental encoding (Healey, 2018; Nairne et al., 2017).[3] The difference in the shapes of the lag-CRPs across conditions will be discussed in further detail in a later section, where we introduce a new analysis to disentangle the influence of recency versus context reinstatement and conduct statistical tests on the size of the TCE. For now, we simply note that the lag-CRP analysis addresses the first of the model evaluation criteria: There appears to be a TCE in incidental encoding with the CDFR task.

Can Retrieved Context Theory Simultaneously Fit Recall Levels and the TCE in Incidental Encoding?

To evaluate the second criterion, whether retrieved context theory can produce high recall accompanied by a modest TCE, we fit a version of the temporal context model (Howard & Kahana, 2002; Sederberg et al., 2008) to the data. The temporal context model is a computational model that implements the core mechanisms of retrieved context theory, including context drift, association formation, and context reinstatement. Although we fit one specific implementation of retrieved context theory, the core predictions made by this theory are consistent across implementations. Here, we give a conceptual overview of the model (for a full formal description, see Appendix A).

Model Description

In the temporal context model, two representational layers interact during encoding and recall: the feature layer and the context layer. Nodes on the feature layer represent individual items. For example, if a participant is presented with the word *river*, the node representing *river* will be activated on the feature layer. Once *river* has been activated, it in turn activates its existing semantic and episodic associations. For example, viewing *river* may activate thoughts of the sound of a river, the experience of walking on a riverbank, or the fact that people often go fishing on a river. This set of *river*-related thoughts, which we call context, is represented as an element on the context layer. The first panel of Figure 3 illustrates the activation of *river* on the feature layer and the resulting activation on the context layer.

When the next word, *money*, is presented, its feature representation will become active and completely replace the representation of *river* on the feature layer (the second panel of Figure 3). That is, feature representations are active on the feature layer only while that item is being presented. Once *money's* feature representation is activated, it in turn activates the context associated with *money* on the context layer.

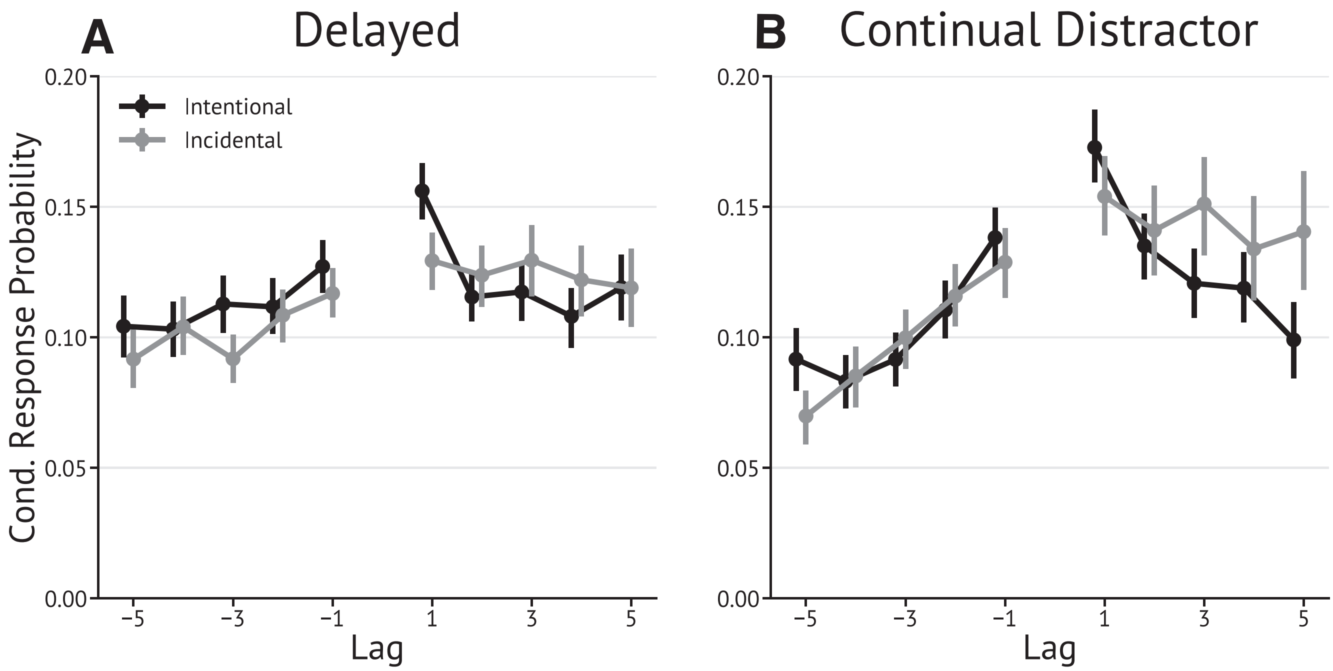
Critically, when *money* appears some of the semantic context of *river* remains active. The context representations of both *river* and *money* blend on the context layer, producing a new state of context—this is context drift. Drift is illustrated in the second panel of Figure 3, where the context representations for both *river* and *money* are active on the context layer. The image of river is slightly smaller than it was in the first panel to represent that as each new word is encoded, earlier representations on the context layer gradually fade. In this way, the context layer serves as a recency-weighted record of the items that have been presented, such that recent items are more strongly represented than earlier items. That is, the context layer carries information about past items and their serial order.

Under the model, episodic memories are formed by creating associations between the feature layer representation of each presented item and the state of the context layer that prevailed when the item was first presented. In most implementations, the formation of these new associations between the item and current mental context occurs before an item activates its context representation (for a discussion, see Sederberg et al., 2008). To continue with our example, when the next word, *cookie*, is presented, the *cookie* feature representation forms a new association to the current mental context, which contains elements of both *money* and *river*. Then, *cookie* activates its element on the context layer. The third panel of Figure 3 displays the state of the model after *cookie's* context representation has been activated.

To model distractor tasks, each distractor period is given its own representation on the context and feature layers. When distractor

[3] In all conditions, the TCE is smaller and more symmetric than is typically observed. The highest values of lag-CRP are typically between 0.3 and 0.5, whereas the highest lag-CRP in these conditions is below 0.2. This may be due to our participants' lack of task experience. Previous work has found that the size of the TCE in participants' recalls increases with practice: In an analysis of data from the PEERS dataset, the TCE was present but small in the first list, peaking at less than 0.2; by the twelfth list, the TCE was much larger (Healey et al., 2019). These results are also consistent with those for intentional and incidental encoding observed in Healey (2018), where participants also studied a single list.

Figure 2
Lag-Conditional Response Probability (Lag-CRP) Curves for Intentional Versus Incidental Encoding



Note. (A) Delayed free recall. (B) Continual distractor free recall. Error bars are bootstrapped 95% confidence intervals.

tasks are presented, they cause context to drift just like a study item, albeit at a different rate (see Appendix A for details). The result is that the context evoked by the previously-studied items fades more than it would without the distractor.

At recall, the current mental context is used as a cue for recalling items. When recall begins, the current state of context is most similar to the contexts associated with items near the end of the list, so the current context provides a strong cue for items at the end of the list. Thus, end-of-list items are typically recalled first. As discussed in the introductory paragraphs, this will naturally produce a (small) TCE.

But the similarity of the end-of-list context to end-of-list items is not the only model mechanism that produces a TCE: Once an item is successfully recalled, its associated context is brought to mind, or reinstated. For example, recalling *cookie* would reinstate the *cookie* context and also the *river–money* context that was active when *cookie* was encoded. This reinstated context is then integrated with the current context, causing context to drift in the same manner as during encoding. The newly updated context is then used to cue the next recall. The TCE arises because the reinstated context is a strong cue for items originally studied nearby in time to the just-recalled item. For example, the *river–money* context reinstated by recalling *cookie* is a strong cue for recalling *money*. Thus, under retrieved context theory, the TCE results largely from context reinstatement.

Simulation Methods

This framework predicts that a significant TCE will occur under incidental encoding because context drift during encoding and reinstatement during recall are both integral to episodic memory formation. It also predicts that overall recall will tend to be lower when the TCE is small because both depend on reinstatement. Thus, it is unclear if the framework's predictions will hold for the incidental conditions, in which the TCE is small but overall recall is nevertheless substantial. We attempted to fit the model to both the SPC and lag-CRP of each of the four conditions to test the model's, and thus retrieved context theory's, compatibility with this pattern. For details on our methods for model fitting and a table of best-fit parameter values, see Appendix A.

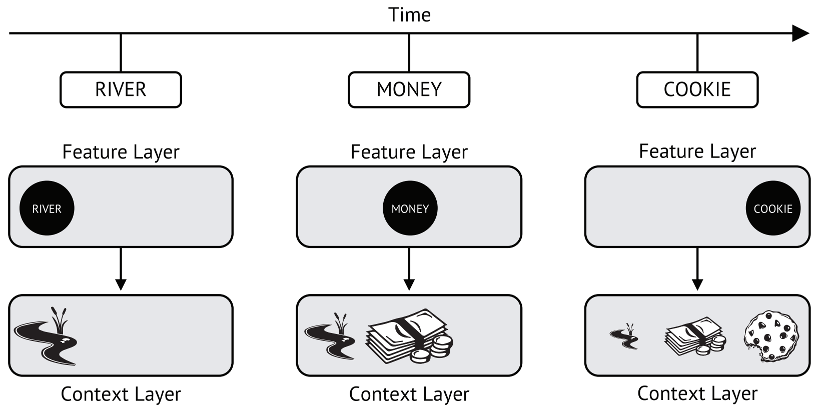
Simulation Results

Simulated data from the best-fitting parameter set for DFR are presented in Figure 4A; best-fit simulated data for CDFR are presented in Figure 4B. The model was able to simultaneously fit the general pattern of recalls and temporal contiguity in each condition. In the model simulations, overall levels of recall are greater in the intentional conditions compared with the incidental conditions, just as in the data.[4] Similarly, the simulated lag-CRPs show evidence of a TCE in all conditions but with a flatter curve for the

[4] Although we did not fit the model directly to probability of first recall (PFR), a valid question is whether the model would be able to fit to this measure of recall initiation. In past work, a similar model was unable to fit both the primacy and recency in PFR curves (Sederberg et al., 2008). We suggest this may stem from a difficulty in fitting the average compared with fitting individual participants. A single participant rarely displays the pattern we see in the PFR average, where there is some primacy but also some recency; rather, they tend to display either recency or primacy, not both (Healey & Kahana, 2014). In the current article, we did not fit the model to the recalls of individual participants because of the nature of an incidental encoding study. Because each participant could contribute only one list of data, there is a great deal of variability at the individual level, which does not facilitate precise model fitting. But in Healey and Kahana (2014), participants completed many lists, and thus it was possible to fit a retrieved context theory model directly to PFR. Their model was able to fit to recall initiation of individual participants in an immediate free recall task.

Figure 3

Visual Representation of the Encoding Period Under Retrieved Context Theory



Note. The identities of items are represented on the feature layer, with one node for each item. Each item has a corresponding node on the context layer which represents that item's contextual associates. When the first item, *river*, is presented, its node becomes active on the feature layer, which then activates the *river*-related node on the context layer. When the next item, *money*, is presented, its representation on the feature layer becomes active, completely replacing the *river* activation on the feature layer. Next, *money*-related context becomes active on the context layer, but it does not entirely replace the *river* context. Instead, the *river* context fades so that the current mental context is a blend of *river* and *money* contexts in which more recent items are more highly activated. The mental context serves as a recency weighted record of the past. This process repeats when *cookie* is presented: Its node on the feature layer becomes active, and its context representation is activated on the context layer, blending with other elements on the context layer to create a new *river*–*money*–*cookie* context.

incidental conditions, especially in the forward direction. This pattern matches that observed in the data. Most critically, in the incidental conditions, the model was able to simultaneously produce relatively high recall and a relatively small TCE.

An obvious next step is to investigate which model parameters account for the differences across conditions. Table A1 shows the best-fit parameter values for each condition. Average parameter values across 30 model fits are presented in Table A2. Many parameter values differ significantly between conditions: Six out of 10 parameters differ between incidental and intentional encoding across distractor conditions (additional parameters differ for only one distractor condition). Many of these differences align with the core mechanisms of retrieved context theory. For example, β_{enc}, which contributes to the size of the recency effect, among other things, is highest in the incidental CDFR condition. This is consistent with the steeper recency effect in this condition. In contrast, β_{rec}, which controls context drift during recall (i.e., reinstatement), is lower in the incidental conditions; when β_{rec} is lower, there is less contiguity, consistent with the smaller TCE in the incidental conditions. We stress that these are not the only parameters that differ across conditions—the question of which model processes account for the difference between intentional and incidental encoding defies a simple summary. It is perhaps unsurprising that expecting a memory test changes many aspects of both encoding and recall. We further investigate the necessity of specific parameters in later sections.

In sum, these model fits show retrieved context theory is consistent with the patterns in both overall recall and temporal contiguity in incidental encoding. But questions remain: Is the TCE in participants' recalls a result of the same recency and reinstatement mechanisms specified in retrieved context theory? Is the model fit quantitatively acceptable in addition to capturing the general shape of the data? We answer these questions in the next sections.

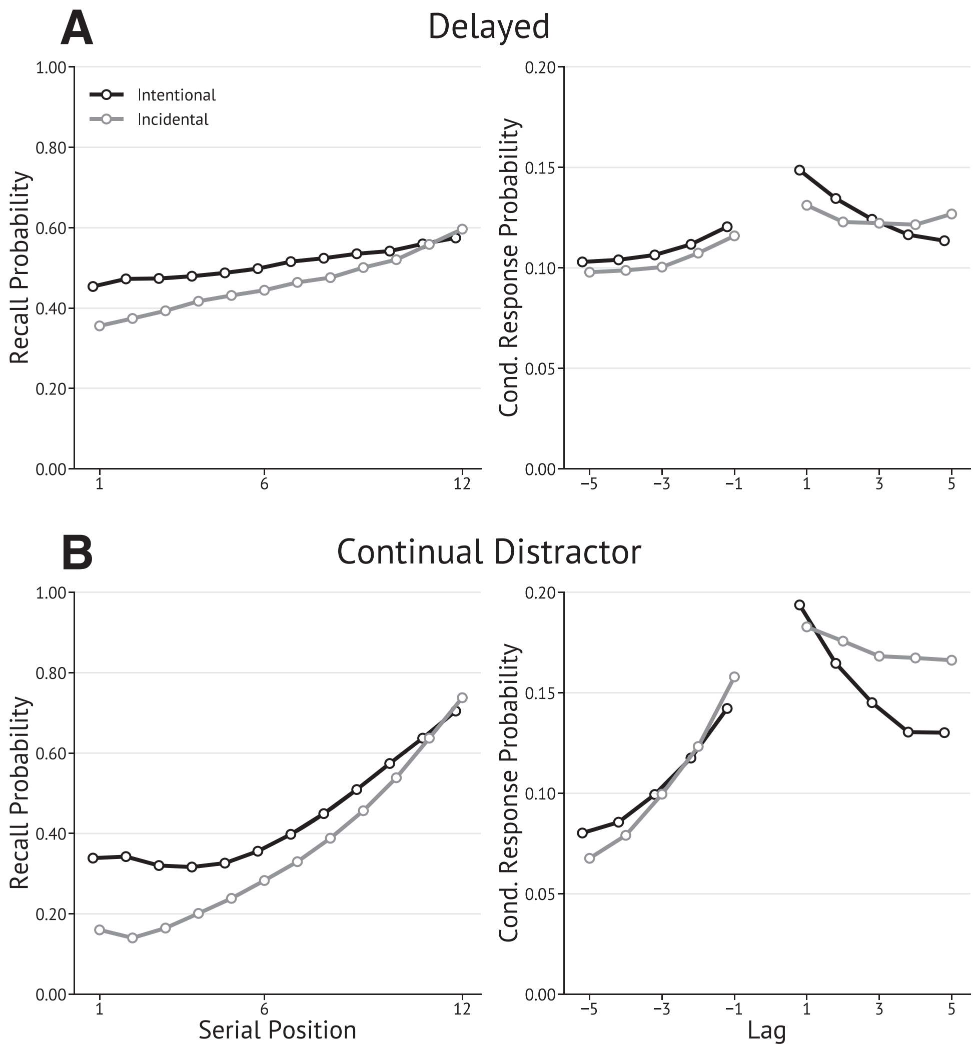
Does Context Reinstatement Occur Under Incidental Encoding?

We have addressed the first two criteria for evaluating retrieved context theory: There is a TCE under incidental encoding which is not limited to short timescales, and the model is able to qualitatively fit both overall recall and temporal contiguity in all conditions. However, as described previously, there are two mechanisms specified by retrieved context theory that work together during recall to generate the TCE: recency and context reinstatement. The context reinstatement mechanism distinguishes retrieved context theory from other models of episodic memory. Yet, it is unclear if reinstatement contributes to the TCE in incidental encoding. Because incidental participants showed more pronounced recency, as well as a reduced TCE, it is possible that the contiguity observed under incidental encoding is due purely to recency.

We demonstrate how recency alone can generate a TCE with an example. Imagine a participant who recalls items from serial positions 10, 11, and 12 from a 12-item list but has no bias for recalling items in temporal order. That is, there is no mechanism (like context reinstatement) to create a bias for short-lag transitions

Figure 4

Best-Fitting Simulated Serial Position Curves (SPCs) and Lag-Conditional Response Probability (Lag-CRP) Curves From the Temporal Context Model



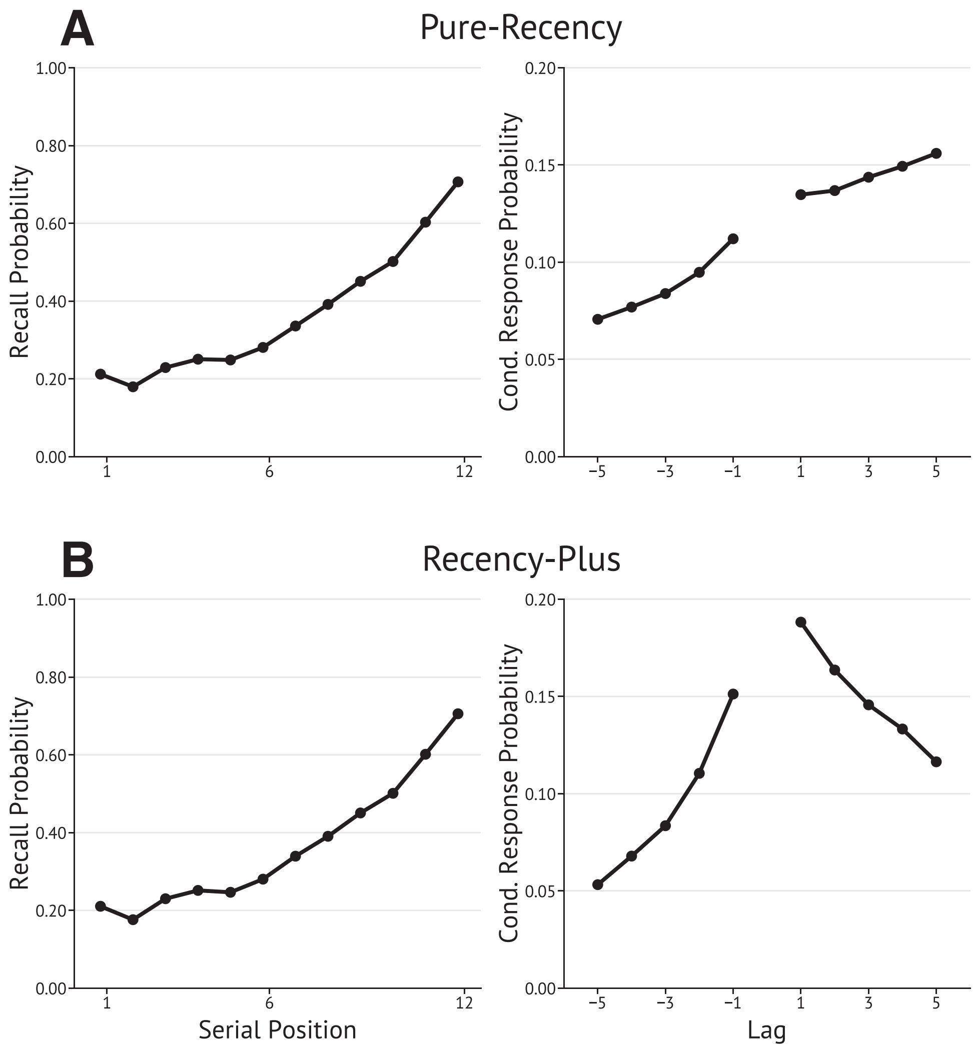
Note. (A) Delayed free recall. (B) Continual distractor free recall. Open points represent simulated data based on the best-fit model.

over long-lag transitions. Instead, the participant would output these items in random order (e.g., 10, 12, 11). Nonetheless, as a result of recalling only the most recent items, only $|lag| = 1$ or $|lag| = 2$ transitions are possible. Thus, this participant's recalls will appear to have a TCE when measured with lag-CRP: They make near lags more often than far lags by necessity because far lags are impossible given their recalls. This is directly a result of having recalled only recent items and is distinct from a mechanism like context reinstatement, where one recalled item directly cues other items studied nearby in the list. In other words, this participant displays a TCE due purely to recency.

To further illustrate, we simulated data from a pure-recency participant using a simple model. For each simulated list, we decided whether the item from each serial position would be recalled by randomly drawing from a binomial distribution where the probability of success was set to the recall probability of the corresponding serial position in the SPC of the incidental CDFR condition (the condition with the strongest recency effect). This gave us a set of n items that were recalled from this list. To determine the order in which the simulated participant recalled these n items, we simply randomly shuffled the items. This produced a recall sequence with strong recency but where items were recalled independently of each other. Thus, any contiguity in this simulation is due to the shape of the SPC, particularly the strong recency effect. The results of these simulations are presented in Figure 5A. The lag-CRP for the pure-recency model appears to have a TCE, at least in the backward direction. Indeed, this pure-recency lag-CRP with its steep negative slope for backward lags and slight positive slope

Figure 5

Simulated Serial Position Curves (SPCs) and Lag-Conditional Response Probability (Lag-CRP) Curves



Note. (A) The pure-recency model, where recall order was randomly selected with regard to lag. (B) The recency-plus model, where recall order was selected with a bias for near lags. For each model, we generated recalls for 100,000 simulated participants, each recalling from one list of 12 items.

for forward lags is quite similar to the lag-CRP observed for the incidental conditions in Figure 2.

We can contrast this pure-recency participant with another participant who recalls exactly the same items but who *does* have a bias toward recalling those items in temporal order. For example, they may have first recalled Item 12, which cued Item 11, which in turn cued Item 10. Again, this participant recalls few items and has a strong recency effect, but they also show a preference for shorter lags within the lags that are possible given their recalls. Even though $|lag| = 1$ and $|lag| = 2$ were possible, this participant made only transitions of $lag = -1$. Like the pure-recency participant, this participant's lag-CRP would show a TCE, but one that reflects recency plus an additional bias toward short lags. We simulated data for this type of "recency-plus" participant with another simple model. The n items recalled for each list were selected with the same procedure used in the pure-recency model, but they were output in a nonrandom order.

Specifically, the first recall, output i, was selected randomly from among the n items. For the next recall, output j, we computed the possible lags $(j - i)$ for the i to j transition and selected the next recall by drawing from a weighted distribution that heavily favored near lags. To create this distribution all possible $|lag| = 1$ transitions were given a weight of .9, all $|lag| = 2$ transitions a weight of .6, all $|lag| = 3$ transitions a weight of .4, all $|lag| = 4$ transitions a weight of .3, and $|lag| \geq 5$ transitions a weight of .2. The weight for each possible lag was then converted to a probability by dividing the sum of the weights across all possible lags. In other words, there was both a strong recency effect *and* a bias for near lags. The lag-CRP for simulated data in Figure 5B displays a TCE with a steep slope for both backward and forward lags.

Taken together, the pure-recency and recency-plus simulations demonstrate the lag-CRP cannot conclusively establish whether contiguity is due only to recency. Even in simulated data with minimal noise, both the pure-recency and the recency-plus models show a gradient across backward lags. The telltale sign of the recency-only model is that there is a slight preference for longer forward transitions. In real data, variability could easily disguise this sign. Therefore, we introduce new measures to test if the TCE observed in incidental CDFR is generated, at least in part, by additional mechanisms such as context reinstatement, as predicted by retrieved context theory.

Temporal Bias Scores

Temporal bias scores measure the level of temporal contiguity in a set of recalls relative to what would be expected if the same items were recalled in random order (Uitvlugt & Healey, 2019). As a result, they are able to disentangle the effects of recency (or other deviations of the SPC from a uniform distribution, like primacy) and other mechanisms, like context reinstatement. If temporal bias scores indicate there is no difference between contiguity expected due to chance and the observed TCE, then any temporal contiguity observed in the lag-CRP must be due to recency (or other aspects of the SPC). For each lag, temporal bias is calculated by finding the difference between the number of times a participant made a transition of that lag and the number of times the participant would be expected to make transitions of that lag if items were recalled in random order. This expected count is calculated by taking the items actually recalled by a participant, randomly permuting the order of recalls, and recording which lags are made within the new permuted order. Repeating this random shuffling procedure many times provides a distribution of expected lag counts, and we define the expected count as the average of this distribution. Temporal bias scores can then be calculated as the difference from chance normalized by chance:

$$\text{temporal bias} = \frac{\text{actual count} - \text{expected count}}{\text{expected count}}. \tag{1}$$

A temporal bias score of zero indicates a lag occurs exactly as often as would be expected if previous recalls have no influence on the next recall: A participant that recalls items in random order should have temporal bias scores of zero for all lags. A score above zero for a given lag indicates the lag occurred more often than expected; there is a positive bias for this lag. A temporal bias score below zero for a given lag indicates the lag occurred less in actual recalls than was expected due to chance. A participant with a strong TCE due to a nonrecency mechanism (e.g., context reinstatement) would have positive temporal bias scores for short lags and negative temporal bias scores for far lags.

Temporal bias curves for our simple pure-recency and recency-plus simulations are presented in Figure 6 (to illustrate how lag-CRP and temporal bias scores differ in a standard free recall task, see Appendix B for a comparison in two archival data sets). The temporal bias scores for the pure-recency model are near zero for all lags, indicating presentation order has no impact on output order beyond the influence of recency on which items are recalled. This is consistent with the data generation process: The order of recalls was indeed chosen randomly. In the recency-plus model, where data were generated with a bias for near lags, the temporal bias curve shows a clear positive bias for near lags that decreases for farther |*lags*|. That is, the temporal bias scores accurately detected that there was a TCE above and beyond the influence of recency.

Context Reinstatement in the Data

Is the TCE observed in our incidental conditions due to recency alone? The temporal bias curves for the actual data (presented in Figure 7) provide a clear answer: There is a TCE in all conditions beyond the influence of recency. In the DFR conditions, the temporal bias scores for the intentional and incidental groups show a positive bias for near lags and a negative bias for far lags. The groups differ only at $lag = +1$, where there is a much stronger positive bias in the intentional condition. Similarly, in both CDFR conditions there is a positive bias for near lags, with the greatest bias for $lag = +1$ in the intentional condition. The incidental CDFR condition displays a positive bias for backward lags from −1 to −5, and most of the forward lags occur less often than expected. This indicates that although there is a TCE in the incidental CDFR condition, these participants also had a bias toward backward lags over forward lags.

To complement temporal bias curves, we can use temporal factor scores to provide a single-number measure of the size of the TCE. These scores allow for a statistical test for the presence of the TCE: whether the effect is significantly larger than expected by chance (Polyn et al., 2009; Sederberg et al., 2011). These scores are calculated by taking the absolute value of the lag of each transition made by a participant and finding its percentile within the distribution of the $|lags|$ that were possible for that transition: A higher score indicates more temporal contiguity. Typically, a score of .5 would indicate no bias. However, temporal factor scores can be compared with a chance level calculated with the same procedure used in calculating temporal bias scores. The order of recalls is permuted many times, and a temporal factor score is calculated for each permuted order (Healey, 2018; Polyn et al., 2011). This allows us to correct for the effects of recency. A temporal factor score above chance indicates a significant TCE due to mechanisms beyond recency.

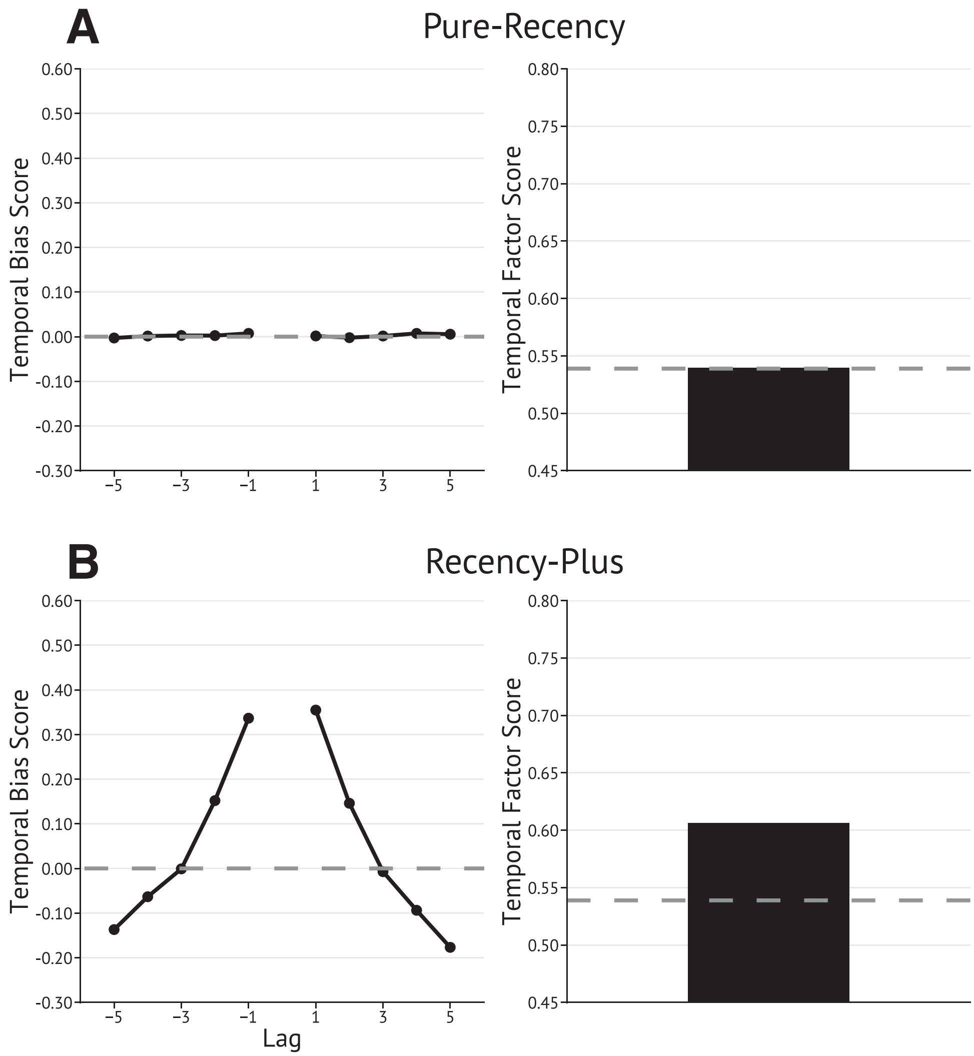
Applying temporal factor scores to our simple recency models (second column of Figure 6) shows they accurately detect that there is no TCE beyond the effects of recency in the pure-recency data, but there is a nonrecency TCE in the recency-plus data. That is, temporal factor scores can detect the difference between a TCE due only to recency and one to which context reinstatement or similar mechanisms also contribute. When applied to the actual data, temporal factor scores reveal a significant TCE in all conditions. The second column of Figure 7 shows that temporal factor scores for all groups, including both incidental groups, are well above chance. In summary, the temporal bias and temporal factor scores for the incidental conditions indicate that contiguity in incidental encoding is consistent with the operation of both recency and context reinstatement.

Context Reinstatement in Retrieved Context Theory

The temporal context model was able to fit the lag-CRP, but as demonstrated with the pure-recency and recency-plus simulations, the lag-CRP analysis does not distinguish between a TCE due to recency and a TCE due to other mechanisms. Therefore, it is still an

Figure 6

Simulated Temporal Bias Curves and Temporal Factor Scores



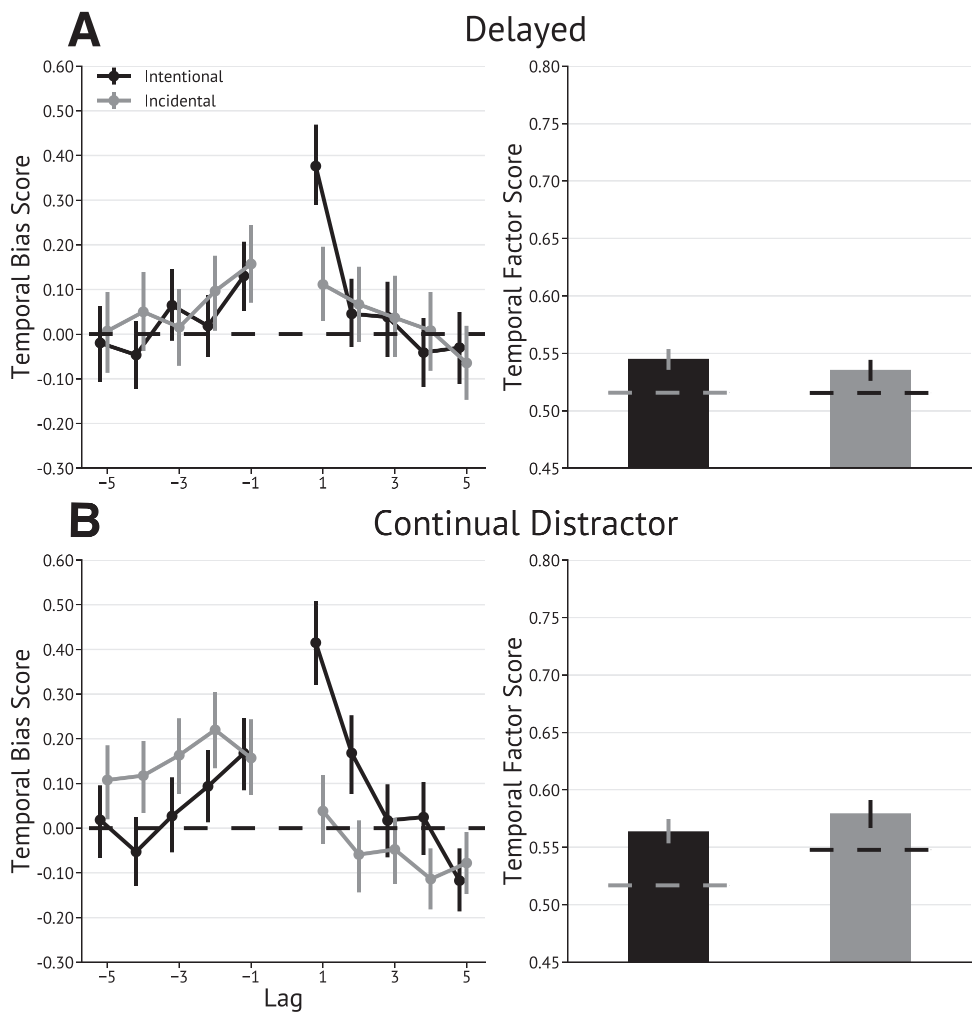
Note. (A) The pure-recency model, where recall order was randomly selected. (B) The recency-plus model, where recall order was selected with bias for near lags. For each model, we generated recalls for 100,000 simulated participants, each recalling from one list of 12 items. Temporal bias scores were calculated for each lag by counting the number of times a participant actually made a transition of that lag and the number of times such a transition was expected by chance. The expected count was determined by permuting the order of the recalled items 500 times and counting how many times each lag occurred for each permutation. Temporal bias is the difference between the actual count and average expected count, divided by the average expected count. The dotted line for the temporal bias scores indicates a score of zero. Chance for temporal factor scores, represented with the dotted line, was determined by calculating temporal factor scores for each of 100 random permutations of the order of recalls and getting the average of this distribution.

open question whether the model was utilizing the core mechanisms of context drift and context reinstatement in generating temporal contiguity, as predicted by retrieved context theory, or if it did so by using other mechanisms.

This is a particular concern in incidental encoding where the TCE is smaller and there is greater recency in participants' recalls. These features may make it possible for the model to match participants' level of contiguity using only the recency mechanism or through flexibility provided by other parameters. Thus, we conducted a series of simulations to determine if context drift and reinstatement truly are critical to the model's ability to account for contiguity in the incidental conditions.

We began by fitting the full version of the model (described in Appendix A) to both SPCs and temporal bias scores, which measure the TCE once the effects of recency have been accounted for. We then compared these full model fits to submodels in which context drift and context reinstatement were eliminated by setting the parameters that control these mechanisms to zero. If these restricted models can fit the

Figure 7
Temporal Bias and Temporal Factor Scores for Intentional Versus Incidental Encoding



Note. (A) Delayed free recall. (B) Continual distractor free recall. Temporal bias scores were calculated for each lag by counting the number of times a participant actually made a transition of that lag and the number of times such a transition would be expected by chance. The expected count was determined by permuting the order of the recalled items 500 times and counting how many times each lag occurred for each permutation. Temporal bias is the difference between the actual count and average expected count, divided by the average expected count. The dotted line for the temporal bias scores indicates a score of zero. Chance for temporal factor scores was determined by calculating temporal factor scores for each of 100 random permutations of the order of recalls and getting the average of this distribution. Chance temporal factor scores are represented with a dotted line. Error bars are bootstrapped 95% confidence intervals.

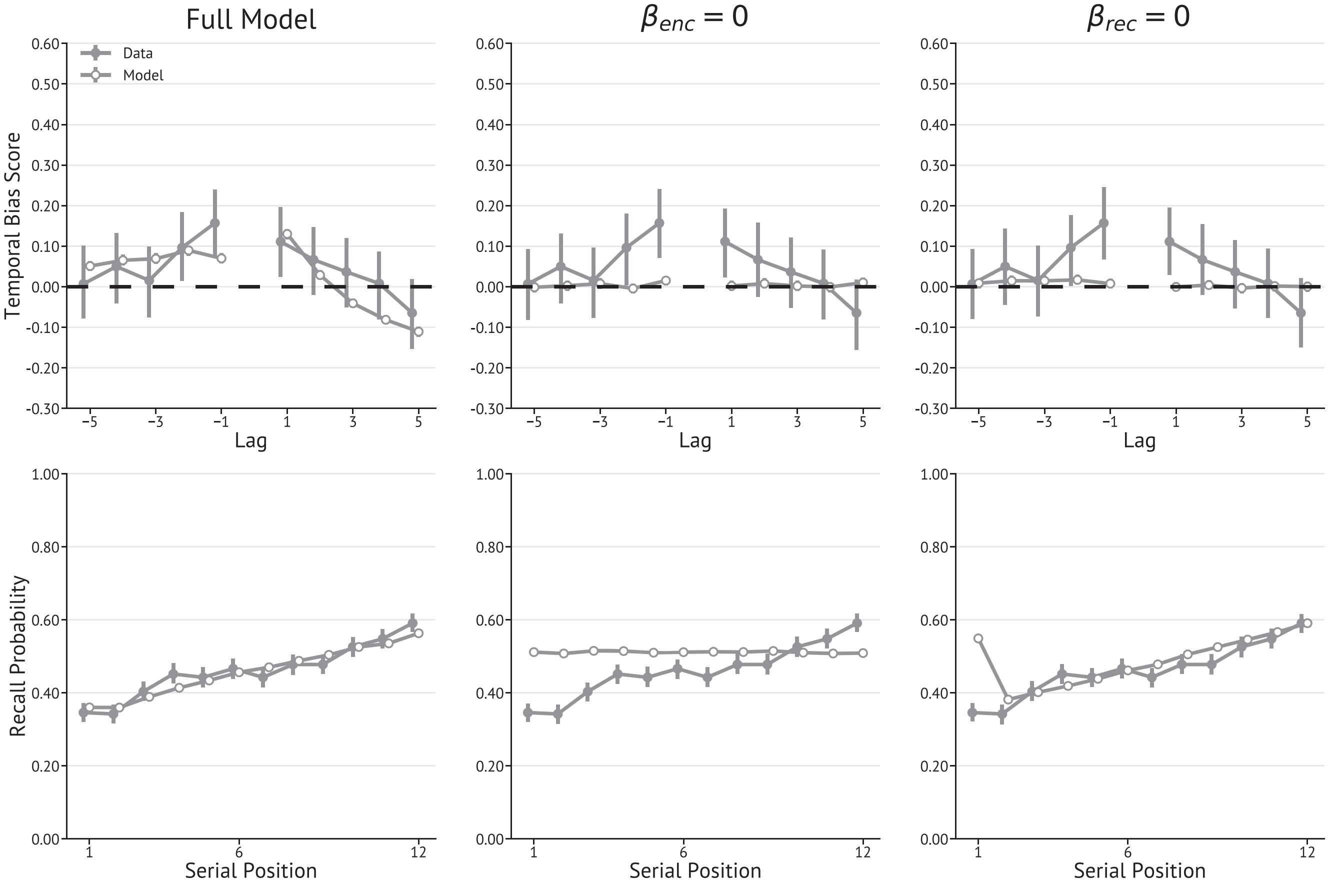
data as well as the full model, it would indicate that contiguity can be generated in the model through flexibility provided by other, theoretically unimportant, parameters (e.g., the decision parameter τ used at recall). If instead a version of the model in which a theoretically important parameter is restricted to zero produces an inferior fit to the full model, in which all parameters are allowed to freely vary, this would provide evidence that the model's fit is dependent on specific, theoretically important, cognitive mechanisms.

Full Model

The temporal context model was fit to SPCs and temporal bias scores for each incidental condition. The full version of the model provided good fits to both overall recall and temporal contiguity generated beyond the effects of recency for both the incidental DFR (see Figure 8) and incidental CDFR (see Figure 9) conditions. Parameter values are presented in Table A3. The model was able to capture the temporal bias for near lags, as well as the general shape of the SPC, for both incidental conditions. The model was also able to approximate the positive bias for backward lags observed in the incidental CDFR condition. This indicates that, consistent with the data, the temporal context model can generate temporal contiguity through both context drift and reinstatement, while still generating recalls with high levels of recency.

Figure 8

Comparison of Temporal Context Model Simulations to the Data for the Incidental Delayed Free Recall Condition



Note. The top row shows temporal bias curves, and the bottom row shows serial position curves (SPCs). Each column shows the best-fitting simulated data for each version of the model. In the full model, presented in the first column, all parameters were allowed to freely vary. In the second column, context drift during encoding was prevented by setting $\beta_{enc} = 0$. In the third column, context drift during recall was prevented by setting $\beta_{rec} = 0$. Open points represent simulated data based on the best-fit model. Error bars are bootstrapped 95% confidence intervals.

Submodel 1: Eliminating Context Drift During Encoding

The rate of context drift during encoding is controlled by β_{enc}, which, among other things, contributes to the size of the recency effect. When β_{enc} is higher, items from the end of the list are more likely to be recalled first. In contrast, when $\beta_{enc} = 0$, context is unable to drift during the encoding period, and all items form associations with the same state of context; each item is an equally good cue for all other items. Therefore, we would expect the TCE to be abolished.

When fitting the submodel to the DFR condition, we prevented context drift during encoding by simply setting β_{enc} to zero. For the CDFR condition, however, we must also consider how the distractor task that intervenes between each item influences drift during encoding. Any context drift caused by the distractor task between items n and $n + 1$ would cause the two items to be associated with slightly different states of context, allowing a TCE to reemerge. Thus, when fitting the submodel to the CDFR condition, we set both β_{enc} and $\beta_{distract}$, which controls the rate of context drift during a distractor task, to be zero.

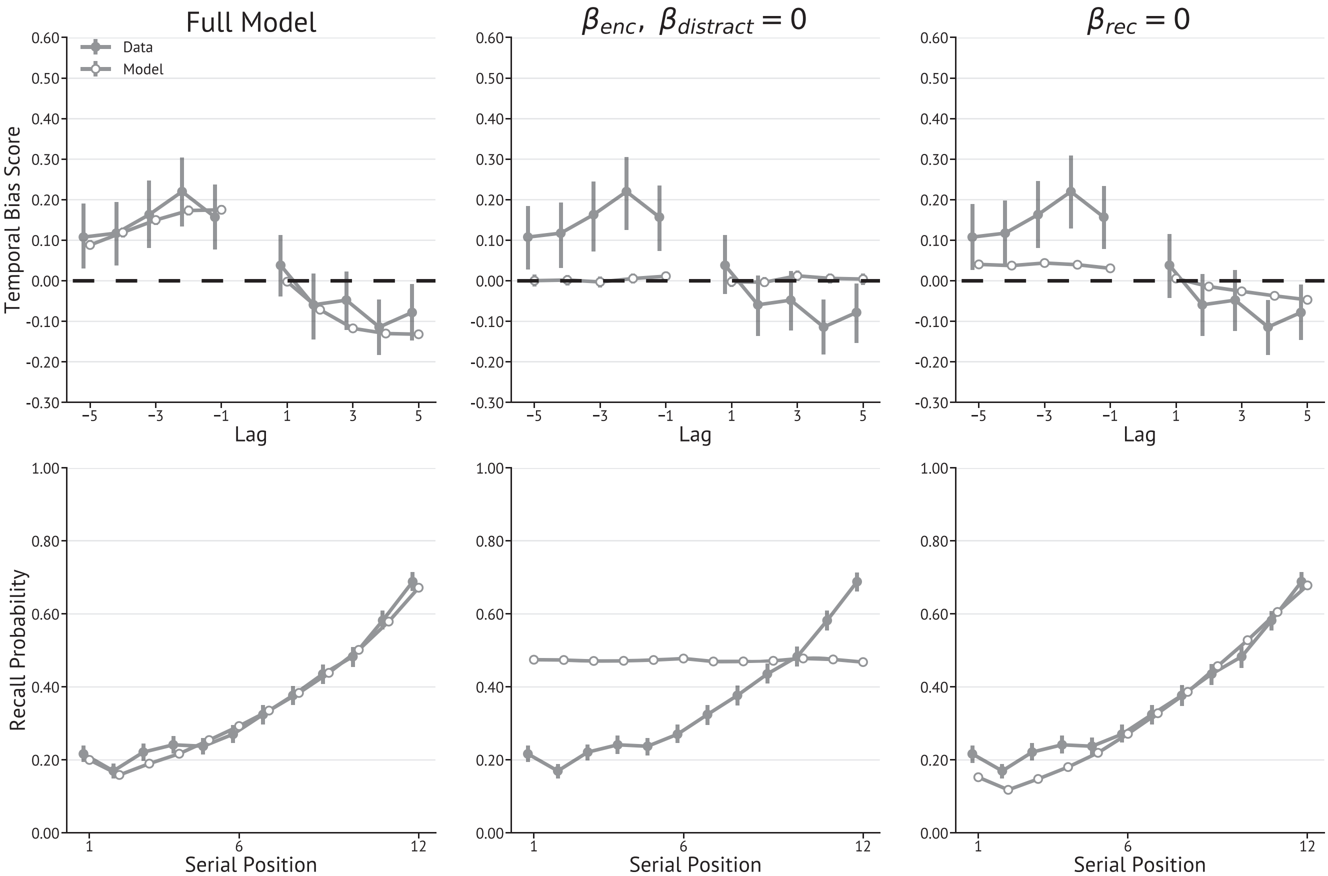
The second columns of Figures 8 and 9 show that the model was not able to fit to temporal bias scores when context drift was eliminated during incidental encoding. In DFR (second column of Figure 8), there was a nearly flat TCE when $\beta_{enc} = 0$. In CDFR, the model was similarly unable to fit to both the SPC and temporal bias scores at the same time (second column of Figure 9). See Table A4 for a list of parameter values for this submodel fit.

Submodel 2: Eliminating Context Reinstatement During Recall

To determine if the context reinstatement mechanism is necessary, the β_{rec} parameter can be set to zero. β_{rec} determines the extent to which a just-recalled item's context is reinstated and incorporated into the cue for the next recall; when β_{rec} is lower there tends to be less contiguity. If $\beta_{rec} = 0$, then context does not drift during retrieval, and no context reinstatement occurs. In principle, the submodel should be unable to fit the temporal bias scores; however, if the model is overflexible, it

Figure 9

Comparison of Temporal Context Model Simulations to the Data for the Incidental Continual Distractor Free Recall Condition



Note. The top row shows temporal bias curves, and the bottom row shows serial position curves (SPCs). Each column shows the best-fitting simulated data for each version of the model. In the full model, presented in the first column, all parameters were allowed to freely vary. In the second column, context drift during the encoding period was prevented by setting β_{enc} and $\beta_{distract} = 0$. In the third column, context drift during recall was prevented by setting $\beta_{rec} = 0$. Open points represent simulated data based on the best-fit model. Error bars are bootstrapped 95% confidence intervals.

may be able to produce a reasonable fit using the remaining mechanisms.

Confirming the importance of context reinstatement, the model produced poor fits when reinstatement was eliminated, with a nearly flat contiguity effect for both encoding conditions (see third columns of Figures 8 and 9 for DFR and CDFR, respectively). Parameter values are presented in Table A5.

The failure of Submodels 1 and 2 to fit the behavioral data indicates context drift is necessary at both encoding and retrieval. These simulations also address the concern of overflexibility. A model of cognitive processes that is overflexible and can fit any potential pattern of results is not falsifiable and therefore is not useful as an explanation of human cognition. The failure of these submodels demonstrates that the success of the full model is not due to overflexibility. That is, the context drift parameters which retrieved context theory predicts are central to memory formation are, indeed, necessary for model fitting.

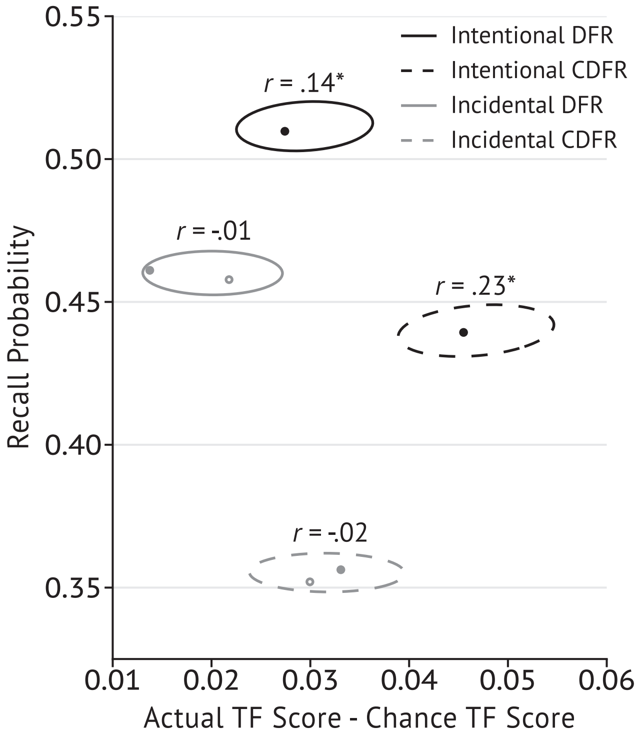
Correlations in the Data and Model

Qualitatively, the fit of the temporal context model to both lag-CRP and temporal bias curves in the data demonstrates that the patterns of recall and contiguity in each condition are consistent with the predictions of retrieved context theory. However, we have not yet tested whether the fits are adequate. We can now do so using the measures of temporal bias. Given that the model must be able to capture recall and the TCE simultaneously, we asked whether the models' overall recall and temporal factor scores (minus chance) fall within the range of these measures in the real data.

For each condition, Figure 10 shows a 95% confidence ellipse on the joint distribution of overall recall and temporal factor scores (minus chance) in the actual data.[5] For all conditions, the means of the simulated data, based on the best-fit model to the SPCs and lag-CRPs for each condition, are plotted as points. For the incidental conditions, we also plot the means for the full model fits to SPCs

[5] A confidence ellipse is simply a two-dimensional 95% confidence interval (CI) based on the variances of the two measures rotated by the covariance of the two measures. Just as a univariate CI means that 95% of exact replications would produce a confidence interval that includes the true univariate mean, the confidence ellipse means that 95% of exact replications would produce an ellipse that includes the bivariate mean.

Figure 10
Confidence Ellipses on the Mean of the Joint Distribution of Overall Recall and Temporal Contiguity With Model Predictions



Note. For the actual data, we show 95% confidence ellipses on the mean of the joint distribution of overall recall and temporal factor (minus chance). For the model, we show the means of the same variables from the simulated model data as a point for each condition. The means for the model fits to the SPCs and lag-CRPs of each condition are presented as solid points. The means for the model fits to the SPCs and temporal bias scores for the incidental conditions are presented as open points. Chance for temporal factor scores was determined by calculating temporal factor scores for each of 100 random permutations of the order of recalls and getting the average of this distribution. The r value above each ellipse is the Pearson correlation between recall probability and the temporal factor minus chance scores for that condition.
* Indicates that the correlation is significant at $\alpha < .05$.

and temporal bias scores (as noted above, in fitting to temporal bias scores, we considered only the theoretically critical incidental conditions and did not fit the intentional conditions). Note that in all cases, the model was not fit directly to the measures plotted here; temporal factor scores for the model were calculated from data that was generated based on best-fit parameters to the SPC and lag-CRP or temporal bias curves. Nonetheless, the model predictions fall within the condition's confidence ellipse for every condition. That is, the model predictions do not deviate significantly from the data.

Discussion

We tested whether the temporal context model, a specific computational implementation of retrieved context theory, is compatible with recent findings that the TCE is dramatically reduced under incidental encoding. The theory assumes the TCE results from fundamental properties of episodic memory, such that temporal information is automatically encoded, and thus predicts a TCE should be observed regardless of encoding intentionality. We evaluated the ability of the model to account for the patterns observed in the data based on three criteria: (a) there should be a TCE under incidental encoding at multiple timescales, (b) the model should be able to simultaneously fit both overall recall and the TCE, and (c) the nature of the incidental TCE should be consistent with the contextual dynamics mechanisms proposed by retrieved context theory.

To address the first criterion, we tested for the presence of a TCE under incidental encoding using lag-CRP curves. A TCE was found in all groups, including when encoding was incidental and study events were separated by a filled distractor. To address the second criterion, we fit the temporal context model to both the SPC and lag-CRP curves for each condition. The model fit was consistent with the data. To address the final criterion, we developed novel analytical techniques designed to disentangle the influences of recency and context reinstatement. They demonstrated that mechanisms beyond recency are involved in generating the TCE in both intentional and incidental encoding, consistent with the operation of a context reinstatement mechanism, as suggested by retrieved context theory. Additional model simulations in which the temporal context model was fit to this new measure of context reinstatement provided evidence that drift during encoding and context reinstatement at recall are critical components of retrieved context theory's ability to account for recall dynamics. In sum, the observed TCE is consistent with retrieved context theory's prediction of a significant TCE regardless of encoding intentionality, timescale, or the magnitude of recency.

Relation to Other Theories

We have argued that our results confirm several key predictions of retrieved context theory and lend support to the claim that the TCE is generated in part by a context reinstatement mechanism. Some mechanisms proposed by other theories would have difficulty producing a TCE under the conditions of the present experiment, particularly under incidental encoding in CDFR. For instance, any theory which assumes that the TCE is a result of deliberate encoding strategies (see Healey, 2018; Healey et al., 2019; Hintzman, 2016) is unable to explain a TCE of any size under incidental encoding. That is, control processes alone do not predict a TCE in incidental encoding.

Similarly, binding in a short-term buffer (e.g., Davelaar et al., 2005; Raaijmakers & Shiffrin, 1981) cannot easily explain a TCE in CDFR because the short-term buffer account predicts bonds should form between items only when they co-occupy the buffer (Phillips et al., 1967). A distractor task should fill the buffer between each item presentation and thus prevent items from co-occupying the buffer and forming temporal bonds. Thus, work showing a TCE in CDFR has been taken as evidence against the short-term buffer account (Howard & Kahana, 1999). Yet, because previous work with CDFR has used intentional encoding, one could argue that the use of deliberate control processes limits the effectiveness of the distractor task and allows items from adjacent serial positions to co-occupy the buffer. For example, the distractor may not be sufficiently demanding and therefore not fully push previously studied items out of the buffer or prevent rehearsal. Or even if we assume the distractor is sufficiently demanding and it is impossible to rehearse during the distractor, a participant may still strategically use part of an item's presentation time to rehearse previous items. However, in our study, a TCE was observed in CDFR when participants were unaware their memory would be tested and thus had no reason to engage in such control processes.

Although our findings are consistent with retrieved context theory and seem quite inconsistent with some proposed contiguity-generating mechanisms, it is important to note that our findings do not uniquely support retrieved context theory. Other models of episodic memory may also be able to account for these results. A model with separate short and long-term memory stores may not be able to account for these findings with a short-term buffer mechanism but can nonetheless account for the ubiquity of the TCE if its implementation of long-term memory includes contextual dynamics similar to those in retrieved context theory (as many dual-store models do, e.g., Davelaar et al., 2005; Lehman & Malmberg, 2013; Mensink & Raaijmakers, 1989). Our results simply indicate that dual-store models' ability to explain the TCE is not dependent on the short-term memory store. In addition, models that include a representation of time, such as the scale-independent memory, perception, and learning model (SIMPLE; Brown et al., 2007) or a hierarchical chunking mechanism (Farrell, 2012), may be consistent with the presence of the TCE in both incidental and intentional encoding and the observed timescale similarity. However, detailed simulations would be required to determine if such models can quantitatively fit the observed patterns.

Implications for Retrieved Context Theory

Our focus here has been whether the retrieved context theory is compatible with the nature of the TCE under incidental encoding. Specifically, does the small size of the TCE falsify the theory's predictions? We have shown that it does not. A model based on retrieved context theory was able to simultaneously generate a small TCE and substantial recall for both short (DFR) and longer (CDFR) timescales. These model fits to SPCs and lag-CRPs still left open the possibility that only the recency mechanism operates in recall following incidental encoding and not the context reinstatement mechanism that is the heart of the theory—this would have severely limited the explanatory scope of retrieved context theory. We tested this with temporal bias curves and temporal factor scores, which eliminate the influence of factors such as recency on the TCE and found that mechanisms beyond recency are involved in the generation of the TCE in both intentional and incidental encoding. Moreover, with a series of model fits to temporal bias scores, we demonstrated that context drift was necessary during encoding and context reinstatement was necessary during retrieval for the incidental conditions. Overall, our findings suggest that the memory system naturally encodes information about temporal order and that retrieved context theory provides a good description of the nature of that information. But some open questions remain.

In our simulations, we used a search algorithm to determine how to parameterize the model for conditions that differed in their level of temporal contiguity. How do participants adapt to different task conditions? We suggest that control processes are critical in determining the degree to which temporal information is encoded and how it influences recall, perhaps by tuning the parameters of the memory system based on task instructions.

Control Processes Can Increase the TCE

In support of this idea, there is evidence that participants who are more likely or more able to engage in control processes display a greater TCE, with more pronounced forward asymmetry, than other participants in typical free recall tasks. Spillers and Unsworth (2011) administered a free recall test to participants with high versus low working memory scores and found that participants with high working memory scores displayed a larger TCE with greater forward asymmetry than the low working memory group; this difference was particularly pronounced at $lag = +1$. Cognitive aging, which is known to impair cognitive control (Hasher et al., 2007; West, 1996) has a similar effect on the TCE: The TCE is smaller and more symmetric for older compared with younger participants (Diamond & Levine, 2019; Healey & Kahana, 2016; Kahana et al., 2002). Moreover, participants with clinical conditions that can impact cognitive control, such as schizophrenia (Polyn et al., 2015; Sahakyan & Kwapil, 2018) or high trait worry (Pajkossy et al., 2017), often have a reduced TCE compared with controls (for an example of a clinical condition increasing the TCE, see Gibson et al., 2019).

These patterns of differences in temporal contiguity, particularly for $lag = +1$, parallel what we observed here in the incidental encoding conditions, where the TCE was stronger and the forward asymmetry was more pronounced in the intentional than incidental encoding conditions. Once the influence of recency in our data was accounted for using temporal bias scores, the most pronounced difference between the intentional and incidental conditions was at the $lag = +1$ (see Figure 7). A similarly pronounced difference in asymmetry is also present in the lag-CRPs of well-practiced participants from archival data in Figure B1, where the $lag = +1$ transitions are much more likely than any other lags.

In all these cases, including the present data, participants with a greater likelihood of engaging control processes, either due to task demands or individual differences, displayed this strong bias for $+1$ lags. That is, high-control participants utilize temporal information to a greater degree and, as a result, display a larger TCE. This is consistent with individual differences studies that have found a positive correlation between temporal contiguity and both overall recall and measures of cognitive ability (Healey et al., 2019; Sederberg et al., 2010; Spillers & Unsworth, 2011). Indeed, there is evidence that both the size of the TCE and the degree of forward asymmetry account for unique variance in recall ability for recall of autobiographical memories.

Control Processes Can Also Decrease the TCE

However, these studies where high cognitive control is linked to a strong TCE all involve cases in which remembering and using temporal information facilitates task performance. When remembering temporal information could actually be detrimental to performance, the relationship between high cognitive ability and the use of temporal information may be reversed. For example, Osth and Fox (2019) found the TCE was absent for a paired associate recognition task in which utilizing temporal information would lead to impaired task performance (but see Davis et al., 2008). Here too, individual differences influence the effects of task: Older adults, who likely have less ability to inhibit the influence of temporal information, do show evidence of a TCE in a paired associates task (Campbell et al., 2014). In free recall tasks designed to merely shift participants' attention away from utilizing temporal information, participants also display a smaller TCE. Hong et al. (2019) found that combining multiple factors that tend to reduce the usefulness of temporal information, such as testing participants on a long list with semantic associations, can eliminate the TCE. Similarly, when Healey and Uitvlugt (2019) modified free recall instructions to ask participants to focus on similarities in meaning between items rather than the order of items during study, they found the TCE was reduced to near zero. In addition, the correlation between the strength of the TCE and overall recall was much smaller in standard lists composed of relatively unrelated words when participants were instructed to focus on semantic similarities. When lists were composed of related words, the correlation between recall and the TCE actually was negative.

We observed a similar pattern in our results: The correlations between recall and the TCE (see Figure 10) were significantly higher for intentional encoding than for incidental encoding for DFR (Fisher's $z = 3.95$, $p < .001$) and CDFR (Fisher's $z = 6.75$, $p < .001$). In both cases, decoupling of recall from temporal contiguity occurred in situations where participants were not intentionally using temporally-based control processes.

In summary, when temporal cues are unhelpful or detrimental to performance, control processes may diminish the TCE, and those with stronger control processing may display a smaller TCE. However, for tasks where the use of order information tends to improve performance, participants with the strongest ability to engage in control processes typically do utilize temporal information to guide recall, whereas those with impaired cognitive control make less use of temporal information, resulting in a lower TCE. This suggests that incorporating control processes into retrieved context theory is an important next step in theory development.

Conclusions

The results of this study support the retrieved context theory account of episodic memory and temporal contiguity. Consistent with the predictions of retrieved context theory, a significant TCE was observed, regardless of encoding intentionality or the timescale of item presentation. A model based on the theory was able to simultaneously fit patterns of high overall recall but modest temporal contiguity in the data. Finally, we demonstrated that the observed temporal contiguity was not a result of recency alone, consistent with the theory's qualitative claim. By fitting constrained submodels, we demonstrated that context drift during recall as a result of context reinstatement, as well as context drift during encoding, are a critical components of memory search. A notable observation, however, is the difference in the size and shape of the TCE between encoding conditions. Specifically, participants in the intentional encoding conditions tended to have a higher probability of making $|lag| = 1$ transitions.

The fact that the TCE is larger under intentional encoding may be due in large part to the use of strategic control processes to tune parameters of the memory system. A reduced TCE and a decreased correlation between overall recall and the TCE is generally observed when capacity for engaging in control processes is impaired, such as when participants are unaware their memory will be tested, they have cognitive impairments, or they are instructed to use an alternate strategy. Considering control processes as a parameter tuning mechanism may improve the ability of retrieved context theory to explain a range of findings.

References

Averell, L., Prince, M., & Heathcote, A. (2016). Fundamental causes of systematic and random variability in recognition memory. *Journal of Memory and Language*, *88*, 51–69. https://doi.org/10.1016/j.jml.2015.12.010

Bhatarah, P., Ward, G., & Tan, L. (2006). Examining the relationship between free recall and immediate serial recall: The effect of concurrent task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 215–229. https://doi.org/10.1037/0278-7393.32.2.215

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*(2), 173–189. https://doi.org/10.1016/0010-0285(74)90009-7

Bradley, M. M., & Glenberg, A. M. (1983). Strengthening associations: Duration, attention, or relations? *Journal of Verbal Learning and Verbal Behavior*, *22*(6), 650–666. https://doi.org/10.1016/S0022-5371(83)90385-7

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Campbell, K. L., & Hasher, L. (2018). Hyper-binding only apparent under fully implicit test conditions. *Psychology and Aging*, *33*(1), 176–181. https://doi.org/10.1037/pag0000216

Campbell, K. L., Trelle, A., & Hasher, L. (2014). Hyper-binding across time: Age differences in the effect of temporal proximity on paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 293–299. https://doi.org/10.1037/a0034109

Caplan, J. B., Glaholt, M., & McIntosh, A. R. (2006). Linking associative and list memory: Pairs versus triples. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *32*(6), 1244–1265. https://doi.org/10.1037/0278-7393.32.6.1244

Cortis Mack, C., Cinel, C., Davies, N., Harding, M., & Ward, G. (2017). Serial position, output order, and list length effects for words presented on smartphones over very long intervals. *Journal of Memory and Language*, *97*, 61–80. https://doi.org/10.1016/j.jml.2017.07.009

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

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. https://doi.org/10.3758/pbr.15.1.64

Diamond, N., & Levine, B. (2019). Linking details to temporal context reinstatement in recall of real-world experiences. *PsyArXiv*. https://doi.org/10.31234/osf.io/e2unt

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. https://doi.org/10.1016/j.neuron.2014.01.042

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

Gibson, B. S., Healey, M. K., & Gondoli, D. M. (2019). Modeling the effect of attention-deficit hyperactivity disorder on episodic memory. *Journal of Abnormal Psychology*, *128*(2), 173–183. https://doi.org/10.1037/abn0000393

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*(4), 355–369. https://doi.org/10.1037/0278-7393.6.4.355

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). Oxford University Press.

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., 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. https://doi.org/10.1037/a0036306

Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General*, *143*(2), 575–596. https://doi.org/10.1037/a0033715

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., & Kahana, M. J. (2020). Age-related changes in the dynamics of memory encoding processes provide a biomarker of successful aging. *PLoS ONE*, *15*(1), e0227274. https://doi.org/10.1371/journal.pone.0227274

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., & Uitvlugt, M. G. (2019). The role of control processes in temporal and semantic contiguity. *Memory & Cognition*, *47*(4), 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

Hogan, R. M. (1975). Interitem encoding and directed search in free recall. *Memory & Cognition*, *3*(2), 197–209. https://doi.org/10.3758/BF03212898

Hong, M. K., Fazio, L. K., & Polyn, S. M. (2019, November). *Examining factors that eliminate contiguity in free recall* [Poster presentation]. 60th Annual Meeting of the Psychonomic Society, Montreal, Quebec, Canada.

Howard, M. W. (2004). Scaling behavior in the temporal context model. *Journal of Mathematical Psychology*, *48*(4), 230–238. https://doi.org/10.1016/j.jmp.2004.03.004

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*(4), 923–941. https://doi.org/10.1037/0278-7393.25.4.923

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

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

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109. https://doi.org/10.3758/BF03197276

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 530–540. https://doi.org/10.1037/0278-7393.28.3.530

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation*, *61*, 237–284. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

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 & Cognition*, *42*(2), 202–214. https://doi.org/10.1037/xlm0000172

Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology*, *34*(5-6), 419–426. https://doi.org/10.1080/002075999399774

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., & Kahana, M. J. (2014). Compound cuing in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 12–24. https://doi.org/10.1037/a0033698

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

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*(5), 649–653. https://doi.org/10.1016/S0022-5371(72)80049-5

Mensink, G.-J. M., & Raaijmakers, J. G. W. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology*, *33*(2), 172–186. https://doi.org/10.1016/0022-2496(89)90029-1

Miller, J. F., Kahana, M. J., & Weidemann, C. T. (2012). Recall termination in free recall. *Memory & Cognition*, *40*(4), 540–550. https://doi.org/10.3758/s13421-011-0178-9

Moreton, B. J., & Ward, G. (2010). Time scale similarity and long-term memory for autobiographical events. *Psychonomic Bulletin & Review*, *17*(4), 510–515. https://doi.org/10.3758/PBR.17.4.510

Morton, N. W., & Polyn, S. M. (2016). A predictive framework for evaluating models of semantic organization in free recall. *Journal of Memory and Language*, *86*, 119–140. https://doi.org/10.1016/j.jml.2015.10.002

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482–488. https://doi.org/10.1037/h0045106

Murdock, B. B. (1965). Effects of a subsidiary task on short-term memory. *British Journal of Psychology*, *56*(4), 413–419. https://doi.org/10.1111/j.2044-8295.1965.tb00983.x

Murdock, B. B. (1974). *Human memory: Theory and data*. Lawrence Erlbaum.

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. https://doi.org/10.1016/j.jml.2016.10.009

Neath, I. (1993). Contextual and distinctive processes and the serial position function. *Journal of Memory and Language*, *32*(6), 820–840. https://doi.org/10.1006/jmla.1993.1041

Osth, A., & Fox, J. (2019). Are associations formed across pairs? a test of learning by temporal contiguity in associative recognition. *Psychonomic Bulletin & Review*, *26*(5), 1659–1656. https://doi.org/10.3758/s13423-019-01616-7

Pajkossy, P., Keresztes, A., & Racsmány, M. (2017). The interplay of trait worry and trait anxiety in determining episodic retrieval: The role of cognitive control. *The Quarterly Journal of Experimental Psychology*, *70*(11), 2234–2250. https://doi.org/10.1080/17470218.2016.1230142

Palombo, D. J., Di Lascio, J. M., Howard, M. W., & Verfaellie, M. (2019). Medial temporal lobe amnesia is associated with a deficit in recovering temporal context. *Journal of Cognitive Neuroscience*, *31*(2), 236–248. https://doi.org/10.1162/jocn_a_01344

Phillips, J. L., Shiffrin, R. J., & Atkinson, R. C. (1967). The effects of list length on short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *6*(3), 303–311. https://doi.org/10.1016/S0022-5371(67)80117-8

Polyn, S. M., Erlikhman, G., & Kahana, M. J. (2011). Semantic cuing and the scale-insensitivity of recency and contiguity. *Journal Experimental Psychology: Learning, Memory & Cognition*, *37*(3), 766–775. https://doi.org/10.1037/a0022475

Polyn, S. M., McCluey, J. D., Morton, N. W., Woolard, A. A., Luksik, A. S., & Heckers, S. (2015). Temporal context and the organisational impairment of memory search in schizophrenia. *Cognitive Neuropsychiatry*, *20*(4), 296–310. https://doi.org/10.1080/13546805.2015.1031372

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

Postman, L. (1971). Organization and interference. *Psychological Review*, *78*(4), 290–302. https://doi.org/10.1037/h0031031

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. https://doi.org/10.1037/0033-295X.88.2.93

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

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. https://doi.org/10.1016/B978-0-12-407187-2.00004-6

Sahakyan, L., & Kwapil, T. R. (2018). Moving beyond summary scores: Decomposing free recall performance to understand episodic memory deficits in schizotypy. *Journal of Experimental Psychology: General*, *147*(12), 1919–1930. https://doi.org/10.1037/xge0000401

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 & Cognition*, *40*(1), 86–93. https://doi.org/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*(11), 898–904. https://doi.org/10.1111/j.1467-9280.2005.01634.x

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., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893–912. https://doi.org/10.1037/a0013396

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. https://doi.org/10.3758/MC.38.6.689

Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal–contextual retrieval from episodic memory. *Journal Experimental Psychology: Learning, Memory & Cognition*, *37*(6), 1532–1539. https://doi.org/10.1037/a0024852

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1589–1626. https://doi.org/10.1037/0278-7393.26.6.1589

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. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory and Language*, *59*(2), 223–236. https://doi.org/10.1016/j.jml.2008.04.002

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), 768–780. https://doi.org/10.1037/pag0000049

West, R. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*(2), 272–292. https://doi.org/10.1037/0033-2909.120.2.272

(*Appendices follow*)

Appendix A

Model Details

There are many models based on retrieved context theory (e.g., Healey & Kahana, 2016; Howard & Kahana, 2002; Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008). All share the underlying mechanisms of context drift and reinstatement. Here, we describe the implementation used in the current article, a version of the temporal context model (Howard & Kahana, 2002; Sederberg et al., 2008).

Model Structure

In the model, two types of cognitive representations interact: the feature representation (*F*), a high-dimensional space in which the features of the current list item are activated, and the context representation (*C*), a corresponding space in which the current state of context is activated. Activation of specific items or contexts within each of these representational spaces is defined as a vector, (*f*) and (*c*) respectively. These vectors have one element for each list item (12 items for the current study), one element for each delay period (one for the DFR and 12 for the CDFR conditions), and one element to represent the state of context prior to presentation of the first item.

Each association matrix is a weighted sum of a preexperimental component (M^{FC}_{pre} and M^{CF}_{pre}) that reflects longstanding semantic relationships and an experimental component (M^{FC}_{exp} and M^{CF}_{exp}) that reflects new learning that occurs during the experiment. Because we were not interested in studying semantic associations here and because the lists were generated randomly without consideration of semantics, we followed the practice of previous work and initialized M^{FC}_{pre} and M^{CF}_{pre} as identity matrices (i.e., items are associated only with their own context element and vice versa).

Encoding

Studying an item from serial position *i* activates the corresponding features, $\mathbf{f}_i$, which in turn retrieve the context states to which those features have previously been associated:

$$\mathbf{c}_i^{\mathrm{IN}} = \frac{M^{FC}\mathbf{f}_i}{||M^{FC}\mathbf{f}_i||}. \tag{2}$$

This retrieved context, $\mathbf{c}_i^{\mathrm{IN}}$, which is normalized to have a length of one, is incorporated into the context representation by adding it to the current context vector $\mathbf{c}_{i-1}$. The context vector is continuously maintained at unit length. Therefore, when a new state of context is added to the existing state, the two vectors, $\mathbf{c}_{i-1}$ and $\mathbf{c}_i^{\mathrm{IN}}$ must be scaled so their sum has a length of one:

$$\mathbf{c}_i = \rho_i \mathbf{c}_{i-1} + \beta \mathbf{c}_i^{\mathrm{IN}}. \tag{3}$$

where β is a model parameter governing how quickly context changes, and ρ_i is chosen such that $||\mathbf{c}_i|| = 1$:

$$\rho_i = \sqrt{1 + \beta^2[(\mathbf{c}_{i-1} \cdot \mathbf{c}_i^{\mathrm{IN}})^2 - 1]} - \beta(\mathbf{c}_{i-1} \cdot \mathbf{c}_i^{\mathrm{IN}}). \tag{4}$$

Because context is always of unit length it can be thought of as a point on the surface of a (hyper) sphere, with β determining how far along the surface of the sphere it travels with each newly presented item and $\mathbf{c}_i^{\mathrm{IN}}$ determining the direction of travel. During encoding, β is set to an encoding-specific value, β_{enc}.

At the start of an experimental session, the experimental associations are initialized to zero. As each new item is presented, new experimental associations are formed, both between the item's feature representation and the current state of context (stored in M^{FC}_{exp}) and between the current state of context and the item's feature representation (stored in M^{CF}_{exp}). These associations are formed according to a Hebbian outer-product learning rule:

$$\begin{aligned} \Delta M^{FC}_{exp} &= \mathbf{c}_{i-1}\mathbf{f}_i^{\top} \\ \Delta M^{CF}_{exp} &= \mathbf{f}_i\mathbf{c}_{i-1}^{\top}\phi_i, \end{aligned} \tag{5}$$

ϕ_i simulates increased attention to beginning-of-list items, producing a primacy effect, by scaling the magnitude of context-to-feature associations across the list:

$$\phi_i = \phi_s e^{-\phi_d(i-1)} + 1, \tag{6}$$

where ϕ_s and ϕ_d are model parameters. See Sederberg et al. (2008) for a more complete discussion.

Newly formed experimental associations are incorporated with preexperimental associations. The balance between new and existing associations is controlled by parameters γ_{FC} and γ_{CF}:

(*Appendices continue*)

Table A1

Best-Fit Parameter Values for the Fits of the Temporal Context Model to the Data of Each Condition

Parameter	Intentional DFR	Incidental DFR	Intentional CDFR	Incidental CDFR
ϕ_s	9.521	6.325	8.989	10.854
ϕ_d	1.024	5.384	0.872	4.818
γ_{fc}	0.158	0.093	0.291	0.988
γ_{cf}	0.569	0.288	0.775	0.814
β_{enc}	0.860	0.944	0.769	0.945
β_{rec}	0.814	0.316	0.319	0.010
$\beta_{distract}$	0.367	0.192	0.435	0.701
θ_s	0.402	0.512	0.408	0.281
θ_r	0.035	0.009	0.055	0.140
τ	0.193	0.113	0.554	0.220
RMSD	0.0081	0.0097	0.0095	0.0093

Note. The model was simultaneously fit to the serial position curve and lag-conditional response probability curve. DFR = delayed free recall; CDFR = continual distractor free recall; RMSD = root-mean-square deviation.

$$M^{FC} = (1-\gamma_{FC})M^{FC}_{pre} + \gamma_{FC}M^{FC}_{exp}$$
$$M^{CF} = (1-\gamma_{CF})M^{CF}_{pre} + \gamma_{CF}M^{CF}_{exp}. \tag{7}$$

For any distractor, whether it occurs between items or at the end of the list, the context is updated in the same way context was updated during encoding (Equation 4), but with a different drift rate parameter, $\beta_{distract}$.

Recall

The recall period proceeds as a series of retrieval attempts closely following the implementation used by Morton and Polyn (2016). At each retrieval attempt, the model either successfully retrieves an item or fails. After a failure, no further retrieval attempts are made. The probability of failing to retrieve an item (i.e., stopping recall) starts out low for the first recall attempt (i.e., output position) and increases exponentially with each output position:

$$P(stop, j) = \theta_s e^{j\theta_r}. \tag{8}$$

where j is the output position, θ_s is a parameter which determines the scaling of the exponential function, and θ_r is a parameter which controls the rate at which the probability of stopping approaches 1.

If recall does not stop at a particular output position, the current contextual state is used to cue retrieval via the M^{CF} associations:

$$\mathbf{a} = M^{CF}\mathbf{c}_t, \tag{9}$$

The resulting **a** gives the degree of support, or activation, for each item in the list. These activations are then used to assign each item a probability of being selected for recall according to:

Table A2

Average (SEM in Parentheses) Best-Fit Parameter Values for 30 Fits of the Temporal Context Model to the Data of Each Condition

Parameter	Intentional DFR	Incidental DFR	Intentional CDFR	Incidental CDFR
ϕ_s	9.827 (0.447)	2.052 (0.367)	10.457 (0.398)	12.125 (0.343)
ϕ_d	1.28 (0.051)	6.148 (0.451)	0.782 (0.016)	5.986 (0.421)
γ_{fc}	0.734 (0.009)	0.163 (0.017)	0.287 (0.020)	0.016 (0.002)
γ_{cf}	0.464 (0.040)	0.420 (0.042)	0.479 (0.042)	0.563 (0.062)
β_{enc}	0.227 (0.010)	0.132 (0.018)	0.320 (0.018)	0.786 (0.031)
β_{rec}	0.800 (0.025)	0.506 (0.033)	0.828 (0.024)	0.714 (0.031)
$\beta_{distract}$	0.910 (0.004)	0.930 (0.005)	0.715 (0.026)	0.735 (0.055)
θ_s	0.377 (0.009)	0.399 (0.013)	0.245 (0.008)	0.358 (0.019)
θ_r	0.047 (0.004)	0.054 (0.006)	0.049 (0.003)	0.108 (0.009)
τ	0.155 (0.004)	0.110 (0.005)	0.529 (0.018)	0.231 (0.005)
RMSD	0.0086 (0.0301)	0.0101 (0.0291)	0.0106 (0.0394)	0.0101 (0.0444)

Note. The model was simultaneously fit to the serial position curve and lag-conditional response probability curve. DFR = delayed free recall; CDFR = continual distractor free recall; RMSD = root-mean-square deviation.

(*Appendices continue*)

Table A3
Best-Fit Parameter Values for the Fits of the Full Temporal Context Model to the Serial Position Curve and Temporal Bias Scores of Each Condition

Parameter	Incidental DFR	Incidental CDFR
ϕ_s	7.400	13.661
ϕ_d	8.311	12.571
γ_{fc}	0.529	0.041
γ_{cf}	0.614	0.959
β_{enc}	0.139	0.499
β_{rec}	0.393	0.797
$\beta_{distract}$	0.848	0.390
θ_s	0.436	0.116
θ_r	0.037	0.210
τ	0.206	0.337
RMSD	0.030	0.0253

Note. DFR = delayed free recall; CDFR = continual distractor free recall; RMSD = root-mean-square deviation.

$$P(i) = (1 - P(stop))\frac{\mathbf{a}_i^\tau}{\sum_k^N \mathbf{a}_k^\tau}, \tag{10}$$

where τ is a parameter that determines how sensitive the model is to differences among items in level of support—when τ is large, the model strongly prefers the item with the highest activation on **a**; when it is small, less well-supported items have a greater chance of winning. In computing $P(i)$, each element of **a** is set to a minimum value of 10^{-7} to ensure no item is assigned a zero-recall probability.

Table A4
Best-Fit Parameter Values for the Fits of the Temporal Context Model to the Serial Position Curve and Temporal Bias Scores of Each Condition When Context Drift is Eliminated During Encoding

Parameter	Incidental DFR	Incidental CDFR
ϕ_s	0.157	8.417
ϕ_d	1.395	1.953
γ_{fc}	0.866	0.670
γ_{cf}	0.016	0.022
β_{enc}	0	0
β_{rec}	0.264	0.163
$\beta_{distract}$	0.084	0
θ_s	0.119	0.087
θ_r	0.219	0.294
τ	29.191	23.146
RMSD	0.085	0.177

Note. Context drift during encoding was eliminated for delayed free recall (DFR) by setting β_{enc}, the parameter that controls the rate of context drift during encoding, to zero. In continual distractor free recall (CDFR), both $\beta_{distract}$ and β_{enc} were set to zero to prevent any drift during or between item presentations. RMSD = root-mean-square deviation.

Table A5
Best-Fit Parameter Values for the Fits of the Temporal Context Model to the Serial Position Curve and Temporal Bias Scores of Each Condition When Context Drift is Eliminated During Recall by Setting β_{rec} = 0

Parameter	Incidental DFR	Incidental CDFR
ϕ_s	19.460	4.482
ϕ_d	5.831	8.339
γ_{fc}	0.411	0.571
γ_{cf}	0.804	0.302
β_{enc}	0.683	0.723
β_{rec}	0	0
$\beta_{distract}$	0.917	0.129
θ_s	0.373	0.442
θ_r	0.054	0.068
τ	0.209	0.603
RMSD	0.0716	0.0616

Note. DFR = delayed free recall; CDFR = continual distractor free recall; RMSD = root-mean-square deviation.

The recalled item's representation is reactivated on the feature layer, allowing the model to retrieve the contextual state associated with the item. Context is updated using the same mechanism used during the study period (separate parameters, β_{enc}, $\beta_{distract}$, and β_{rec} control the rate of context drift during encoding, the distractor task, and retrieval). This updated context is then used as a cue for the next recall. The cycle of cue-recall-update context-cue continues until the model fails to recall an item (Equation 8).

To fit the model for a given condition, we attempted to minimize the root-mean-square deviation (RMSD) between the condition's across-participant average in the actual data and the model's simulated data. There were k = 10 free parameters in this model. At each generation, 3,000 simulated participants, each with a different set of parameter values, studied and recalled three lists. We ran this entire procedure 30 times for each condition. The best fitting parameter values and the RMSD values across the 30 model fits for each condition are listed in Table A1. The average parameter values and standard error of the mean (*SEM*) across the 30 model fits are presented in Table A2. To generate simulated data for the figures, we used the best-fitting parameter sets to simulate recalls for 60,000 simulated participants per condition (each studying one list).

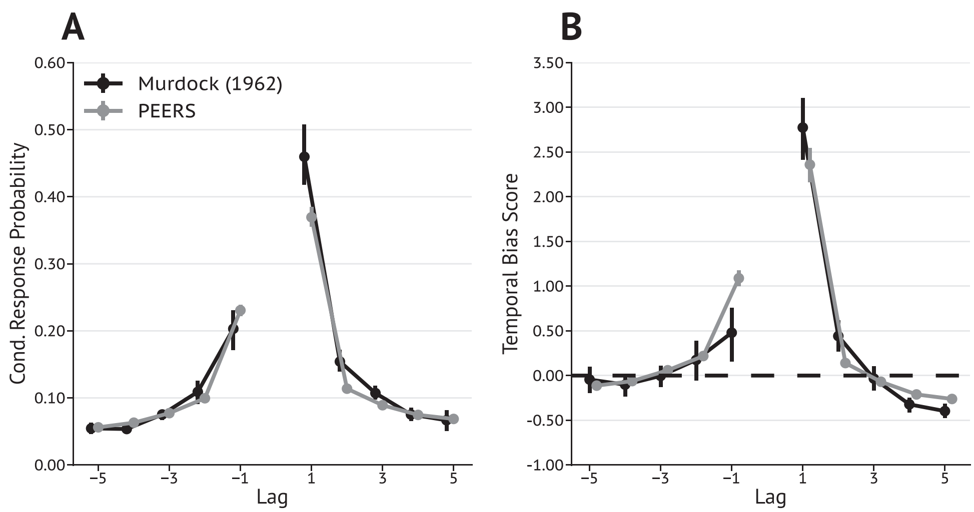
(*Appendices continue*)

Appendix B

Temporal Bias Scores in Archival Data Sets

Figure B1

Lag Conditional Response Probabilities (Lag-CRPs) and Temporal Bias Scores in Archival Data Sets



Note. (A) Lag-CRP and (B) temporal bias curves for two large archival datasets. The black line represents data from the list length 15, 2-s presentation rate condition of Murdock (1962). The gray line represents data from Experiment 1 of the Penn Electrophysiology of Encoding and Retrieval Study (PEERS). Temporal bias scores were calculated for each lag by counting the number of times a participant actually made a transition of that lag and the number of times such a transition would be expected by chance. The expected count was determined by permuting the order of the recalled items 500 times and counting how many times each lag occurred for each permutation. Temporal bias is the difference between the actual count and average expected count, divided by the average expected count. The dotted line for the temporal bias scores indicates a score of zero, where there is no difference between the actual and expected counts. Error bars indicate bootstrapped 95% confidence intervals.

We applied the temporal bias analysis to data sets where a strong TCE has been observed to show what temporal bias curves look like in data with a typical TCE. In Figure B1, the lag-CRP and temporal bias curve for two data sets are presented: the list length 15, 2-s presentation rate condition from Murdock (1962) and the Penn Electrophysiology of Encoding and Retrieval Study Experiment 1 (Healey et al., 2019). In both of these data sets, the participants were well-practiced and as a result display a typical lag-CRP, with particularly high scores for *lag* = +1 and lower scores for farther lags. When participants are experienced with a list learning task, they tend to have a stronger TCE (Healey et al., 2019). As such, it is unsurprising that the data sets represented in Figure B1 display a TCE that is strong even without the influences of recency, as evidenced by the strong bias for near lags in temporal bias scores.

Received December 12, 2019
Revision received July 17, 2020
Accepted November 20, 2020 ■