

Improving Memory Search Through Model-Based Cue Selection



Charlotte A. Cornell¹, Kenneth A. Norman²,
Thomas L. Griffiths^{2,3}, and Qiong Zhang^{1,4}

¹Department of Psychology, Rutgers University–New Brunswick; ²Department of Psychology, Princeton University; ³Department of Computer Science, Princeton University; and

⁴Department of Computer Science, Rutgers University–New Brunswick

Psychological Science
2024, Vol. 35(1) 55–71
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09567976231215298
www.psychologicalscience.org/PS



Abstract

We often use cues from our environment when we get stuck searching our memories, but prior research has failed to show benefits of cuing with other, randomly selected list items during memory search. What accounts for this discrepancy? We proposed that cues' content critically determines their effectiveness and sought to select the right cues by building a computational model of how cues affect memory search. Participants ($N = 195$ young adults from the United States) recalled significantly more items when receiving our model's best (vs. worst) cue. Our model provides an account of why some cues better aid recall: Effective cues activate contexts most similar to the remaining items' contexts, facilitating recall in an unsearched area of memory. We discuss our contributions in relation to prominent theories about the effect of external cues.

Keywords

cues, memory search, free recall, part-set cuing, computational modeling, open data, open materials, preregistered

Received 2/6/23; Revision accepted 10/30/23

Imagine reminiscing about a vacation you took with family or friends. You share stories about the trip and search your memory to recall specific events and adventures. Although you may recall a fair amount on your own, you may not remember everything and get stuck at some point. Someone may chime in with another experience, helping you continue recalling the vacation. External reminders like this should support memory because retrieval cues can often make previously inaccessible information accessible (e.g., Tulving, 1983). However, several laboratory studies have failed to find a benefit of cuing when individuals freely search their memory (e.g., Allen, 1969; Basden et al., 1977; Rundus, 1973; Slamecka, 1968; Sloman et al., 1991). In these studies, participants performed a free-recall task (Murdock, 1962; Roberts, 1972; Standing, 1973) in which they studied lists of items and—after each list—were asked to recall as many items as possible in any order. In some of the studies, some participants received a random subset of list items as retrieval cues before starting to recall, whereas other participants did not;

surprisingly, the cued participants recalled a smaller proportion of the remaining items than the uncued participants did (Basden et al., 1977; Rundus, 1973; Slamecka, 1968; Sloman et al., 1991). In other free-recall studies, researchers waited until recall was stuck before presenting other, randomly selected list items as cues: Slamecka (1968, Experiment 5) presented half of the not-yet-recalled list items, and Allen (1969, Experiment 1) presented a sixth of the list items, regardless of whether those items had been recalled; neither study observed a benefit of cuing. Taken together, this body of work shows a negative effect of randomly selected cues (at worst) or no benefit (at best). The goal of the present work was to examine the lack of benefit from cuing and explore how cues can be provided to improve memory search.

Corresponding Author:

Qiong Zhang, Rutgers University–New Brunswick, Department of Psychology

Email: qiong.z@rutgers.edu

Whereas in the studies cited above, cues were randomly selected, we argue that the right cues must be selected to facilitate recall. Indeed, researchers who have tested the effect of the content of cues have observed cues to benefit the retrieval process, facilitating access to previously inaccessible memories, specifically when the cues shared high similarity with to-be-recalled items (Basden, 1973; Hudson & Austin, 1970; Kroeger et al., 2019; Roediger, 1973; Tulving & Pearlstone, 1966). For example, participants who received category names before recalling a categorized list remembered more items than uncued participants (Tulving & Pearlstone, 1966). In uncategorized lists, participants who received items from odd serial positions as cues after repeatedly studying the list recalled more than uncued participants (Basden, 1973). That is, these studies obtained a benefit of cues by selecting either cues with high semantic similarity to list items or cues studied nearby in time to list items (Basden, 1973; Hudson & Austin, 1970; Roediger, 1973; Tulving & Pearlstone, 1966). However, these studies could not provide the most effective cues because they did not quantify the exact amount of similarity among items nor did they simultaneously integrate the contribution of semantic versus temporal information. Furthermore, in a free-recall task in which participants make multiple retrieval attempts, considering the similarity between cues and remaining items alone is not sufficient because the effect of cues on the full trajectory of the recall sequence must also be considered.

In the present work, we provide a principled way for selecting retrieval cues when recall gets stuck during free recall, by formally estimating items' semantic and temporal representations in a memory space and mathematically describing the dynamics of how items are encoded and later recalled from this space. To achieve this, we extended an existing model of memory search, the context maintenance and retrieval (CMR) model, which organizes memories into a latent context space. The model posits that the current location in the context space is where new information is encoded during study and is what drives the next recall during retrieval. CMR has been shown to capture a range of behavioral patterns in standard free-recall tasks (Howard & Kahana, 2002a; Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008). However, it has not been extended to situations with external cues. We thus proposed to extend CMR to model how cue presentation affects memory search after initial recall ends (a process that we refer to as cued memory search). Importantly, when one's current context location can no longer drive further recalls and an external cue is presented, we proposed that there is an additional cognitive process involved.

Statement of Relevance

People often use information from their environment when they need help remembering things. Here, we sought to develop an automated way of generating useful reminders when memory search gets stuck. To do this, we built a computational model of memory search that predicted the effectiveness of specific cues, and we integrated it into our live experiment. Our model was able to successfully select cues that were more (vs. less) helpful by predicting how memories would be organized into a "memory space" and then choosing cues that activated parts of this space containing not-yet-retrieved memories. These results provide new insights into how to restart memory when recall fails, and they provide a theoretical foundation for future systems that enhance human performance by selecting effective retrieval cues.

In this additional process, one's current context location is replaced with the cue's context location, and this updated context location guides future recall attempts. The most effective cue is the one that leads to the most recalls (determined by model simulations) after the context location updates.

To test the effectiveness of our proposed model in selecting cues, we followed experiments that presented cues after recall ended (Allen, 1969; Slamecka, 1968). We let participants begin a free-recall task, and when participants had difficulty remembering items on their own, they could self-request the presentation of a single, not-yet-retrieved list item (see Fig. 1). The key innovation of our paradigm is that we integrated the model into a live experiment to predict in real time (accounting for initially recalled items on a trial) the effect of possible retrieval cues in continuing recall of the remaining items. This allowed us to present our model's best or worst cues in addition to random cues (see Fig. 1). We predicted that using the model to select effective cues would help participants recall additional items in response to the cue.

In the remainder of this article, we first introduce our model of cued memory search and its predictions for the effect of retrieval cues. We then test our model's ability to select cues in real time. We demonstrate that cues facilitate further recall and that our model can both capture the observed effect of cues and provide an account as to why some cues are better at aiding recall.

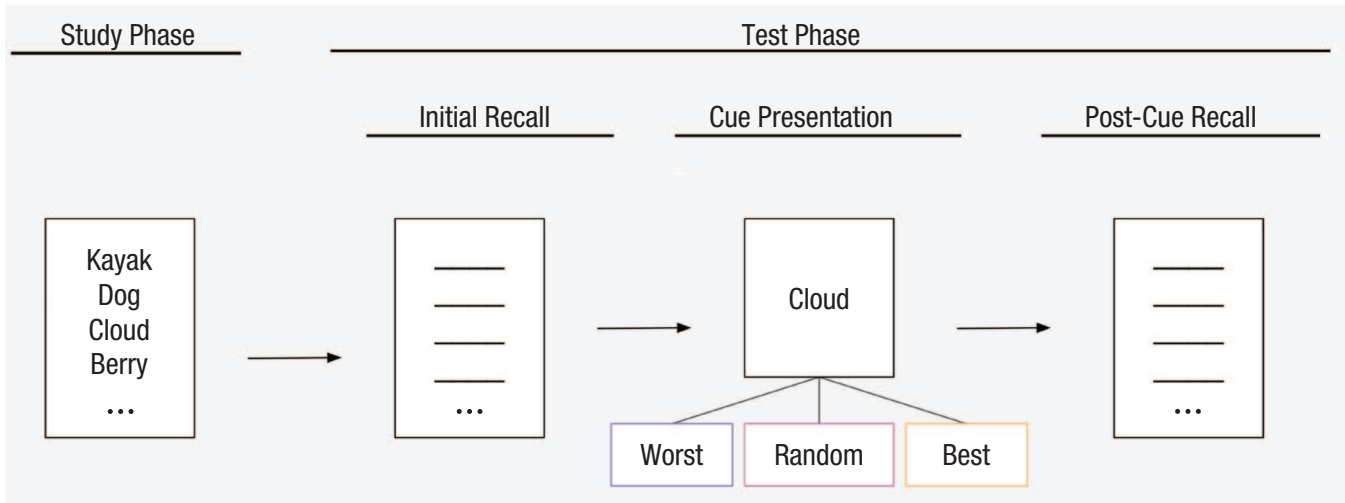


Fig. 1. Our experimental design. Participants first study a list of words and then immediately begin the test phase. During initial recall, participants recall as many items as they can. When they cannot recall anymore, they can self-request a cue (selected by our model from the remaining list items) and continue recall in a post-cue recall phase. To select cues, we developed a model of cued memory search that predicts the effect of cues on memory performance: We hypothesized that memory performance given the model's best cue will be better than a randomly selected cue, which in turn will be better than the model's worst cue.

Method

A model of cued memory search

To build a model of cued memory search, in this section, we reviewed the CMR model, which was developed to account for behavioral patterns in the free-recall task without the presence of external cues (CMR; Lohanas et al., 2015; Polyn et al., 2009; for related work, see Howard & Kahana, 2002a). Our implementation of CMR matches the version used by Polyn et al. (2009) with a simplified retrieval rule following Zhang et al. (2023). Next, we propose an extension to the CMR model that accounts for the effect of cue presentation on memory search after recall ends. Last but not least, we introduce our proposed model's predictions about which cues are more beneficial than others as well as how the predictions can be used to deliver external cues to improve individuals' memory search performance in real time.

The CMR model. As participants study a list of items one after another in a free-recall task, CMR proposes that their context slowly drifts toward the memory representations of recently encountered experiences. The state of the context at time t is given by

$$c_t = \rho c_{t-1} + \beta c^{IN}, \quad (1)$$

where c^{IN} is the retrieved context induced by an encountered experience, $\beta \in [0, 1]$ is a parameter determining the rate at which context drifts toward the new experience, and ρ is a scalar ensuring $\|c_t\| = 1$. When

an item is presented in the study list, it activates its preexperimental context c^{IN} :

$$c^{IN} = M_{pre}^{FC} f_t, \quad (2)$$

where M_{pre}^{FC} represents item-to-context associations that existed prior to the experiment (initialized as an identity matrix, under the simplifying assumption that an item is associated only with its own context; see Polyn et al., 2009), and f_t is a binary vector that is all zeros except at the presented item's position. Therefore, $M_{pre}^{FC} f_t$ is the context previously associated with the presented item. In addition to these fixed preexperimental item-to-context associations held in M_{pre}^{FC} , there are also experimental item-to-context and context-to-item associations held in M_{exp}^{FC} and M_{exp}^{CF} that capture new learning in the experiment. These matrices are initialized to zero and are updated during the study phase. Specifically, when an item is presented, a new association is formed between the presented item and the current context state via the Hebbian outer-product learning rule:

$$\Delta M_{exp}^{FC} = \Delta M_{exp}^{CF} = f_t c_{t-1}^T. \quad (3)$$

The model assumes that an item is always associated with the current context successfully (without encoding noise or encoding failure). The overall effect of having context drift toward the retrieved context of presented items in Equations 1 and 2, together with associative learning in Equation 3, is that each item is embedded at a location in the context space corresponding to the representations of other recently encountered items.

During recall, context continues to drift following the same process during study according to Equation 1, but with c^{IN} expressed differently. During study, an item activates its preexperimental context only when it is presented; however, at recall, the retrieved context of a recalled item activates both its preexperimental context ($M_{pre}^{FC}f_i$) and its experimental context formed during study ($M_{exp}^{FC}f_i$). The extent of retrieving an item’s preexperimental versus experimental context is determined by a parameter, $\gamma_{fc} \in [0,1]$:

$$c_{rec}^{IN} = (1 - \gamma_{fc})M_{pre}^{FC}f_i + \gamma_{fc}M_{exp}^{FC}f_i. \quad (4)$$

When context drifts toward this retrieved context, which items are likely to be recalled? The support (or activation) f^{IN} at time t for recalling different items depends on both how much the current context c_t matches with items’ experimental contexts (stored in M_{exp}^{CF}) as well as items’ preexperimental contexts (stored in M_{pre}^{CF} ; see description below). The relative activation of these associations is determined by a parameter, $\gamma_{cf} \in [0,1]$, such that

$$f^{IN} = \gamma_{cf}\phi_t M_{exp}^{CF}c_t + (1 - \gamma_{cf})M_{pre}^{CF}c_t. \quad (5)$$

Here, in addition to M_{exp}^{CF} being scaled by γ_{cf} , it is also scaled by ϕ_t to simulate increased attention to beginning-of-list items, allowing the model to capture the primacy effect. Specifically, $\phi_t = \phi_s e^{-\phi_s(i-1)} + 1$, which creates a primacy gradient by determining the magnitude of context-to-item associations according to item i ’s serial position in the study list (Lohnas et al., 2015; Polyn et al., 2009). M_{pre}^{CF} , representing preexperimental context-to-item associations, is initialized as an identity matrix such as M_{pre}^{FC} . Each element in M_{pre}^{CF} , with indices m and n , is additionally incremented by a semantic association between items m and n , determined by taking the cosine similarity of the two items’ GloVe model embeddings (derived on the basis of the words’ co-occurrence in large text corpora; Pennington et al., 2014) and scaling it by a parameter s_{cf} , to capture semantic clustering effects observed at recall. Put together, items’ activations in f^{IN} are determined not only by the experimental associations formed during study but also by attentional scaling (to capture the primacy gradient) and by the approximated semantic associations held by individuals prior to the experiment.

Finally, to be able to fully simulate recall patterns based on items’ support in f^{IN} , the model also needs a retrieval rule and a stopping rule. We used the softmax function as the retrieval rule, $p_i = \frac{e^{kf_i^{IN}}}{\sum_j e^{kf_j^{IN}}}$, where

f_i^{IN} is the support to retrieve item i and the parameter k determines the amount of noise during retrieval.

When an item is retrieved, the context state drifts; cuing with this updated context state supports the retrieval of new items. This retrieval process continues until determined by the stopping rule: The probability of stopping at each time point is expressed as $p_{stop} = e^{-\epsilon_d f_{nr}^{IN} / f_r^{IN}}$, where f_r^{IN} indicates the summed support for already-recalled items, f_{nr}^{IN} indicates the summed support for not-yet-recalled items, and ϵ_d is a scaling factor (Kragel et al., 2015; Zhang et al., 2023).

Overall, because items studied nearby in the list are tied to similar context states during encoding, subsequent recalls are likely to be nearby items in the study list; they are additionally likely to be items that are semantically similar to the current context. Intuitively, consider the vacation example again: Adventures that occurred nearby in time are related in memory, and the act of recalling one event leads to the likelihood of recalling a similar event because of the shared timeline and/or semantics.

Our proposed model: extending CMR for cued memory search.

Whereas CMR can account for behavioral patterns in a standard free-recall task, our proposed model aimed to account for recall patterns in our paradigm in Figure 1, where—after studying a list of items and recalling as much as possible—participants could self-request a cue and then continue their recalls in a post-cue recall phase. There were two major assumptions for our model:

1. When a cue is presented after initial recall ends, the current context state is no longer an effective retrieval cue, and therefore context is updated to fully match the induced context of the cue. That is,

$$c_t = c_{cue}, \quad (6)$$

which is equivalent to letting $\beta = 1$, where $c^{IN} = c_{cue}$ in Equation 1.

2. The remaining processes that govern memory search (characterized by parameters in CMR) are identical between the initial recall phase and the post-cue recall phase.

In other words, the presentation of a cue only temporarily alters the current context state (Assumption 1) but not any fundamental aspects of how memory search proceeds (Assumption 2). Specifically, our model first lets an initial recall session proceed in the same way as a standard free-recall task captured by a CMR model. It then simulates a cue by assuming that, when the end-of-recall context state is no longer useful, the current context is set to the context induced by the presented cue. Following Equation 4 of the CMR model,

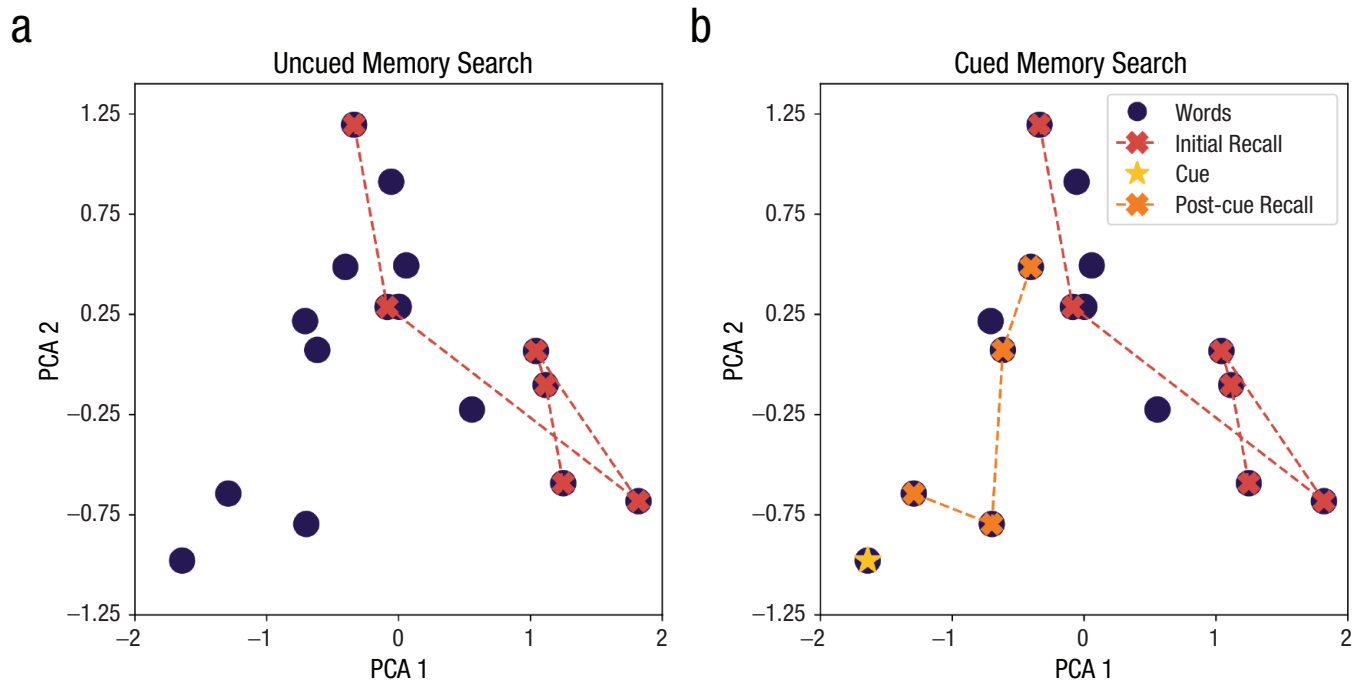


Fig. 2. Models of uncued and cued memory search. (a) Uncued memory search during the initial recall session, simulated by a standard context maintenance and retrieval model. (b) Memory search in our model extension. It captures initial recall; by subsequently using a remaining item in an unsearched context area as a retrieval cue, our model can simulate post-cue recall that continues from the cue. PCA = principal component analysis.

the cue activates both its preexperimental context ($M_{pre}^{FC}f_{cue}$) and its experimental context ($M_{exp}^{FC}f_{cue}$) as determined by the same γ_{fc} parameter:

$$c_{cue} = (1 - \gamma_{fc})M_{pre}^{FC}f_{cue} + \gamma_{fc}M_{exp}^{FC}f_{cue}. \quad (7)$$

The cue can perfectly access its experimental context, following the assumption in CMR that items are always successfully associated with their corresponding encoding contexts. After the current context is updated to the cue's context, a post-cue recall phase proceeds with the same process, behavior, and parameters in a CMR model as the initial recall phase. That is, the model attempts to retrieve items from the remaining subset of list items on that trial, and context continues to drift toward any additionally retrieved items just like before the cue. As in the vacation example when no more experiences can be remembered, your ending context cue (largely composed of the most recently recalled event) is not an adequate reminder of the remaining events. When you listen to a friend chime in with a memory you have not mentioned, your internal context is set to the context induced by hearing your friend's memory. Then after the cue, you attempt to remember more vacation memories just as before being prompted with a remaining experience.

Figure 2 visualizes our model extension compared with the CMR model in a simulated trial. It displays the locations where items are encoded in the context space (reduced to two dimensions using principal component analysis) and their recall trajectories. Figure 2a simulates the initial recall phase of an individual who recalled six words as captured by a CMR model of free recall. Figure 2b shows that, by then presenting a remaining item in an unsearched area of the context space, our model continues memory search from the cue's context and recalls an additional four words.

Model-based cue selection in real time. To be consistent with the paradigms that provide cues at the end of recall (Allen, 1969, Experiment 1; Slamecka, 1968, Experiment 5), we considered only items in the study list as potential cues; more specifically, because participants will have already recalled some list items, we considered only the remaining words as potential cues (as in Experiment 5 by Slamecka, 1968). Unlike those studies that provided a random subset of cues, we used single cues to precisely evaluate the effect of individual cues in the model. Among these remaining words, which cue should we deliver to the participants when requested? We estimated the performance associated with different cues within a trial by simulating our model's post-cue recall session for each

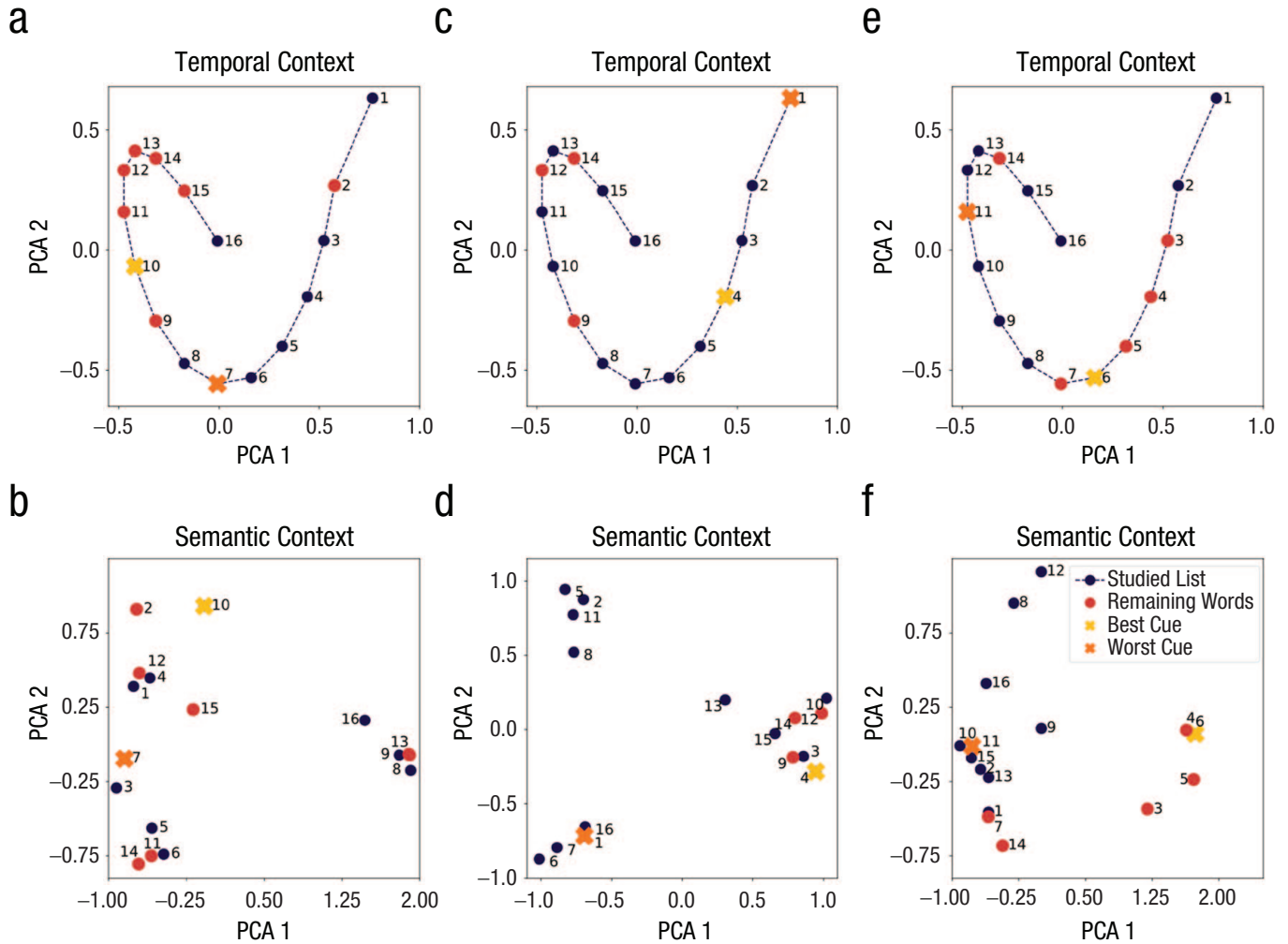


Fig. 3. Contextual similarity of best and worst cues to remaining words. Each column presents an example trial with the temporal (top) and semantic (bottom) representations of a list's context space. Items are labeled with their serial positions in both spaces. In each case, the best cue can access a cluster of remaining words, whereas the worst cue tends to be temporally and semantically dissimilar to the remaining words in the context space. The left column is an example trial in which the best cue facilitates recall of temporally similar words (a) more than semantically similar words (b) because a cluster of temporally similar words remains. The middle column is an example trial in which there is no temporal benefit (c) but the best cue has a semantic benefit (d). The right column is an example trial in which the best cue has both a temporal (e) and semantic (f) benefit. PCA = principal component analysis.

possible cue on that trial. The best retrieval cue is the item with the highest mean recall gain, and the worst retrieval cue is the item with the lowest mean recall gain according to our model simulations. Because recall is driven by items' similarities to the current state of the internal context, the benefit of a cue typically follows from its amount of contextual overlap with the remaining words. Figure 3 visualizes a few trials using principal component analysis of the list items' context representations, comparing the best and worst cues' locations in the temporal and semantic context spaces of the list. As depicted, the best cues are temporally and/or semantically closer to a cluster of remaining words, whereas the worst cues have less context similarity to the remaining words.

For example, sometimes remaining words were studied nearby in time but were from different semantic categories. Figure 3a shows a trial in which the best cue had a similar temporal context to the remaining words, unlike the worst cue, which was encoded earlier in the study phase; in this example, there was no remaining semantic cluster for any cue to access (see Fig. 3b). In another case, the remaining words were semantically similar but were studied at various points in the list. The best cue was within the semantic cluster, but there was no temporal cluster; the worst cue was contextually further from the remaining words in both spaces (see Figs. 3c and 3d). In the most beneficial case, a cue activated a temporally and semantically similar context to the remaining words

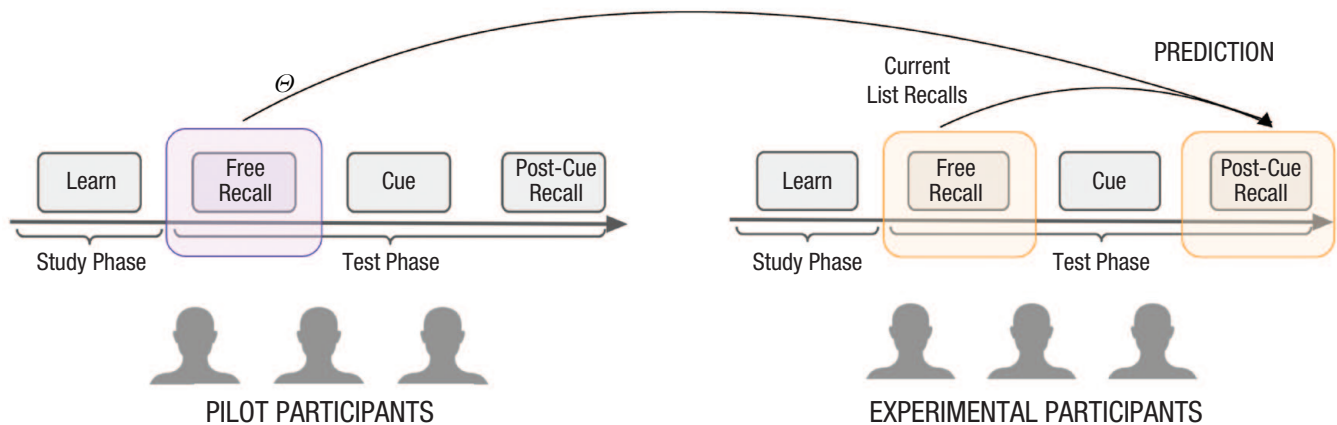


Fig. 4. An illustration of the model fitting and prediction process. We first fitted context maintenance and retrieval to pilot data to obtain a parameter set Θ that could capture pilot participants' free-recall patterns. We preregistered these parameters and then used them in our live experiment in combination with the experimental participants' recalls to predict post-cue recall behavior, that is, demonstrating the ability of the model to generalize over (1) a different group of subjects and (2) a different period of recall.

(see Figs. 3e and 3f). The best cue belonged to temporal and semantic clusters of the remaining items, whereas the worst cue was contextually distant in both cases. Consider the vacation example again: If a friend mentions an event from a day from which you had yet to recall experiences, the event's temporal similarity can prompt retrieval of the day's experiences. If the event was semantically similar to remaining memories, such as a dinner one night, this helps your retrieval of other meals during the trip. Our model predicts that some items are more effective than other items as retrieval cues because their induced contexts are more similar to the contexts of the remaining items.

We integrated this proposed model into our experiment to predict the effect of all possible retrieval cues on each trial and select cues on the basis of their simulated performances. Figure 4 visualizes the process used to achieve this. Importantly, choosing an effective cue in real time (when data are being collected) requires us to know the parameters of CMR in advance. To achieve this, we made an additional Assumption 3: Parameters that govern memory search in CMR are the same in a pilot experiment and the actual experiment we analyzed. We first collected pilot data to fit the model parameters. Because we assumed that memory search after recall termination and cue presentation continues in a similar manner as uncued memory search, we estimated parameters with the pilot data set's initial recall behavior. We obtained a parameter set by using Bayesian optimization to search CMR's parameter space to minimize the normalized root mean square error between the pilot participants' initial recall on cued trials and CMR's simulated recall on those trials across four free-recall behavior patterns: (a) the serial

position curve, (b) the probability of first recall, (c) the conditional response probability, and (d) the semantic similarity probability. Further details about the pilot data set, fitting process, and parameter values can be found in Appendix A. Given our fit model from pilot data (with model parameters preregistered), we integrated it into our experiment to collect this study's data. This real-time system allowed us to use our model to predict, on a trial-by-trial basis, which retrieval cues would benefit recall the most. Achieving this will also demonstrate the ability of the model to generalize over (a) a different group of subjects and (b) a different period of recall.

Open practices statement

The CMR model parameters, experiment, and analyses were preregistered (https://aspredicted.org/16R_Q8Q), and all data (https://osf.io/wh5pg/?view_only=c1862a7008d84fdf9fd221cbe5ad1919) and codes are publicly accessible (https://github.com/charliecornell/self_paced_cues).

Experimental method

Participants. We recruited 200 participants (ages 18–25 years) with Prolific, and they were compensated for taking part in this study. The sample was of convenience, and its size had more than 95% power on the basis of power analysis conducted on pilot data. All participants were fluent English speakers from the United States with at least a 95% approval rating on Prolific, and they gave informed consent. Following our preregistered exclusion criteria, we removed five participants from all analyses for

not attending to at least nine of 12 study phases (i.e., not passing the size-judgment task during encoding described below with at least 50% accuracy). As also preregistered, we further excluded 816 of the 2,340 total trials from all analyses because (a) a cue was not presented, either because 15 or 16 words were initially recalled or a cue was not requested, or (b) a cue was requested less than 10 s into initial recall, where participants might have given up too quickly during the initial recall phase when a cue was available. This left 172 participants with at least one cued trial for the following analyses. This study was approved by the institutional review board of Rutgers, The State University of New Jersey (study ID: Pro2021001945).

Materials. The stimulus set consisted of 326 words and 31 semantic categories, which is a subset of the word pool in the study by Polyn et al. (2011) after removing categories for which different subjects could have vastly different experiences (e.g., college names). Each list had 16 words from four distinct categories with four words per category. Each participant's set of lists was randomly generated, and the list and word order were randomly shuffled. The experiment was implemented in psiTurk (Gureckis et al., 2016) and jsPsych (de Leeuw, 2015).

Design. Each participant completed 14 memory trials (the first two trials being practice rounds and not included in the analysis). For each trial, participants studied 16 words that were sequentially presented on their computer screen for 2 s with a 1-s delay between each word. As each word appeared, participants made a size judgment about the word (bigger or smaller than a shoebox) by pressing either "Q" or "P" on a keyboard to maintain their attention on the encoding task. Directly following the study phase, participants were given 90 s to recall as many words from the list as they could in any order by typing the words into a textbox on their screen.

During recall, participants determined when they could be shown one remaining word from the list by clicking a "Remind Me" button on their screen (note that they were not obliged to click the button). Participants were told to use this word as a reminder of the remaining list items. For the two practice rounds and four experimental trials, the cue was selected randomly. For each of the other eight experimental trials, the best cue or the worst cue was selected on the basis of model simulations of the post-cue recall performance, which accounted for the items already recalled during the initial recall session of the trial. Because our model predicts variability between runs, the post-cue recall session was simulated in real time for 35 repetitions to obtain the mean performance for each possible retrieval cue. Four trials used our model's best cue, and the other four displayed our model's worst cue. The order of the

12 worst, random, and best trials was randomly determined for each participant. There was a 3-s buffer between the button click of "Remind Me" and cue presentation to allow our model time to simulate post-cue recall for the trial. The selected cue then appeared for 2 s followed by a 1-s delay. Participants were able to use any remaining time in the 90-s time frame to continue recall (the cue time not included). However, if participants requested a cue after 70 s into the initial recall period, they were given 20 s for recall. The experiment took approximately 40 min in total.

Results

Initial recall behavior

For the 1,524 cued trials, participants persisted in their initial recall for an average of 44.0 s ($SD = 17.2$) and recalled a mean of 8.47 words before requesting a cue ($SD = 2.92$). During this recall phase, participants displayed typical free-recall behaviors (see Fig. 5). These recall patterns include serial position effects (primacy—enhanced recall of items from the start of the list—and recency—enhanced recall of items from the end of the list; Murdock, 1962) as well as contiguity effects (semantic clustering—items from the same semantic category recalled successively—and temporal clustering—items studied in nearby serial positions recalled successively; Howard & Kahana, 2002b; Kahana, 1996). The temporal contiguity effect is bidirectional; list items studied recently before or after the just-recalled item are more likely retrieved, with a greater likelihood for items studied after (i.e., forward asymmetry; Kahana, 1996). CMR formally explains these recall regularities as a consequence of the drifting internal context representation becoming associated with each studied item, which is then used to guide memory search.

Regarding serial position effects, Figures 5a and 5b display the serial position curve and probability of first recall: Individuals are more likely to retrieve, and start recall with, items presented at the beginning of the list (primacy) and the end of the list (recency). Regarding contiguity effects, Figures 5c and 5d present the conditional response probability plot (computed by dividing the number of times that a transition to each lag is actually made by the number of times that it could have been made for each serial position; Kahana, 1996) and the semantic similarity plot (computed by finding the average cosine similarity between every pair of recalled items at different lags for their output positions). The decaying probability by lag on these plots supports that individuals were more likely to successively retrieve items from nearby serial positions and items that shared semantic features. These figures illustrate two things:

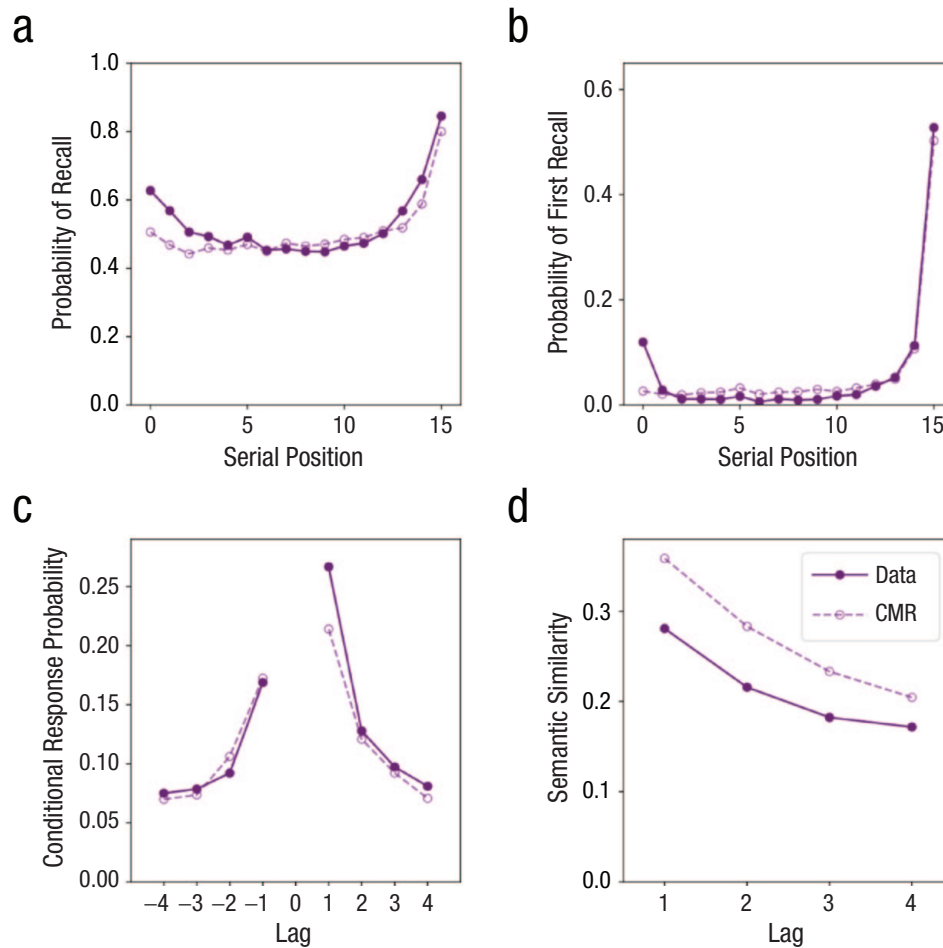


Fig. 5. Initial recall behavior in the data and context maintenance and retrieval (CMR). The participants displayed typical free-recall patterns during the initial recall session, and a CMR model that was previously fitted to a different set of participants was able to capture these behaviors. The parameter set was $\beta_{enc} = 0.850$, $\beta_{rec} = 0.828$, $\gamma_{fc} = 0.341$, $\gamma_{cf} = 0.316$, $s_{cf} = 1.406$, $\phi_s = 4.371$, $\phi_d = 2.226$, $k = 6.140$, and $\epsilon_d = 1.846$. These behavioral patterns are: (a) the serial position curve, (b) the first recall probability, (c) the conditional response lag probability, and (d) the semantic lag similarity.

One, participants showed typical free-recall behaviors during their initial recall, and two, these free-recall behaviors align well with predictions from a CMR model obtained prior to the collection of the data (with its predetermined parameter set fitted to our pilot sample's initial recall behavior). The model slightly overpredicts the semantic similarity effect in Figure 5d; this might be attributable to our use of GloVe embeddings to capture semantic similarity—these embeddings are known to be imperfect approximations of semantic structure in individual participants (see Polyn et al., 2009). We observed stronger temporal contiguity (see Fig. 5c) compared with prior work on categorized free recall (Polyn et al., 2011). This might have occurred either because the structure of our lists differed from those used by Polyn et al. (2011)—semantic categories could repeat across lists, and our lists had four semantic

categories with four words each, as opposed to three semantic categories with eight words each, which could have decreased category salience—or because the size-judgment task during encoding reduced opportunities to rehearse list items (see Ward & Tan, 2023).

Post-cue recall behavior

Our model captured initial recall patterns. Next, we considered how well it captured post-cue recall patterns. Figures 6a and 6b present temporal and semantic clustering patterns in the data and our model as conditional response and semantic similarity plots. Participants displayed typical contiguity effects during post-cue recall in which temporal and semantic similarity supported adjacent recalls. Whereas the data curve at Lag +4 increased from Lag +3 on the semantic similarity plot,

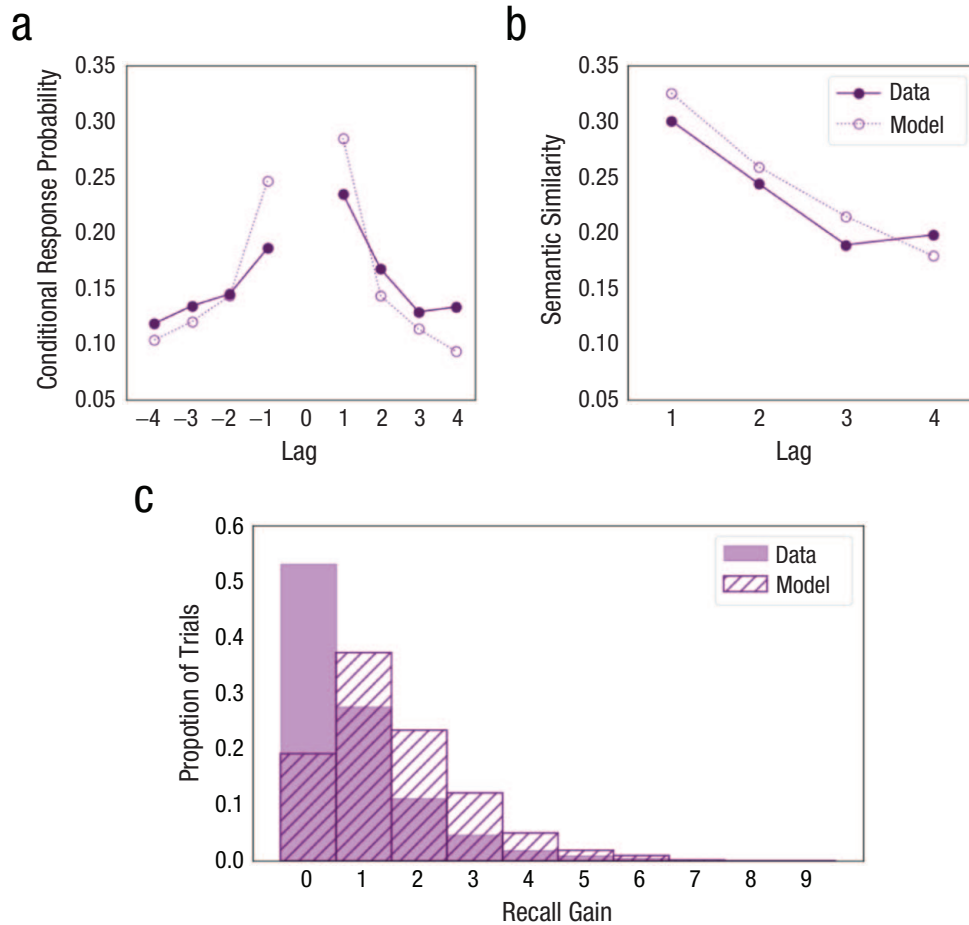


Fig. 6. Effect of cues on subsequent recalls. (a) The conditional response lag probability plot and (b) the semantic similarity lag plot display the temporal and semantic clustering of participants' recalls. Our model's ability to capture these behaviors with its parameters prefitted to initial recall behavior supports the claim that participants used similar retrieval processes during post-cue recall and initial recall. (c) Frequency distribution of recall gain in the data and model simulations whereby we observed a facilitative effect from cues as predicted by our model.

only 8% of trials had a recall gain of four or more words. Our model's behavioral fit to initial free-recall patterns was able to predict similar post-cue recall behavior. Thus, post-cue recall behavior provides support for the following claims: (a) Participants had similar recall behavior both before and after a cue, both of which were captured by CMR's retrieval process, and (b) post-cue recalls can be simulated by assuming that—when the current context state is no longer useful for retrieval—the current context is set to the cue's context, which functions as a retrieval cue from which post-cue recall can continue, with a similar retrieval process as in initial recall.

Next, we tested the effect of retrieval cues on post-cue recall performance. We found that cues significantly facilitated further recall: Participants' total recall on a trial ($M = 9.27$, 95% confidence interval, or CI = [9.12,

$9.41]$) was significantly greater than their initial recall before a cue ($M = 8.47$, 95% CI = [8.33, 8.62]), $t(1523) = -27.50$, $p < .001$, Cohen's $d = -0.27$, 95% CI = [-0.35, -0.20]. We observed an average of 0.80 ($SD = 1.13$) additional words recalled. This recall improvement supports our model's predicted benefit from cues, with the corresponding trial simulations averaging 1.58 additional words recalled ($SD = 1.23$). The model's overprediction of mean recall gain may be a consequence of its overprediction of semantic similarity effects (see Fig. 5d). Figure 6c shows the recall gain distributions in the data and our model.

Effectiveness of model-based cue selection

In addition to observing a facilitative effect from cues as predicted by our model, some retrieval cues are

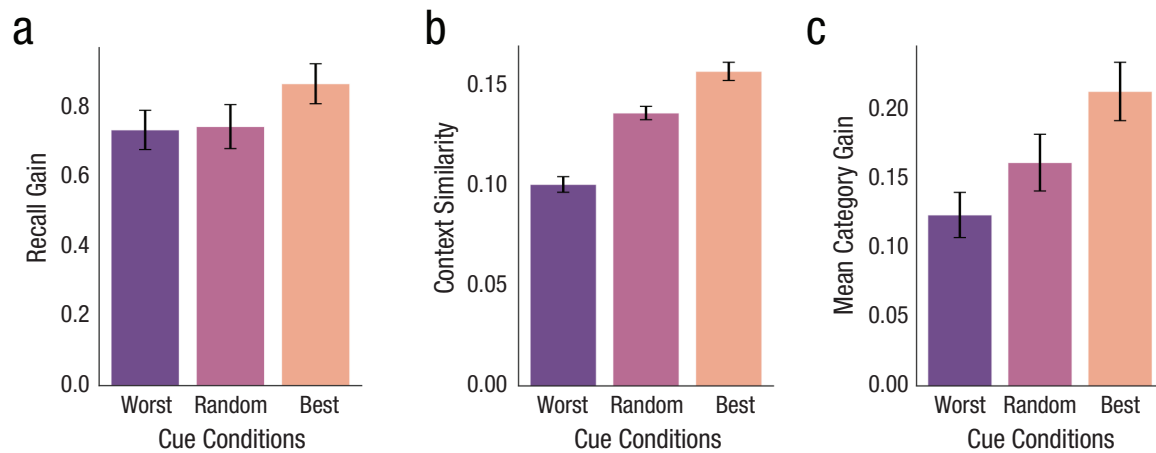


Fig. 7. Effect of our model's cue selection. The *x*-axes represent worst versus random versus best cue conditions within participant trial sets. The *y*-axes display nontransformed values. Error bars represent the standard error of the mean. (a) Mean recall gain by cue condition within subjects. Recall gain increased by cue condition and was significantly greater under best cues than worst cues, supporting our model's ability to effectively select retrieval cues. Across all participants, empirical and simulated recall gains were significantly correlated while controlling for the number of remaining words for each trial (see text). (b) Mean cosine similarity between the cue's context and the remaining words' contexts, as formalized by context maintenance and retrieval, by cue condition within subjects. Context similarity increased by cue condition and was significantly greater for best cues than for worst cues, supporting that cues' benefit corresponds with their similarity to the remaining words. (c) Mean number of additional categories recalled by cue condition. Category gain increased by cue condition, supporting the claim that effective cues tap into unsearched areas of memory.

predicted to be more effective than others. We tested the partial correlation between the log-transformed empirical and simulated recall gains, controlling for the number of words remaining on each trial because it correlates with further recall in the model and data. The association was significant and positive, $\rho(1508) = .130$, $p < .001$: Participants recalled more on trials in which our model predicted a larger benefit.

We further tested our model's ability to select cues by designing our study to have three conditions within the participants' set of trials: worst, random, and best cues. Cues delivered in these conditions were not predetermined prior to the experiment but were selected in real time on each trial after the initial recall period ended, by determining the performance associated with each remaining word from the list with model simulations. The cue with the highest mean recall gain in the simulations was selected as the best cue; the cue with the lowest mean recall gain was selected as the worst cue. The following analyses consider only participants, and their corresponding trial simulations, who were presented a cue on at least one best, one random, and one worst trial ($n = 141$). We log-transformed both participants' and the model's recall amounts for the following analyses. Participants demonstrated a significant effect of cues within their trial sets (see Fig. 7a): Worst cues ($M = 0.73$, 95% CI = [0.62, 0.84]) led to significantly less mean recall gain than best cues ($M = 0.86$,

95% CI = [0.75, 0.97]), $t(140) = -3.38$, $p = .001$, Cohen's $d = -0.34$, 95% CI = [-0.57, -0.10], and random cues' mean recall gain ($M = 0.74$, $SD = 0.75$) was between worst and best cues. The probability of these observations occurring by chance was significantly low ($p < .001$; this was determined by a permutation test in which we randomly shuffled the labels for best, random, and worst conditions 1,000 times). The differential effect between best and worst cues and increasing performance by cue condition demonstrate our model's ability to effectively predict the influence of retrieval cues on memory search.

Beyond our preregistered analyses, we explored whether a retrieval cue's contextual overlap with the contexts of the remaining words contributes to our model's selection of effective retrieval cues. We computed context similarity, operationalized as the average cosine similarity between the cue's context at encoding and every remaining word's context at encoding in CMR. The mean context similarity between the cue and remaining words on best-cue trials within a participant's set of 12 trials ($M = 0.16$, 95% CI = [0.15, 0.16]) was significantly greater than on worst-cue trials ($M = 0.10$, 95% CI = [0.09, 0.11]), $t(140) = -11.55$, $p < .001$, Cohen's $d = -1.12$, 95% CI = [-1.37, -0.87], and the mean context similarity on random-cue trials ($M = 0.14$, $SD = 0.04$) was in between the worst and best trials (see Fig. 7b). The probability of observing these effects (a monotonically

increasing relationship by cue condition, from worst to random to best, and a mean difference between best and worst cues' context similarities as large as the observed difference) by chance was significantly low ($p = .003$; the permutation test followed the same criteria as the prior tests). That is, the more our model predicted a retrieval cue to benefit recall compared with other possible cues, the greater context similarity the cue was estimated to have to the remaining words. Put another way, effective cues were those that tapped into an unsearched area of the context space, closest to the remaining items.

Because we used categorized lists, the ability of cues to tap into an unsearched area of memory can also be seen in measures of category recall. On 40% of trials, at least one of four categories was not reached before a cue. Within participants' trial sets, significantly more categories were recalled after a cue ($M = 3.64$, 95% CI = [3.57, 3.71]) than before a cue ($M = 3.49$, 95% CI = [3.42, 3.56]), $t(171) = -11.43$, $p < .001$, Cohen's $d = -0.32$, 95% CI = [-0.53, -0.10]; an additional 0.15 categories were reached on average ($SD = 0.17$). Moreover, Figure 7c shows that best cues led to a larger increment in the number of categories recalled ($M = 0.21$, 95% CI = [0.17, 0.25]) compared with worst cues ($M = 0.12$, 95% CI = [0.09, 0.16]), $t(140) = -3.61$, $p < .001$, Cohen's $d = -0.40$, 95% CI = [-0.63, -0.16], with random cues in between ($M = 0.16$, $SD = 0.24$). The probability of observing these effects by chance was significantly low ($p < .001$; the permutation test followed the same criteria as prior tests).

To gain further insight into the model's cue selection, we examined how the best and worst cues related to simple semantic and temporal heuristics that could be used for cue selection (e.g., selecting cues from as-yet-unrecalled categories or temporal clusters of as-yet-unrecalled items). We found that the model's cue choices were aligned with most of these heuristics (e.g., best cues were more likely than worst cues to come from an as-yet-unrecalled category; see Appendix B for details). So what is the value of our model of cued memory search, when its cue choices are largely aligned with simple semantic and temporal heuristics? Importantly, although individual heuristics were aligned (to varying degrees) with the model's behavior, several challenges could arise if one were to actually try to use these heuristics for cue selection: If two cues were equally suitable according to the available heuristics, or none of the available cues were suitable, or if different heuristics (e.g., semantic vs. temporal) favored different cues, how could one select a cue? The advantage of our model is that it provides a principled and quantitatively precise way of selecting cues in any (free-recall) circumstance that could emerge, sparing us the

need to derive ad hoc—and likely suboptimal—ways of agglomerating heuristics.

General Discussion

Contrary to intuition, previous free-recall studies found that providing participants with a random subset of list items does not benefit memory performance (Allen, 1969; Basden et al., 1977; Rundus, 1973; Slamecka, 1968; Sloman et al., 1991). We hypothesized that the content of cues would modulate this effect; specifically, the effectiveness of cues should depend on how well the cues match the contextual features of remaining memories. To test this, we built a computational model of cued memory search, and we used model simulations to select which cues to present at the end of recall. We found that participants recalled significantly more items on trials in which they received our model's best (vs. worst) retrieval cue. The results indicate that presenting participants with a remaining word from the studied list as a retrieval cue reactivates its encoding context, facilitating the recall of similar words. We now turn to the broader implications of these results.

Other accounts of external cues

A related paradigm in the external cuing literature is part-set cuing. In this paradigm, some participants receive a random subset of list items as retrieval cues at the start of a recall task, whereas others receive no cues. These studies have found a negative cuing effect in which cued participants recalled fewer of the remaining items than noncued participants (Basden et al., 1977; Rundus, 1973; Slamecka, 1968; Sloman et al., 1991; for reviews, see Nickerson, 1984; Pepe et al., 2023). How are theoretical accounts of the effect of external cues in this paradigm related to our proposed account of cued memory search?

Current accounts of part-set cuing propose that cues can disrupt, inhibit, or compete with retrieval attempts. Some explanations propose that the covert retrieval of cue items either strengthens the cues' memory traces and blocks retrieval of noncue items (retrieval competition; Rundus, 1973) or weakens the memory traces of noncue items (retrieval inhibition; Bäuml & Aslan, 2004). Another hypothesis is that cues force a recall order inconsistent with an individual's retrieval plan (strategy disruption; Basden & Basden, 1995; Basden et al., 1977). Recent work also theorized that after a prolonged retention interval, a forget cue, or an imagination task, the study context is no longer active; in this case, part-set cues reactivate the study context, which (in turn) benefits recall instead of harming it (context reactivation hypothesis; Bäuml &

Samenieh, 2012; Bäuml & Schlichting, 2014; Goernert & Larson, 1994). Some research also proposed a combination of these accounts, whereby the presence of the mechanisms depends on how strongly items are encoded and the time between study and test (Bäuml & Aslan, 2006; reviewed by Lehmer & Bäuml, 2018).

Our formal account of cued memory search proposes that the context retrieved by the external cue becomes the current context, after which retrieval proceeds as usual. This is consistent with the context reactivation hypothesis, as we showed that reactivating the cue's encoding context prompts additional recalls in an unsearched area of memory. Whereas the context reactivation account has been concerned with study context being activated or not during recall (Bäuml & Samenieh, 2012; Bäuml & Schlichting, 2014; Goernert & Larson, 1994; Lehmer & Bäuml, 2018), we provide a more detailed picture by proposing which part of the study context is activated and the effect of that activation. The strategy disruption theory assumes that one's retrieval plan becomes disrupted by cues; in our paradigm, initial recall ended before cuing, so there was no remaining retrieval plan that a cue could disrupt. Similarly, a recent study showed that allowing participants to self-request random part-set cues did not reduce their total recall, compared with uncued participants (Wallner & Bäuml, 2021). Regarding the competition and inhibition theories of part-set cuing, we do not assume that items' representations are strengthened or weakened when a cue is presented. Still, our results do not discount the potential of these mechanisms in cued memory tasks; rather, we highlight the integral role of temporal and semantic context for external cues in memory search. It remains a future direction to explore under what cuing circumstances disruption, competition, and inhibition accounts may play a role in cued memory search.

Modeling cued memory search

Given our model's ability to capture cuing effects in our paradigm, our computational work offers unique contributions to modeling cued memory search. First, our model can account for detailed patterns of post-cue recall behavior, in addition to the overall number of recalled items. Participants demonstrated temporal contiguity and semantic clustering after an external cue that our model captured. Further, memory facilitation was larger for cues with greater context similarity to the remaining words (as estimated using CMR's formal construction of the context space). This context-based recall behavior supports our model's proposition that post-cue recall proceeds in the same fashion as pre-cue

recall, after the context state updates on the basis of the external cue.

Second, whereas traditional paradigms used randomly selected items as cues, either from the entire list (Allen, 1969, Experiment 1) or from the subset of not-yet-recalled list items (Slamecka, 1968, Experiment 5), we formally accounted for the already-searched space of memory when choosing cues, and we estimated the effect of presenting not-yet-recalled items on one's future recall trajectory based on these model estimates. This effort allowed us to understand why some cues benefit memory performance more than others. We hypothesized that the most useful retrieval cues activate an area in memory that is contextually similar to the remaining items. This was supported by our finding that participants performed better given cues that our model selected to be the best (vs. worst) retrieval cues. These findings show that our model can be of practical use for exploring memory search paradigms that use external cues.

In addition to these novel analyses, we provided a strong test of our proposed model by demonstrating its ability to make predictions that generalize across periods of recall data (pre-cue vs. post-cue) and across groups of participants. Models with a large number of parameters have faced criticism in the past (Meyer & Kornblum, 1993) because "designing independent tests of the model may be difficult, since its ten parameters and numerous countervailing processes make unambiguous predictions hard to come by" (Roediger & Neely, 1982). In other words, there is always danger in fitting a complicated model (with a large set of parameters) to capture a simple set of behavioral patterns because one can always introduce new processes or alter parameter values during post hoc theorizing. To eliminate these concerns, our model accounts for post-cue recall behavior without ever being fitted to these data by assuming that the primary memory search process operates the same under uncued and cued conditions. Our model also generalized to a different group of participants; we committed to a predetermined set of model parameters (preregistered before the current study) previously fitted to a different set of participants. The alignment between the effectiveness of cues predicted by the model and the recall gain measured over the participants provides a strong test of the robustness of our proposed model of cued memory search.

In the present study, we fitted our model to aggregate pilot data to select useful cues in real time. Future work could collect more free-recall data from individual participants; this would make it possible to fit the model to individuals' specific recall patterns, which (in turn) could improve cue selection for these individuals. This research could help with the design of memory

interventions for individuals in educational settings or those with memory impairments. In addition to conferring “in-the-moment” benefits (i.e., recalling more items), systems that generate effective retrieval cues could confer long-lasting benefits on memory: It is well known that successfully recalling a memory is one of the best ways to ensure its future accessibility (the testing effect; e.g., Rowland, 2014). It stands to reason, then, that unsticking recall might boost subsequent retention of the additional items that are recalled; this topic could be investigated in future work.

Conclusion

In contrast to previous studies that randomly selected cues and failed to facilitate recall, we showed that the contextual content of cues modulates their effect on recall. We extended the CMR model of free recall to capture an external cue’s effect on memory search. These modeling efforts open promising directions for empirically and computationally exploring other external cuing phenomena, which can provide further insight into what mechanisms may critically underlie cuing effects in memory tasks. Given how commonly external cues are used in everyday memory search, our work offers a theoretical foundation for building future systems that can assist one’s memory.

Appendix A

Parameter Estimation in Context Maintenance and Retrieval

We recruited 53 pilot participants (ages 18–25 years) with Prolific. The materials and experimental design were the same with a few exceptions: One, participants completed 10 (not 12) experimental trials. Two, participants were shown only random cues with no 3-s delay (because no buffer window was needed to run our model’s post-cue recall session). Five subjects and 155 of the 480 trials were excluded from the fitting procedure following the same exclusion criteria. We fitted context maintenance and retrieval (CMR) to the initial recall behavior on cued trials across four sets of behavioral patterns: (a) serial position curve, (b) first recall probability, (c) conditional response probability at Lags $-4/+4$, and (d) semantic similarity probability at Lags $-4/+4$. To model these behaviors, we obtained a set of parameters that minimized the difference between the empirical values and the simulated values as a function of the parameters, expressed as the normalized root mean square error across the four behavioral sets. We used 200 iterations of Bayesian optimization, after 400 random initializations, to obtain the parameter set. The fit parameter set was $\beta_{enc} = 0.850$, $\beta_{rec} = 0.828$, $\gamma_{fc} = 0.341$, $\gamma_{cf} = 0.316$, $s_{cf} = 1.406$, $\phi_s = 4.371$, $\phi_d = 2.226$, $k = 6.140$,

Table A1. Cue Selection Heuristics

Name	Effective cue definition
Semantic 0	In an unaccessed category
Semantic 1	In a category with most remaining items
Temporal N	In a group of at least N consecutively studied remaining items
Primacy/Recency N	In a group of at least N consecutively studied remaining items and within first or last three serial positions

Table A2. Our Model’s Best and Worst Cues and Simple Heuristics

Heuristic	Best cues		Worst cues		Performance			
	M	SD	M	SD	t	df	p	Cohen’s d [95% CI]
Semantic 0	.73	.32	.25	.37	9.00	86	< .001***	1.38 [1.05, 1.71]
Semantic 1	.76	.24	.42	.30	11.95	148	< .001***	1.25 [1.00, 1.50]
Temporal 2	.89	.20	.59	.33	10.04	144	< .001***	1.11 [0.86, 1.35]
Temporal 3	.82	.28	.49	.35	8.05	123	< .001***	1.04 [0.78, 1.31]
Temporal 4	.79	.33	.40	.37	7.54	99	< .001***	1.10 [0.80, 1.40]
Primacy/Recency 2	.40	.38	.30	.34	1.96	114	.052	0.27 [0.01, 0.53]
Primacy/Recency 3	.50	.39	.31	.37	2.90	68	.005**	0.49 [0.15, 0.83]

Note: CI = confidence interval.

** $p < .01$. *** $p < .001$.

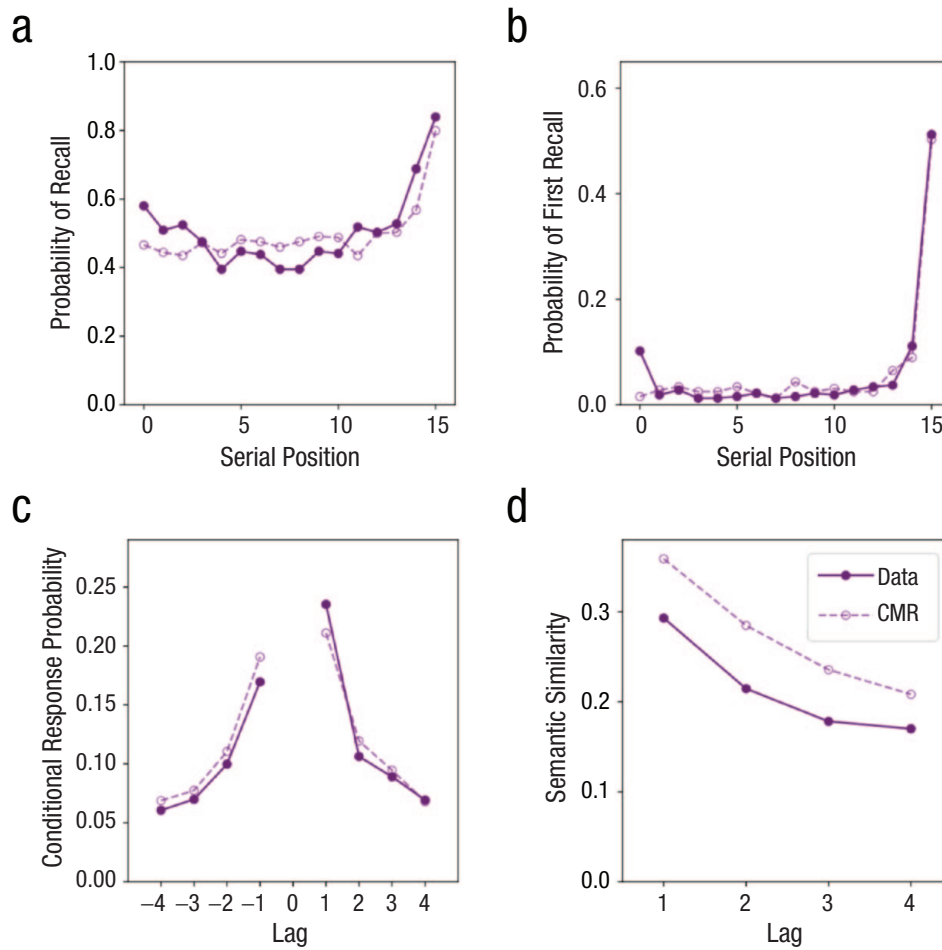


Fig. A1. Behavioral fit of context maintenance and retrieval (CMR) to the pilot data’s initial recall across four free-recall patterns. We used this behavior to model post-cue recall during the study’s live experiment and analyses.

and $\epsilon_d = 1.846$. Figure A1 displays CMR’s behavior compared with the pilot data.

Appendix B

Cue Selection Heuristics

To better understand how our model chose cues, we compared our model’s selection of best and worst cues with heuristics that assessed whether cue belonged to an as-yet-unrecalled category or to a temporal cluster of as-yet-unrecalled items (see Table A1). Specifically, for each candidate heuristic, we computed the proportions of best and worst model-selected cues that satisfied that heuristic. We then tested the mean difference within participants’ trial sets between the proportion of best cues that satisfied the heuristic and the proportion of worst cues that satisfied the heuristic (see Table A2).

Participants were included only if they had at least one best cue and at least one worst cue presented.

Our model’s selection of best and worst cues generally aligned with the semantic and temporal heuristics that we tested: Best cues were more likely than worst cues to be within an unaccessed category (Semantic 0); the category with the most remaining words (Semantic 1); a group of two, three, or four consecutively studied remaining items (Temporal N); or a group of three consecutively studied remaining items within the first or last three serial positions (Primacy/Recency 3). The difference was marginally nonsignificant ($p = .052$) for the Primacy/Recency 2 heuristic, which assessed whether a cue was nearby at least one remaining item in the first or last three serial positions. These results provide converging support that our model can uncover both semantic and temporal ways of selecting cues via its ability to organize memories into a single context space.

Transparency

Action Editor: Karen Rodrigue

Editor: Jennifer L. Tackett

Author Contribution(s)

Charlotte A. Cornell: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; Writing – review & editing.

Kenneth A. Norman: Conceptualization; Writing – review & editing.

Thomas L. Griffiths: Conceptualization; Writing – review & editing.

Qiong Zhang: Conceptualization; Funding acquisition; Investigation; Methodology; Supervision; Writing – original draft; Writing – review & editing.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by a start-up fund from Rutgers University–New Brunswick and a grant from the National Science Foundation (BCS-2316716), both awarded to Q. Zhang.

Open Practices

This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



References

- Allen, M. M. (1969). Cueing and retrieval in free recall. *Journal of Experimental Psychology*, 81(1), 29–35.
- Basden, D. R. (1973). Cued and uncued free recall of unrelated words following interpolated learning. *Journal of Experimental Psychology*, 98(2), 429–431.
- Basden, D. R., & Basden, B. H. (1995). Some tests of the strategy disruption interpretation of part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1656–1669.
- Basden, D. R., Basden, B. H., & Galloway, B. C. (1977). Inhibition with part-list cuing: Some tests of the item strength hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, 3(1), 100–108.
- Bäuml, K.-H. T., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition*, 32(4), 610–617.
- Bäuml, K.-H. T., & Aslan, A. (2006). Part-list cuing can be transient and lasting: The role of encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(1), 33–43.
- Bäuml, K.-H. T., & Samenih, A. (2012). Selective memory retrieval can impair and improve retrieval of other memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(2), 488–494.
- Bäuml, K.-H. T., & Schlichting, A. (2014). Memory retrieval as a self-propagating process. *Cognition*, 132(1), 16–21.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1–12.
- Goernert, P. N., & Larson, M. E. (1994). The initiation and release of retrieval inhibition. *The Journal of General Psychology*, 121(1), 61–66.
- Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern, D., Hamrick, J. B., & Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, 48(3), 829–842.
- Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3), 269–299.
- Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, 46(1), 85–98.
- Hudson, R. L., & Austin, J. B. (1970). Effect of context and category name on the recall of categorized word lists. *Journal of Experimental Psychology*, 86(1), 43–47.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24(1), 103–109.
- Kragel, J. E., Morton, N. W., & Polyn, S. M. (2015). Neural activity in the medial temporal lobe reveals the fidelity of mental time travel. *Journal of Neuroscience*, 35(7), 2914–2926.
- Kroeger, M. E., Hueng, N. L., Curry, S. D., Copeland, M. B., & Kelley, M. R. (2019). On the composition of part-set cues. *Quarterly Journal of Experimental Psychology*, 72(10), 2365–2370.
- Lehmer, E.-M., & Bäuml, K.-H. T. (2018). The many faces of part-list cuing—Evidence for the interplay between detrimental and beneficial mechanisms. *Frontiers in Psychology*, 9, Article 701. <https://doi.org/10.3389/fpsyg.2018.00701>
- Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, 122(2), 337–363.
- Meyer, D. E., & Kornblum, S. (1993). *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (Vol. 14). MIT Press.
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5), 482–488.
- Nickerson, R. S. (1984). Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory & Cognition*, 12(6), 531–552.
- Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1532–1543). Association for Computational Linguistics.
- Pepe, N. W., Moyer, A., Peña, T., & Rajaram, S. (2023). Deceitful hints: A meta-analytic review of the part-list cuing impairment in recall. *Psychonomic Bulletin & Review*, 30(4), 1243–1272.
- Polyn, S. M., Erlichman, G., & Kahana, M. J. (2011). Semantic cuing and the scale insensitivity of recency and

- contiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 766–775.
- 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.
- Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology*, 92(3), 365–372.
- Roediger, H. L. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior*, 12(6), 644–657.
- Roediger, H. L., & Neely, J. H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 36(2), 213–242.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463.
- Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior*, 12(1), 43–50.
- 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.
- Slamecka, N. J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology*, 76(4, Pt. 1), 504–513.
- Sloman, S. A., Bower, G. H., & Rohrer, D. (1991). Congruency effects in part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 974–982.
- Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, 25(2), 207–222.
- Tulving, E. (1983). *Elements of episodic memory*. Clarendon Press.
- Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5(4), 381–391.
- Wallner, L., & Bäuml, K.-H. T. (2021). Self-paced part-list cuing. *Psychonomic Bulletin & Review*, 28(6), 2012–2018.
- Ward, G., & Tan, L. (2023). The role of rehearsal and reminding in the recall of categorized word lists. *Cognitive Psychology*, 143, Article 101563. <https://doi.org/10.1016/j.cogpsych.2023.101563>
- Zhang, Q., Griffiths, T. L., & Norman, K. A. (2023). Optimal policies for free recall. *Psychological Review*, 130(4), 1104–1124.