12-15-2022

Evaluating Evidence of Episodic Tendencies During Semantic Fluency

Rebecca Anne Wilder
Syracuse University

Abstract

While semantic organization has been widely observed in episodic memory tasks, episodic organization has yet to be observed in the semantic fluency task, due to structural differences between test paradigms. Episodic memory effects require an opportunity for target information to first be learned and later retrieved. Semantic fluency tasks, which are designed to measure retrieval capacity for facts, are typically limited to a single test-phase format. In semantic fluency tasks, participants are presented with a semantic retrieval cues (i.e. category prompts) and asked to list as many items as they can think of that fit the classification. The repeated fluency paradigm, a variant of semantic fluency, presents category prompts to participants multiple times throughout the experiment, which provides participants with a sufficient opportunity to repeat words and reference temporal features (e.g. serial order that words were listed) across different trials of the same category.
This study used the repeated fluency task to evaluate whether repetition, contiguity, and temporal interval effects would be present during semantic retrieval. Conventional episodic comparisons for these effects were contrasted to measurements of word-typicality and semantic similarity to ensure that any observed effects could not be entirely accounted for by common semantic explanations. Results from this experiment found significant evidence that repetition and contiguity effects were present in the data. Further, neither of these effects could be accounted for by word-typicality or semantic similarity. Although the temporal interval analysis showed no significant differences in the overall magnitude of repetition effects, one of two contiguity comparisons found significant differences across temporal intervals. Taken together, these results suggest that the delineation between episodic and semantic retrievals becomes less defined when a semantic retrieval cue is presented more than once.
Keywords: Episodic memory, semantic memory, semantic fluency, repetition effects, contiguity effects.

Evaluating Evidence of Episodic Tendencies During Semantic Fluency

by

Rebecca Wilder

B.A., University of California, Irvine, 2018

Thesis

Submitted in partial fulfillment of the requirements for the degree of
Master of Science in Psychology

Syracuse University
December 2022

Contents

Introduction

The integrity of semantic memory, relating to memory for facts, concepts, and general knowledge, is commonly measured through the semantic fluency task. In semantic fluency tasks, participants are presented with a generic category prompt (e.g. *List Vegetables*) and are asked to list as many category members as they can think of. Semantic fluency tasks are commonly applied to clinical diagnostic procedures, particularly for conditions like Alzheimer's disease and semantic dementia (Henry, Crawford, & Phillips, 2004; Saranpää, Kivisaari, Krumm, & Salmelin, 2022). Research using the repeated fluency paradigm, a variation of semantic fluency that presents category prompts multiple times throughout the procedure, has found evidence of practice effects occurring at shorter intervals (Cooper, Lacritz, Weiner, Rosenberg, & Cullum, 2004). One possible account for these practice effects is that individuals may be relying on both their episodic and semantic memory during a single task (Greene, 1989). A patient may be primed to extract certain words or potentially even think back, explicitly to what they may have said during a previous procedure. For example, someone may be primed to list the word *CORN* during a semantic fluency task, if they had previously listed that word before. If the examinations are close enough, they may also be able to remember the word they listed before and after *CORN*. This calls into question the validity of outcomes from patients who have completed more than one semantic fluency task in a short time-span.

Concerns about practice effects during semantic fluency are supported by evidence of semantic organization in conventional episodic memory tasks, such as free recall. Free recall is a common episodic memory paradigm in which participants are presented with one or more lists of stimuli to learn and later retrieve, during the test-phase of the procedure(Wixted & Rohrer, 1994). Unlike cued recall or recognition memory tasks, free recall does not provide participants with specific retrieval cues, allows participants to recall information in any order, and commonly uses words as stimuli. Past work as demonstrated a tendency for semantically similar information to be grouped together at testing(Howard & Kahana, 2002; Polyn, Norman, & Kahana, 2009). To illustrate, *CHAIR, TELEVISION*, and *PHONE* may be recalled more closely together than expected, given the serial order that stimuli were encoded in, due to overlapping semantic associations between both words (i.e. *household items*. The Deese-Roediger-McDermott (DRM) paradigm, a free recall task that uses thematically similar stimuli, found that inter-item associations tends to result in semantically meaningful intrusions (Deese, 1959; Roediger & McDermott, 1995). For example, when stimuli lists that included the words *BED, AWAKE, TIRED,* and *DREAM*, participants would often recall the word *SLEEP*, despite that word not being presented. Although free recall is generally thought to execute searches for different qualitatively different information than semantic fluency, these findings suggest that our memory for facts and concepts likely facilitates our ability to extract episodic memories.

Different Ways To Measure Episodic and Semantic Memory

While there is robust evidence that episodic free recall data reflects a degree of semantic organization, less is known about the degree to which episodic tendencies are present during semantic retrieval. This is due to the fact that episodic memory tasks require an opportunity for a stimulus to be encoded and a distinct opportunity to retrieve the previously encoded information (Bower, 2000). Conversely, semantic memory tasks, such as semantic fluency, follow a single test-phase format and evaluate one's ability to generate members of a given category. While these procedures do not provide an opportunity to perceptually encode information, retrieval-based models of learning suggest that the act of retrieving information prompts an encoding event (Karpicke, Lehman, & Aue, 2014; Prince, Daselaar, & Cabeza, 2005). Accordingly, the present study used the repeated fluency task to provide provides an opportunity for items to be generated during the initial presentation of a category and repeated during the second presentation of a category. If practice effects were, in fact, an expression of episodic reliance, then one could expect to see response tendencies that are commonly observed in episodic recall tasks. Here, I outline three prominent effects in free recall tasks [1] (repetition, contiguity, and temporal interval effects) and test for evidence of these effects when semantic fluency trials are repeated at short intervals. Repetition effects state that items with additional encoding opportunities should be retrieved at higher rates, contiguity effects state that items encoded in close proximity tend to be retrieved in close proximity, and temporal interval effects state that probability of retrieval, as well as other episodic memory effects (e.g. repetition or contiguity effects) should attenuate as the interval between encoding and retrieval increases.

The repetition effects analysis included two comparisons that assessed the degree to which words were being repeated across trials and whether repeated words were more likely to be clustered together on the second test-phase. Contiguity analyses looked only at adjacently repeated words on the second test-phase (i.e. repeat-repeat transitions) and assessed if repeat-repeat transitions demonstrated a forward-order directionality bias, visually resembled a conventional Lag-CRP curve, and expressed temporal clustering. A directionality bias, here, refers to an increased probability for generating adjacently repeated words in the same order, across test-phases. Lag-CRP curves are a common method for visualizing contiguity effects in episodic memory tasks and maintain a distinct shape (see Figure 1), where the probabilities of recalling any two items vary as a function

[1] free recall is an episodic memory task, in which participants are presented with one or more lists of words and asked to later remember those words, without the aid of retrieval cues

of the number of words encoded between both items. Signs of temporal clustering on the second category test-phase trials would be . Temporal interval analyses assessed whether the number of intervening trials between a category's first and second test-phase influenced the magnitude of repetition and contiguity effects. I hypothesized first that semantic fluency data will demonstrate evidence of episodic organization. will be demonstrated through repetition, contiguity, and temporal interval effects in the repeated fluency task. Further, I also hypothesized that any observed effects would not merely be able to be accounted for by semantic organization alone.

Background

Semantic Associations Influence Episodic Free Recall

Previous research has found overlapping semantic effects during episodic memory tests, indicating that episodic memories are not only susceptible to temporal and state-dependent modulation, but are also sensitive to semantic priming and cueing (Guerin & Miller, 2008). Semantic associations may involve shared attributes (e.g. same color), similar applications (e.g. household appliances), or membership to a common subgroup (e.g. fruits), whereas temporal associations reference the serial encoding proximity of two adjacently retrieved items. The sensitivity of episodic retrievals to semantic cueing can also expressed through semantically meaningful intrusions. For example, on tasks where subjects are asked to read and later recall short-story passages, failure to retrieve precise elements of the story was frequently accompanied by conceptually synonymous substitutes, formally termed "thematic surrogates" (Dooling & Lachman, 1971; Howe, 1970). Retrieval errors may also reference overlapping themes or semantic features of a stimuli list, which is common to see on tasks such as the Deese-Roediger-Mcdermott paradigm (Deese, 1959; Roediger & McDermott, 1995). Taken together, these findings suggest that episodic retrievals, particularly those that occur in free recall data may be organized both by temporal and semantic features.

Semantic Fluency

Semantic and episodic retrieval are generally thought to extract qualitatively different types of information and are measured in different ways (Tulving, 1972). A prominent method of assessing the integrity of semantic memory is through the semantic fluency paradigm. Populations with impaired semantic memory, such as those with Alzheimer's disease, tend to list fewer items during semantic fluency tasks (Henry et al., 2004). Cognitively healthy performance on these tasks is hallmarked by a tendency for semantic retrievals to prompt searches for meaningfully associated items (Troyer, Moscovitch, &

Winocur, 1997). For example, when listing items that are commonly found in a grocery store, writing down *APPLES* may prompt a subsequent search for the word *ORANGES*, as these two products are frequently associated together. Common words, referring to concepts that are highly represented in everyday language, also tend to be generated more frequently than uncommon words (Morrison & Ellis, 1995). On the same list, a typical word e.g. *APPLES* is more likely to be listed than e.g. *CAPERS*. Additionally, individuals tend to cluster, or produce words within a given subgroup of a category during semantic fluency tasks (Hills, Jones, & Todd, 2012; Troyer et al., 1997). For example, words that relate to fruits, vegetables, and beverages would all be considered grocery store items, but may be grouped within their respective clusters.

The Repeated Fluency Task

Unlike traditional semantic fluency procedures, the repeated fluency paradigm offers multiple testing opportunities for each category prompt (Zemla & Austerweil, 2018). Retrieval-based learning models argue that the act of recall prompts both a mental representation and an encoding event for the target information (Karpicke et al., 2014; McNamara & Diwadkar, 1996). The multi-test format lends itself to assessing the degree to which common episodic memory effects are present in semantic fluency data. Namely, a category's initial semantic fluency trial prompts an encoding event for all generated items, while the same category's repeated semantic fluency trial offers a distinct retrieval opportunity. While this task allows participants to repeat items across category trials, they are not explicitly told to do so. Instead, participants are only instructed to list items that belong to a category and not list the same word more than once per trial. Each category's first test-phase provides an opportunity for words to be initially generated and the second test-phase allows for the observation of inherently episodic tendencies, such as whether words are repeated, the order of retrieval on the second test-phase, distance between repeated words across both category test-phases.

Repetition Effects: Probability of Repetitions Versus Word-Typicality and Repetition Clustering Analysis

Repetition effects refer to a tendency for items with additional encoding opportunities to, later, have an increased probability of retrieval at testing (Hebb, 1961; Madigan, 1969). These effects are thought to be caused by reactivation of a memory trace, upon presentation of a retrieval cue for a given episode (Habib, 2016; Raaijmakers, 2003; Tulving & Pearlstone, 1966). Semantic cues, conversely, do not reference particular features of an episode. Instead, searches for these memories are prompted by references to overlapping semantic associations (Collins & Loftus, 1975). It is possible, however, that word retrievals can be primed by both semantic and temporal associations. A semantic account of the effect would attribute instances of repetition to word-typicality, similarly to the APPLES and CAPERS example given earlier. Namely, a tendency for participants to repeat certain words across both test-phases trials of the repeated fluency paradigm could potentially be explained by the typicality of said words. For example, a word like *BANANA* could potentially be repeated because it received an additional representation on the first category test-phase or simply because this word is frequently used in day to day language. A word like *KIWANO*, otherwise known as the Horned-Cucumber, is far less likely to be repeated out of semantic necessity. Taken together, a word that is more likely to be repeated across both test-phases of the repeated fluency task than to be listed at all on the second test-phase at all, represents a marker for repetition effects during semantic retrieval. Looking beyond whether words tended to be repeated across test-phases, a next step would be to look at what cues repeated words. Namely, a tendency for repeated words to be clustered together could indicate that the act of generating repeated words on second category test-phases references features of a previously encoded episode (i.e. the first time that a participant was presented with that semantic cue). Conversely, it could also be the case that clustered repetitions were retrieved under a purely semantic retrieval strategy, *but* that these searches are also sensitive to additional representations of information, referring here to both the category prompt and retrieval of previously listed words.

Contiguity Effects: Directionality, Lag CRP, and Temporal Clustering

The contiguity effect describes an increased probability for items that were encoded in nearby serial positions to be retrieved in nearby output positions (Kahana, 1996). This effect is thought to be an expression of retrievals cueing searches for items that were encoded in close proximity (Sederberg, Howard, & Kahana, 2008). Further contiguity effects also express a degree of directionality bias, where words are more likely to be recalled in the order that they were encoded than in reverse order. A popular method of visualizing contiguity effects in free recall is by calculating a curve of conditional response probabilities, dependent on the encoding distance between two adjacently retrieved words (i.e. lag). Further, contiguity effects can be quantified and empirically tested by calculating temporal clustering scores. As contiguity comparisons measure effects that are related to the temporal order and spacing of encoding events in free recall, comparisons for this analysis only considered adjacently repeated words on the second test-phase trials. This study defines adjacently retrieved repetitions as repeat-repeat transitions.

Directionality. While contiguity effects are reciprocal, meaning that retrievals may cue searches for either adjacent item, they also tend to be asymmetric. Particularly, it is more likely for repeated transitions to be retrieved in forward-order, i.e. encoding order, than backwards order (Howard & Kahana, 2002). This is demonstrated in Figure 1, where the conditional response probability of items with a +1 lag is greater than those with a -1 lag. Conversely, associations are often treated symmetric in most models of semantic similarity, despite the fact that certain associations *can* demonstrate asymmetry (Pakhomov, Hemmy, & Lim, 2012). For example, the idiom *"it's raining cats and dogs!"* may result in a bias towards generating the word *"DOG"* after the word *"CAT"*. Another concern may be that more typical words tend to be generated earlier in semantic fluency trials, which could have, in part, accounted for the observed forward-order asymmetrical bias in the data. While it is not possible to entirely disentangle the effects of semantic versus episodic , failure to observe a forward-order directionality bias would suggest that the order of repeat-repeat transitions most likely not primed by episodic memories.

Lag-CRP and Contiguity Curves. One way to differentiate between potential contiguity effects and semantic priming would be to assess the degree to which repeat-repeat transitions reference the overall, serial order of retrievals from the first category test-phase. In free recall, contiguity effects are hallmarked by a distinct curve of conditional response probabilities for the distance observed between each set of adjacently retrieved words (Lag-CRP) (Howard & Kahana, 2002; Howard, Youker, & Venkatadass, 2008). This curve is demonstrated in Figure 1, which was reprinted from (Kahana, 1996). In episodic memory tests, lag is defined as the encoding distance between two adjacently retrieved words and Lag-CRP refers to the conditional probabilities of adjacent retrievals occurring at each lag. Lag is a directional measurement, calculated by subtracting the serial position of the second output word from the serial position of the first output word (Kahana, 1996). For example, if *WORD D* was presented in the fourth serial position and *WORD F* was presented in the sixth serial position, a participant recalled *WORD D* → *WORD F*, this transition would have a lag of +2. Conversely, if the participant had recalled the words as *WORD F* → *WORD D*, the transition would have a lag of -2. Lag-CRP is calculated as the number of times a particular transition occurred (e.g. a lag of +2 occurred seven times in the data), divided by the total possible number of transitions that could have occurred at that lag (e.g. a total of ten possible transitions could have occurred at a lag of +2). This manuscript defines transitions with positive lags as forward-order transitions and transitions with negative lags as backward-order transitions. In episodic recall tasks, contiguity is denoted by two distinct curves that reflect increased probabilities as transitions approach a lag of zero[2]. Transitions with positive lags also tend

[2] When a lag occurs between two words a lag of zero is not possible, but represents a useful marker between positive and negative lags

to have higher conditional response probabilities than transitions with negative lags. A final characteristic is that conditional response probabilities, as well as differences between forward and backward-order transitions, attenuate as the absolute value of lag increases. This is shown in Figure 1, where conditional response probabilities spike at smaller lags and flatten out at larger lags.

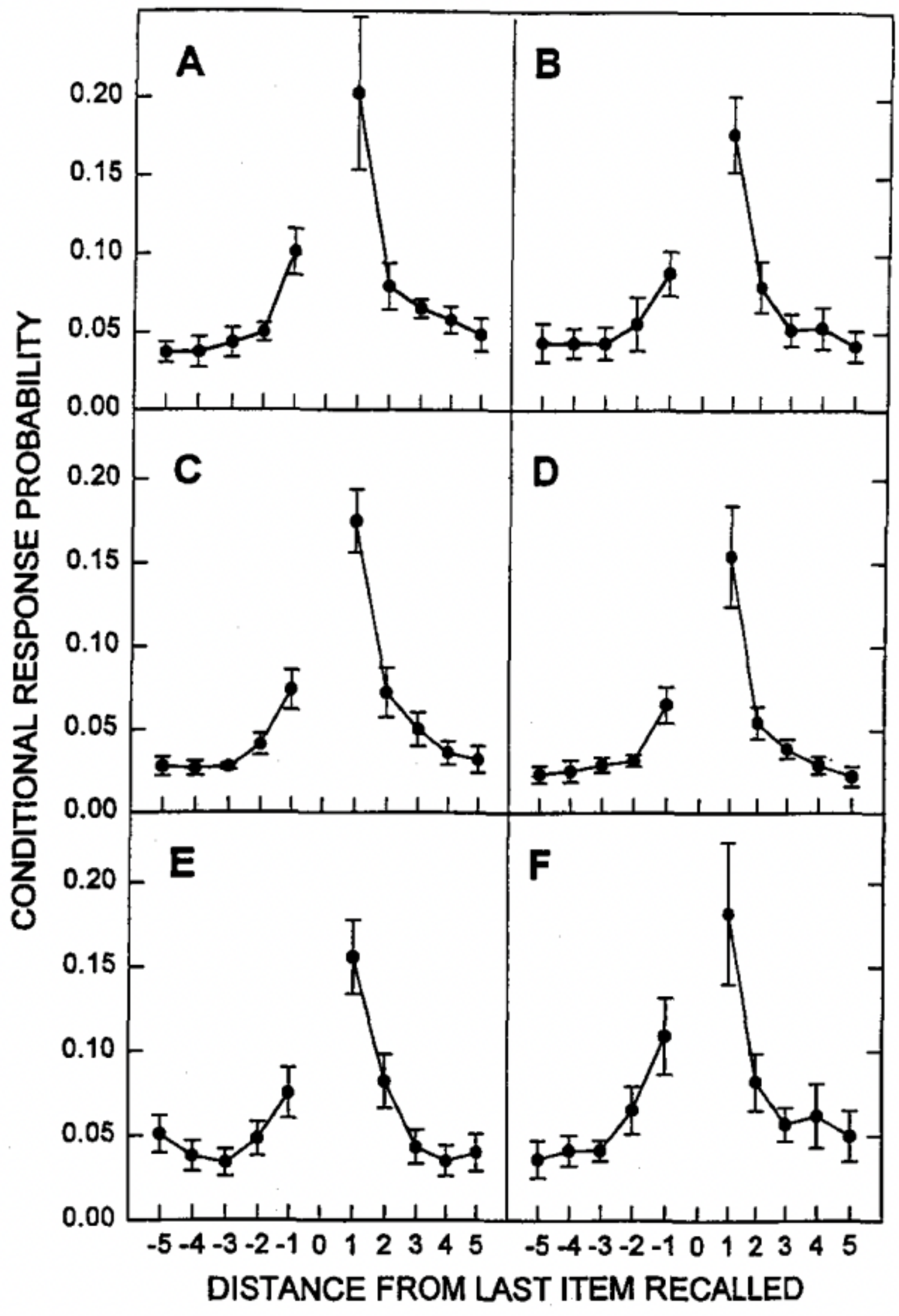


Figure 1. Six panel figure of contiguity curves from (Kahana, 1996). These figures were created using data from six different free recall experiments, differing in stimuli presentation format (i.e. auditorily or visually presenting stimuli), duration, and list-length (i.e. 20 words or 30 words presented). This figure was included to demonstrate the robustness n of contiguity effects, such that there is little observable deviation across any panels, depsite each experiment having different conditions. The x-axis on each panel represents the number of presented stimuli between adjacently recalled words. Lags can be either positive or negative, with positive lags reflecting transitions that adhered to the serial order that items were presented in and negative lags reflecting transitions that were recalled in the reverse order. The y-axis represents the probability of two words being adjacently retrieved at each lag.

Temporal Clustering Scores. While contiguity effects are commonly visualized through Lag-CRP curves, they can also be quantified and empirically tested by using temporal clustering scores (Polyn et al., 2009). Clustering scores are expressed as percentiles and are calculated as the proportion of possible lags that could have occurred and were greater than the lag that actually occurred for a given transition (Polyn et al., 2009; Sadeh, Moran, & Stern, 2019). Temporal clustering scores range from 0 to 1, with a score of 0.5 representing a lag that was smaller than 50% of all possible transitions that could have occured. This study calculated temporal clustering scores for all repeat-repeat transitions, with lag being calculated as the difference in serial positions of both words from the category's first test-phase. Additionally, temporal clustering scores may be averaged across trials or by participants to indicate the degree of temporal clustering observed for a particular category or between subjects.

Semantic Similarity Measurements. Because the order of retrievals on semantic fluency tasks tend to reflect meaningful semantic associations between words, a reasonable next step would be to contrast temporal clustering scores with estimates of semantic similarity. This experiment estimated semantic similarity between words using a model of semantic similarity (Word2Vec). Word2Vec calculates the distance between each item, and is pretrained on embeddings from (Mikolov, Chen, Corrado, & Dean, 2014). Word embeddings refer to semantic representations that allow similar words to have similar representations. In this study, the Word2Vec model was pretrained on pages from Wikipedia. This variant is of Word2Vec is commonly referred to as Wikipedia2Vec (Yamada et al., 2020). I chose to use a variation of Word2Vec, as opposed to other models of semantic similarity (e.g. BEAGLE Jones & Mewhort, 2007) due to the base model's ability to incorporate pretrained embeddings from large corpora of text (e.g. Google News, Wikipedia) and recent adoption as a tool for modeling semantic representations in patients with Alzheimer's disease (Saranpää et al., 2022).

In a scenario where contiguity effects can be entirely accounted for by semantic similarity, it would be highly unlikely to see temporal clustering for semantically dissimilar transitions. This is due to the fact that, while dissimilar transitions do occur during semantic fluency procedures, it is far more common for transitions to be semantically similar (Jones & Mewhort, 2007). Therefore, the observation of temporal clustering, among semantically dissimilar transitions, would suggest that any observed temporal contiguity could not be accounted for by primed semantic associations alone.

Method

Participants

The study sampled 49 cognitively healthy, young adults from Syracuse University. Each participant who completed the task was awarded one hour of course credit through the university's SONA research participation pool. Subjects ranged in age between 18 to 23 years old, with an average age of 19 years old. Approximately 40% of the participants identified as female (n= 19), 60% as male (n= 29), and less than 1% identifying as neither male nor female (n= 1). Of the 49 individuals who completed the study, roughly 73% identified English as their native language (n= 35), 20% as having a different native language (n= 10), and 8% who did not disclose a native language (n=4). There were no significant differences between the number of words listed by native and non native speakers on the first ($t(43)$= -1.1, 95% CI = [-4.64, 1.6], p= 0.32) or second ($t(43)$= -0.71, 95% CI = [-4.29, 1.95], p= 0.48) category test-phase trials. Additionally, no significant differences were found between the number of repeated words that were listed by native and non-native english speakers ($t(43)$= -0.12, 95% CI = [-0.05, 0.07], p= 0.91). Performance differences (i.e. the number of words listed on each semantic fluency trial) between native and nonnative english speakers is shown on Figure 11. Figure 12 shows the differences in the average probability of repeating a word, across different categories, between native and non-native english speakers. Figures 11 and 12 can be found in the manuscript's Appendix. Additional details about semantic fluency performance and the proportion of repeated words for each category are shown on Tables 3 and 4, which can also be found in the manuscript's Appendix.

Repeated Fluency Task

Participants completed a repeated fluency task, which sampled 12 distinct category prompts[3]. The intervals between each category's first and second test-phase were spaced between either two, four, or six intervening fluency trials. This spacing sequence is demonstrated in Figure 2. For each trial, subjects were presented with a category prompt and had 90 seconds to generate as many items as they could think of that belonged to the category. All trials were followed by a 30 second distractor task, where subjects counted backwards by 3's from randomly sampled numbers. The starting values for each distractor task ranged between values of 500 to 9,000 (e.g. count backwards by 3's from 5,676).

[3] While all 12 categories were presented to subjects, only the first 11 categories for each subject were used in data analyses due to an experimental coding error. The category for the second to final trial was only presented once, whereas the final category trial was presented three times. The trial from the single presentation category was removed and only the first two presentations of the category presented three times were used.

To ensure that each category was equally represented between participants at all three temporal intervals, the first category of the task was determined by participant id while the order for temporal intervals remained fixed. All participants began with a category at temporal interval six, followed by a category at temporal interval four, and a category at temporal interval two. While the category sequence itself remained fixed, the starting category was cycled forward by one iteration for each participant. This sequence is also shown in Figure 2, where the first participant (panel 1) begins on Category A, the second participant begins on Category B, and the third participant begins on Category C.

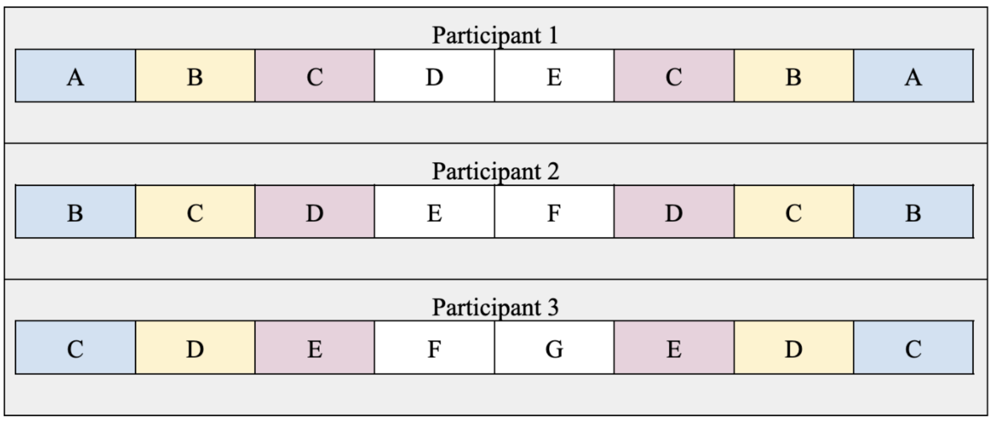


Figure 2. Demonstration of a category spacing sequence across all three temporal intervals. Each letter represents a different category trial, whereas the shading of each box represents the category's temporal interval. Boxes shaded in blue have a temporal interval of six, boxes shaded in yellow have a temporal interval of four, and boxes shaded in red have a temporal interval of two.

Results

Repetition Effects

This set of analyses evaluated 1. the probability of words being repeated across category test-phases and 2. whether repetitions tended to be clustered together (i.e. greater probability of occurring successively, rather than interspersed between new words) on the second test-phase trials.

Probability of Repetitions Versus Word-Typicality. This analysis found that conditional probabilities $[p(i \in List2 | i \in List1)]$ for each of the 948 repeated words exceeded the expected value that was predicted by their corresponding typicality estimates $[(\Sigma(word_i \in List2) - 1)/N]$.

Results from this analysis are shown in Figure 3). Scores along the y-axis indicate an item's probability of being listed on the second test-phase of a category, given that the word was also listed on the first test-phase. both typicality scores that fall along the x-axis of the figure and conditional probabilities of repeating a word fall along the y-axis. An identity line represents a null threshold, in which a word's conditional probability of being repeated is equal to its typicality estimate. The size of each marker represents the volume of words that overlapped in both typicality estimates and probabilities of being repeated. A Welch Two-Sample *t*-test found that the probability of repeating a word was significantly higher than the typicality scores for each repeated word [$t(52.8)= 69.67$, sd= 0.29, 95% CI = [0.64, 0.68], $p < 0.001$].

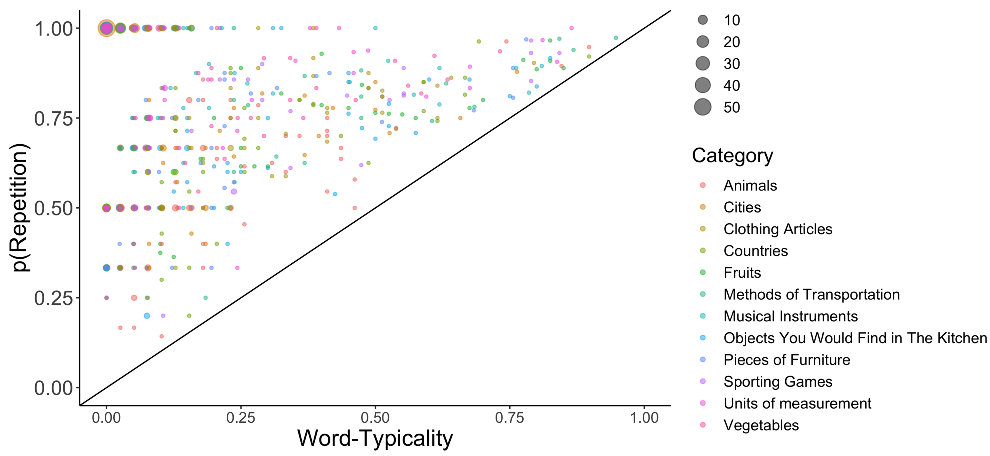


Figure 3. Conditional probability of repeating an item, as a function of word-typicality. Values along the y-axis of the figure represent the probability that a word will be listed on the second test-phase, given that it was also listed on the first test-phase. Values along the x-axis reflect typicality scores for each word (i.e. what was the probability of all other participants listing that word on the second test-phase). The size of each marker represents the number of words that had overlapping probabilities of being repeated and typicality scores.

Repetition Clustering Analysis. The next comparison in this set of analyses considered retrieval probabilities for one of four possible transition types: A repeated word that followed a repeated word (repeat-repeat transition), a new word that followed a repeated word (new-repeat transition), a repeated word that followed a new word (repeat-new transition), and a new word that followed a new word (new-new transition). Conditional probabilities were calculated as the number of times a particular transition occurred within a list, divided by the number of possible transitions that could have occurred. For transition types where the current retrieval follows a repeated item, the denominator considers the total number of transitions where either a new or a repeated item followed a repeated item. For transition types where a word follows a new item, the denominator considers all transitions where either a repeated item followed a new word or where a new item followed a new item. Conditional probability formulas for each transition type are shown in Table 1.

Transition probabilities were averaged across both participants and category trials before being compared to a distribution of bootstrapped values. Distributions of bootstrapped probabilities, for each transition type, were created by calculating the same conditional transition probabilities across 10,000 simulations of the data. Each simulation re-sampled the order of second test-phase trials to create a randomized, null distribution. These findings are visually represented in Figures 4 and 5, where each panel contains an averaged probability for each transition type (obtained from participant data) and distribution of

Transition Type	Formula
p(Repeat | Repeat)	$\frac{\Sigma(\text{Repeat} \rightarrow \text{Repeat})}{\Sigma(\text{Repeat} \rightarrow \text{Repeat}) + \Sigma(\text{Repeat} \rightarrow \text{New})}$
p(New | Repeat)	$\frac{\Sigma(\text{Repeat} \rightarrow \text{New})}{\Sigma(\text{Repeat} \rightarrow \text{Repeat}) + \Sigma(\text{Repeat} \rightarrow \text{New})}$
p(Repeat | New)	$\frac{\Sigma(\text{New} \rightarrow \text{Repeat})}{\Sigma(\text{New} \rightarrow \text{Repeat}) + \Sigma(\text{New} \rightarrow \text{New})}$
p(New | New)	$\frac{\Sigma(\text{New} \rightarrow \text{New})}{\Sigma(\text{New} \rightarrow \text{Repeat}) + \Sigma(\text{New} \rightarrow \text{New})}$

Table 1
Formulas for calculating conditional probabilities for each of the four transition types in this comparison.

probabilities from bootstrapped samples. Figure 4 contains transition probabilities for retrievals that were preceded by repeated words and Figure 5 contains transition probabilities for retrievals that were preceded by new words. Averaged transition probabilities from participant data are represented by red, vertical lines on each panel, whereas bootstrapped probabilities are represented by smoothed histograms. Values along the y-axis reflect the density of the probabilities that fall along the x-axis.

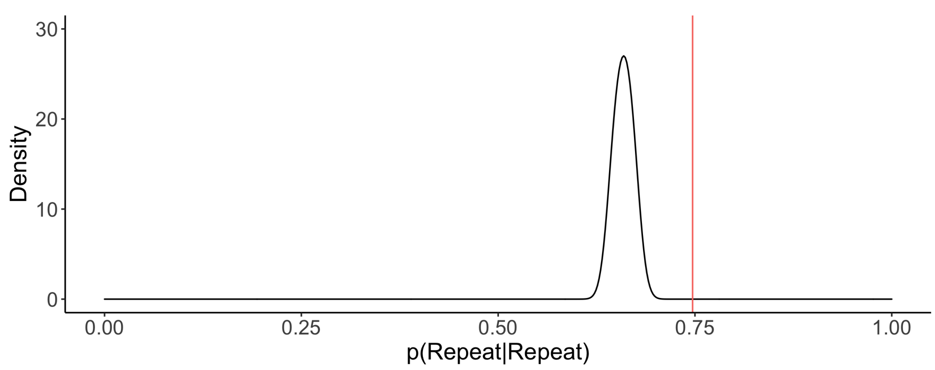


Figure 4. Transition probabilities for repeat-repeat and new-repeat transition types. Each panel calculates the probability of a word being new or repeated, when the previously retrieved word was repeated. The top panel refers to the probability of a repeated word being listed when the prior retrieval was a repetition and the bottom refers to the probability of a new word being listed when the prior retrieval was a repetition

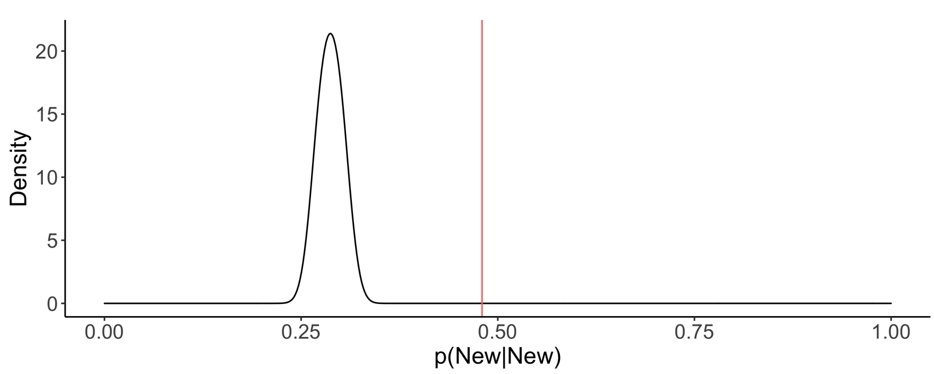


Figure 5. Two panel figure of transition probabilities for repeat-new and new-new transition types. Each panel calculates the probability of a word being new or repeated, when the previously retrieved word was new. The top panel refers to the probability of a repeated word being listed when the prior retrieval was a new word and the bottom refers to the probability of a new word being listed when the prior retrieval was a new word

Transition Type	p(Repeat ǀ Repeat)	p(Repeat ǀ New)	p(New ǀ Repeat)	p(New ǀ New)
Participant Average	0.749	0.513	0.251	0.487
Distribution Boundaries	[0.621, 0.722]	[0.632, 0.770]	[0.278 , 0.379]	[0.23, 0.368]

Table 2
Table of conditional probabilities and boundaries for bootstrapped distribution of probabilities. Each transition type in this analysis has a complimentary pair, i.e. the values for p(Repeat|Repeat) and p(New|Repeat) should sum to one, as well as the values for p(Repeat|New) and p(New|New).

The probability of generating a repeated word, when the prior retrieval was also a repeated word(0.75) and generating a new word, when the prior retrieval was also a new word (0.49), both exceeded the upper boundary of the null distribution. The probabilities for generating a repeated word after a new word (0.51) and a new word after a repeated word (0.251) both fell below the lower boundaries of the null distribution. The conditional probabilities for each transition type, averaged across participants and category, and boundaries of the corresponding bootstrapped distribution can be found on Table 2.

Contiguity Effects: Directionality, Lag-CRP, and Temporal Clustering Analysis

Directionality Bias Analysis. This comparison looked at repeat-repeat transitions, which accounted for 58% of all transitions that occurred on second test phase trials. A binomial test was used to determine whether both words in a repeat-repeat transition had a significantly higher probability of being generated in the same order across both test-phases. Forward-order transitions were defined as repeat-repeat transitions that adhered to the retrieval order of words from the first test-phase. Backward-order transitions referred to repeat-repeat transitions where both items were retrieved out of order from the first test-phase. A significantly higher proportion of repeat-repeat transitions were recalled in forward-order than backward-order (proportion of forward-order transitions= 0.6, N= 4227, 95% CI= [0.59, 0.62], p= 0.001). Further details regarding transition directionality for each category can be found in the manuscript's Appendix section on Table 5.

Contiguity Effects and Lag-CRP Analaysis. The next comparison evaluated evidence of contiguity effects across repeat-repeat transitions. The lag between both words, for each repeat-repeat transition, was calculated by subtracting the serial position of the second word from the serial position of the first word. A positive lag indicates a forward-order transition, whereas a negative lag indicates a backward-order transition. Next, the number of instances that a transition occurred at a particular lag was calculated across both participants and categories. These values were then divided by the total number of possible lags that could have occurred after the first word in each transition. The number of possible lags was determined by subtracting the serial positions of the remaining items (i.e. words from the first test-phase that had not yet been listed on the second test-phase). Counts for possible lags were similarly calculated across participants and categories. The resulting conditional response probabilities (Lag-CRP) were then plotted in Figure 6 to replicate a conventional contiguity curve. The curve shown in Figure 6 visually resembles the episodic contiguity curve example from Figure 1 and maintains several key features. CRP values increased as the absolute value of a transition's lag approached 0. Forward-order transitions had higher Lag-CRP values than backward-order transitions at shorter lags (i.e. where the absolute value of lag was less than five). The Lag-CRP curve flattens out as the absolute value of lag increases, with differences between forward and backward-order transitions dissipating at larger lags (i.e. where the absolute value of lag was greater than or equal to five).

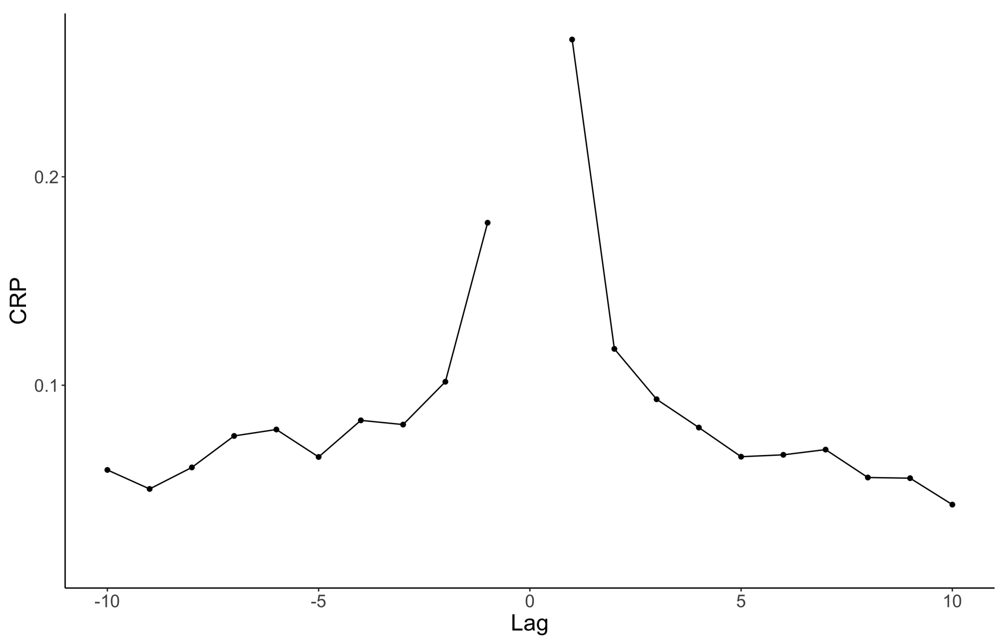


Figure 6. Conditional response probabilities as a function of the lag between two items in a repeat-repeat transition. Lag-CRP values fall along the y-axis of the figure, whereas the range of lags across all transitions fall along the x-axis of the figure. The x-axis for this figure was truncated from a range of lags between -23 to -25, to lags between -10 and 10 for visualization purposes.

Temporal Clustering Analysis. This analysis empirically tested contiguity effects by calculating temporal clustering scores for each repeat-repeat transition and contrasting those values to estimates of semantic similarity between both words in the transition. The temporal clustering analysis used the same method as Lag-CRP to calculate the lag and possible lags for each transition. These values were then ranked and a percentile was obtained by calculating the proportion of possible lags (i.e. could have occurred after the first item) that were greater than the given transition's actual lag. Transitions with a score of over 0.5 were defined as temporally clustered, given that the lag between both words was smaller than 50% of all possible transitions that could have occurred. Semantic similarity scores for each repeat-repeat transition were first obtained through Wikipedia2Vec and then scaled by category. A transition with a scaled rating greater than 0 (i.e. the average similarity rating for transitions in that category) was defined as semantically similar. Results, shown in Figure 7, are divided into four quadrants. Each quadrant represents one of four possible classifications a transition could have belonged to, dependent on its similarity ratings and temporal clustering scores. Panels in the upper-half of the figure contain temporally clustered transitions, whereas panels in the lower-half of the figure contain transitions that were not temporally clustered. Quadrants on the right-half of the figure include semantically similar transitions and contain transitions that were excluded from empirical testing. This was due to difficulty in disambiguating what drives a semantically similar and temporally clustered transition. The quadrants on the left-half of the figure contain semantically dissimilar transitions. Values in the upper quadrant were considered to be temporally clustered and values in the lower quadrant were not considered to be temporally clustered. The average temporal clustering scores for each category can be found on Table 6 in the manuscript's Appendix. All categories had an average temporal clustering score that exceeded the 0.5 threshold for this comparison, with the least clustered category (*Pieces of Furniture*) having a temporal clustering score of 0.59 and the most clustered category (*Animals*) having a temporal clustering score of 0.76.

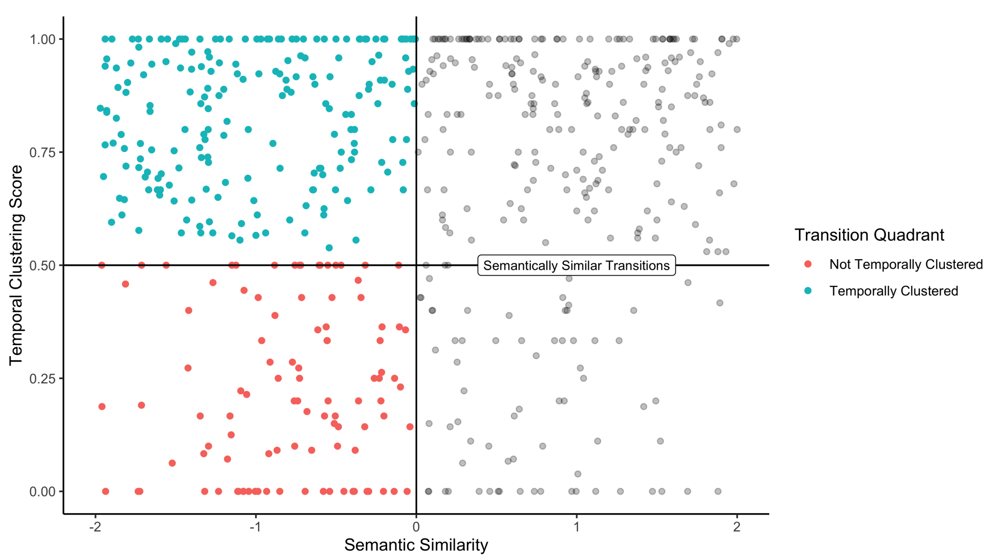


Figure 7. Repeat-repeat transitions grouped by one of four classifications. The horizontal line indicates this analysis's threshold for temporally clustered transitions, where transitions in the upper-half of the figure were considered to be clustered and those in the lower-half were not. The vertical line of the figure represents the threshold for semantic similarity, where transitions to the right of the line were considered to be semantically similar and items to the left of the line were not. Semantically similar transitions (shaded in gray) were excluded from the statistical testing in this analysis, due to the ambiguous nature of semantically similar and temporally clustered transitions.

Semantically dissimilar transitions were considered to have the lowest chance of being cued by overt semantic associations and occurring purely as a function of semantic priming. Accordingly, only semantically dissimilar transitions were considered during empirical testing. A two-sample t-Test was used to determine if semantically dissimilar transitions were significantly more likely to be considered temporally clustered. Results indicated that the number of semantically dissimilar and temporally clustered transitions was significantly greater than the number of semantically dissimilar and not temporally clustered transitions ($t(814)= 3.2956$, p = 0.0013, CI 95% [-0.21102770, -0.05260344]).

Temporal Interval Analysis

The temporal interval analysis was used to determine whether increasing the number of trials between category test-phases moderated the magnitude of repetition and contiguity effects.

Temporal Interval Analysis: Repetition Effects. This comparison evaluated whether the probability of repeating words significantly differed across all three temporal intervals. A One-Way Analysis of Variance (ANOVA) found that the probability of repeating words did not significantly differ across any temporal interval ($F(2, 1385)= 0.47$, $p= 0.62$). The three boxplots, shown in 8, contain the range of conditional probabilities for repeating a word at each temporal interval.

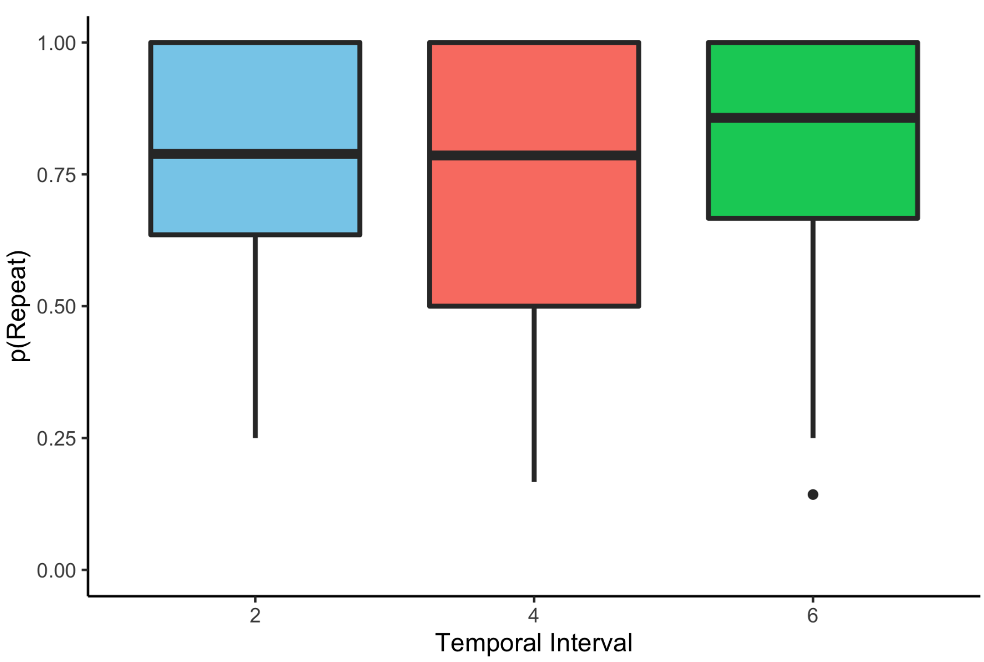


Figure 8. Conditional probability of repeating an item as a function of temporal intervals. Each boxplot represents the range of conditional probabilities, as well as the median value (denoted by the horizontal line), across all three temporal interval levels. The values along the x-axis refer to how many trials intervened the initial and repeated presentation of a category, while the values along the y-axis represent the probability of an item being listed on the second test-phase of a category, given that it was also listed on the first test-phase.

Temporal Interval Analysis: Contiguity Curve and Lag-CRP. The next comparison recalculated Lag-CRP values for three different subsets of the data, which corresponded to the temporal interval that each repeat-repeat transition occurred in. The curve shown in Figure 9 replicates the contiguity curve shown in Figure 6, with the exception that each line represents Lag-CRP at a different temporal interval. A One-Way

Anova found no significant differences in conditional response probabilities between any of the three temporal intervals (F(2, 4224)= 0.02, p= 0.83). The curves across all three temporal intervals also visually resemble Figure 6. A follow up comparison assessed

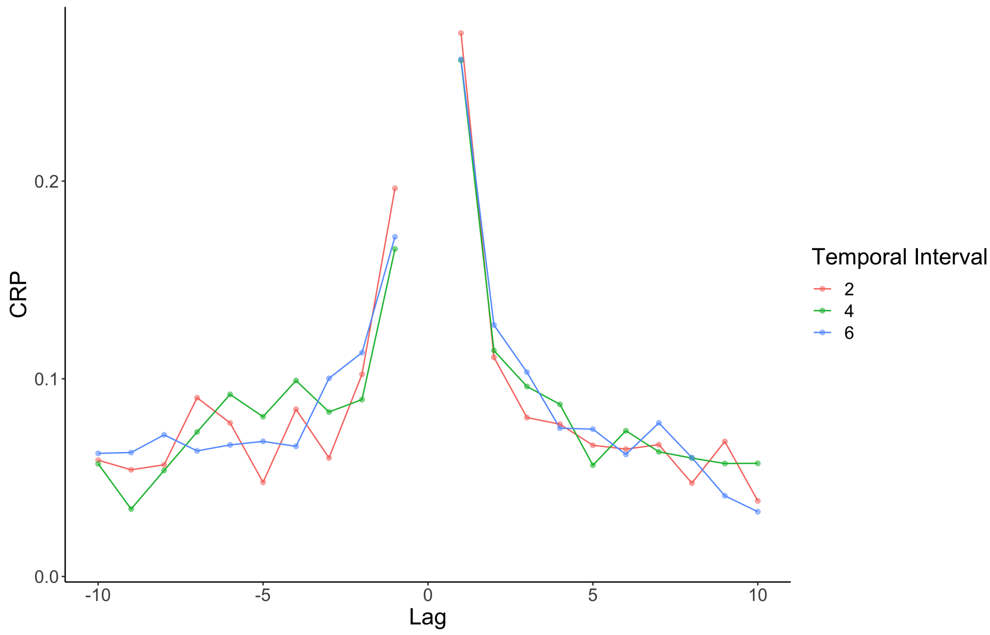


Figure 9. Lag-CRP as a function of temporal interval. The three lines in this figure represent three distinct intervals that trials could have been spaced between, i.e. two intervening trials, four intervening trials, and six intervening trials.

whether in Lag-CRP significantly differed across temporal intervals, only considering the smallest possible lags (i.e. lag of -1 and lag of +1). A two-sample t-test found that conditional response probabilities, when lag was equal to +1, were significantly higher in temporal interval two categories than temporal interval four categories (t(1986)= 2.47, p= 0.013) and temporal interval six categories (t(1978)= 1.9, p= 0.049). No significant differences were found between temporal intervals four and six, when lag was equal to +1 (t(1953)= 1.29, p= 0.2). When lag was equal to -1, Lag-CRP was significantly higher for temporal interval two than temporal interval four (t(2470) = 2.00, p = 0.047) and (t(2468)= 1.89, p= 0.043). No significant differences were found between temporal intervals four and six, when lag was equal to -1 (t(2490)= 0.13, p= 0.89). A barplot with the average Lag-CRPs for each temporal interval is shown in Figure 10. Each bar represents a temporal interval, with the left panel showing average CRP's when lag is equal to -1 and the right panel showing CRP's when lag is equal to +1.

Discussion

Repetition Effects

Probability of Repetitions Versus Word-Typicality. In a scenario where repetitions were simply occurring by chance, or due to certain words being highly represented under a category (e.g. the word *DOG* for the category prompt *Animals*), the conditional probability of repeating a word should not exceed that word's general

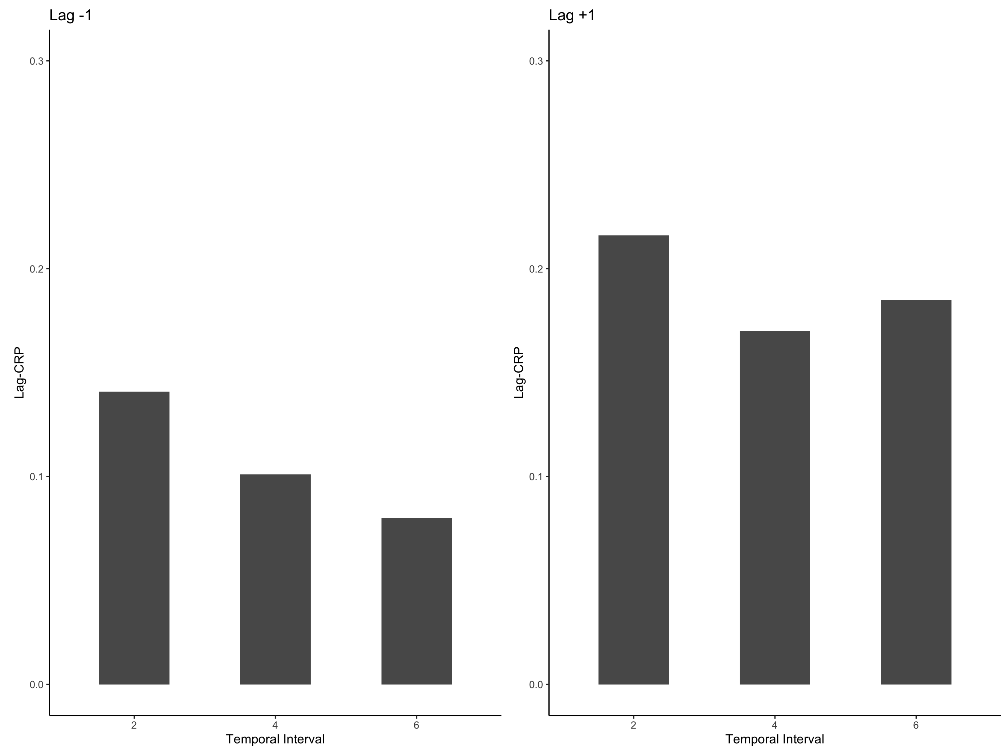


Figure 10. Temporal interval analysis: Lag-CRP for small lags. Each bar represents the average CRP (y-axis) across temporal intervals (x-axis) The left panel shows average CRPs when lag is equal to +1 and the right panel shows average CRPs when lag is equal to -1.

probability of being listed on the category's second test-phase (i.e. word-typicality scores). The conditional probabilities shown in Figure 3 exceeding the expected typicality threshold for each repeated word suggests that items were being repeated across test-phases more frequently than chance and to a degree that could not be accounted for by an obvious semantic explanation. Further this comparison revealed that low typicality items had among the highest probabilities of being repeated across test-phases, which could be interpreted as one of two possible outcomes. The first interpretation is that participants retrieved information both semantically and episodically on second test-phase trials. A potential strategy for second test-phase trials would be to think back to words that were listed on the first test-phase of a category and switch to semantic retrievals when episodically retrieving information becomes too cumbersome. While there is not a definitive method of parsing between episodically and semantically retrieved repetitions, the observation of less typical words having among the highest conditional probabilities of being repeated represents a plausible marker for episodic recalls occurring during the repeated fluency task. Another interpretation of this finding could simply be that additional references to information modulate episodic and semantic retrievals in a similar way. Whether represented as an episode or semantic knowledge, this finding suggests that initially recalling information increases its retrievability in the future.

Repetition Clustering. In a scenario where participants were repeating words but still engaging in a purely semantic retrieval strategy, one could expect to see repeated and new words interspersed, with no clear indications of clustering. The tendency for participants to respectively group new and repeated words together on second test-phase trials similarly, while not necessarily evidence of episodic recalls, could be interpreted in one of two ways. The first is that participants were simply engaging in the same semantic retrieval strategy across both category test-phases. In this case participants may have repeated words that were highly represented under the given category prompt, which subsequently cued the same semantic associates from the first test-phase trial. An alternative explanation could be generating a new word on the second test-phase of a category cues the retrieval of semantic associates, whereas the retrieval of a repeated word on the first test-phase trial episodically cues the retrieval of other previously listed words.

Contiguity Effects

Directionality. While observing a forward-order directionality bias does not, on its own, represent a definitive marker for temporal ordering effects, the absence of an ordinal bias effectively rule out the possibility of observing episodic tendencies in both subsequent contiguity analyses (i.e. the Lag-CRP curve and Temporal Clustering scores). The bias towards forward order transitions introduces a potential episodic account for future tendencies in the data. Specifically, the fact that the majority of Repeat-Repeat transitions were recalled in forward-order could point to the possibility of individuals thinking back to their original response lists or being episodically primed to retrieve words in the order that they were generated in, on the first category test-phase. So, while inconclusive as an independent measure, the forward-order directionality bias provided a foundation for future comparisons and, concurrently, added context to the battery of contiguity analyses.

Lag-CRP Curve. There is not an obvious semantic explanation to account for the striking resemblance between the contiguity curve calculated from repeat-repeat transitions in this experiment (Figure 6) and episodic contiguity curves (Figure 1). However, while not likely, it is possible that the contiguity curve in Figure 6 consisted of purely semantic item generations and that, when presented with an additional retrieval opportunity, semantically retrieved repeat-repeat transitions appear visually similar to episodic recall data. One possible explanation could be that repeat-repeat transitions on the second test-phase trials were reflecting the semantic similarity, rather than referencing features of an encoded episode. An episodic account of this tendency would indicate that words from the first test-phase trials, particularly those retrieved in close proximity, later served as episodic retrieval cues for each other (Hintzman, 2016). This is further supported by forward-order transitions (i.e. transitions with positive lags) maintaining an advantage over backward-order transitions, indicating that both lag and temporal order of retrievals on the first test-phase trials influenced the organization of retrievals on the second test-phase trials.

Temporal Clustering Analysis. Given that semantic retrievals are often organized by overlapping concepts and shared attributes, a semantic account of contiguity effects would suggest that repeat-repeat transitions with smaller lags simply reflect semantic priming between conceptually similar words (e.g. $DOG \rightarrow CAT$) across both test-phase trials. Findings from the temporal clustering analysis, however, undermine this account. Semantically dissimilar transitions were likely to occur across both test-phase trials, but under a purely semantic retrieval strategy, it would be highly unlikely for semantically dissimilar transitions to also be temporally clustered. This is due to the fact temporal clustering effects require both dissimilar words to be generated in close proximity across both category test-phases. Using an example from the data, the repeat-repeat transition $CUCUMBER \rightarrow CHICKPEA$ was classified as semantically dissimilar, with a rating of -1.2. and temporal clustering score of 0.88. Under a purely semantic retrieval strategy, one would expect to see a greater proportion of semantically dissimilar transitions be classified as not temporally clustered. As demonstrated in 7, significantly more dissimilar transitions had temporal clustering scores that were greater than 0.5 and, thus, were considered to be temporally clustered. Further, temporal clustering scores were averaged across both categories and participants, with the averages for each category and each participant exceeding 0.5. This indicates that that repeat-repeat transitions consistently demonstrated temporal clustering effects. While it is still not possible to infer whether these retrievals were episodic or semantic in nature, the larger aggregates of data suggest that repeat-repeat transitions (particularly those that were classified as semantically dissimilar) were referencing temporal features of the first test-phase trials.

Temporal Interval Analysis

While a cursory glance at the temporal interval analysis may initially point to temporal intervals not modulating the organization of second test-phase retrievals, this is not necessarily the case. Neither the probability of repeating words nor Lag-CRP values significantly differed across temporal intervals. However, this may be more representative of structural differences between free recall and repeated fluency tasks than the influence of temporal intervals. This is supported by two critical findings from the Lag-CRP comparisons. The first is that, when only considering repeat-repeat transitions, the contiguity curve from Figure 6 is maintained across all three temporal intervals, shown in 9. While not impossible, it is highly unlikely that the distinct shape and directionality bias would be maintained across all intervals, strictly as a function of primed semantic associates. This outcome seems more representative of the findings from (Kahana, 1996), shown in Figure 1, where contiguity effects were observed across varying list lengths, presentation formats, and exposure duration. The second critical finding is that Lag-CRP values were significantly higher in temporal interval two than temporal intervals four and six, when only considering short lags (i.e. lags of +1 and lags of -1). An important consideration is that this comparison only looked at transitions that were repeated adjacently across both category test-phases, meaning that these transitions were the most temporally clustered in the data and were unaffected by additional noise from new item generations on second test-phase trials. Taken together, the initial null findings for differences in the probability of repeating a word and Lag-CRP values seem to speak more to differences between repeated fluency and free recall than the temporal modulation of retrievals.

Limitations

This study may have been potentially limited by three key factors: structural differences between free recall and semantic fluency tasks, semantic fluency performance for non-native english speakers, and category sequencing error on the repeated fluency task. Unlike a free recall task, presenting all participants with the same set of stimuli, the list of generated words on each category's first test-phase differed across participants. This was a particular concern when analyzing temporal interval effects, due to the lack of control that experimenters had over the list length and retrieval content of semantic fluency trials. For example, participants who consistently listed more words per trial would inherently have greater a higher number of intervening events that increased with each temporal interval. Conversely, this limitation could potentially explain idiosyncrasies in the temporal interval analyses. For example, there were no differences across temporal intervals when evaluating the probability of repeating a word or Lag-CRP across repeat-repeat transitions; However, the conditional response probabilities for smaller lags (i.e. lags of +1 and lags of -1) were significantly higher for temporal interval two categories than temporal interval four or six categories.

Another key difference between the repeated fluency and free recall paradigm involves participant instruction. In a free recall task, participants are explicitly asked to study and recall stimuli, whereas in the repeated fluency task, participants were only instructed to list items that fit within a given category prompt. This limited the number of inferences that could be drawn from observing episodic tendencies. Without explicit control over the content of each category's first test-phase and clear directive as to the type of retrievals to execute on each category's second test-phase, it was not possible to determine whether any words were episodically or semantically retrieved. Despite an inability to characterize these effects as evidence of episodic recalls during semantic fluency, large aggregates of data still point to notable overlap between episodic and semantic retrieval tendencies.

While the inclusion of both native and non-native english speakers could have potentially influenced outcomes on this task, performance differences between both groups suggest that this is unlikely. One consideration, however, is that comparisons involving estimates of word-typicality and semantic similarity may have been negatively skewed. The semantic meaning of words can differ both across translation and cultural norms (e.g. fruits that are common in the United States may not be considered common in other countries). Certain exemplars for a category, such as the word *Banana* for the category *Fruits*, may have received lower typicality scores than if the study had only sampled native english speakers. Given that neither semantic fluency performance (i.e. number of words listed per trial) nor that words were repeated significantly differed between either group, the inclusion of nonnative speakers was not likely to have affected the overall results of this study.

One final limitation was caused by a coding error in the repeated fluency task. This error dealt with the sequencing of categories and resulted in all participants receiving an additional presentation of the final category prompt (i.e. three trials), but only a single presentation of the second to last category prompt. This was addressed in analyses by excluding data from each participant's final trial (i.e. the third presentation of the final category), as well as excluding data from the category that was only presented once. Analyses used data from 11 out of the 12 category prompts, excluding the second to last trial for each participant. The experiment's trial sequence ensured that all categories were still equally represented across participants and temporal intervals. Additionally, all participants completed at least one test-phase for each semantic fluency category and each category that was analyzed had two test-phase trials.

Conclusion

One interpretation of these results is that repetition and contiguity effects can occur during semantic fluency, but were not previously observable under a single test-phase format. It could be the case that the additional representations a word received, from being initially generated, enhanced its representation under the category on the second test-phase. Similarly, cueing words from the first test-phase may have strengthened the semantic association between both words on second test-phase trials. Another interpretation of these results is that participants were switching between episodic and semantic recalls during second test-phase trials. These switches could have occurred strategically, in order to optimize the number of words that were listed on each trial, or unintentionally, when one retrieval type cues another retrieval type. A scenario in which these switches occur intentionally would likely be expressed as an individual beginning with one retrieval type and switching to the other when the current strategy becomes too effortful. Conversely, unintentional switches could occur if retrieving information in one way spontaneously cues an individual to retrieve information in another way. For example, a participant could have started a trial by listing semantically retrieved information and, upon repeating a previously listed word, been reminded of other words that were listed on the first test-phase.

While the repeated fluency paradigm was originally designed to model semantic representations and estimate individual semantic networks (Zemla & Austerweil, 2018), these findings suggest that repeated presentations of a semantic cue does not necessarily elicit purely semantic responses each time. At the very least, evidence of repetition and contiguity effects indicate that the responses from second test-phase trials are organized differently than responses from a conventional semantic fluency test. Further this implicates other applications of semantic fluency, such as its use for diagnostic criteria for dementia, Alzheimer's disease, and Mild Cognitive Impairment, (Saranpää et al., 2022). Diagnosing one of the described conditions often requires multiple screenings, in which different neuropsychological batteries may use similar or identical semantic fluency tests. It could be the case that certain words and primed associates are more accessible to patients, due to repeated exposure and temporal associations from previous exams. It is also possible that outcomes from multiple administrations of semantic fluency may be interpreted as purely semantic in nature, when in actuality patients may be executing a combination of episodic and semantic retrievals. Consequently, the repeated fluency task, as applied in this experiment, offers key insights on how semantic retrieval changes across multiple representations of the same cue. While episodic and semantic retrieval are often described as qualitatively different(Quillian, 1967; Tulving, 1972), findings from this experiment suggest that the delineation between both variants of long-term memory is not as clear.

Supplementary Materials

A link to supplementary materials can be found at https://osf.io/3ngsk/. Here you can find a step-by-step tutorial for each analysis in this document, a sample data used for the tutorial, and the full dataset that was collected and analyzed in this experiment .

- **Analyses Tutorial:** wilder_masters_appendix.pdf
- **Tutorial Dataset:** snafu_example.csv
- **Full Dataset:** results_cleaned.csv

Appendix

Analysis Formulas

Category	First-Test Phase	Second Test-Phase
Vegetables	10.56	17.74
Pieces of Furniture	10.00	11.29
Clothing Articles	16.18	11.9
Cities	15.68	19.05
Countries	18.69	16.97
Fruits	14.15	14.65
Animals	19.44	12.22
Methods of Transportation	12.05	12.34
Units of measurement	10.54	21.23
Sporting Games	12.79	13.63
Musical Instruments	11.17	11.21
Objects You Would Find in The Kitchen	17.70	18.50

Table 3
The average number of words that were generated on each category. The center column reflects the average number of words that were generated on first test-phase trials and the right column reflects the number of words that were generated on second test-phase trials. Values were averaged first by participants and then by trial

Category	First Test-Phase	Second Test-Phase
Vegetables	0.778	0.659
Pieces of Furniture	0.719	0.645
Clothing Articles	0.713	0.72
Cities	0.647	0.685
Countries	0.704	0.596
Fruits	0.752	0.726
Animals	0.641	0.716
Methods of Transportation	0.774	0.744
Units of measurement	0.802	0.572
Sporting Games	0.691	0.633
Musical Instruments	0.794	0.81
Objects You Would Find in The Kitchen	0.643	0.61

Table 4
The proportion of words that were repeated on each test-phase trial. The center column reflects the proportion of repeated words for first test-phase trials, (i.e. how many words went on to be repeated that were generated on first test-phase trials?), the right column shows the proportion of words that were repeated on second test-phase trials i.e. how many words were repetitions on the second presentation of the category prompt?. Values were averaged first by participants and then by trial

Supplementary Tables.

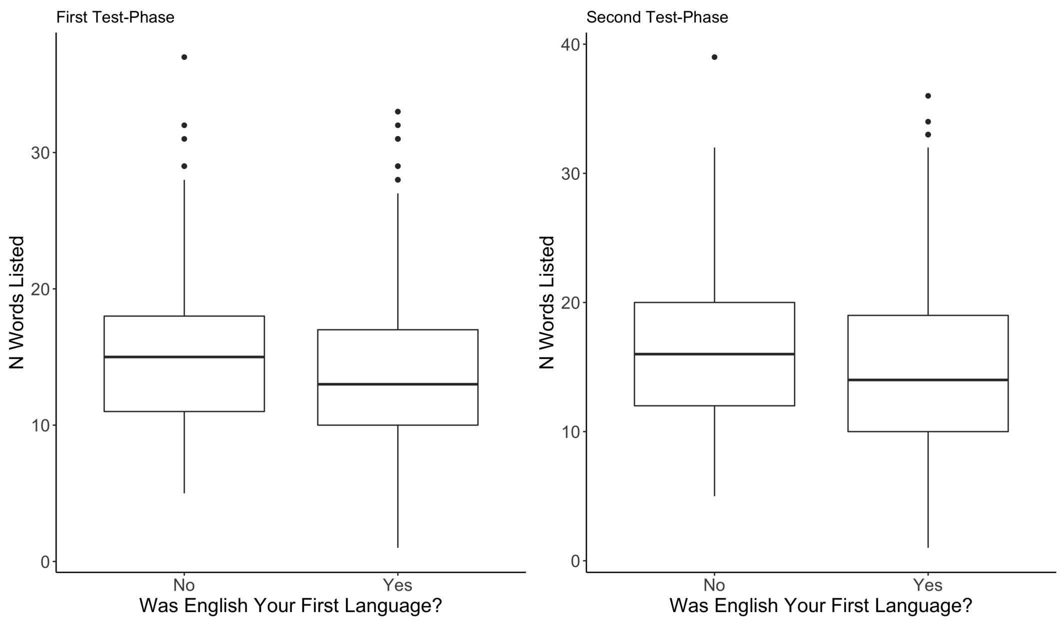


Figure 11. Performance differences between native and nonnative english speakers. The left panel shows the average number of words that was listed by each group on the first test-phase trials, whereas the right panel shows the average number of listed by each group on the second test-phase trials. Each panel contains two boxplots, one for non-native english speakers and one for native english speakers. Boxplots contain the average number of words listed across each category for the first (left panel) and second (right panel) test-phases. No significant differences were found between groups.

Figure 12. The probability of repeating words across native and non-native english speakers. Similarly to 11, each boxplot represents the average probability of repeating words across different categories. There were no significant differences between the probabilities of native speakers and non-native english speakers repeating words

Category	Forward-Order Transitions	Backward-Order Transitions
Animals	0.604	0.396
Cities	0.614	0.386
Clothing Articles	0.631	0.369
Countries	0.588	0.412
Fruits	0.612	0.388
Methods of Transportation	0.611	0.389
Musical Instruments	0.572	0.428
Objects You Would Find in The Kitchen	0.592	0.41
Pieces of Furniture	0.558	0.442
Sporting Games	0.64	0.36
Units of measurement	0.59	0.41
Vegetables	0.576	0.424

Table 5
Proportion of forward-order and backward-order transitions by category. This table only considers repeat-repeat transitions, referring to adjacently output words on the second test-phase that were also listed on the first test-phase.

Category	Temporal Clustering Score
Animals	0.701
Cities	0.729
Clothing Articles	0.752
Countries	0.760
Fruits	0.716
Methods of Transportation	0.630
Musical Instruments	0.682
Objects You Would Find in The Kitchen	0.709
Pieces of Furniture	0.592
Sporting Games	0.661
Units of measurement	0.669
Vegetables	0.680

Table 6
Temporal clustering scores for each category were achieved by averaging the scores of repeat-repeat transitions on second test-phase trials. Temporal clustering scores, on average, were greatest for the 'Countries' category (0.76) and smallest for the 'Pieces of Furniture' category (0.59).

References

Bower, G. (2000). A brief history of memory research. *Oxford Handbook of Memory*, 3–32.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428.

Cooper, D., Lacritz, L. H., Weiner, M., Rosenberg, R., & Cullum, C. (2004). Category fluency in mild cognitive impairment: reduced effect of practice in test-retest conditions. *Alzheimer Disease & Associated Disorders*, *18*(3), 120–122.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17-24.

Dooling, D. L., & Lachman, R. (1971). Effects of academic experience on student performance. In *1971 conference of the american psychological association.* San Francisco, CA.

Greene, R. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 371-386.

Guerin, S., & Miller, M. (2008). Semantic organization of study materials has opposite effects on recognition and recall. *Psychonomic Bulletin and Review*, *12Z*(2), 302–308.

Habib, R. (2016). The effect of semantic context on the neural correlates of retrieval success and the subsequent processing of retrieved words. *Behavioral Neuroscience*, *44*(1), 109-119.

Hebb, D. (1961). Distinctive features of learning in the higher animal. *Brain Mechanisms and Learning*, *1*(3), 8-27.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia*, *42*(9), 1212–1222. doi: 10.1016/j.neuropsychologia.2004.02.001

Hills, T., Jones, M., & Todd, P. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431–441.

Hintzman, D. L. (2016). Is memory organized by temporal contiguity? *Memory & Cognition*, *44*(3), 365–375.

Howard, M., & Kahana, M. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*(1), 85–98.

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

Howe, M. (1970). Repeated presentation and recall of meaningful prose. *Journal of Educational Psychology*, *61*(3), 214–219.

Jones, M., & Mewhort, D. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1-37. doi:

10.1037/0033-295X.114.1.1

Kahana, M. (1996). Associative retrieval processes in free recall. *Memory and Cognition*, *12*(24), 103-109.

Karpicke, J., Lehman, M., & Aue, W. (2014). Retrieval-based learning: An episodic context account. In *Psychology of learning and motivation* (Vol. 61, pp. 237–284). Academic Press.

Madigan, S. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*(6), 828–835.

McNamara, T., & Diwadkar, V. (1996). The context of memory retrieval. *Journal of Memory and Language*, *35*(6), 877–892.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2014). Efficient estimation of word representations in vector space. *ArXiv*(4), 155-162.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 112-123. doi: 116–133-21

Pakhomov, S. V., Hemmy, L. S., & Lim, K. O. (2012). Automated semantic indices related to cognitive function and rate of cognitive decline. *Neuropsychologia*, *50*(9), 2165–2175.

Polyn, S., Norman, K., & Kahana, M. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *12*(2), 43-52.

Prince, S., Daselaar, S., & Cabeza, R. (2005). Neural correlates of relational memory: successful encoding and retrieval of semantic and perceptual associations. *Journal of Neuroscience*, *25*(5), 1203–1210.

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410–430.

Raaijmakers, J. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*(3), 431–452.

Roediger, H., & McDermott, K. (1995). Creating false memories: Remembering words not presented in lists. *Psychonomic Bulletin & Review*, *21*(4), 803–814.

Sadeh, T., Moran, R., & Stern, Y. (2019). The dynamics of remember/know judgments in free recall: An examination of item- and context-information over time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *57*, 981–992. doi: 10.1037/xlm0000664

Saranpää, A. M., Kivisaari, S. L., Krumm, S., & Salmelin, R. (2022). Moving in semantic space in prodromal and very early Alzheimer's disease: An item-level characterisation of the semantic fluency task. *Frontiers in Psychology*, *13*(491), 491-501.

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. doi: 10.1037/a0013396

Troyer, A., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138-147.

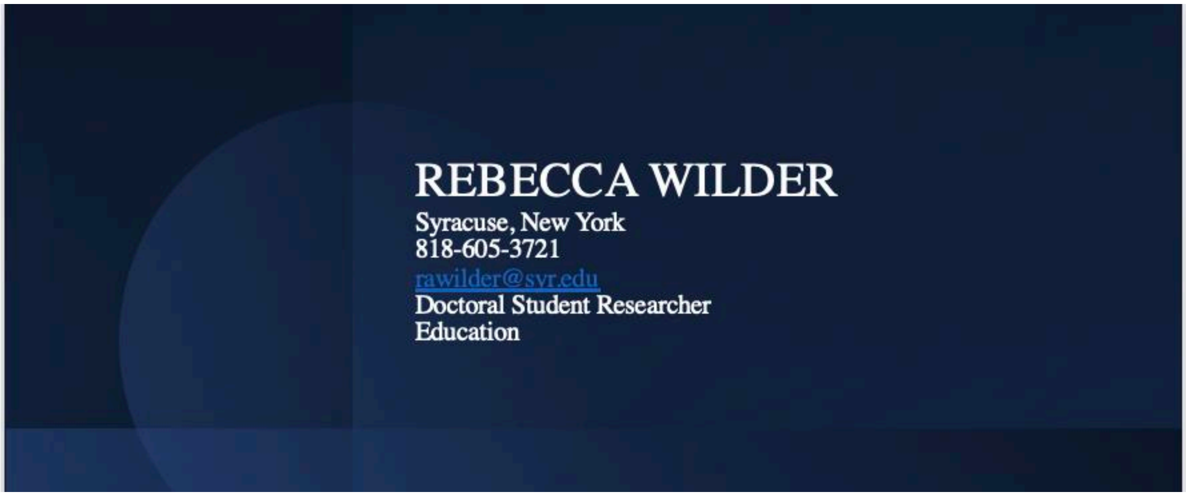
Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*, 393–403.

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.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, *1*(1), 89–106. doi: 10.3758/bf03200763

Yamada, I., Asai, A., Sakuma, J., Shindo, H., Takeda, H., Takefuji, Y., & Matsumoto, Y. (2020). Wikipedia2vec: An efficient tool for learning distributed representations of wikipedia articles. *ACM Transactions on Information Systems (TOIS)*, *38*(4), 18-29.

Zemla, J. C., & Austerweil, J. L. (2018). Estimating semantic networks of groups and individuals from fluency data. *Computational Brain & Behavior*, *1*(1), 36–58.



Education

Doctorate in Philosophy (Ph.D.), Cognitive Psychology

Syracuse University
Interdisciplinary Graduate Neuroscience Concentration
- Declared 2021

Master of Science (M.S.)
- Anticipated 2022

Doctorate in Philosophy (Ph.D.)
- Anticipated 2025

2018 - 2019

Post-Baccalaureate in Psychological Science

University of California, Irvine
Research-focused certificate program
Research Labs
- Yassa Translational Neurobiology Lab
- Bornstein Computational Cognitive Neuroscience Lab (CCNL)

2013 – 2018

Bachelor of Arts (B.A.) Psychological Science

University of California, Irvine

Honor Roll

- Dean's List: Winter 2015
- Dean's List: Fall 2016
- Dean's List: Winter 2017
- Dean's List: Spring 2017
- Dean's List: Fall 2017
- Dean's List: Winter 2018
- Dean's List: Spring 2018

Fellowship

- Sinai Scholar Alumni, 2018 (University of California, Irvine)
- InterAnalytic Couples Therapy, Human Sexuality, & Sex Therapy Fellowship, 2018 (American Association of Couples and Sex Therapists; University of California, Los Angeles)

Undergraduate Research Assistantships

- Yassa Translational Neurobiology Lab
- Piff Psychology Lab

Research and Professional Experience

2021-Present

Graduate Student Researcher

Zemla Lab

Master's Thesis

Evaluating evidence of episodic reliance during semantic retrieval tasks.

- Proposed

2021-2022

Teaching Assistant

Syracuse University

Cognitive Psychology (PSY 322; NEU 322)

2020-2021

Graduate Student Researcher

Lohnas Lab

First Year Project

Evaluating potential modulators during initial recall that influence long-term retrievability in final free recall.

- Completed

2020-2021

Teaching Assistant

Syracuse University

Introduction to Psychology (PSY 205)
2020-2021

Introduction to Cognitive Psychology (PSY 322)
2021-2022

Statistical Methods II (PSY 252)
2022-2023

Research Assistant

Yassa Translational Neurobiology Lab
Primary Projects
- Memory Orthogoanlization and Mnemonic Devices Task
- Emotional Pattern Separation Performance Differences Associated With Major Depressive Disorder

General Responsibilities
- Neuroimaging (fMRI Scanning)
- Neuroimaging and behavioral data analysis
- Assistance with neuropsychological inventory administration and scoring
- Data cleaning

2018-2019

Research Specialist

Bornstein Computation & Cognitive Neuroscience Lab
Primary Projects
- *Contextual Bandit*: Reinstatement of reward values from past decisions bias future choices
- Intertemporal Choice: The influence of stressors and perceived payout probability on model-based planning and delayed-reward discounting

General Responsibilities
- Programming experiment
- Coding analyses scripts
- Training research assistants
- Drafting IRB documents

2016-2017

Research Assistant

Piff Psychology Lab
Lab Projects
The Role of Emotional Valence & Social Drive in Smartphone Application Usage

Conference Presentations

- Wilder, R.A. (2019, May) Emotional Pattern Separation Performance Differences Associated with Major Depressive Disorder. Poster presented at the annual Associations for Psychological Science Convention, Washington D.C.
- Wilder, R. A. & Jones, N. M. (2019, February) The Role of Emotional Valence & Social Drive in Smartphone Application Usage. Poster presented at the annual Society for Personality & Social Psychology conference, Portland, OR
- Wilder, R.A. (2018, May) Memory & Mood: Emotional Pattern Separation Performance Differences Associated with Major Depressive Disorder. Poster presented at the annual Undergraduate Research Opportunity Program conference, Irvine, CA.
- Wilder, R. A., Philbrick, G. M., Le, K., Jones, N. M. (2017, May). Social media application use and belongingness. Poster presented at the annual Undergraduate Research Opportunity Program conference, Irvine, CA

Volunteer Work and Extra Curricular

2020-Present
Member
Society for Neuroscience

2021-Present
Member
Psychology Action Committee (PAC)
Syracuse University
2018Present
Volunteer
Habitat for Humanity & The ReStore

- Syracuse
- Irvine

20182019
Sinai Scholar Alumni
Sinai Scholars
20092013
Volunteer
Calabasas Environmental Commission