American Psychological Association

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

2021, Vol. 47, No. 5, 765-784
http://dx.doi.org/10.1037/xlm0000964

Predicting Recall of Words and Lists

Ada Aka, Tung D. Phan, and Michael J. Kahana
University of Pennsylvania

For more than a half-century, lists of words have served as the memoranda of choice in studies of human memory. To better understand why some words and lists are easier to recall than others, we estimated multivariate models of word and list recall. In each of the 23 sessions, subjects ($N = 98$) studied and recalled the same set of 576 words, presented in 24 study-test lists. Fitting a statistical model to these data revealed positive effects of animacy, contextual diversity, valence, arousal, concreteness, and semantic structure on recall of individual words. We next asked whether a similar approach would allow us to account for list-level variability in recall performance. Here we hypothesized that semantically coherent lists would be most memorable. Consistent with this prediction, we found that semantic similarity, weighted by temporal distance, was a strong positive predictor of list-level recall. Additionally, we found significant effects of average contextual diversity, valence, animacy, and concreteness on list-level recall. Our findings extend previous models of item-level recall and show that aggregate measures of item recallability also account for variability in list-level performance.

Keywords: episodic memory, list recall, semantic associations, word properties, word recall

Ever since Ebbinghaus introduced the consonant-vowel-consonant (CVC) in his seminal experimental analysis of serial learning, students of memory have sought to measure, manipulate, and control the mnemonic difficulty of list materials. Glaze (1928) demonstrated that CVCs varied substantially in their meaningfulness, and Hull (1933) showed that these differences predicted the ease of learning individual items. The fact that CVCs varied considerably in their memorability, and that such variation was often idiosyncratic to individual subjects, was one reason that postwar-era scholars abandoned CVCs in favor of common words as the memoranda of choice in studies of learning and memory. Using words, researchers could draw upon a much larger sample of memoranda; and although words also vary in their memorability, they exhibit less interpretive ambiguity across individuals and require less response learning than do CVCs. Nonetheless, it is of vital importance that researchers have good models for estimating the mnemonic difficulty of learning both individual words and entire lists. The goal of the present study is to construct parallel statistical models to account for variability in free recall of both words and lists (as measures of list-level performance are common) and to assess the stability of these models at the level of individual subjects.

Earlier studies establish the importance of several variables as predictors of word-level recall. Below we discuss findings involving seven specific word properties: concreteness, contextual diversity, word length, valence, arousal, meaningfulness, and animacy. In the case of concreteness,[1] prior work has demonstrated that concrete words exhibit a mnemonic advantage over abstract words in a variety of memory paradigms, including free recall (Dukes & Bastian, 1966; Hamilton & Rajaram, 2001; Paivio, 1967), item recognition (Gorman, 1961), paired-associates (Epstein, Rock, & Zuckerman, 1960), and immediate serial recall (Walker & Hulme, 1999).

By contrast, traditional word frequency[2] exhibits a mixed pattern of results, with low-frequency words possessing an advantage on recognition tests (Gorman, 1961) and high-frequency words possessing an advantage in free recall of pure lists (Hall, 1954). In free recall of mixed lists, both low- and high-frequency words exhibit superior recall to words of midfrequencies (Lohnas & Kahana, 2013). Recent work has demonstrated that contextual diversity, the number of contexts in which a word has appeared, is a better measure to predict word-naming and lexical decision times (Adelman, Brown, & Quesada, 2006; Brysbaert & New, 2009). Distinguishable effects of contextual diversity and word frequency has been shown in recognition memory (Steyvers & Malmberg, 2003), serial recall (Parmentier, Comesaña, & Soares, 2017), and

This article was published Online First October 22, 2020.

Ada Aka, Department of Psychology and Department of Marketing, University of Pennsylvania; Tung D. Phan and Michael J. Kahana, Department of Psychology, University of Pennsylvania..

Tung D. Phan is now at Amazon, Seattle, Washington.

The authors gratefully acknowledge support through National Institutes of Health Grant MH55687. We thank Adam Broitman, Logan O'Sullivan and Nicole Kratz, Patrick Crutchley, Elizabeth Crutchley, and Kylie Hower for assistance with data collection, Jonathan Miller and Isaac Pedisich for assistance with designing and programming the experiment, and Rivka Cohen, Nora Herweg, and Christoph Weidemann for very helpful discussions.

Correspondence concerning this article should be addressed to Ada Aka, Department of Psychology, University of Pennsylvania, 425 South University Avenue, Suite 201, Philadelphia, PA 19104. E-mail: adaaka@wharton.upenn.edu

[1] *Concreteness* is defined in terms of directness of reference to sense experience (Paivio et al., 1968).

[2] *Word frequency* refers to the estimate of the frequency usage in the English language.

cued recall (Criss, Aue, & Smith, 2011). In light of this work, contextual diversity has became a commonly used measure for frequency.

Researchers have found word length[3] to influence memorability of studied items. In the case of immediate serial recall, short words boast a substantial mnemonic advantage (Baddeley, Thomson, & Buchanan, 1975), whereas in free recall results have been mixed (Hulme, Suprenant, Bireta, Stuart, & Neath, 2004; Katkov, Romani, & Tsodyks, 2014; Tehan & Tolan, 2007).

Previous work has also demonstrated a positive relationship between the emotionality[4] of a word and its memorability. In free recall, emotional words (positively or negatively valent or arousing) are generally better remembered than neutral words (Dolcos, LaBar, & Cabeza, 2004; Kensinger & Corkin, 2003; LaBar & Cabeza, 2006). More recently, some researchers challenged the understanding of the relationship between emotionality and memory. Talmi & Moscovitch (2004) highlighted the role of emotionality as a potential organizing principle, and Hunt, Trammel, & Krumrei-Mancuso (2015) demonstrated that emotion may impair memory for overall meaning for items. More specifically, in Hunt et al.'s (2015) study, emotion impaired recall of the semantically related list but not the unrelated list.

Whereas the previous five measures pertain to the intrinsic properties of words, one might expect the similarity relations among words to be particularly predictive of item recall. The highly cue-dependent nature of recall (Kahana, 1996) and the substantial evidence for semantic organization in recall (Howard & Kahana, 2002b; Klein, Addis, & Kahana, 2005) indicate that remembering one word will tend to retrieve related words from memory. More specifically, the semantic similarity effect highlights the importance of semantically related items for recall performance, whereas the temporal contiguity effect demonstrates how recalled items often cue the recall of other items presented in adjacent (temporal) serial positions in the list.

In a classic study, Noble (1952) defined the meaningfulness (*m*) of an item as the number of strong associations that a given word possesses with other words, as measured using a free association task (Noble, 1952; Paivio, Yuille, & Madigan, 1968). The positive relationship between meaningfulness and recall performance has been demonstrated in multiple studies (Christian, Bickley, Tarka, & Clayton, 1978; Paivio et al., 1968). More recently, Nelson and colleagues systematically investigated the effects of associative networks on performance in a wide range of memory tasks. In cued recall tasks, they found that increasing the number of semantic associations to either the cue or the target word resulted in lower recall rates. In free recall, however, they found less consistent results (Bruza, Kitto, Nelson, & McEvoy, 2009; Nelson & McEvoy, 1979; Nelson, McEvoy, & Pointer, 2003; Nelson, Schreiber, & McEvoy, 1992).

Whereas these previous studies focused on characterizing the effects of each of these variables in isolation, Rubin and Friendly (1986) took a multivariate approach to predict the mnemonic difficulty of individual words. They considered measures of orthography, pronunciability, imagery, concreteness, meaningfulness, availability, familiarity, frequency of occurrence, goodness, and emotionality. Rubin and Friendly's (1986) results demonstrated that free recall of 925 nouns can be best predicted based on the words' availability, imagery, and emotionality. Additionally, contradicting the findings from traditional literature, Rubin and Friendly questioned the role of meaningfulness, frequency, and pronunciability.

Most recently, Nairne and colleagues (2013, 2017) have shown that human memory is tuned to process animacy-related information. Nairne interprets this phenomenon as reflecting the adaptive value of remembering information that promotes survival and reproductive success (Nairne et al., 2013). After including animacy as a predictor variable and reanalyzing the Rubin and Friendly (1986) data, Nairne et al. (2013) identified animacy as one of the most important predictors of recall. Based on this relationship between animacy and memory, we have added animacy to our predictor variables to investigate its influence on word and list-level recall performance.

Whereas Rubin and Friendly (1986) and Nairne et al. (2013) sought to predict the memorial difficulty of individual words, here we sought to model both the difficulty of individual words and of entire word lists. We also sought to reexamine the role of word meaningfulness when defined using modern computational linguistic metrics of word similarity (e.g., word2vec) and taking into account the semantic and temporal relationships among items. To examine the role of meaningfulness on both word and list memory we created a new meaningfulness measure: average semantic relatedness between a target word and all other items in its list weighted by the serial position (temporal) lag distance between each item pair. By weighting semantic similarity, we build on previous work showing that semantic and temporal similarity interact (positively) to predict successful recall (e.g., Howard & Kahana, 2002b). Finally, by conducting our analyses of word and list difficulty in a unique multisession experiment in which each of 98 subjects saw the same pool of 576 words (24 words $\times$ 24 lists per session) in each of the 23 sessions, we were able to evaluate our multivariate model of at the level of individual subjects. This latter feature of our approach allows us to establish the stability of our word and list-recall models across individual subjects.

Method

The data reported here comes from Experiment 4 of the IRB-approved Penn Electrophysiology of Encoding and Retrieval Study (PEERS). The primary goal of PEERS is to assemble a large public database on the electrophysiological correlates of memory encoding and retrieval. Data from Experiments 1–3 have been reported in several prior publications (e.g., Healey & Kahana, 2014, 2016; Lohnas & Kahana, 2013, 2014; Lohnas, Polyn, & Kahana, 2015), and a subset of data from Experiment 4 has been reported in Kahana, Aggarwal, and Phan (2018). Subjects consisted of 98 young adults (ages 18–35) who were recruited from among the students and staff at the University of Pennsylvania and neighboring institutions. All subjects were right-handed and native English speakers.

[3] *Word length* refers how many letters the word consists.

[4] Two classical emotionality measures exist: *Valence* refers to the degree of pleasantness of the word, and *arousal* refers to the degree to which a physiological reaction is elicited by the word.

Subjects performed a delayed free recall experiment consisting of 23 experimental sessions.[5] Each session consisted of 24 trials, with each trial containing a list of 24 words, presented one at a time on a computer screen. A random half of the lists (excluding the first list) were preceded by a 24-s, distractor-filled delay, and all lists were followed by a 24-s distractor period. A free recall test followed the postlist distractor on each list.

Each trial began with a 10-s countdown, which was displayed onscreen. Subjects were permitted to pause and resume this countdown at any time by pressing a key. After the countdown was complete, a fixation cross appeared on the screen for 1,500 ms. For trials without a prelist distractor, the fixation cross was immediately followed by the presentation of the first word. For trials with a prelist distractor, this fixation cross was instead followed by a 24-s distractor period. After the distractor period, the screen went blank for a jittered 800–1,200 ms (uniformly distributed), after which the first word was presented. Each word was presented on the screen in white text on a black background for 1,600 ms, and was followed by a jittered interstimulus interval of 800–1,200 ms (uniformly distributed). Following the interstimulus interval after the final word in each list, subjects performed a distractor task for 24 s. This postlist distractor task was followed by a 1,200- to 1,400-ms (uniformly distributed) delay, after which a tone sounded and a row of asterisks appeared onscreen for 500 ms, indicating the start of the free recall period. subjects were given 75 s to recall aloud as many of the words from the current list as possible, in any order. A fixation cross was displayed onscreen for the duration of the recall period followed by a blank screen was displayed for 1,000 ms, after which the 10-s countdown for the next list began. Subjects were also given a short break (approximately 5 min) after every eight lists in a session.

Both the prelist and postlist distractor tasks consisted of answering math problems of the form $A + B + C = ?$, where A, B, and C were positive, single-digit integers. Math problems were displayed onscreen one at a time in white text on a black background, and subjects were instructed to type the answer to each equation as quickly and accurately as possible. New problems continued to appear until the full 24 s had elapsed, at which point the final problem was immediately removed from the screen. Subjects were given a monetary bonus based on the speed and accuracy of their responses.

Each session required $24 \times 24 = 576$ words. The word pool for this experiment thus consisted of a 576-words. Each of these 576 words appeared exactly once in each experimental session (24 lists $\times$ 24 items), so each subject saw the same set of words 23 times. Within each session, words were randomly assigned to lists following certain constraints on semantic similarity, as described in our earlier PEERS papers. With this experimental design of our multisession study, we attempted to wash out all the idiosyncratic reasons why certain words might be poorly or well remembered such as whether a word falls into a favorable or unfavorable list position. It should be noted that words in the pool did not have extreme values along dimensions of word frequency, concreteness, and emotional valence as these words are usually omitted from experiments with controlled word pools and we wanted to create a word pool similar to those used in other memory studies.

All previously published raw behavioral data from the PEERS studies, as well as the new data reported in the present article, may be freely obtained from the authors' website, http://memory.psych.upenn.edu.

Variables

We created parallel statistical models to predict word-level and list-level recall performance. Based on previous work, we identified six properties of words that would be expected to predict recall performance: concreteness, contextual diversity, word length, emotional valence, arousal, and animacy. In addition, we included a variant of Noble's (1952) classic "meaningfulness" index using word2vec (Mikolov, Chen, Corrado, & Dean, 2013), which we elaborate below. Whereas one value for each word is used in the word-recall model, an average value computed using each of 24 list words' properties is used in the list-recall model. For the word-level recall model we included a session number variable and for the list-recall model we included variables of trial number and session number to account for proactive interference and practice effects. (Please see the Appendix for words in the word pool along with each word's average recall probability and associated predictor variables.)

Concreteness

Concreteness measures of the 568 words in the word pool are obtained from Brysbaert, Warriner, and Kuperman (2014). These authors collected concreteness ratings through an Internet crowdsourcing website by asking subjects to indicate how concrete the meaning of each word is (i.e., can be experienced directly from one of the five senses) using a 5-point rating scale going from abstract to concrete. For example, whereas the word *apple* is a concrete word that has a concreteness value of 5, the word *patient* is an abstract word that has a concreteness value of 2.5.

Eight words in our word-pool do not have a concreteness value reported in Brysbaert et al. (2014), thus we ran our own norming study ($N = 38$) using Amazon Mechanical Turk to collect the missing concreteness ratings. All instructions and methods used in this norming study were identical to those used by Brysbaert et al. (2014).

Contextual Diversity

Contextual diversity measures are obtained through the SUBTLX-US database (Brysbaert & New, 2009) where contextual diversity (SUBTL-CD) is defined as the percent of the films the word appears. In our word pool, for example, whereas *world* is a word with high contextual diversity, *scallop* is a word with low contextual diversity. All words in our word pool has a corresponding contextual diversity measure in the database.

Word Length

Word length is calculated by counting the number of letters in each word. In our word pool, *playground* is the longest word with 10 letters, whereas *ox* is the shortest word with two letters.

[5] Subjects participated in a 24th experimental session during which they studied lists composed of both old words (drawn from the pool of 576) and new words matched on the word attributes. Because the focus of this article is on recallability of words under constant conditions, our analyses do not include data from this last session.

Valence

Emotional valence is the degree of pleasantness of the word. Our word pool's emotional valence ratings are obtained from Warriner, Kuperman, and Brysbaert (2013). Authors collected ratings through an Internet crowd-sourcing website by asking subjects to give each word a numerical value from 1 to 9, with 1 being unpleasant and 9 being pleasant. For example, whereas *lover* is the word with the highest emotional valence (8.05), *virus* is the word with the lowest emotional valence (1.71). In our analyses, the bivariate scale of valence scores (1 to 9) are converted to a bipolar scale (0 to 4), forming a pure emotional intensity scale that is not conditional on positivity/negativity. This is supported by previous work in which memory benefit for emotional valence was not found to be significantly different for positive and negative emotional lists (Palmer & Dodson, 2009).

Arousal

Arousal is the degree to which a physiological reaction is elicited by the word. Arousal ratings are obtained from the study mentioned above (Warriner et al., 2013). Each word has a numerical arousal value from 1 to 9. For example, whereas *lover* is the most arousing word (7.45), *pail* is the least arousing word (2.24).

Meaningfulness

We define meaningfulness as the average semantic relatedness between a target word and all other items in its list weighted by the serial position (temporal) lag distance between each item pair. This measure is an extended version of the Noble (1952). In earlier studies (e.g., Paivio et al., 1968; Toglia & Battig, 1978), researchers measured meaningfulness as the number of free associations produced to a given item within a fixed interval (e.g., 60 s). In this study, to compute meaningfulness measure, we used a pretrained word embedding model, word2vec, a modern corpus-based computational method (Mikolov et al., 2013) that is applied to Google News articles with more than 100 billion words. In the word2vec space, there are three million words and phrases each having 300 dimensional vectors. Word2vec has been validated in prior work on human memory and cognition (e.g., Bhatia, 2016; Bian, Gao, & Liu, 2014) and is a suitable alternative which is trained on extensive amounts of natural language data from the recent Google News articles. All of our word pool items were represented in word2vec.

More specifically, to compute meaningfulness for a given word, we take the cosine-theta semantic similarity vector distance between a target word and 23 other words in its list, and we also weigh each semantic similarity measure by the absolute temporal lag distance between the words' serial positions. Then, we average these weighted similarities to get a single meaningfulness measure for each presentation of each word.

Animacy

Animacy is defined as whether a word is living (animate) or nonliving (inanimate). Two independent raters evaluated each of the study words for their animacy. Raters separately gave a 0 for inanimate and a 1 for animate words. Interrater reliability between the raters was Cohen's kappa $> .90$ for the animacy ratings. According to the evaluations, 165 of the 576 words in the study pool were animate.

Two additional variables that were considered in our list-recall model were trial number and session number. These variables were added as we think they may influence the list-recall performance:

Trial Number

Trial number within a session ranging from 1 to 24.

Session Number

Session number that the data comes from for each subject, ranging from 0 to 22.

To address interpretive problems arising from multicollinearity, we regressed each variable in our model on all other variables with which it had a moderate-to-high correlation. We then replaced the original variable with the residuals of this regression model (see Table 1). As an example, consider variable Animacy. It is moderately correlated with the variables Concreteness, Valence, and Arousal. We regressed Animacy on Concreteness, Valence, and Arousal and used the residuals of this regression model as the residualized Animacy variable. Following this residualization process, we confirmed that none of the correlation coefficients among our variables exceeded $r > .15$ (see Tables 2 and 3).

Table 2 reports correlations among the raw variables as well as the residualized variables in the word-level recall model. Table 3 reports the same correlation matrices from the list-level model. We use these reidentified variables for our linear mixed effects models. As the tables show, our residualization method removed all of the strong correlations among the remaining variables.

Results

A key feature of the present experiment is that subjects studied and attempted to recall the same set of 576 words in each of the 23 daily sessions. As such, a given word appeared in 23 randomly determined lists and serial positions for each of the 98 subjects, resulting in a total of 2,254 occurrences of that word across the sample. This allowed us to quantify the recallability of each word with a high degree of precision. Figure 1 shows each word's average recall probability sorted from lowest to highest. Clearly subjects found some words to be very difficult to recall while other words came to mind easily: the

Table 1
Newly Defined Variables

Variable redefined	Variable(s) regressed
Word length	Contextual diversity
Valence	Contextual diversity
Arousal	Concreteness, valence
Animacy	Concreteness, valence, arousal

Note. To address interpretive problems arising from multicollinearity, we regressed each variable in our model on all other variables with which it had a moderate-to-high correlation. We then replaced the original variable with the residuals of this regression model.

Table 2
Correlation Matrix Predictors of Interest for Word-Level Recall Models

Precictor	Concreteness	Contextual diversity	Length	Valence	Arousal	M List	Animacy
Word-level recall predictors before residualization							
Contextual diversity	−0.10	*	*	*	*	*	*
Word length	−0.02	−0.16	*	*	*	*	*
Valence	−0.13	0.18	0.00	*	*	*	*
Arousal	−0.14	0.07	0.04	0.37	*	*	*
Meaningfulness	0.00	0.00	0.00	0.00	0.00	*	*
Animacy	−0.21	0.06	0.01	0.21	0.18	0.00	*
Recall probability	0.01***	0.02***	0.00	0.02***	0.02***	0.00*	0.04***
Word-level recall predictors after residualization							
Contextual diversity	−0.10	*	*	*	*	*	*
Word length	−0.04	0.00	*	*	*	*	*
Valence	−0.11	0.00	0.03	*	*	*	*
Arousal	0.00	0.00	0.04	0.00	*	*	*
Meaningfulness	0.00	0.00	0.00	0.00	0.00	*	*
Animacy	−0.05	0.02	0.00	0.05	0.02	0.00	*

Note. The predictor variables appear to be weakly correlated or not correlated at all after residualization. Univariate correlations between each variable (before residualization) and recall probability are reported.
* $p < .05$. *** $p < .001$.

word *survey* exhibited the lowest recall probability (35%), whereas the word *boyfriend* exhibited the highest recall probability (68%). We first identify predictors of individual word recall and then turn our attention to predictors of list-level recall.

We developed two parallel regression models to predict word and list-recall performance. We included seven predictor variables hypothesized to account for variability in recall performance: word concreteness, contextual diversity, word length, emotional valence, arousal, and animacy, and finally meaningfulness. We also included session number (and in the case of the list-recall model, trial number) as additional predictor variables. Technical definitions of each of these variables appear in Method section along with references to their use in the prior literature. One variable, meaningfulness, appears for the first time in the present report. To compute meaningfulness, we averaged semantic relatedness between a target word and all other items in its list weighted by the serial position (temporal) lag distance between each item pair.

We first fit the word-level recall model to data from each of the 98 subjects. Figure 2A illustrates results for each of the predictor variables. Each dot indicates the β value obtained by fitting the logistic regression model to data from one subject. The bars represent the overall population effects calculated by taking the mean of subject-specific β values for variables in the model. Filled circles indicate those β values that exceed our false discovery rate-corrected significance threshold ($p < .05$). As may be seen from the distributions of significant coefficients, some variables exhibited consistent positive or negative effects across subjects (e.g., meaningfulness, animacy), whereas the word-length variable exhibited mixed effects, with some subjects

Table 3
Correlation Matrix Predictors of Interest for List-Recall Model

Predictor	Concreteness	Contextual diversity	Length	Valence	Arousal	M List	Animacy
List-level recall predictors before residualization							
Contextual diversity	−0.15	*	*	*	*	*	*
Word length	−0.03	−0.15	*	*	*	*	*
Valence	−0.15	0.18	0.00	*	*	*	*
Arousal	−0.16	0.08	0.04	0.39	*	*	*
Meaningfulness	0.09	−0.03	0.02	0.04	0.00	*	*
Animacy	−0.22	0.07	0.01	0.23	0.21	0.04	*
Recall probability	0.00	0.03***	−0.01	0.02***	0.01*	0.01*	0.02***
List-level recall predictors after residualization							
Contextual diversity	−0.15	*	*	*	*	*	*
Word length	−0.05	0.00	*	*	*	*	*
Valence	−0.13	0.00	0.03	*	*	*	*
Arousal	0.00	−0.01	0.04	0.00	*	*	*
Meaningfulness	0.09	−0.03	0.01	0.04	0.00	*	*
Animacy	0.00	0.01	0.00	0.00	0.00	0.05	*

Note. The predictor variables appear to be weakly correlated or not correlated at all after residualization. Univariate correlations between each variable (before residualization) and recall probability are reported.
* $p < .05$. *** $p < .001$.

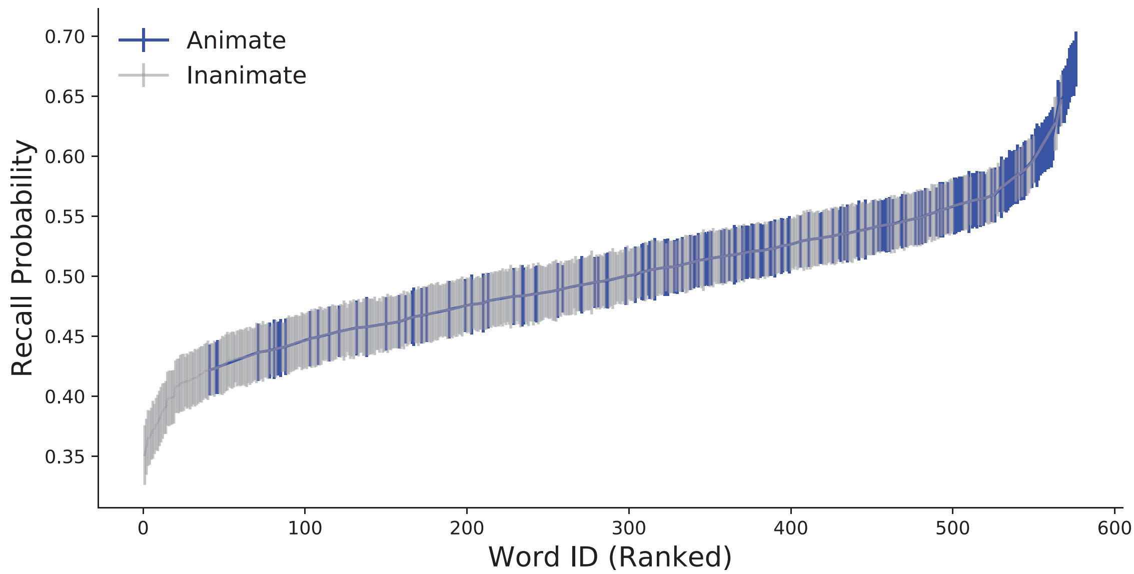

Figure 1. Variability in free recall of words and average recall probability of each word. Average recall probabilities are sorted from lowest to highest and plotted along with the standard error confidence band around the average values. Blue (dark gray) and gray markers indicate animate and inanimate words, respectively. See the online article for the color version of this figure.

having significant positive coefficients and others having significant negative coefficients.

To model these data both across subjects and trials/sessions, we used a linear mixed effects approach, as shown in Table 4 (Bates, Mächler, Bolker, & Walker, 2015). To reduce the potential effects of multicollinearity, we residualized each variable upon any other variable with which it had a moderate-to-high correlation (see Method) and used these redefined variables in our subsequent analyses. We accounted for the individual differences in the predictors using random effects for subjects. The subject-level random effects of a predictor were treated as deviations from the fixed effect (population effect). We logit-transformed the response variable (probability of recall) to remove the range restrictions of probability outcomes.

Results of our word-level recall model revealed significant positive effects of animacy, contextual diversity, emotional valence (either positive or negative), arousal, concreteness, and semantic structure on recall of individual words (see Table 4). Animacy, contextual diversity, valence, and arousal distinguished themselves as the most predictive characteristics for item-level recall. In each case, the direction of the effect aligns with the terminology used; subjects more easily remembered words that were animate, contextually diverse, emotionally arousing, and/or emotionally valenced. Although meaningfulness also positively predicted word-level recall, this effect was substantially smaller than any of the other significant variables.

Because we often evaluate memory using measures of list-level performance, both in pure and applied settings, we also sought to model variability in memory performance at the list level. Such a model would allow us to construct lists that subjects would find easier or harder to remember. Because recalling a list is a dynamic, path-dependent, process, averaging predicted recall for individual words would not accurately represent recall performance at the list level. Further, comparisons between list-level and word-level models may uncover organizational principles that uniquely support list recall and that would be missed using an item-level analysis.

For the list-level recall model we found strong positive effects of average contextual diversity, valence, animacy, concreteness, and meaningfulness (see Table 4). Whereas several of these findings align with our word-level recall analysis, the predictive power of these variables differed somewhat across models. Specifically, whereas meaningfulness barely predicted recall it appeared to more strongly predict list level recall. This aligned with our hypothesis that semantically coherent lists would be more easily recalled than less semantically coherent lists.

Discussion

We asked how the properties of words influence their memorability. To answer this question we developed linear fixed effects models to account for variability in both item and list-level recall. Our model included five standard predictor variables (concreteness, contextual diversity, word length, emotional valence, and arousal) and the new "meaningfulness" measure designed capture a word's semantic relatedness to other words in the target list weighted by temporal relationships and the recently identified animacy measure. In the item-level model, we assessed how these predictor variables accounted for variability in recall probability across the 576 items seen by each subject in our study. In the list-level model, we evaluated how the same predictor variables, averaged across all items in a given list, accounted for that list's average recall probability. We applied this model to all 54,096 lists seen by the 98 subjects in our study and included covariates to account for practice and interference effects.

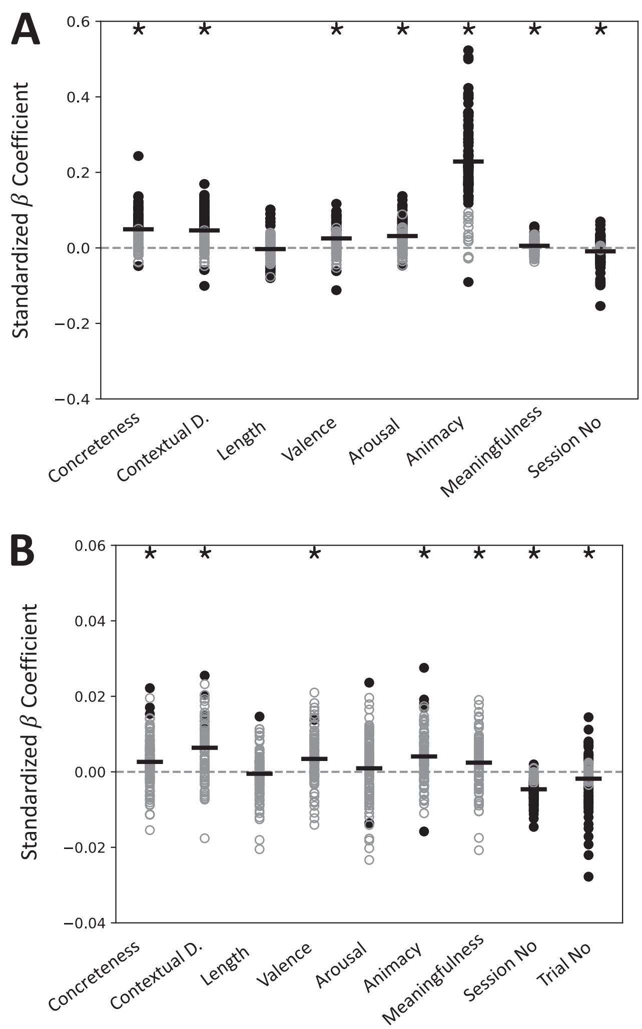

Figure 2. Distributions of β values for each predictor variable in the word- and list-recall models when fit to each subject separately. (A) Word-recall model. (B) List-recall model. Each circle on the plots denotes the normalized logistic regression coefficient for a single subject, with filled circles indicating coefficients that met a false discovery rate-correct $p < .05$ significance criterion. Single asterisks indicate whether the beta coefficients were significantly different than zero.

Results from the item-level model replicated previous findings regarding the positive influence of concreteness, contextual diversity, emotional valence, arousal, and animacy on recall (Dolcos et al., 2004; Dukes & Bastian, 1966; Kensinger & Corkin, 2003; LaBar & Cabeza, 2006; Nairne et al., 2013; Parmentier et al., 2017). The present study extended these previous results by showing that four of these variables, namely concreteness, contextual diversity, valence, and animacy, also strongly predict variability in list-level recall. One of the primary goals of the present study was to determine the influence of semantic structure on the recallability of items and lists. Using the word2vec model of semantic space (Mikolov et al., 2013), we were able to assess an item's semantic relatedness to each of the other list items. In addition, we captured the interaction between temporal and semantic proximity (Howard & Kahana, 2002a) by weighting semantic relatedness as a decreasing power function of the temporal distance between item pairs.

Our analyses revealed that subjects tend to recall words that have high semantic associations to their list neighbors and that subjects tend to recall a higher proportion of list items when the list exhibits temporally coherent semantic organization (i.e., semantically related items occurring in nearby list positions). It is easy to imagine how such associations would facilitate recall through semantic elaboration during encoding and through semantic cueing recall. If the words *flower* and *rabbit* appear in close proximity, then studying *rabbit* will remind subjects of *flower*, leading the two words to become more strongly associated. These associations, in turn, will facilitate cue dependent retrieval. Lists will benefit to the extent that they possess words with coherent semantic structure. Models of recall that assume a similarity-driven cue-dependent retrieval process will predict both effects. Models that further posit a role for temporal organization, either through temporal context, interitem associations, or chunking mechanisms (e.g., Polyn, Norman, & Kahana, 2009), should predict an additional benefit when similar items appear in close proximity during study.

In a contemporaneous report, Lau, Goh, and Yap (2018) also examined predictors of item-level recall performance. As in the present study, they found recall probability to be significantly positively correlated with word frequency, arousal, and a measure of semantic density that is close to our measure of meaningfulness. Both studies found consistent effects despite several major methodological differences, such as the use of naive versus practiced subjects, the immediate versus delayed nature of recall, and the statistics of item characteristics in the pools being used. Relatedly, Cox, Hemmer, Aue, and Criss (2018)

Table 4
Fixed Effects of Variables Predicting Probability of Word-Level and List-Level Recall in Multivariate Analyses

Predictor	*M* β	*SE* β
Predictors of word-level recall model		
Concreteness	0.03***	0.004
Contextual diversity	0.06***	0.005
Word length	−0.003	0.003
Valence	0.05***	0.004
Arousal	0.04***	0.004
Animacy	0.09***	0.006
Meaningfulness	0.005*	0.005
Session number	−0.009***	0.0003
Predictors of list-level recall model		
Concreteness	0.002*	0.0008
Contextual diversity	0.008***	0.001
Word length	−0.0004	0.0008
Valence	0.005***	0.0008
Arousal	0.001	0.0009
Animacy	0.004***	0.0008
Meaningfulness	0.002**	0.0008
Session number	−0.002***	0.0001
Trial number	−0.005***	0.0001

Note. Word length, valence, arousal, and animacy variables are residualized variables.
* $p < .05$. ** $p < .01$. *** $p < .001$.

highlighted the importance of semantic features and multiple word properties in a large-scale study using hierarchical Bayesian techniques. They examined how individual performance was correlated between a variety of memory tasks and how item-level information supports the memory performance.

Whereas both the present work and earlier studies considered how word properties predict recall of individual items, here we also considered how these predictor variables could account for overall levels of list recall. Given that we usually assess subject's performance at the list level, understanding the predictors of list-level recall can have important practical value in both designing experiments and optimizing neuropsychological measures of memory function.

Our analysis of list-level recall revealed significant positive contributions of concreteness, contextual diversity, valence, and animacy. These results parallel those of the word-recall model. In addition, we also found a significant positive contribution of meaningfulness, which is an aggregate measure of interitem similarities weighted by temporal lag. Because each recalled item serves as a cue for subsequent recalls, lists with semantically related items (i.e., high values of meaningfulness) yield higher levels of recall. These list level results align with previous work by Nelson and colleagues (Nelson et al., 2003) showing how semantic associative networks can be a source of positive transfer in cued-recall tasks. Here we extend these findings to the setting of delayed free recall.

The present study demonstrates significant positive effects of word animacy on both item-level and list-level recall (see Table 4). Previous work has shown positive effects of animacy on word-level recall (Nairne et al., 2013), but the present study provides additional information by showing that these effects appear robust at the list level even when controlling for the semantic similarities among list items (our measure of meaningfulness). Thus, whereas one might have suspected that subjects would organize list items according to their animacy, and that such organization would support recall, the benefits of animacy for list recall persist even when the list-level model includes an index of semantic organization. This result aligns with a recent report by VanArsdall, Nairne, Pandeirada, and Cogdill (2017), who found little support for category clustering using an embedded list technique that includes animate words, inanimate words, and filler words. Importantly, however, their studies documented strong and persistent animacy advantages despite the lack of category clustering.

We would like to address two potential limitations regarding our study's predictor variable choice. First, as the average valence scores reported in the Appendix illustrate, our findings primarily relate to words with positive valence as our word pool was specifically designed to avoid words with very strong negative connotations (e.g., *death*, *funeral*). This decision was made because we included older adults in a parallel study using the same materials (not reported in this article). Second, although stronger measures for semantic similarity of list items such as taxonomic relatedness and situational relatedness exist, we purposefully created our meaningfulness measure to capture both semantic and temporal associations among items in a straightforward manner.

An important future direction will be to relate our aggregate measure of meaningfulness to the predictions of memory models that simulate the dynamics of recall as a function of the semantic and temporal structure of lists (Farrell, 2012; Healey & Kahana, 2014; Lohnas et al., 2015) These models should make specific, testable predictions about how aggregate measures such as meaningfulness arise from semantic structure both within and across lists. At a coarse level, our findings appear consistent with models in which increased within-list semantic relatedness enhances recall by facilitating cue-dependent retrieval. Lists whose words have strong semantic relatedness to neighboring items should engender benefits as cue-dependent retrieval favors items that share temporal and semantic features with the just-recalled item.

Conclusion

For more than a century, students of memory have turned to common words as the memoranda of choice in their experiments. As such, understanding how word properties relate to their memorability has attracted considerable attention (Rubin & Friendly, 1986; Schlosberg & Woodworth, 1954). In this article we use parallel models to systematically examine the influence of different word properties in item- and list-level recall. In addition, we also exploit powerful new methods from natural language processing for measuring meaningfulness and looking at its influence both at the level of individual words and entire lists. In extending the analysis of psycholinguistic and semantic factors in predicting recall from the item-level to the list-level, our models can help to optimize experimental design to better control variability in list-level recall performance, both for more accurate assessment of individual differences and experimental influences on recall performance.

References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17,* 814–823. http://dx.doi.org/10.1111/j.1467-9280.2006.01787.x

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575–589. http://dx.doi.org/10.1016/S0022-5371(75)80045-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Bhatia, S. (2016). *Vector space semantic models predict subjective probability judgments for real-world events*. Retrieved from https://www.sas.upenn.edu/~bhatiasu/Bhatia%202016%20CogSci%20PP.pdf

Bian, J., Gao, B., & Liu, T. Y. (2014). Knowledge-powered deep learning for word embedding. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8724 LNAI,* 132–148. http://dx.doi.org/10.1007/978-3-662-44848-9_9

Bruza, P., Kitto, K., Nelson, D., & McEvoy, C. (2009). Extracting spooky-activation-at-a-distance from considerations of entanglement. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5494,* 71–83. http://dx.doi.org/10.1007/978-3-642-00834-4_8

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English.

Behavior Research Methods, 41, 977–990. http://dx.doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46,* 904–911. http://dx.doi.org/10.3758/s13428-013-0403-5

Christian, J., Bickley, W., Tarka, M., & Clayton, K. (1978). Measures of free recall of 900 English nouns: Correlations with imagery, concreteness, meaningfulness, and frequency. *Memory & Cognition, 6,* 379–390. http://dx.doi.org/10.3758/BF03197470

Cox, G. E., Hemmer, P., Aue, W. R., & Criss, A. H. (2018). Information and processes underlying semantic and episodic memory across tasks, items, and individuals. *Journal of Experimental Psychology: General, 147,* 545–590.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language, 64,* 119–132. http://dx.doi.org/10.1016/j.jml.2010.10.001

Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. *Neuron, 42,* 855–863. http://dx.doi.org/10.1016/S0896-6273(04)00289-2

Dukes, W. F., & Bastian, J. (1966). Recall of abstract and concrete words equated for meaningfulness. *Journal of Verbal Learning and Verbal Behavior, 5,* 455–458. http://dx.doi.org/10.1016/S0022-5371(66)80060-9

Epstein, W., Rock, I., & Zuckerman, C. B. (1960). Meaning and familiarity in associative learning. *Psychological Monographs: General and Applied, 74,* 22.

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review, 119,* 223–271. http://dx.doi.org/10.1037/a0027371

Glaze, J. A. (1928). The association value of non-sense syllables. *Pedagogical Seminary and Journal of Genetic Psychology, 35,* 255–269. http://dx.doi.org/10.1080/08856559.1928.10532156

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology, 61,* 23–29. http://dx.doi.org/10.1037/h0040561

Hall, J. F. (1954). Learning as a function of word-frequency. *The American Journal of Psychology, 67,* 138–140. http://dx.doi.org/10.2307/1418830

Hamilton, M., & Rajaram, S. (2001). The concreteness effect in implicit and explicit memory tests. *Journal of Memory and Language, 44,* 96–117. http://dx.doi.org/10.1006/jmla.2000.2749

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

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review, 123,* 23–69. http://dx.doi.org/10.1037/rev0000015

Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299.

Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language, 46,* 85–98. http://dx.doi.org/10.1006/jmla.2001.2798

Hull, C. (1933). The meaningfulness of 320 selected nonsense syllables. *The American Journal of Psychology, 45,* 730–734.

Hulme, C., Suprenant, A. M., Bireta, T. J., Stuart, G., & Neath, I. (2004). Abolishing the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 98–106. http://dx.doi.org/10.1037/0278-7393.30.1.98

Hunt, C., Trammel, J., & Krumrei-Mancuso, E. J. (2015). The interaction of semantic relatedness and emotion: Why emotion may not lead to better memory. *Psi Chi Journal of Psychological Research, 20,* 45–51.

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

Kahana, M. J., Aggarwal, E. V., & Phan, T. D. (2018). The variability puzzle in human memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 1857–1863.

Katkov, M., Romani, S., & Tsodyks, M. (2014). Word length effect in free recall of randomly assembled word lists. *Frontiers in Computational Neuroscience, 8,* 129. http://dx.doi.org/10.3389/fncom.2014.00129

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31,* 1169–1180. http://dx.doi.org/10.3758/BF03195800

Klein, K. A., Addis, K. M., & Kahana, M. J. (2005). A comparative analysis of serial and free recall. *Memory & Cognition, 33,* 833–839. http://dx.doi.org/10.3758/BF03193078

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience, 7,* 54–64. http://dx.doi.org/10.1038/nrn1825

Lau, M. C., Goh, W. D., & Yap, M. J. (2018). An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach. *Quarterly Journal of Experimental Psychology, 71,* 2207–2222. http://dx.doi.org/10.1177/1747021817739834

Lohnas, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1943–1946. http://dx.doi.org/10.1037/a0033669

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

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Intralist and interlist effects in free recall. *Psychological Review, 122,* 337–363. http://dx.doi.org/10.1037/a0039036

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality*. http://dx.doi.org/10.1162/jmlr.2003.3.4-5.951

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science, 26,* 22–27.

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science*. Advance online publication. http://dx.doi.org/10.1177/0956797613480803

Nelson, D. L., & McEvoy, C. L. (1979). Encoding context and set size. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 292–314. http://dx.doi.org/10.1037/0278-7393.5.3.292

Nelson, D. L., McEvoy, C. L., & Pointer, L. (2003). Spreading activation or spooky action at a distance? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29,* 42–52. http://dx.doi.org/10.1037//0278-7393.29.1.42

Nelson, D. L., Schreiber, T. A., & McEvoy, C. L. (1992). Processing implicit and explicit representations. *Psychological Review, 99,* 322–348. http://dx.doi.org/10.1037/0033-295X.99.2.322

Noble, C. E. (1952). An analysis of meaning. *Psychological Review, 59,* 421–430. http://dx.doi.org/10.1037/h0054087

Paivio, A. (1967). Paired-associate learning and free recall of nouns as a function of concreteness, specificity, imagery, and meaningfulness. *Psychological Reports, 20,* 239–245.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76,* 1–25. http://dx.doi.org/10.1037/h0025327

Palmer, J. E., & Dodson, C. S. (2009). Investigating the mechanisms fuelling reduced false recall of emotional material. *Cognition and Emotion, 23,* 238–259.

Parmentier, F. B., Comesaña, M., & Soares, A. P. (2017). Disentangling the effects of word frequency and contextual diversity on serial recall performance. *Quarterly Journal of Experimental Psychology, 70,* 1–17. http://dx.doi.org/10.1080/17470218.2015.1105268

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116,* 129.

Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory & Cognition, 14,* 79–94. http://dx.doi.org/10.3758/BF03209231

Schlosberg, H., & Woodworth, R. S. (1954). *Experimental psychology*. New York, NY: Holt.

Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29,* 760–766. http://dx.doi.org/10.1037/0278-7393.29.5.760

Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition, 32,* 742–751. http://dx.doi.org/10.3758/BF03195864

Tehan, G., & Tolan, G. A. (2007). Word length effects in long-term memory. *Journal of Memory and Language, 56,* 35–48. http://dx.doi.org/10.1016/j.jml.2006.08.015

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Oxford, UK: Erlbaum.

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N., & Cogdill, M. (2017). A categorical recall strategy does not explain animacy effects in episodic memory. *The Quarterly Journal of Experimental Psychology, 70,* 761–771.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1256–1271. http://dx.doi.org/10.1037/0278-7393.25.5.1256

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45,* 1191–1207. http://dx.doi.org/10.3758/s13428-012-0314-x

Appendix

Model Equations and Experimental Stimuli

Word-Recall Model

$$\text{Rec}_{i,e,j} \sim \beta_{0,j} + \beta_{1,j}C + \beta_{2,j}CD + \beta_{3,j}L + \beta_{4,j}V + \beta_{5,j}Ar + \beta_{6,j}M + \beta_{7,j}An + \beta_{8,j}S + \epsilon_{i,e,j} \quad (1)$$

where, $\beta_{k,j} \sim N(\beta_k, \sigma_k^2)$, $k = 0, \ldots, 8$, $j = 1, \ldots, N$, Rec is 0 or 1 (whether a word is recalled or not) and is linked to the predictors with a logistic function.

List-Recall Model

$$\text{logit}(\text{PRec}_{i,e,j}) \sim \beta_{0,j} + \beta_{1,j}C + \beta_{2,j}CD + \beta_{3,j}L + \beta_{4,j}V + \beta_{5,j}Ar + \beta_{6,j}M + \beta_{7,j}An + \beta_{8,j}S + \beta_{9,j}T + \epsilon_{i,e,j} \quad (2)$$

where, $\beta_{k,j} \sim N(\beta_k, \sigma_k^2)$, $k = 0, \ldots, 9$, $j = 1, \ldots, N$.

Subscript i denotes subject, e denotes session, and j denotes item. C denotes Concreteness, CD denotes Contextual Diversity, L denotes Length, V denotes Valence, Ar denotes Arousal, M denotes Meaningfulness, An denotes Animacy, S denotes Session No, T denotes Trial No.

Word	Recall prob. (R)	Concreteness (C)	Contextual diversity (D)	Valence (V)	Arousal (Ar)	Animacy (An)
ACTOR	0.55	4.57	6.84	6.15	4.35	1
ACTRESS	0.61	4.54	4.52	5.42	5.43	1
AGENT	0.48	3.61	16.68	5.23	3.43	1
AIRPLANE	0.54	4.96	4.22	5.25	5.62	0
AIRPORT	0.56	4.87	11.71	6	5.5	0
ANKLE	0.57	4.81	3.41	5.4	3.11	0
ANTLER	0.48	4.86	0.15	3.21	5.32	0
APPLE	0.57	5	8.21	6.62	3.52	0
APRON	0.52	4.87	1.28	5.8	2.9	0
ARM	0.59	4.96	23.19	5.44	3.44	0
ARMY	0.57	4.7	17.95	4.65	4.49	0
ASIA	0.65	3.89	2.19	3.52	3.48	0
ATLAS	0.48	4.79	0.42	5.95	2.5	0
ATOM	0.49	3.34	1.10	5.74	4.29	0
AUTHOR	0.48	4.26	2.99	6.33	2.73	1
AWARD	0.48	4.14	4.54	7.86	5.85	0
BABY	0.61	5	60.66	6.67	4.97	1
BACKBONE	0.45	4.19	0.93	5.16	4.05	0
BACON	0.56	4.9	4.76	7.52	4.16	0
BADGE	0.45	4.93	5.52	5.24	4.4	0
BALLOON	0.46	4.92	3.11	6.84	3.9	0
BANJO	0.46	4.9	0.60	6.35	3.53	0
BANK	0.50	4.78	18.96	6	4.19	0
BANKER	0.53	4.43	1.91	4.89	3.38	1
BANQUET	0.53	4	1.93	6.11	4.57	0
BARLEY	0.45	4.59	0.39	4.95	3.62	0
BARREL	0.48	4.86	4.45	4.92	3.43	0
BASEMENT	0.55	4.89	8.09	4.81	3.33	0
BATHTUB	0.48	4.92	2.67	6.26	3.63	0
BEAKER	0.47	4.72	0.19	5.5	3.79	0
BEAST	0.50	4.63	6.58	4.42	5.83	1
BEAVER	0.54	4.68	1.51	5	4.05	1
BEEF	0.57	4.74	7.67	6.11	4.37	0
BELLY	0.53	4.8	6.94	4.37	3.75	0
BIKE	0.46	5	6.84	6.1	3.62	0
BINDER	0.42	4.89	0.26	5.16	3.17	0
BISON	0.59	4.68	0.12	4.53	3.77	1
BLACKBOARD	0.51	4.72	0.54	4.95	4	0
BLADE	0.53	4.93	4.67	3.9	4.52	0

(*Appendix continues*)

Appendix (*continued*)

Word	Recall prob. (R)	Concreteness (C)	Contextual diversity (D)	Valence (V)	Arousal (Ar)	Animacy (An)
BLENDER	0.45	5	0.70	5.16	4.05	0
BLOCKADE	0.47	4.25	0.44	3.45	3.4	0
BLOUSE	0.54	4.96	2.23	5.73	3.24	0
BLUEPRINT	0.44	4.77	0.57	4.95	3.71	0
BOARD	0.47	4.57	20.05	5.33	3.52	0
BODY	0.50	4.79	47.22	5.95	4.62	0
BOUQUET	0.48	4.74	1.39	6.67	3.33	0
BOX	0.54	4.9	28.05	5.33	2.67	0
BOYFRIEND	0.68	4.59	21.80	7.06	4.9	1
BRACES	0.43	5	1.37	6.48	3.96	0
BRAKE	0.38	4.44	2.34	4.9	3.82	0
BRANCH	0.43	4.9	4.46	5.15	2.67	0
BRANDY	0.51	4.81	3.42	5.67	3.86	0
BREAST	0.57	4.89	3.77	6.64	5.39	0
BRICK	0.49	4.83	3.04	4.65	2.53	0
BRIEFCASE	0.49	4.86	3.03	5.2	3.59	0
BROOK	0.46	4.43	0.83	7	3.33	0
BROTHER	0.65	4.43	45.24	6.18	4.48	1
BUBBLE	0.43	4.6	3.48	6.43	4.19	0
BUCKET	0.45	4.96	4.66	4.55	2.96	0
BUG	0.52	5	6.95	3.45	6.06	1
BUGGY	0.50	4.18	0.88	4.65	4.04	0
BULLET	0.51	4.83	11.99	3.45	5.89	0
BUNNY	0.55	4.97	4.05	7.3	3.86	1
BUREAU	0.48	4.04	4.21	4.7	3.74	0
BURGLAR	0.52	4.44	1.61	2.67	5.32	1
BUTCHER	0.55	4.65	3.47	4.4	4.15	1
CABBAGE	0.53	4.75	1.30	4.6	2.91	0
CABIN	0.54	4.92	5.47	5.9	3.74	0
CAFE	0.49	4.96	2.00	4.8	3.48	0
CAMEL	0.49	4.93	1.65	5.29	3.1	1
CANAL	0.47	4.68	2.18	5.71	4.05	0
CANDY	0.47	4.83	10.75	7.27	5.03	0
CANYON	0.51	4.81	2.63	5.5	3.9	0
CAPTIVE	0.45	3.03	1.16	3.27	4.88	1
CARRIAGE	0.53	4.86	2.87	6.1	2.52	0
CARROT	0.52	5	1.47	5.79	3.91	0
CASHEW	0.48	4.92	0.08	7.51	6.59	0
CASHIER	0.47	4.89	1.38	5.1	3.45	1
CASKET	0.53	4.86	1.20	2.42	4.82	0
CATCHER	0.44	4.44	1.10	5.39	3.26	1
CATTLE	0.54	4.64	3.15	5.42	2.64	1
CEILING	0.52	4.85	4.17	5.39	2.75	0
CELLAR	0.51	4.68	2.80	4.7	3.14	0
CHAMPAGNE	0.56	4.82	10.09	6.86	3.8	0
CHAPEL	0.50	4.6	2.07	6.67	2.45	0
CHAUFFEUR	0.58	4.43	1.76	5.42	4.7	1
CHEMIST	0.54	4.24	0.72	4.95	3.95	1
CHEST	0.51	4.93	14.03	5.18	4.95	0
CHILD	0.58	4.78	36.56	7.2	5.33	1
CHIPMUNK	0.52	4.97	0.35	7.33	3.8	1
CHURCH	0.54	4.9	15.98	5.21	3.63	0
CIGAR	0.43	4.93	4.61	4.4	4.27	0
CITRUS	0.51	4.21	0.24	6.36	4.16	0
CLAM	0.52	4.89	1.88	4.7	3.36	1
CLAMP	0.41	4.53	1.90	4.6	5.05	0

(*Appendix continues*)

Appendix (*continued*)

Word	Recall prob. (R)	Concreteness (C)	Contextual diversity (D)	Valence (V)	Arousal (Ar)	Animacy (An)
CLIMBER	0.49	4.5	0.35	6	4.45	1
CLOCK	0.37	5	19.54	5.65	3.35	0
CLOTHES	0.48	4.76	33.26	6.77	3.14	0
CLOUD	0.47	4.54	4.63	6.2	2.81	0
COBRA	0.51	5	0.73	4.42	5.71	1
COCKTAIL	0.47	4.4	5.11	6.95	5.6	0
COCOON	0.46	4.83	0.55	5.22	2.85	0
COD	0.60	4.61	0.83	5.21	3.95	1
COFFEE	0.49	4.81	36.23	7	5.1	0
COIN	0.49	4.89	3.42	6.55	3.13	0
COLLEGE	0.59	4.62	22.90	6.44	4	0
COLONEL	0.56	3.89	7.65	5.18	3.9	1
COMET	0.47	4.67	0.74	6.9	4.8	0
COMPASS	0.39	4.66	1.48	5.75	2.85	0
CONCERT	0.46	4.35	4.97	7	5.17	0
CONTRACT	0.41	4.15	9.99	5.1	3.95	0
CONVICT	0.55	4.11	2.49	2.28	4.95	1
COOK	0.53	4.32	16.32	7.12	4.33	1
COOKBOOK	0.52	4.9	0.37	6.45	3.45	0
CORAL	0.44	4.4	0.68	6.42	3.18	1
COSTUME	0.43	4.57	4.71	6.05	4.78	0
COTTAGE	0.52	4.85	1.96	6.63	2.95	0
COUCH	0.51	4.71	9.44	6.52	3.4	0
COUNTRY	0.50	4.17	39.32	6.14	3.71	0
COUNTY	0.42	4.04	10.97	5.18	3.4	0
COURSE	0.42	3.82	79.47	5.5	3.67	0
COUSIN	0.57	3.7	12.41	6.11	2.6	1
COWBOY	0.56	4.72	5.75	5.43	4.43	1
CRAB	0.52	4.9	2.24	5.81	4.13	1
CRATER	0.47	4.61	0.74	5.15	4.84	0
CRAYON	0.45	4.87	0.24	5.76	2.91	0
CREATURE	0.46	4.07	7.83	6.06	4.77	1
CREVICE	0.56	4.43	0.23	4.67	4.58	0
CRIB	0.55	4.86	2.53	6.43	4.26	0
CRICKET	0.42	4.77	1.12	5.71	3.22	1
CRITIC	0.44	3.55	1.50	4.1	4.25	1
CROSS	0.46	4.44	20.27	5.67	3.05	0
CROWN	0.55	4.81	4.33	6	4.52	0
CRUTCH	0.42	4.5	0.63	3.64	3.67	0
CUPBOARD	0.47	4.79	1.18	4.81	3.52	0
CURTAIN	0.47	4.82	3.95	5.36	3.62	0
CUSTARD	0.48	4.85	0.48	5.45	3.5	0
CYCLONE	0.48	4.48	0.21	3.47	5.09	0
DAISY	0.52	5	2.12	7.48	3.95	1
DANCER	0.52	4.75	5.95	6.64	4.52	1
DANDRUFF	0.43	4.79	0.38	3.05	4.41	0
DASHBOARD	0.48	4.61	0.54	5.25	3.15	0
DAUGHTER	0.64	4.79	35.68	6.73	5	1
DENIM	0.53	4.77	0.31	6	3.67	0
DENTIST	0.56	4.93	3.51	3.84	4.37	1
DIME	0.47	4.85	5.64	5.58	3.52	0
DINER	0.52	4.82	3.83	6.75	4.04	1
DIVER	0.50	4.69	0.63	5.66	5.42	1
DOLPHIN	0.57	4.96	0.86	6.67	3	1
DONKEY	0.53	5	1.93	6.29	2.9	1
DONOR	0.46	3.54	1.56	6.57	3.76	1
DORM	0.55	4.41	2.23	5.16	4.29	0
DOUGHNUT	0.49	4.96	1.93	7.5	4.5	0
DRAGON	0.54	4.39	3.30	6.68	5.45	1
DRAWING	0.44	4.6	6.72	4.67	3	0

(*Appendix continues*)

Appendix (*continued*)

Word	Recall prob. (R)	Concreteness (C)	Contextual diversity (D)	Valence (V)	Arousal (Ar)	Animacy (An)
DRESS	0.55	4.93	26.06	6.42	4.73	0
DRESSER	0.48	4.96	1.90	5.28	2.58	0
DRILL	0.44	4.4	5.56	4.73	5.11	0
DRINK	0.44	4.76	51.90	6.67	5.19	0
DRIVER	0.52	4.71	16.67	6.39	3.15	1
DRUG	0.42	4.48	13.03	4.11	4.48	0
DUST	0.52	4.4	9.67	3.72	3.45	0
DUSTPAN	0.46	5	0.12	4.67	2.86	0
EAGLE	0.51	5	3.33	6.47	4.57	1
EGYPT	0.60	3.71	1.66	3.08	5.14	0
ELBOW	0.53	5	2.97	5.38	3.2	0
EMPIRE	0.45	3	4.41	5.36	4.59	0
EUROPE	0.59	3.66	9.18	5.35	3.57	0
EXPERT	0.48	2.85	10.04	6.74	4.05	1
EYELASH	0.44	5	0.41	5.45	2.61	0
FARMER	0.56	4.54	3.80	6.14	3.67	1
FEMALE	0.57	4.57	12.36	7.52	5.9	1
FIDDLE	0.43	4.81	1.54	5.05	4.05	0
FILM	0.42	4.71	12.08	6.33	4.1	0
FINGER	0.54	5	15.34	5.8	4.15	0
FIREMAN	0.56	4.8	1.25	6.47	4.52	1
FIREPLACE	0.52	4.68	2.25	5.95	5.2	0
FLAG	0.41	4.79	5.93	6.1	3.74	0
FLASHLIGHT	0.43	5	2.50	6	4.04	0
FLASK	0.48	4.79	0.48	5.5	4.24	0
FLEET	0.49	3.81	2.59	5.4	4.43	0
FLESH	0.47	4.59	9.44	5.2	4.11	0
FLIPPER	0.43	4.26	0.39	5.84	3.05	0
FLOWER	0.51	5	7.50	7.3	3.67	0
FLUTE	0.43	5	0.85	6.29	3.72	0
FOOT	0.54	4.9	24.49	4.68	2.77	0
FOOTBALL	0.51	4.73	9.61	6.52	5.65	0
FOREHEAD	0.50	4.9	4.01	5.04	3.14	0
FOREST	0.54	4.76	5.96	6.68	4.44	0
FOX	0.55	4.97	4.88	5.52	4.36	1
FRAGRANCE	0.50	4.03	0.66	6.67	4.72	0
FRAME	0.43	4.3	6.28	5.32	4.04	0
FRANCE	0.59	3.79	7.46	5.21	3.71	0
FRECKLE	0.46	4.56	0.21	5.53	3.36	0
FREEZER	0.46	4.87	2.55	5.32	2.7	0
FRIAR	0.60	3.88	0.37	5.2	3.26	1
FRIEND	0.57	3.07	73.65	6.79	4.29	1
FRUIT	0.53	4.81	8.52	7	4.09	0
FUNGUS	0.48	4.59	0.92	2.79	4.67	1
GALLON	0.45	3.92	1.10	5.67	3.4	0
GANGSTER	0.55	3.93	2.04	2.59	6.36	1
GARBAGE	0.42	4.69	10.16	2.88	3.84	0
GARDEN	0.55	4.73	10.03	7.25	3.71	0
GARLIC	0.50	4.89	2.29	5.67	4.12	0
GAVEL	0.50	4.88	0.35	4.3	2.72	0
GAZELLE	0.56	4.72	0.41	6.47	4.05	1
GHETTO	0.51	3.82	1.59	3.16	7.05	0
GIFT	0.43	4.56	22.38	7.27	4.64	0
GIRL	0.66	4.85	73.24	7.15	5.23	1
GLASS	0.52	4.82	21.95	5.48	3.14	0
GLOBE	0.46	4.59	2.48	6.15	3.36	0
GLOVE	0.45	4.97	4.01	6.11	3.57	0
GOBLIN	0.51	4.38	0.23	3.16	4.7	1
GRAPE	0.51	5	1.63	6.7	3.5	0
GRAVE	0.52	4.56	10.50	2.4	4.54	0
GRAVEL	0.52	5	0.58	4.42	2.95	0

(*Appendix continues*)

Appendix (*continued*)

Word	Recall prob. (R)	Concreteness (C)	Contextual diversity (D)	Valence (V)	Arousal (Ar)	Animacy (An)
GREASE	0.49	4.61	3.23	3.9	3.62	0
GRILL	0.46	4.86	2.16	6.64	5.26	0
GRIZZLY	0.56	3.17	0.68	4.45	5	1
GROUND	0.46	4.77	27.00	5.28	2.35	0
GUARD	0.51	4.04	18.71	5.89	3.6	1
GUITAR	0.44	4.9	4.07	7.1	4.4	0
GYMNAST	0.51	4.85	0.26	5.22	4.55	1
HAMPER	0.43	4.21	0.72	5.15	3.14	0
HAND	0.53	4.72	66.31	5.9	3.98	0
HANDBAG	0.48	4.93	0.70	4.68	3.94	0
HARP	0.40	4.85	0.97	5.4	3.26	0
HATCHET	0.50	4.93	0.85	4.43	5.14	0
HAWK	0.52	4.93	2.97	6.46	4.83	1
HEADBAND	0.47	5	0.17	5.16	3.33	0
HEART	0.50	4.52	54.63	6.95	5.07	0
HEDGE	0.47	4.54	0.69	5.14	3.39	0
HELMET	0.45	4.92	3.61	5.26	3.71	0
HERO	0.50	3.07	16.46	7.44	6.35	1
HIGHWAY	0.54	4.72	6.87	5.19	4.28	0
HIKER	0.55	4.53	0.11	6.68	4.32	1
HONEY	0.52	4.88	49.89	7.27	4.38	0
HOOD	0.45	4.88	5.66	4.95	3.33	0
HOOK	0.44	4.79	15.32	4	4	0
HORNET	0.45	4.96	0.38	3.37	5.73	1
HORSE	0.58	5	18.80	6.05	4.16	1
HOSTESS	0.53	4.12	1.87	6.7	4.18	1
HOUND	0.53	4.48	1.99	5.3	4.27	1
HUMAN	0.52	4.93	34.94	6.45	3.62	1
HUSBAND	0.65	4.11	40.26	7.41	4.38	1
ICEBERG	0.51	4.96	0.95	5.05	5.12	0
ICING	0.48	4.66	0.67	6.05	4.32	0
IDOL	0.46	3.63	1.22	5.4	4.38	0
IGLOO	0.56	4.73	0.17	3.81	4.5	0
INFANT	0.60	4.93	1.72	6.65	4.1	1
INMATE	0.56	4.19	0.93	2.67	4.73	1
ISLAND	0.54	4.96	8.15	7.18	4.25	0
ITEM	0.40	4.41	5.15	5.29	2.9	0
JAPAN	0.63	4.82	4.05	4.87	3.86	0
JEANS	0.53	5	2.90	5.47	3.95	0
JELLO	0.49	4.18	0.18	2.52	5.57	0
JELLY	0.52	4.93	2.86	5.9	3.63	0
JOURNAL	0.45	4.63	3.23	5.91	3.23	0
JUDGE	0.54	3.75	19.33	3.89	4.5	1
JUGGLER	0.48	4.5	0.23	6.3	4.05	1
JUNGLE	0.50	4.66	5.79	5.7	4.06	0
JURY	0.52	4.64	6.97	4.23	5.2	0
KEEPER	0.44	3	1.93	5.89	3.81	1
KETCHUP	0.55	5	2.27	6.03	3.85	0
KIDNEY	0.50	4.96	2.83	4.9	3.95	0
KITCHEN	0.56	4.97	21.64	6.17	3.52	0
KLEENEX	0.45	4.92	0.82	6.11	3.65	0
KNAPSACK	0.49	4.9	0.32	5.3	3.86	0
KNIFE	0.56	4.9	14.96	4.33	4.86	0
LABEL	0.37	4.46	3.05	4.79	3.38	0
LACE	0.54	4.85	1.62	6.58	3.75	0
LADDER	0.49	5	3.73	5.32	4.09	0
LADY	0.61	4.33	49.23	6.91	4.05	1
LAGOON	0.53	4.5	0.45	6.23	3.56	0
LAKE	0.56	4.88	9.56	7.13	2.64	0
LAMP	0.39	4.97	4.58	5.74	2.71	0
LAPEL	0.51	4.56	0.29	5.74	3.17	0

(*Appendix continues*)

Appendix (*continued*)

Word	Recall prob. (R)	Concreteness (C)	Contextual diversity (D)	Valence (V)	Arousal (Ar)	Animacy (An)
LASER	0.43	4.5	2.67	5.64	5.09	0
LAVA	0.51	4.82	1.04	4.33	5.26	0
LEADER	0.47	3.89	10.66	6.24	4.96	1
LEG	0.57	4.83	18.93	6.22	2.75	0
LEOPARD	0.56	5	1.00	6.43	6.26	1
LETTUCE	0.53	4.97	1.47	5.84	3.64	0
LIGHTNING	0.52	4.59	4.29	5.34	6.75	0
LILY	0.56	4.69	3.12	7.05	2.64	1
LION	0.58	4.96	3.91	5.84	5.29	1
LIPSTICK	0.54	4.9	3.95	6.35	4	0
LIVER	0.51	4.68	5.10	4.19	3.27	0
LIZARD	0.54	4.68	1.37	5.43	5.5	1
LODGE	0.48	4	2.15	6.43	3.33	0
LOFT	0.45	4.32	1.24	5.96	3.42	0
LONDON	0.63	3.92	9.22	7.29	3.88	0
LOVER	0.61	3.68	10.56	8.05	7.45	1
LUGGAGE	0.50	4.83	4.46	5.19	3.75	0
LUMBER	0.45	4.56	0.99	5.65	3.32	0
LUNCH	0.49	4.31	31.44	6.64	3.57	0
MACHINE	0.47	4.25	22.41	5	4.39	0
MAILBOX	0.48	5	1.88	6.05	2.29	0
MAILMAN	0.56	4.57	1.28	5.32	3.32	1
MAMMAL	0.49	4.59	0.62	5.95	3.81	1
MAPLE	0.42	4.46	1.08	6.09	3.77	1
MARINE	0.52	4.25	3.39	5.95	3.44	1
MARKER	0.41	4.62	1.91	5.8	4.05	0
MARKET	0.46	4.7	14.18	6.21	3.55	0
MARROW	0.44	4.48	0.61	4.29	3.65	0
MARS	0.53	4.48	2.49	7.09	5.52	0
MARSH	0.51	4.85	0.72	7.09	5.52	0
MASK	0.45	4.96	5.72	4.81	3.26	0
MATCH	0.40	4.14	18.56	5.61	3.05	0
MATTRESS	0.54	5	2.74	5.74	3.45	0
MEAT	0.59	4.9	15.33	6.62	4.3	0
MEDAL	0.46	4.89	3.53	5.2	5.3	0
MESSAGE	0.44	3.97	28.90	6.18	3.81	0
MILDEW	0.48	4.57	0.14	2.61	4.14	0
MILK	0.52	4.92	15.05	6.74	2.33	0
MISSILE	0.52	4.83	2.47	2.85	5.67	0
MISTER	0.49	3.15	14.64	5.56	3.2	1
MONEY	0.49	4.54	67.32	7.1	6.86	0
MONSTER	0.51	3.72	11.27	2.55	5.55	1
MOP	0.49	4.97	1.80	4.53	3.14	0
MOTEL	0.48	4.93	5.59	5.3	3.55	0
MOTOR	0.41	4.84	5.60	5.64	5.42	0
MUFFIN	0.45	4.78	1.84	7.1	4.05	0
MUMMY	0.54	4.72	1.98	4.81	3.75	0
MUSTARD	0.53	4.93	2.71	4.74	3.39	0
NAPKIN	0.41	4.93	1.63	5.63	3.09	0
NECKLACE	0.48	4.96	3.33	6.85	3.52	0
NEUTRON	0.46	2.69	0.29	6.62	4.47	0
NIGHTGOWN	0.53	4.9	0.93	5.85	3.35	0
NOMAD	0.51	4.1	0.13	4.71	3.17	1
NOTEBOOK	0.46	4.92	1.47	6.05	3.58	0
NOVEL	0.43	4.21	3.54	5.74	3.41	0
NURSE	0.58	4.39	13.10	5.41	4.64	1
OFFICE	0.51	4.93	47.39	4.54	3.05	0
OINTMENT	0.49	4.5	0.75	4.81	2.86	0
OMELET	0.53	4.93	1.00	6.25	3.89	0
ONION	0.53	4.86	1.81	5.37	4.95	0
ORANGE	0.58	4.66	8.15	6.81	4.04	0

(*Appendix continues*)

Appendix (*continued*)

Word	Recall prob. (R)	Concreteness (C)	Contextual diversity (D)	Valence (V)	Arousal (Ar)	Animacy (An)
ORCHID	0.51	4.92	0.56	7	4.11	1
OUTDOORS	0.47	4.61	1.14	6.91	3.94	0
OUTFIT	0.46	4.12	9.99	6.05	4.19	0
OUTLAW	0.52	3.61	0.98	4.1	6	1
OX	0.65	4.86	2.13	4.95	3.82	1
OYSTER	0.56	4.85	1.23	4.81	3.11	1
OZONE	0.46	3.5	0.67	5.05	4.17	0
PACKAGE	0.46	4.72	8.40	5.17	4.73	0
PADDING	0.44	4.52	0.46	5.25	2.76	0
PADDLE	0.44	4.8	1.28	5.38	4.17	0
PAIL	0.54	4.93	0.48	4.5	2.24	0
PALACE	0.55	4.57	5.22	6.1	4.67	0
PANTHER	0.52	4.93	0.55	6.1	5.45	1
PAPER	0.47	4.93	31.31	5.42	3.52	0
PARENT	0.58	4.56	5.66	6.73	4.14	1
PARROT	0.45	5	1.13	6.79	4.65	1
PARSLEY	0.52	4.77	0.39	6.26	2.77	0
PARTNER	0.52	3.53	21.13	7.11	3.7	1
PASSAGE	0.44	3.8	3.47	5.88	3.35	0
PASTA	0.46	4.86	1.66	7.08	3.97	0
PASTRY	0.46	4.97	0.89	6.9	4.95	0
PATIENT	0.46	2.5	17.41	6.71	2.77	1
PATROL	0.48	3.86	5.97	4.04	4.26	0
PEACH	0.55	4.9	2.53	6.83	4.7	0
PEANUT	0.48	4.89	4.32	6.38	3.48	0
PEBBLE	0.50	4.86	0.58	5.72	2.85	0
PECAN	0.46	4.87	0.43	6.63	3.52	0
PEDAL	0.42	4.44	1.01	5.16	3.89	0
PENGUIN	0.54	5	0.72	6.65	4	1
PEPPER	0.48	4.59	3.18	5.63	4.3	0
PERCH	0.50	4.1	0.44	5.35	3.38	0
PERFUME	0.44	4.66	4.66	6.58	4.28	0
PERMIT	0.38	3.43	5.35	5.26	3.76	0
PIANO	0.45	4.9	7.12	6.4	3.61	0
PICNIC	0.53	4.83	4.66	7.11	3.65	0
PICTURE	0.42	4.52	37.89	6.73	3.29	0
PIGEON	0.51	4.71	2.16	5.58	2.95	1
PIGMENT	0.41	4.4	0.15	6.13	3.95	0
PILOT	0.54	4.67	6.76	6	5.6	1
PIMPLE	0.48	4.77	0.75	2.11	3.9	0
PISTOL	0.53	4.89	3.86	3.92	5.79	0
PISTON	0.45	4.81	0.29	4.57	4.85	0
PIZZA	0.54	5	10.03	7.89	4.58	0
PLAID	0.54	4.23	0.86	5.71	3.72	0
PLASTER	0.43	4.59	1.24	5.24	3.9	0
PLATE	0.49	4.77	11.03	4.8	3.18	0
PLAYGROUND	0.55	4.77	2.30	7.14	4.67	0
PLAZA	0.49	4.44	2.07	6.33	3.9	0
PLIERS	0.51	4.93	0.58	4.48	3.55	0
PLUTO	0.51	3.82	0.46	6.32	3.05	0
POCKET	0.45	4.68	15.28	5.67	4.5	0
POET	0.46	4.36	3.27	6.85	2.91	1
POISON	0.52	4.27	8.30	2.16	6.01	0
POLICE	0.61	4.79	40.16	4.59	5.95	1
POPCORN	0.42	5	3.68	7.26	5.23	0
PORK	0.52	4.79	3.77	5	3.8	0
PORTRAIT	0.46	4.9	2.29	6.05	3.2	0
POSSUM	0.49	4.73	0.64	3.74	4	1
POSTAGE	0.42	4.37	0.50	5.09	2.57	0
POWDER	0.45	4.76	6.22	5.26	2.77	0
PREACHER	0.56	4.7	1.73	5.09	4.55	1

(*Appendix continues*)

Appendix (*continued*)

Word	Recall prob. (R)	Concreteness (C)	Contextual diversity (D)	Valence (V)	Arousal (Ar)	Animacy (An)
PRIMATE	0.44	4.5	0.32	5.84	2.95	1
PRINCE	0.60	4.44	8.21	5.44	5.15	1
PRINCESS	0.67	4.72	8.06	7.64	5.42	1
PROTON	0.47	3	0.12	6.26	3.36	0
PUDDING	0.49	4.9	2.38	6.72	3.8	0
PUDDLE	0.47	4.67	1.06	4.3	4.33	0
PUPIL	0.51	4.55	1.38	5.53	3.76	1
PUPPY	0.59	4.78	4.57	7.85	5.84	1
QUAIL	0.54	4.65	0.48	5.43	2.67	1
QUARTER	0.43	4.43	10.93	5.61	3.85	0
QUEEN	0.67	4.45	14.16	6.52	5.05	1
RABBIT	0.53	4.93	5.33	7.21	3.98	1
RACKET	0.38	4.26	3.58	3.95	4.33	0
RADISH	0.48	4.87	0.31	4.7	3.71	0
RAFT	0.48	5	1.56	5.7	4.55	0
RATTLE	0.44	4.07	1.69	4.47	4.48	0
RAZOR	0.48	4.9	3.00	4.9	4.23	0
REBEL	0.44	3.07	1.98	4.37	5.29	1
RECEIPT	0.42	4.86	3.35	5.41	4.5	0
RECORD	0.36	4.15	28.55	5.89	3.3	0
RELISH	0.50	3.3	0.98	4.55	4.45	0
REPORT	0.42	3.92	31.25	4.77	3.52	0
RIFLE	0.56	4.85	4.59	4.3	6.14	0
RIVER	0.54	4.89	14.44	6.72	4.22	0
ROBBER	0.52	4.31	1.79	2.9	6.2	1
ROBIN	0.47	4.61	3.34	6.63	2.64	1
ROBOT	0.44	4.65	2.58	6.18	4.43	0
ROCKET	0.49	4.73	3.43	5.8	5.04	0
ROD	0.53	4.43	3.16	4.95	3.05	0
ROOSTER	0.47	4.75	1.22	5.53	4.57	1
RUG	0.52	4.79	4.22	5	3.24	0
RUST	0.48	4.52	1.20	4.05	3.42	0
SADDLE	0.46	4.85	3.16	4.95	3.1	0
SALAD	0.51	4.97	6.99	6.35	3.78	0
SALMON	0.55	4.81	2.19	6.48	3.87	1
SALT	0.51	4.89	7.30	6.05	4.53	0
SANDWICH	0.51	4.9	8.92	7.18	4.94	0
SAUSAGE	0.55	4.88	2.69	6.32	4.8	0
SCALLOP	0.53	4.61	0.08	5.06	3.8	1
SCALPEL	0.56	4.86	1.42	3.95	4.48	0
SCARECROW	0.46	4.68	0.49	5.19	3.16	0
SCARF	0.46	4.97	1.97	6	2.39	0
SCISSORS	0.47	4.85	2.47	5.03	4.02	0
SCOTCH	0.48	4.55	5.70	5.89	4.2	0
SCRIBBLE	0.43	4.1	0.26	5.21	3.76	0
SCULPTURE	0.46	4.79	1.23	6.5	3.48	0
SEAFOOD	0.54	4.83	0.98	6.45	4.73	0
SEAGULL	0.50	5	0.41	5.27	2.9	1
SEAL	0.51	4.63	5.63	5	2.5	1
SERVANT	0.49	4.64	4.59	4	3.77	1
SERVER	0.52	4.55	1.44	5.35	3.71	1
SHARK	0.54	4.93	3.08	4.02	5.27	1
SHELF	0.47	4.96	3.27	5.62	3.09	0
SHELTER	0.43	4.64	4.47	6.5	3.25	0
SHERIFF	0.58	4.5	7.89	4.44	4.3	1
SHIRT	0.53	4.94	16.69	5.56	2.3	0
SHORTCAKE	0.49	4.41	0.18	6.89	4.14	0
SHORTS	0.53	4.82	4.24	5.95	4.23	0
SHOULDER	0.52	4.93	11.64	5.1	2.96	0
SHOVEL	0.44	4.97	3.00	4.8	3.5	0
SHRUB	0.42	4.92	0.15	4.21	3.4	1

(*Appendix continues*)

Appendix (*continued*)

Word	Recall prob. (R)	Concreteness (C)	Contextual diversity (D)	Valence (V)	Arousal (Ar)	Animacy (An)
SIBLING	0.59	4.37	0.48	7.27	5.16	1
SIDEWALK	0.54	4.96	2.77	5.67	3.65	0
SILK	0.53	4.7	4.11	6.55	3.88	0
SISTER	0.67	4	36.31	7	3.86	1
SKETCH	0.41	4.56	1.90	6.21	4.33	0
SKILLET	0.53	4.73	0.38	5.85	3.24	0
SKIRT	0.56	4.82	4.59	6.14	4.88	0
SLIDE	0.45	4.48	6.78	5.71	3.3	0
SLIME	0.53	4.48	1.26	3.15	4.58	0
SLOPE	0.47	4.07	1.43	5.42	3.59	0
SLUG	0.52	4.64	2.25	3.16	4.9	1
SMOG	0.44	4.14	0.55	2.56	4.3	0
SNACK	0.46	4.36	4.15	6.53	4.11	0
SNAIL	0.50	4.93	0.64	4.52	3.05	1
SNAKE	0.54	5	6.27	4.03	7.24	1
SODA	0.41	4.97	7.95	5.47	4.77	0
SOFTBALL	0.47	4.89	0.79	5.53	4.1	0
SPACE	0.49	3.54	19.96	6.89	3.6	0
SPARROW	0.51	4.85	0.69	6.58	3.78	1
SPHINX	0.57	4.83	0.37	6.05	4.9	0
SPIDER	0.48	4.97	2.75	3.35	6.91	1
SPONGE	0.40	5	2.71	5.45	4.25	0
SPOOL	0.46	4.62	0.21	5.71	3.37	0
SPOON	0.47	4.96	3.22	5.9	3.79	0
SPOUSE	0.58	3.85	0.95	7.44	5.76	1
STAKE	0.44	4.21	7.94	4.78	3.42	0
STALLION	0.56	4.72	1.13	6.35	4.77	1
STAMP	0.40	4.7	2.65	5.8	3.45	0
STAPLE	0.39	4.34	0.61	5	4.48	0
STAR	0.50	4.69	21.71	7.47	5.5	0
STATUE	0.45	4.93	3.78	5.95	2.82	0
STICKER	0.37	4.67	1.25	5.57	4.05	0
STOMACH	0.49	4.89	14.53	4.53	3.76	0
STONE	0.49	4.72	11.18	4.81	3.25	0
STOVE	0.48	4.96	3.36	5.63	3.82	0
STREAM	0.50	4.5	3.51	6.9	4.35	0
STUDENT	0.62	4.92	13.59	6.41	4.25	1
SUBWAY	0.48	4.86	4.03	5.44	4.41	0
SUITCASE	0.45	4.97	4.42	5.25	3.24	0
SUMMIT	0.55	4.21	0.94	5.5	4.35	0
SUNRISE	0.49	4.69	2.72	7.35	4.68	0
SUNSET	0.52	4.54	4.47	7.46	4.68	0
SUPPER	0.43	4.63	7.05	6.72	3.6	0
SURVEY	0.35	4.08	1.82	5.55	3.68	0
SUSPECT	0.50	2.59	15.46	2.39	4.57	1
SWAMP	0.50	4.96	2.43	4.42	3.33	0
SWIMMER	0.53	4.77	1.05	6.26	4.26	1
SWITCH	0.37	4.07	11.92	5.29	3.9	0
SWORD	0.52	4.93	5.51	5.27	5.95	0
TABLE	0.50	4.9	34.44	5.49	3	0
TABLET	0.38	4.82	0.48	6.21	3.65	0
TART	0.47	3.27	1.11	5.43	4.8	0
TAXI	0.52	4.93	7.94	4.79	3.79	0
TEACHER	0.61	4.52	15.56	7.37	2.9	1
TEMPLE	0.50	4.53	4.88	5.3	3.36	0
TERMITE	0.48	4.7	0.29	3.08	4.24	1
THIEF	0.56	4.37	8.39	2.32	6.05	1
THREAD	0.44	4.83	2.50	5.5	3.87	0
THRONE	0.54	4.64	2.77	5.45	5.22	0
TILE	0.49	4.68	0.93	5	2.89	0
TOASTER	0.42	4.9	1.41	5.8	3.85	0

(*Appendix continues*)

Appendix (*continued*)

Word	Recall prob. (R)	Concreteness (C)	Contextual diversity (D)	Valence (V)	Arousal (Ar)	Animacy (An)
TOMBSTONE	0.55	4.71	0.66	3.14	4.86	0
TORTOISE	0.55	4.87	0.32	5.58	3.32	1
TOURIST	0.50	4.59	2.23	5.71	3.57	1
TRACTOR	0.49	5	1.26	5.05	3.73	0
TRANSPLANT	0.54	3.77	1.61	4.6	6.37	0
TREAT	0.41	3.79	21.88	6.84	5	0
TRENCH	0.48	4.46	1.22	4.43	3.22	0
TRIBE	0.46	4.14	2.29	5.63	4.29	0
TROMBONE	0.50	4.9	0.56	5	3.43	0
TROUT	0.54	4.72	1.34	5.62	3.85	1
TRUCK	0.50	4.84	18.54	5.16	3.76	0
TUBA	0.45	4.86	0.38	5.58	3.95	0
TUNNEL	0.51	4.82	5.31	4.48	4.09	0
TURKEY	0.51	4.89	6.82	5.9	3.45	1
TURNIP	0.49	4.79	0.54	4.63	3.32	0
TURTLE	0.51	5	3.09	6.16	2.52	1
TUTU	0.54	4.68	0.41	6.19	4.15	0
TWEEZERS	0.48	4.96	0.52	4.63	4.64	0
TWIG	0.49	4.75	0.74	5.47	3.18	0
TWISTER	0.46	4.44	0.69	4.25	6	0
TYPIST	0.54	4.41	0.35	5.3	4.14	1
ULCER	0.48	4.69	1.03	2.7	4.76	0
UMPIRE	0.50	4.27	0.36	4.19	4.57	1
UNCLE	0.60	4.24	22.23	6.5	4.05	1
VAGRANT	0.56	3.46	0.35	2.63	3.82	1
VALLEY	0.50	4.72	7.15	6.22	2.7	0
VALVE	0.41	4.83	1.65	5.1	3.84	0
VELVET	0.52	4.44	1.84	6.3	4.53	0
VENUS	0.55	4.54	1.57	5.37	4.18	0
VICTIM	0.49	3.59	14.02	2.05	5.37	1
VIKING	0.53	3.53	0.83	5.75	5.63	1
VIRUS	0.45	3.48	3.15	1.71	4.61	1
WAGON	0.46	4.89	5.58	5.21	3.1	0
WAITER	0.53	4.67	5.16	5.05	3.05	1
WAITRESS	0.61	4.56	4.74	5.1	3.5	1
WARDROBE	0.49	4.67	2.86	6.09	5.16	0
WASHER	0.41	4.7	0.95	5.16	2.9	0
WASP	0.46	4.96	0.51	2.71	5.33	1
WHISKERS	0.48	4.89	0.99	6	4.61	0
WHISTLE	0.37	4.42	5.64	5.7	3.94	0
WIDOW	0.54	4.33	4.51	2.28	3.5	1
WIFE	0.67	4.13	57.36	6.7	4.21	1
WINDOW	0.53	4.86	29.11	6.47	3.27	0
WITNESS	0.44	4.07	14.57	5.61	3.67	1
WOMAN	0.64	4.46	70.64	7.09	3.8	1
WORKER	0.47	4.59	4.98	5.95	3.6	1
WORLD	0.48	4.36	73.84	6.46	4.55	0
WRENCH	0.46	4.93	1.51	4.86	3.7	0
WRIST	0.56	4.93	4.41	5.06	3.27	0
XEROX	0.49	3.96	0.37	3	5.48	0
YACHT	0.55	4.97	2.35	5.88	3.98	0
YARN	0.44	4.93	0.72	5.47	2.7	0
YOLK	0.47	4.78	0.23	5.32	3.05	0
ZEBRA	0.56	4.86	0.69	6.47	3.9	1
ZIPPER	0.47	4.83	1.41	5.11	3.73	0
Mean	0.50	4.59	6.99	5.52	4.04	
SD	0.06	0.44	11.79	1.15	0.90	

Received March 23, 2018
Revision received August 13, 2020
Accepted August 13, 2020 ■