Cognition 239 (2023) 105497

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Short communication

Semantic determinants of memorability

Ada Aka [a,*,1], Sudeep Bhatia [b,1], John McCoy [b,1]

[a] Stanford University, United States of America
[b] University of Pennsylvania, United States of America

ARTICLE INFO

Dataset link: https://osf.io/7phj6

Keywords:
Human memory
Semantic representations
Memorability predictions
Psycholinguistics

ABSTRACT

We examine why some words are more memorable than others by using predictive machine learning models applied to word recognition and recall datasets. Our approach provides more accurate out-of-sample predictions for recognition and recall than previous psychological models, and outperforms human participants in new studies of memorability prediction. Our approach's predictive power stems from its ability to capture the semantic determinants of memorability in a data-driven manner. We identify which semantic categories are important for memorability and show that, unlike features such as word frequency that influence recognition and recall differently, the memorability of semantic categories is consistent across recognition and recall. Our paper sheds light on the complex psychological drivers of memorability, and in doing so illustrates the power of machine learning methods for psychological theory development.

1. Introduction

Over the last hundred years, psychologists have used word list memorization tasks to study "what" and "how" people remember. Typically, participants are presented with a list of words and attempt to either recall as many words as possible, or indicate whether they recognize particular words as amongst those previously studied. This research has identified the cognitive processes at play in memory by modeling how changing the composition and order of the presented word list influences recognition and recall (Gillund & Shiffrin, 1984; Metcalfe & Murdock, 1981; Yonelinas, 2002). It has also shown that certain words are more likely to be recognized or recalled than others, independently of the context in which they are presented (Aka, Phan, & Kahana, 2020; Kahana, Aggarwal, & Phan, 2018; Kensinger & Corkin, 2003; Madan, 2021; Rubin & Friendly, 1986).

Most research on what determines word memorability has focused on the role of 'psycholinguistic' word features (Rubin & Friendly, 1986). Prior research has shown that words are more likely to be both recognized and recalled if they are concrete, imageable, emotional, or arousing (Buchanan, Etzel, Adolphs, & Tranel, 2006; Cahill & McGaugh, 1995; Ghatala, 1981; Kensinger & Corkin, 2003; Paivio, Walsh, & Bons, 1994; Schwanenflugel, Akin, & Luh, 1992). Low frequency words are better recognized than high frequency words, but high frequency words are better recalled than low frequency words (Glanzer & Bowles, 1976; Hall, 1954).

Another key determinant of memorability involves semantics — the meaning of words and the themes these words convey. More than half a century ago, Bower (1967) and Underwood (1969) emphasized the importance of semantics by identifying it as one of the fundamental attributes of a memory trace. In particular, semantic attributes were shown to help form semantic categories and facilitate memory search, recognition, and retrieval for both words and images (Bellezza, 1981; Bower, 1970; Hovhannisyan et al., 2021; Mandler, 1968; Tulving & Pearlstone, 1966). For example, as shown by semantic similarity effects, people are more likely to transition among semantically related items in recall (Howard & Kahana, 2002b; Romney, Brewer, & Batchelder, 1993; Xie, Bainbridge, Inati, Baker, & Zaghloul, 2020). In fact, individual semantic features (such as size and usefulness) have been shown to be strong predictors of memorability even in models that control for 'psycholinguistic' word features (such as word frequency) (Aka et al., 2020; Madan, 2021).

Semantics may also result in memory interference (Nelson, Kitto, Galea, McEvoy, & Bruza, 2013). In the Deese–Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995), semantic information facilitates the false recognition and recall of items

* Corresponding author.
E-mail addresses: adaaka@stanford.edu (A. Aka), bhatiasu@sas.upenn.edu (S. Bhatia), jpmccoy@wharton.upenn.edu (J. McCoy).
[1] Funding was received from the National Science Foundation, United States, the Alfred P. Sloan Foundation, and Wharton Risk Management Center's Russell Ackoff Doctoral Student Fellowship. We thank Michael Kahana and members of the Computational Memory Lab for generously sharing the PEERS dataset. We also thank the Wharton Behavioral Lab for their assistance with data collection. Our experiments were pre-registered and our data is accessible at: https://osf.io/7phj6.

https://doi.org/10.1016/j.cognition.2023.105497
Received 19 January 2022; Received in revised form 26 March 2023; Accepted 11 May 2023
Available online 11 July 2023
0010-0277/

related semantically to the studied items. Along similar lines, Zaromb et al. (2006) showed how most recall errors consist of items that are semantically similar to the recalled items from a target list.

Due to its strong influence on behavior, semantic information naturally plays an essential role in many computational models that describe memory processes (Kahana, 2020). First, in the Search of Associative Memory (SAM) retrieval model, an exemplar dual-store model (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981) retrieval depends on both episodic (list-specific) and semantic (pre-existing) associations. In extensions of the SAM model, such as eSAM (Sirotin, Kimball, & Kahana, 2005) and fSAM (Kimball, Smith, & Kahana, 2007), prior semantic knowledge plays an important role.

In a separate cluster of computational models related to Retrieved Context Theory (Healey & Kahana, 2016; Howard & Kahana, 2002a; Lohnas, Polyn, & Kahana, 2015; Polyn, Norman, & Kahana, 2009), an important component is associative matrices representing semantic information. In these models, prior semantic knowledge is obtained using high-dimensional distributed semantic representations of words, in a manner similar to how we derived our semantic embeddings. In recent work, some of the authors of the current paper have shown that the CMR framework, equipped with such semantic representations, provides a good account of free association data (Richie, Aka, & Bhatia, 2022). While the current paper uses the Word2Vec language model, prior work used various sources of semantic information, such as latent semantic analysis (Landauer & Dumais, 1997) and word association space (Steyvers, Shiffrin, & Nelson, 2005). Compared to a variety of other corpus based measures of semantic association, representations derived from word association space were shown to best predict semantic organization effects in a CMR model (Morton & Polyn, 2016).

The research described above, as well as other related work (Anderson & Bower, 1972; Cox & Criss, 2020; Hintzman, 1984; Humphreys, Li, Burt, & Loft, 2020; Murdock, 1995; Shiffrin & Steyvers, 1997) has contributed to the development of models with varying degrees of complexity that investigate the influence of semantics in episodic memory. These models capture a number of correct and false memory effects, including those observed in the DRM paradigm.

In addition to the major role of semantic information on memory processes, prior literature on semantic categories and human memory has also identified multiple distinct categories as being more memorable than others. Some scholars have argued that information that promotes survival and reproductive success is more likely to be remembered, for example, information relating to snakes and spiders (Nairne, Thompson, & Pandeirada, 2007; Öhman & Mineka, 2001) death (Bugaiska, Mermillod, & Bonin, 2015), contamination (Bonin, Thiebaut, Witt, & Méot, 2019), animate concepts (VanArsdall, Nairne, Pandeirada, & Blunt, 2013), and the faces of potential mates (Pandeirada, Fernandes, Vasconcelos, & Nairne, 2017).

Despite introducing many important findings, however, most previous research on memorability has been limited to a small number of pre-determined, low-dimensional features and their factorial manipulations (e.g., studying animacy versus inanimacy). As Cox, Hemmer, Aue, and Criss (2018) have pointed out, effects identified using these approaches are subject to Clark's famous "language-as-fixed-effect" fallacy (1973), since stimuli in such memorability experiments are not random samples. In their paper, the authors circumvent the "language-as-fixed-effect" fallacy and demonstrate how episodic memory, both recall and recognition, primarily depends on semantic information (specifically semantic distinctiveness and concreteness).

In our work, we continue to move further away from the conventional approach and instead acknowledge a much larger number of continuous stimulus dimensions that jointly contribute to memory performance. Rather than relying on a small number of pre-determined features, we use automatically extracted, high-dimensional distributed semantic representations of words (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Such representations have been used to analyze cognitive phenomena such as similarity judgments and categorization (Bhatia, Richie, & Zou, 2019; Floyd, Dalawella, Goldberg, Lew-Williams, & Griffiths, 2021; Günther, Rinaldi, & Marelli, 2019; Jones, Willits, Dennis, & Jones, 2015; Lenci, 2018; Mandera, Keuleers, & Brysbaert, 2017; Peterson, Chen, & Griffiths, 2020) and thus are suited for studying memorability as well. Especially relevant to the current paper, Xie et al. (2020) used such high-dimensional representations to emphasize how words that have a greater semantically similarity to other words have an advantage during retrieval and they may also lead to more intrusions when retrieval is not successful.

We combine large datasets of word memorability (Kahana et al., 2018; Weidemann & Kahana, 2016) with off-the-shelf machine learning techniques to build semantic representation models that make precise quantitative predictions of the probability that an arbitrary word is recognized or recalled in standard memory paradigms. We compare the predictive performance of our models to both predictions made from word features, and to the predictions of human participants. These comparisons provide new evidence to understand the complex psychological determinants of memorability, and they contribute to theory development in a data-driven manner (Agrawal, Peterson, & Griffiths, 2020; Peterson, Bourgin, Agrawal, Reichman, & Griffiths, 2021). For example, we show that although there is a low correlation between recognition and recall probabilities (Brown, 1976; Myers, 1914; Tversky, 1974), which semantic categories are memorable is consistent across recognition and recall.

2. Methods

2.1. Datasets and measures

Recognition Memory Dataset. We used recognition memory data from Experiments 1–3 of the large-scale Penn Electrophysiology of Encoding and Retrieval Study (PEERS), available at http://memory.psych.upenn.edu/Data_Archive (Healey, Crutchley, & Kahana, 2014; Lohnas & Kahana, 2013). Applying the same selection criteria as Weidemann and Kahana (2016) resulted in 171 participants (ages 18–30). Participants each took part in up to 20 experimental sessions, and collectively contributed 3120 sessions of data. In each session, participants studied 12 to 16 lists that each contained 16 words from a pool of 1638 words. At the end of each session, participants completed a recognition memory task by indicating, for each of 320 probe words, whether the word was previously shown in the session. We estimated the recognition probability of each word as the percentage of participants who correctly recognized it (Bainbridge, 2017; Isola, Parikh, Torralba, & Oliva, 2011).

Recall Memory Dataset. We used recall memory data from Experiment 4 of the PEERS dataset (Aka et al., 2020; Kahana et al., 2018). We used data from all 98 participants (ages 18–30), each of whom took part in 23 experimental sessions and collectively contributed 2254 sessions of data. In each session, participants studied 24 lists that each contained 24 words. After the presentation of each list, participants answered simple arithmetic questions for 24 s, and then had 75 s to recall as many words as they could from the just-presented list.

The word pool consisted of 576 nouns. The word pool items were originally sampled from the University of South Florida free association norms dataset (Nelson, McEvoy, & Schreiber, 2004), and are later used in all of the Penn Electrophysiology of Encoding and Retrieval (PEERS) studies. In the Supplementary Materials, we provide the 576 item word pool along with the average word features and their standard deviations (Table SM-8). We estimated the recall probability of each word as the percentage of participants who correctly recalled it.

Additional Memory Tasks to Test Generalization. To assess the generalizability of our model, we used additional data (https://osf.io/dd8kp/) from two memory tasks: single item recognition and free recall (Cox et al., 2018). The Supplementary Materials provides additional information about these tasks.

Human Predictions of Memorability. In two pre-registered (https://osf.io/7phj6), incentivized experiments, participants predicted how recognizable or recallable words were in the memory studies described above. In both the recognition prediction study (N = 479, 241 female, 231 male, 7 non-binary; mean age = 31) and the recall prediction study (N = 486; 267 female, 209 male, 10 non-binary; mean age = 35), participants (US residents) were recruited from Prolific Academic. They were paid USD0.75 (USD7.50/hour) and could earn a performance bonus. The experiments were approved by the University of Pennsylvania Institutional Review Board and all participants provided informed consent.

After reading a description of either the original recognition or recall study described above, participants predicted, on a 100-point scale, the probability of recognition or recall for 24 words randomly sampled from the 576 words used in both the original recognition and recall studies.[2] The most accurate 30% of participants were paid double (USD1.50).

Distributed Semantic Representations. We obtained high-dimensional semantic representations of words from lexical co-occurrence statistics in very large corpora. Such representations encode words as numeric vectors. The vectors of two words that share common context in the training corpora, and are thus likely similar in meaning, are close together. We used the pre-trained Word2Vec model (Mikolov et al., 2013), which has 300-dimensional representations of over 3 million common words and phrases, including all of the words studied in this paper. This model has performed well in previous applications to psychology (Bhatia et al., 2019; Floyd et al., 2021; Peterson et al., 2020).

Psycholinguistic Word Features. We coded six features of all of the words common to the recognition and recall datasets. We used concreteness ratings (on a 5-point scale) collected by Brysbaert, Warriner, and Kuperman (2014) for 566 words, and those collected by Aka et al. (2020) for the remaining 10 words. We used valence ratings and arousal ratings on nine-point scales (Warriner, Kuperman, & Brysbaert, 2013). For word frequency, we used the contextual diversity measure (the number of documents a word appears in) from the SUBTLX-US dataset (Brysbaert & New, 2009). We used binary animacy ratings (Aka et al., 2020), and for word length counted the number of letters in a word. Prior literature has frequently examined these six word features in relation to memorability and has often identified them as influential predictors. However, we do not claim that our psycholinguistic word feature model includes a fully exhaustive set of predictors.

2.2. Predicting memorability

To predict word recognition probabilities, we regressed, using Ridge regression to prevent overfitting given high-dimensional covariates, the recognition probability of each word on either the six psycholinguistic word features described above, or on the 300-dimensional semantic representation of the word. We refer to these as the 'word feature model' and the 'semantic representation model' respectively.

We evaluated our memorability predictions out-of-sample with leave-one-out cross validation. That is, to predict the recognition probability of each word we trained a model on the remaining words in the dataset and evaluated the predicted probability of the held out word. To make predictions for words not in our dataset, we trained a version of our model on all words in the recognition dataset. This trained model can make predictions for almost any arbitrary word. We release our pre-trained model and instructions on how to use it at: https://osf.io/7phj6. In the Supplementary Materials, we repeated our analyses with alternative regression techniques including a logistic regression with a binomial link function applied to trial-level (rather than aggregate-level data) and fit separate models to data from each participant. We implemented analogous models for recall.

[2] As preregistered, participants who failed either of two attention checks (N = 21 for recognition prediction and N = 14 for recall prediction) were excluded.

2.3. Lexicon analysis to interpret the semantic representation model

To interpret the semantic representation models, we used them to predict recognition and recall probabilities for the words in the Linguistic Inquiry and Word Count (LIWC) lexicon (Pennebaker, Francis, & Booth, 2001), the General Inquirer (GI) lexicon (Stone, Dunphy, & Smith, 1966), and the NRC Word-Emotion Association lexicon (Mohammad & Turney, 2013). The LIWC, GI, and the NRC Word-Emotion Association lexicon are very popular multi-construct lexicons that are widely used in psychology and adjacent areas. Together they include almost one hundred psychological constructs and thousands of keywords within these constructs.

The NRC characterizes words in terms of eight basic emotions and positive or negative sentiment. LIWC and GI characterize words in term of psychological constructs. For example, LIWC's 'Social Processes' construct includes words such as 'mate' and 'talk'. Many semantic categories in LIWC and GI are subsumed by higher-level categories, for example in LIWC 'Social Processes' subsumes 'Family', 'Friends', 'Female References' and 'Male References'. Given the large number of categories and the small number of words in some categories, we predominantly use these higher-level categories.[3]

In the Supplementary Materials, we include analyses demonstrating how influential semantic categories for each individual strongly resemble to those recovered using the aggregate-level models. We note that the lexicons used in this manuscript are not the only alternatives. Our methodology to interpret the semantic representation model is flexible, and can accommodate any set of keywords, dictionaries, or lexicons (when keywords have a distributed semantic representation available).

3. Results

The actual word recognition probabilities ranged from 78% to 100% (M = 91%, SD = 4%) and the recall probabilities from 35% to 69% (M = 50%, SD = 6%). These ranges are typical (Graf & Mandler, 1984; Lohnas & Kahana, 2013; Poon & Fozard, 1980; Rubin & Friendly, 1986). Figure SM-1 and Table SM-1 show the distribution of recognition and recall probabilities, and the most and least memorable words. We observed a low correlation between recognition and recall (Spearman's $\rho(575) = .09$, $p = .04$, 95% $CI = [.004, .166]$), which is consistent with prior work (Brown, 1976; Myers, 1914; Tversky, 1974).

3.1. Memorability predictions

Fig. 1 compares actual recognition and recall probabilities to the out-of-sample predictions from the semantic representation models, the word feature models, and participants in the prediction studies. To enable comparisons between recognition and recall, we analyzed model performance on the 576 words that occur in both the recognition and recall datasets. In the Supplementary Materials, we show that the recognition semantics representation model exhibited similar predictive accuracy on all 1062 other words in the recognition dataset, and report the results of other methods for training the semantic representation models. We assessed prediction accuracy using the correlation between actual memorability and out-of-sample predictions, and mean squared error (MSE). The semantic representation model demonstrated high predictive accuracy for both recognition ($r(575) = .50$, $p < .001$, 95% $CI = [.437, .560]$, $MSE = .001$) and recall ($r(575) = .70$, $p < .001$, 95% $CI = [.655, .740]$, $MSE = .002$), compared to the word feature model (recognition: $r(575) = .37$, $p < .001$, 95% $CI = [.297, .438]$, $MSE = .001$; recall: $r(575) = .48$, $p < .001$, 95% $CI = [.419, .545]$, $MSE = .002$) and the aggregated human predictions (recognition: $r(575) = .10$, $p = .012$, 95% $CI = [.023, .185]$, $MSE = 0.086$; recall: $r(575) = .35$, $p < .001$, 95% $CI = [.275, .418]$, $MSE = 0.009$).

[3] Tables SM-3 and SM-4 show, for LIWC and GI, the number of words per category and the subsidiary categories making up the high-level categories.

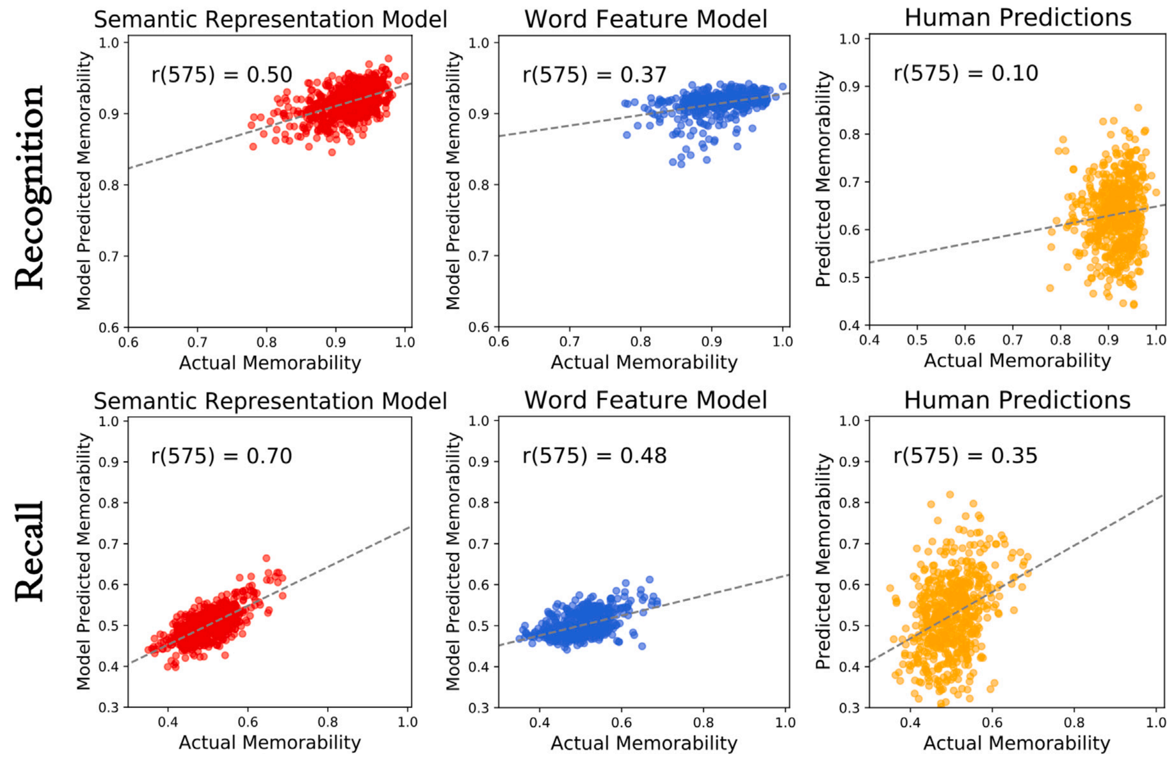

Fig. 1. Actual recognition and recall of the 576 words in the recognition and recall datasets compared to the leave-one-out predictions of the semantic representation model, the word feature model, and aggregated human predictions. The gray dashed line shows the best linear fit.

By a Steiger test for dependent correlations (see Supplementary Materials for details), the semantic representation model outperformed the word feature model (recognition: $z = 3.37$, $p < .001$; recall: $z = 7.44$, $p < .001$) and human predictions (recognition: $z = 7.81$, $p < .001$; recall: $z = 9.62$, $p < .001$). A comparison of squared errors also showed that the semantic representation model outperformed the word feature model (recognition: $t(575) = -2.89$, $p = .004$, paired-sample; recall: $t(575) = -6.61$, $p < .001$) and human predictions (recognition: $t(575) = -45.76$, $p < .001$; recall: $t(575) = -12.59$, $p < .001$).

In the Supplementary Materials, we go beyond the aggregate-level semantic representation model fit to average memorability scores from all participants and present similar models applied to trial-level (and individual-level) data. We first show that our results persist with alternative regression techniques, including a logistic regression with a binomial link function applied to trial-level data (rather than aggregate-level data) (Figure SM-2). We then demonstrate the predictive performance of our approach when a separate model is fit to data from each participant, as opposed to one model fit to aggregate-level data from all participants (Figure SM-4). While the trial and individual-level models have relatively more modest out-of-sample predictive accuracies, these results nevertheless provide additional evidence regarding the semantic representation model's power and promise, not only in explaining aggregate-level memorability but also in evaluating trial-level data and participant-level variability.

To determine a theoretical ceiling for our models, we also calculated the split-half reliability for the recognition task and the recall task using participant data. Specifically, we computed the average correlation across 1000 random half-splits of the participant data for recognition and recall. The split-half reliability for recognition was 0.683 (explaining 47% of the variance), and 0.917 for recall (explaining 84% of the variance).

3.2. Information overlap across models

To examine the information overlap between the words features model and the semantic representation model, we regressed actual memorability on the predictions of the word feature model and the semantic representation model together, and computed squared semi-partial correlations. The semantic representation model uniquely explained 15% of the variance of actual recognition, but the word feature model uniquely explained only 3%. The semantic representation model uniquely explained 27% of the variance of actual recall, 16 times more than the word feature model, which uniquely explained only 1.6%.

We also evaluated a combined model that used both semantic representations and word features. Word features did not add information to the semantic representations when predicting recognition, and only slightly increased accuracy when predicting recall (see Supplementary Materials for details).

3.3. Tests of generalizability

To further test using distributed semantic representations to understand human memory processes, we trained new versions of the semantic representation model on data for two additional memory tasks: free recall and single item recognition (Cox et al., 2018), which are described in the Supplementary Materials. As shown in Fig. 2, this results in accurate out-of-sample predictions.

To investigate the generalizability of the particular semantic representation model that we had previously trained, we used the semantic representation model trained on PEERS recall data to predict the Cox et al. free recall data. This resulted in accurate predictions ($r(923) = .45$, $p < .001$, 95% $CI = [.385, .512]$), suggesting that the model generalizes. Similarly, a semantic representation model trained on the Cox et al. free recall data predicts the PEERS recall data ($r(575) = .54$, $p < .001$, 95% $CI = [.497, .588]$). The single item recognition data is not as suitable for this analysis as the Cox et al. and PEERS recognition tasks differ in a number of ways, for example, whether the words were encoded in pairs or in isolation. Nevertheless, using the semantic representation model we had trained on PEERS recognition data, we were able to predict the Cox et al. single recognition data ($r(923) = .11$, $p < .001$, 95% $CI = [.046, .173]$). The performance of the

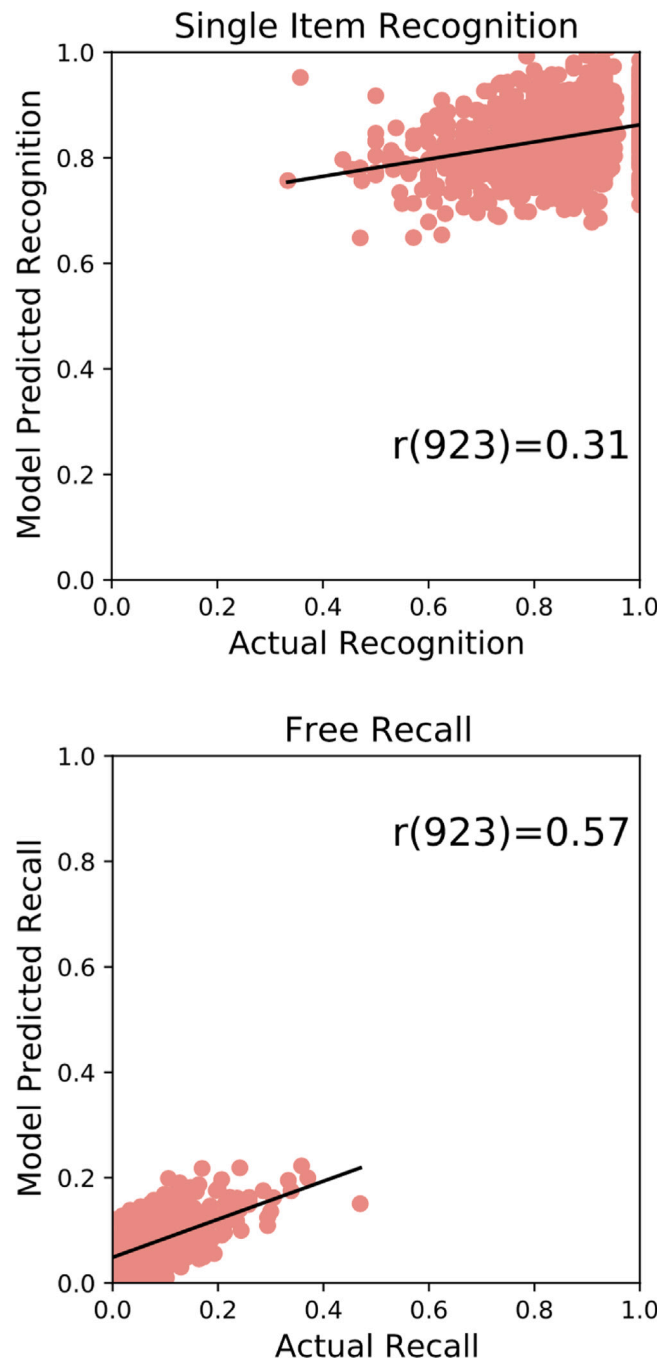

Fig. 2. Actual memorability against predictions (leave-one-out, cross validated) for two additional memory tasks (Cox et al., 2018). The black line denotes the best linear fit.

semantic representation model trained on Cox et al. single recognition data was relatively less reliable when predicting the PEERS recognition data ($r(575) = .08$, $p = .06$, 95% $CI = [.00, .161]$), likely due to the aforementioned task-related differences mentioned above.

In Section 3.4., we also tested generalization on the Cox et al. (2018) data across tasks, since it involves the same participants completing multiple tasks.

3.4. Memorability of semantic categories

Table 1 shows, for all pairs of dichotomous categories in LIWC, GI, and NRC, how the memorability predictions differ, and thus suggests semantic factors that contribute to the memorability of a word. For example, words with a negative sentiment, as coded by the NRC lexicon, are more likely to be recognized and recalled compared to words with a positive sentiment.

To assess the influence of different semantic categories on memory, we regressed, for each word in the LIWC and GI lexicons belonging to only a single higher-level semantic category, the predictions of the semantic representation model on category membership. Category membership of LIWC higher-level semantic categories predicted recall ($R^2 = .22$, $F(13, 3571) = 82$, $p < .001$) and recognition ($R^2 = .17$, $F(13, 3571) = 57$, $p < .001$). Table 2 shows average predicted recognition and recall by semantic category, with categories compared using Tukey-Kramer HSD to account for multiple comparisons and unequal sample sizes. For example, semantic categories such as 'Informal language' and 'Death' are memorable in terms of both recognition and recall. The Supplementary Materials describes robustness checks that do not assume single-category membership, and that use LIWC's subsidiary categories and higher-level GI categories (Table SM-6 and Table SM-7).

We used existing lexicons to study the memorability of different semantic categories by analyzing the predictions of our semantic representation model. The model can also be applied to other words, including those coded along various dimensions. We will later discuss the advantages of this, but, as one example, we applied the model to a humor norms dataset (Engelthaler & Hills, 2018) (see Supplementary Materials) and found a moderate positive correlation between humor and predicted recognition ($r(4988) = .33$, $p < .001$, 95% $CI = [.305, .355]$) and recall ($r(4988) = .15$, $p < .001$, 95% $CI = [.123, .177]$).

3.4.1. Recognition versus recall

An important question is the relationship between memorable semantic categories for recognition versus recall, given that the low correlation between recognition and recall (Spearman's $\rho(575) = .09$, $p = .04$, 95% $CI = [.004, .166]$) may suggest different semantic categories at play. As Table 2 indicates, however, the semantic categories predicted to be highly recognizable are similar to those predicted to be highly recallable. There is a high correlation between recognition and recall predictions across both the LIWC high level psychological categories ($\rho(13) = .57$, $p < .001$, 95% $CI = [.057, .845]$) and the LIWC low level psychological categories ($\rho(39) = .75$, $p < .001$, 95% $CI = [.572, .86]$). Words in the LIWC lexicon do not simply have similar recognition and recall predictions: while the correlation between recognition and recall across individual words in LIWC is larger than in the PEERS dataset, it is considerably smaller ($\rho(13880) = .30$, $p < .001$, 95% $CI = [.285, .315]$) than the correlations on semantic categories.

The high correlation across semantic categories between recognition and recall does not preclude a low correlation across words between recognition and recall, if other factors have a different relationship with recognition and recall. To illustrate how other factors differentially relate to recognition and recall, Table 3 compares standardized beta coefficients from a multiple regression (using ordinary least squares to obtain unbiased parameter estimates) on word features for recognition ($R^2 = .16$, $F(6, 569) = 17.76$, $p < .001$) and recall ($R^2 = .25$, $F(6, 569) = 31.26$, $p < .001$). For example, consistent with previous work (Glanzer & Bowles, 1976; Hall, 1954), the effect of word frequency is negative for recognition and positive for recall.

Since the Cox et al. dataset involves the same set of participants completing both the single recognition and free recall tasks, we also examined this dataset. Here, the correlation between the two tasks (Spearman's $\rho(923) = .22$, $p < .001$, 95% $CI = [0.161, 0.283]$) was much higher compared to the PEERS dataset. Thus, we expected the semantic representation model we had trained on Cox et al. free recall data to better predict the Cox et al. single recognition data and vice versa. As hypothesized, this resulted in accurate predictions ($r(923) = .30$, $p < .001$, 95% $CI = [0.238, 0.356]$ and $r(923) = .37$, $p < .001$, 95% $CI = [0.313, 0.424]$ respectively), providing additional evidence that the model generalizes. We report the predictive accuracy from all of these models in Table SM-2.

Table 1
Comparisons of predicted memorability for all pairs of dichotomous categories in LIWC, GI, and NRC. These comparisons show, for each dichotomous category, how the recognition and recall predictions differ from one another, and shed light on the semantic determinants that contribute to the memorability of a word.

Dichotomy	Lexicon	Recall	Recognition
Valence Negative vs. Positive	GI	Negative valence higher ($t(2301) = 2.98,\ p = .029$)	Negative valence higher ($t(2301) = 7.65,\ p < .001$)
	NRC	No significant difference ($t(5544) = 0.35,\ p = .71$)	Negative valence higher ($t(5544) = 10.01,\ p < .001$)
Sentiment Negative vs. Positive	NRC	Negative sentiment higher ($t(2336) = 3.36,\ p = .0008$)	Negative sentiment higher ($t(2336) = 10.65,\ p < .001$)
	LIWC	No significant difference ($t(1299) = 1.96,\ p = .0503$)	No significant difference ($t(1299) = 1.83,\ p = .0681$)
Pain vs. Pleasure	GI	Pain higher ($t(236) = 2.03,\ p = .0439$)	Pain higher ($t(236) = 2.70,\ p = .0074$)
Vice vs. Virtue	GI	No significant difference ($t(896) = 1.76,\ p = .0792$)	Vice higher ($t(896) = 6.57,\ p < .001$)
Osgood dimension Weak vs. Strong	GI	Weak higher ($t(1562) = 3.97,\ p < .001$)	Weak higher ($t(1562) = 6.27,\ p < .001$)
Osgood dimension Passive vs. Active	GI	Passive higher ($t(1552) = 4.31,\ p < .001$)	Passive higher ($t(1552) = 5.61,\ p < .001$)
Estimation words Under vs. Over	GI	Underestimation higher ($t(591) = 2.87,\ p = .0042$)	Underestimation higher ($t(591) = 2.01,\ p < .001$)

Table 2
Recognition and recall predictions of the semantic representation model across LIWC higher-level psychological categories. LIWC categories that do not share a letter in the comparisons column are significantly different (at the 0.05 level) by a Tukey-Kramer HSD test. Also shown are the average recognition and recall predictions for each category.

Recognition			Recall		
LIWC	Comparisons	Mn.	LIWC	Comparisons	Mn.
Informal Language (damn, hm)	A	0.94	Social Processes (mate, talk)	A	0.53
Death (bury, kill)	B	0.92	Informal Language (damn, hm)	A B	0.52
Affective Processes (happy, cried)	B	0.92	Religion (altar, church)	B C	0.51
Biological Processes (eat, blood)	B	0.92	Death (bury, kill)	C D E	0.50
Religion (altar, church)	B C	0.92	Home (kitchen, landlord)	D E F	0.50
Perceptual Processes (look, heard)	B C	0.91	Biological Processes (eat, blood)	D	0.49
Drives (superior, success)	C D	0.91	Drives (superior, success)	D E	0.49
Social Processes (mate, talk)	C D E	0.91	Leisure (cook, chat)	E F G	0.48
Leisure (cook, chat)	D E F	0.91	Work (job, majors)	E F G	0.48
Money (cash, audit)	D E F G	0.90	Affective Processes (happy, cried)	F G	0.48
Time orientations (will, is)	F G	0.90	Perceptual Processes (look, heard)	G H	0.48
Home (kitchen, landlord)	E F G	0.90	Time orientations (will, is)	G H	0.48
Cognitive Processes (cause, know)	F G	0.90	Money (cash, audit)	H I	0.47
Work (job, majors)	G	0.90	Cognitive Processes (cause, know)	I	0.46

Table 3
A comparison of standardized beta coefficients from multiple regressions of recognition and recall on word features using the PEERS datasets.

	Recognition		Recall	
	Standardized beta	95% CI	Standardized beta	95% CI
Word frequency	−0.35	[−0.426, −0.267]	0.17	[0.096, 0.247]
Animacy	0.00	[−0.079, 0.078]	0.4	[0.324, 0.472]
Valence	−0.12	[−0.199, −0.042]	0.08	[0.004, 0.153]
Concreteness	0.03	[−0.048, 0.110]	0.17	[0.094, 0.243]
Arousal	0.10	[0.018, 0.174]	0.16	[0.085, 0.233]
Word length	−0.05	[−0.131, 0.023]	−0.01	[−0.086, 0.060]

4. Discussion

We have shown that high-dimensional distributed semantic representations provide more accurate out-of-sample predictions of word recognition and recall than psycholinguistic word features in terms of correlations, mean squared errors, and unique variance explained. It has been suggested that psychology should pay more attention to prediction, relative to explanation, than it has historically (Yarkoni & Westfall, 2017), and an advantage of our semantic representation model is that it makes precise predictions for words outside of our original dataset. Model predictions can be used to gain insight into the psychological phenomena being predicted (Agrawal et al., 2020; Peterson et al., 2021), for example, in this case, to examine the role of particular semantic categories across recognition and recall.

Model predictions also allow one to generate hypotheses, provide evidence for existing hypotheses, and assist stimuli selection. To illustrate one such use, we applied our semantic representation model to words normed for humor, and found a moderate correlation between humor and both recognition and recall. Our results provide an additional source of evidence for conclusions from previous work on humor and memorability (Cline & Kellaris, 2007; Summerfelt, Lippman, & Hyman, 2010). If this previous work did not exist, our results would suggest a new hypothesis to test. We have released our pre-trained recognition and recall memorability models and provide instructions on using them to obtain predictions. We hope that these models have the potential to benefit other future research projects.

By using the semantic representation model to predict the memorability of words in established lexicons, we identified semantic dimensions that help explain which words are more memorable. For example, consistent with a pattern of results in psychology that "bad is stronger than good" (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), recognition and recall predictions tended to be higher for negative words than positive words along various dimensions. Some semantic categories we identified as high in recognition and recall, such as words relating to 'Death', accord with previous work (Bugaiska et al., 2015), and provide a new kind of evidence for hypotheses about the memorability of survival information. Other relationships we identified have not been previously discussed, for example informal words have high predicted recognition and recall, and words relating to 'Cognitive Processes' have low predicted recognition and recall.

The most important result of our analysis of semantic categories is that while the correlation across words between recognition and recall is low, both in our datasets and as discussed in previous work (Brown, 1976; Myers, 1914; Tversky, 1974), we showed that the more and less memorable semantic categories are consistent across recognition and recall. These findings provide additional empirical support for work demonstrating the consistency of influential semantic information across recall and recognition (Cox et al., 2018) and extend it by providing information about the specific semantic categories that influence memorability. This result suggests future work to investigate the differences, and common mechanisms, of recognition and recall. For example, one could use our semantic representation model to select stimuli for recognition and recall tasks that we predict will result in large differences, and examine the effects of various retrieval cues on memorability of recognition and recall.

We had people predict word memorability as an additional baseline for the semantic representation models, but our human prediction experiments also inform a debate concerning metamemory. The metamemory literature has focused on people predicting their own memory performance, rather than that of others, and contains mixed results on peoples' abilities to make such predictions. Some studies demonstrate how people make above-chance accuracy predictions for subsequent memory performance (Lovelace & Marsh, 1985; Mazzoni, Cornoldi, & Marchitelli, 1990; Nelson, 1988), whereas others show incorrect predictions (Devolder, Brigham, & Pressley, 1990; Wixted, 1992) due to factors including overweighting ease of processing (Kornell, Rhodes, Castel, & Tauber, 2011; Rhodes & Castel, 2008). In our studies, which use a much larger set of words than is typical in metamemory studies, the correlation between actual memorability and the average prediction of human participants was small for recognition, and medium sized for recall, and less than half that of the semantic representation model. This suggests, consistent with the second set of studies above, that people may have relatively little metacognitive sense of what is memorable. People may over or underestimate the role of psycholinguistic features such as word frequency, or may have inaccurate beliefs about the influence of word meaning on memorability. As an additional alternative, people may also be considering words in isolation when predicting their memorability. However, actual memorability performance depends on both the quality of individual memory representations and their ability to be distinguished from other interfering representations and contextual factors.[4] Future research could explore these possibilities to gain a better understanding of metacognitive processes related to memorability.

Our analysis of the semantic representation model indicates that it explains roughly 25% of the variance in recognition, and 49% in recall, compared to split-half reliabilities for recognition (47% of the variance), and recall (84% of the variance). This comparison makes clear that while our semantic representation model explains a great deal of variance for both recognition and recall, there are, unsurprisingly, many additional determinants of memorability that we do not expect it to capture. For example, fluctuations in attentional states or other encoding and retrieval related dynamics may complement semantic determinants to account for variability in memorability.

As discussed above, the semantic representation model demonstrated a higher predictive accuracy for recall than for recognition. There are multiple possible reasons this may be the case. First, while the ranges of word recognition and recall probabilities we used are typical of the literature, recall probabilities cover a larger range than recognition (recall: 34% range, recognition: 22% range). It is also possible that recall processes may simply be less noisy and more predictable across individuals, or that semantic information is more reliable for predicting recall than recognition. Exploring these possibilities would potentially make for interesting future work.

In his pioneering study of memory, Ebbinghaus used nonsense syllables for their "very lack of meaning" (Ebbinghaus, 1885), but demonstrated the complexity of even such syllables by comparing his difficulty memorizing them to his difficulty memorizing cantos of Byron's "Don Juan". While we have begun to model how meaning influences memorability, much more progress is required to tackle Byron's satire.

CRediT authorship contribution statement

Ada Aka: Developed the study concept, Study design, Testing, Data curation, Initial data analyses, Data analysis and interpretation, Drafted the manuscript. **Sudeep Bhatia:** Study design, Data analysis and interpretation, Provided critical revisions. **John McCoy:** Study design, Data analysis and interpretation, Provided critical revisions.

Data availability

Our data and analysis code is available: https://osf.io/7phj6.

Acknowledgment

All authors approved the version of the manuscript to be published.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105497.

References

Agrawal, M., Peterson, J. C., & Griffiths, T. L. (2020). Scaling up psychology via scientific regret minimization. *Proceedings of the National Academy of Sciences*, *117*(16), 8825–8835.

Aka, A., Phan, T. D., & Kahana, M. J. (2020). Predicting recall of words and lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97.

Bainbridge, W. A. (2017). The memorability of people: Intrinsic memorability across transformations of a person's face. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 706.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370.

Bellezza, F. S. (1981). Mnemonic devices: Classification, characteristics, and criteria. *Review of Educational Research*, *51*(2), 247–275.

Bhatia, S., Richie, R., & Zou, W. (2019). Distributed semantic representations for modeling human judgment. *Current Opinion in Behavioral Sciences*, *29*, 31–36.

[4] We thank our reviewer for suggesting this alternative possibility.

Bonin, P., Thiebaut, G., Witt, A., & Méot, A. (2019). Contamination is "good" for your memory! further evidence for the adaptive view of memory. *Evolutionary Psychological Science*, *5*(3), 300–316.

Bower, G. (1967). A multicomponent theory of the memory trace. In *Psychology of learning and motivation, Vol. 1* (pp. 229–325). Elsevier.

Bower, G. H. (1970). Organizational factors in memory. *Cognitive Psychology*, *1*(1), 18–46.

Brown, J. (1976). An analysis of recognition and recall and of problems in their comparison.

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*(4), 977–990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods*, *46*(3), 904–911. http://dx.doi.org/10.3758/s13428-013-0403-5.

Buchanan, T. W., Etzel, J. A., Adolphs, R., & Tranel, D. (2006). The influence of autonomic arousal and semantic relatedness on memory for emotional words. *International Journal of Psychophysiology*, *61*(1), 26–33.

Bugaiska, A., Mermillod, M., & Bonin, P. (2015). Does the thought of death contribute to the memory benefit of encoding with a survival scenario? *Memory*, *23*(2), 213–232.

Cahill, L., & McGaugh, J. L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition*, *4*(4), 410–421.

Cline, T. W., & Kellaris, J. J. (2007). The influence of humor strength and humor—message relatedness on ad memorability: A dual process model. *Journal of Advertising*, *36*(1), 55–67.

Cox, G. E., & Criss, A. H. (2020). Similarity leads to correlated processing: A dynamic model of encoding and recognition of episodic associations. *Psychological Review*, *127*(5), 792.

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*(4), 545.

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

Devolder, P. A., Brigham, M. C., & Pressley, M. (1990). Memory performance awareness in younger and older adults. *Psychology and Aging*, *5*(2), 291.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. Teachers College Press, Translated by Henry A. Ruger & Clara E. Bussenius (1913).

Engelthaler, T., & Hills, T. T. (2018). Humor norms for 4,997 english words. *Behavior Research Methods*, *50*(3), 1116–1124.

Floyd, S., Dalawella, K., Goldberg, A., Lew-Williams, C., & Griffiths, T. (2021). Modeling rules and similarity in colexification. In *Proceedings of the annual meeting of the cognitive science society, Vol. 43*.

Ghatala, E. S. (1981). Imagery and associative processing effects on incidental memory. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(1), 47.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(1), 21.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior*, *23*(5), 553–568.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033.

Hall, J. F. (1954). Learning as a function of word-frequency. *The American Journal of Psychology*, *67*(1), 138–140. http://dx.doi.org/10.2307/1418830.

Healey, M. K., Crutchley, P., & Kahana, M. J. (2014). Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology: General*, *143*(4), 1553–1569. http://dx.doi.org/10.1037/a0036306.

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, *123*(1), 23.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96–101.

Hovhannisyan, M., Clarke, A., Geib, B. R., Cicchinelli, R., Monge, Z., Worth, T., et al. (2021). The visual and semantic features that predict object memory: Concept property norms for 1,000 object images. *Memory & Cognition*, *49*, 712–731.

Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

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

Humphreys, M. S., Li, Y. R., Burt, J. S., & Loft, S. (2020). How semantic processing affects recognition memory. *Journal of Memory and Language*, *113*, Article 104109.

Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011). *Understanding the intrinsic memorability of images*: *Tech. Rep.*, Massachusetts Inst of Tech Cambridge.

Jones, M. N., Willits, J., Dennis, S., & Jones, M. (2015). Models of semantic memory. In *Oxford handbook of mathematical and computational psychology* (pp. 232–254). Oxford University Press New York.

Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology*, *71*, 107.

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*(12), 1857.

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, *31*(8), 1169–1180.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review*, *114*(4), 954.

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*(6), 787–794.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211.

Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, *4*, 151–171.

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*(6), 1943.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, *122*(2), 337–363. http://dx.doi.org/10.1037/a0039036.

Lovelace, E. A., & Marsh, G. R. (1985). Prediction and evaluation of memory performance by young and old adults. *Journal of Gerontology*, *40*(2), 192–197.

Madan, C. R. (2021). Exploring word memorability: How well do different word properties explain item free-recall probability? *Psychonomic Bulletin & Review*, *28*(2), 583–595.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: a review and empirical validation. *Journal of Memory and Language*, *92*, 57–78.

Mandler, G. (1968). Organized recall: individual functions. *Psychonomic Science*, *13*(4), 235–236.

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition*, *18*(2), 196–204.

Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, *20*(2), 161–189.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. arXiv preprint arXiv:1310.4546.

Mohammad, S. M., & Turney, P. D. (2013). *Nrc emotion lexicon, Vol. 2*. Canada: National Research Council.

Morton, N. W., & Polyn, S. M. (2016). A predictive framework for evaluating models of semantic organization in free recall. *Journal of Memory and Language*, *86*, 119–140.

Murdock, B. B. (1995). Human memory in the twenty-first century. *The Science of the Mind*, 109–122.

Myers, G. C. (1914). A comparative study of recognition and recall. *Psychological Review*, *21*(6), 442.

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. (2007). Adaptive memory: survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 263.

Nelson, T. O. (1988). Predictive accuracy of the feeling of knowing across different criterion tasks and across different subject populations and individuals.

Nelson, D. L., Kitto, K., Galea, D., McEvoy, C. L., & Bruza, P. D. (2013). How activation, entanglement, and searching a semantic network contribute to event memory. *Memory & Cognition*, *41*(6), 797–819.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483.

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and why? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1196.

Pandeirada, J. N., Fernandes, N. L., Vasconcelos, M., & Nairne, J. S. (2017). Adaptive memory: remembering potential mates. *Evolutionary Psychology*, *15*(4), Article 1474704917742807.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, *71*(2001), 2001.

Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D., & Griffiths, T. L. (2021). Using large-scale experiments and machine learning to discover theories of human decision-making. *Science*, *372*(6547), 1209–1214.

Peterson, J. C., Chen, D., & Griffiths, T. L. (2020). Parallelograms revisited: Exploring the limitations of vector space models for simple analogies. *Cognition*, *205*, Article 104440.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129.

Poon, L. W., & Fozard, J. L. (1980). Age and word frequency effects in continuous recognition memory. *Journal of Gerontology*, *35*(1), 77–86.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*(4), 615.

Richie, R., Aka, A., & Bhatia, S. (2022). Free association in a neural network. *Psychological Review*.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science*, *4*(1), 28–34.

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*(1), 79–94.

Schwanenflugel, P. J., Akin, C., & Luh, W.-M. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, *20*(1), 96–104.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166.

Sirotin, Y. B., Kimball, D. R., & Kahana, M. J. (2005). Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin & Review*, *12*(5), 787–805.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2005). Word association spaces for predicting semantic similarity effects in episodic memory.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The general inquirer: A computer approach to content analysis.

Summerfelt, H., Lippman, L., & Hyman, I. E., Jr. (2010). The effect of humor on memory: Constrained by the pun. *The Journal of General Psychology*, *137*(4), 376–394.

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.

Tversky, B. (1974). Eye fixations in prediction of recognition and recall. *Memory & Cognition*, *2*(2), 275–278.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, *76*(6), 559.

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N., & Blunt, J. R. (2013). Adaptive memory: Animacy processing produces mnemonic advantages. *Experimental Psychology*, *60*(3), 172–178. http://dx.doi.org/10.1027/1618-3169/a000186.

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

Weidemann, C. T., & Kahana, M. J. (2016). Assessing recognition memory using confidence ratings and response times. *Royal Society Open Science*, *3*(4), Article 150670.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 681.

Xie, W., Bainbridge, W. A., Inati, S. K., Baker, C. I., & Zaghloul, K. A. (2020). Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe. *Nature Human Behaviour*, *4*(9), 937–948.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517.

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., et al. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 792.