



# How does attribute ambiguity improve memory?

C. J. Brainerd<sup>1</sup> · M. Chang<sup>1</sup> · D. M. Bialer<sup>1</sup> · X. Liu<sup>1</sup>

Accepted: 30 June 2022 / Published online: 26 July 2022  
© The Psychonomic Society, Inc. 2022

## Abstract

The memory effects of semantic attributes (e.g., concreteness, familiarity, valence) have long been studied by manipulating their average perceived intensities, as quantified in word rating norms. The semantic ambiguity hypothesis specifies that the uncertainty as well as the intensity of semantic attributes is processed when words are encoded. Testing that hypothesis requires a normed measure of ambiguity, so that ambiguity and intensity can be manipulated independently. The standard deviation (*SD*) of intensity ratings has been used for that purpose, which has produced three characteristic ambiguity effects. Owing to the recency of such research, fundamental questions remain about the validity of this method of measuring ambiguity and about its process-level effects on memory. In a validity experiment, we found that the rating *SDs* of six semantic attributes (arousal, concreteness, familiarity, meaningfulness, negative valence, positive valence) passed tests of concurrent and predictive validity. In three memory experiments, we found that manipulating rating *SDs* had a specific effect on retrieval: It influenced subjects' ability to use reconstructive retrieval to recall words. That pattern was predicted by the current theoretical explanation of how ambiguity benefits memory.

**Keywords** Semantic attributes · Attribute ambiguity · Fuzzy-trace theory · Dual-retrieval model · Reconstructive retrieval

## Introduction

The semantic ambiguity hypothesis is the proposal that (a) uncertainty in the perceived intensity of common semantic attributes (e.g., concreteness, familiarity, valence) affects how they are processed as words are encoded, and (b) those processing differences have downstream effects on episodic memory. To study this hypothesis, a standardized method of measuring and manipulating attribute ambiguity is needed. Normed variability in the perceived intensity of semantic attributes has been used for that purpose in recent experimentation. The primary objective of the research that we report here was to resolve two unanswered questions about this method. First, is it valid – that is, is there any independent evidence that it taps the uncertainty of those perceptions? Second, if ambiguity-driven differences in encoding have downstream memory effects, which specific retrieval processes capture those effects?

## The semantic ambiguity hypothesis

Published word norms for semantic attributes have been used to study the semantic ambiguity hypothesis. Such norms are widely used tools in the larger memory and cognition literature. That literature contains several such norms, in which thousands of words have been judged for the perceived intensity of specific attributes – with judgments usually being made on unimodal (lower → higher) numerical scales. Published rating norms provide the means (*Ms*) and standard deviations (*SDs*) of intensity ratings of individual words. Sixteen examples are displayed in Table 1, along with an illustrative norming project for each attribute.

It has often been found that varying the *M* intensity ratings of semantic attributes affects the content that people process when they encode words (e.g., Einstein & McDaniel, 1990; Gomes et al., 2013; Henriques & Davidson, 2000; Kounios & Holcomb, 1994), as well as the accuracy of subsequent memory for those words (for reviews, see Bookbinder & Brainerd, 2016; Juhasz, 2005; Kensinger, 2009; Marschark et al., 1987). The semantic ambiguity hypothesis posits that varying the *SDs* of those same ratings also affects the content that people process when they encode words

✉ C. J. Brainerd  
cb299@cornell.edu

<sup>1</sup> Department of Psychology and Human Neuroscience  
Institute, Cornell University, Ithaca, NY, USA

**Table 1** Common semantic attributes of memory items and illustrative norming projects

Attribute	Definition and illustrative norming projects
Age-of-acquisition	When did you first learn a word, between early childhood to adolescence (Schock et al., 2012)
Anger	How angry does a word make you feel, from not at all to strongly (Stadthagen-González et al., 2018)
Body-object interaction	How easily can the human body physically interact with a word's referents, from not at all to easily (Pexman et al., 2019)
Categorization	How well does a word fit in some larger taxonomic category or class, from not at all to easily (Toglia & Battig, 1978)
Concreteness	To what extent does a word refer to real objects, persons, places, or things that can be seen, heard, felt, smelled, or tasted, ranging from not all to definitely (Paivio et al., 1968)
Disgust	How disgusted does a word make you feel, from not at all to strongly (Ferré et al., 2017)
Familiarity	How often has a word been previously experienced, from rarely to very often (Juhász et al., 2015)
Fear	To what extent does a word stimulate feelings of fear, from not at all to strongly (Ferré et al., 2017)
Happiness	How happy does a word make you feel, from not at all to strongly (Hinojosa et al., 2016)
Humor	To what extent does a word stimulate feelings of amusement, from clearly unfunny to very funny (Engelthaler & Hills, 2018)
Imagery	To what extent does a word stimulate visual or auditory images of things or events, from not at all to easily (Paivio et al., 1968)
Meaningfulness	To what extent does a word retrieve associated words, from not at all to easily (Underwood & Schulz, 1960)
Number of features	How many specific features does a word retrieve, from few or none to many (Toglia & Battig, 1978)
Pleasantness	To what extent does a word stimulate feelings that range from very unpleasant to very pleasant (Toglia & Battig, 1978)
Sadness	How sad does a word make you feel, from not at all to strongly (Stevenson et al., 2007)
Semantic size	How large is a word's physical or symbolic referent, from small to large (Scott et al., 2019)

and the accuracy of subsequent memory (Brainerd, 2018; Brainerd et al., 2021a).

Empirically, we know that not only does an attribute's perceived intensity differ for different words [e.g., *mother* has a stronger positive valence than *father*,  $M_s = 8.38$  and  $7.08$ , respectively, in the Bradley & Lang, 1999 norms], but so does the degree of uncertainty that attaches to those perceptions – even among words whose perceived intensities are the same on average. For instance, the  $M$  valence ratings of *freedom* and *talent* are the same ( $7.6$  in the Bradley-Lang norms), but people disagree far more in their valence perceptions of *freedom* than their perceptions of *talent* ( $SD_s = 2.0$  and  $1.3$ , respectively). The hypothesis that such differences affect how people process semantic attributes during encoding is counterintuitive from the perspective of the traditional idea (e.g., Toglia & Battig, 1978) that rating  $SD_s$  are merely indexes of measurement error (unreliability in the measurement of perceived intensity).

Nevertheless, some early support for this hypothesis can be found in research by Mattek et al. (2017) on perceived valence. Mattek et al. evaluated the prediction that if ambiguity affects how valence is encoded, correlations between  $M$  intensity ratings of valence and arousal, which are expected to be high on theoretical grounds (see Kuppens et al., 2013), will be low when valence ambiguity is high. They found that, indeed, such correlations varied dramatically for items whose valence was uncertain versus items that were clearly positive or negative. Brainerd (2018) then used valence rating  $SD_s$  to measure graded changes in the ambiguity of

perceived valence. In emotional word and picture norms, he found that the strength of such valence-arousal correlations declined linearly as rating  $SD_s$  increased.

The ambiguity hypothesis has been extended by generating further predictions about valence ambiguity and showing that those same predictions hold for other semantic attributes. Two such predictions about valence ambiguity were discussed by Brainerd, et al., (2021a) and then extended to other attributes by Brainerd et al. (2021b). The predictions were that (a) words' rating  $SD_s$  should be inverted-U functions of their  $M$  intensity ratings for an attribute, and (b) words with higher rating  $SD_s$  should be easier to remember in recall experiments. Both predictions follow from fuzzy-trace theory's (FTT) account of how subjects perceive the intensity of semantic attributes (e.g., Brainerd et al., 2017). Concerning the first, FTT posits that ambiguity is low when a word retrieves few attribute features (low intensity) or many features (high intensity) but is high when it retrieves a moderate number (intermediate intensity). Concerning the other prediction, FTT posits that when manipulations make semantic attributes more ambiguous, controlled processing increases, which yields richer and more detailed representations that redound to the benefit of later recall (see also, Bugg, 2014).

Brainerd et al. (2021a) confirmed both predictions for valence, and Brainerd et al. (2021a) confirmed them for some other semantic attributes. In the latter research, it was found that all 16 attributes in Table 1 exhibited inverted-U relations between  $M$  intensity ratings and rating  $SD_s$ . Those

results included various controls that ruled out the possibility that the inverted Us could be artifacts of scale boundary effects (i.e., that variance at the low and high endpoints can only go in one direction, back toward the mean, and hence must be lower than for other scale numbers). Three examples appear in Fig. 1. It was also found that words with higher *SDs* yielded better recall than words with lower *SDs* when the *M* intensity ratings and rating *SDs* of selected attributes were manipulated factorially.

## The present experiments

The current situation with respect to the ambiguity hypothesis is this. On the one hand, rating *SDs* are known to be

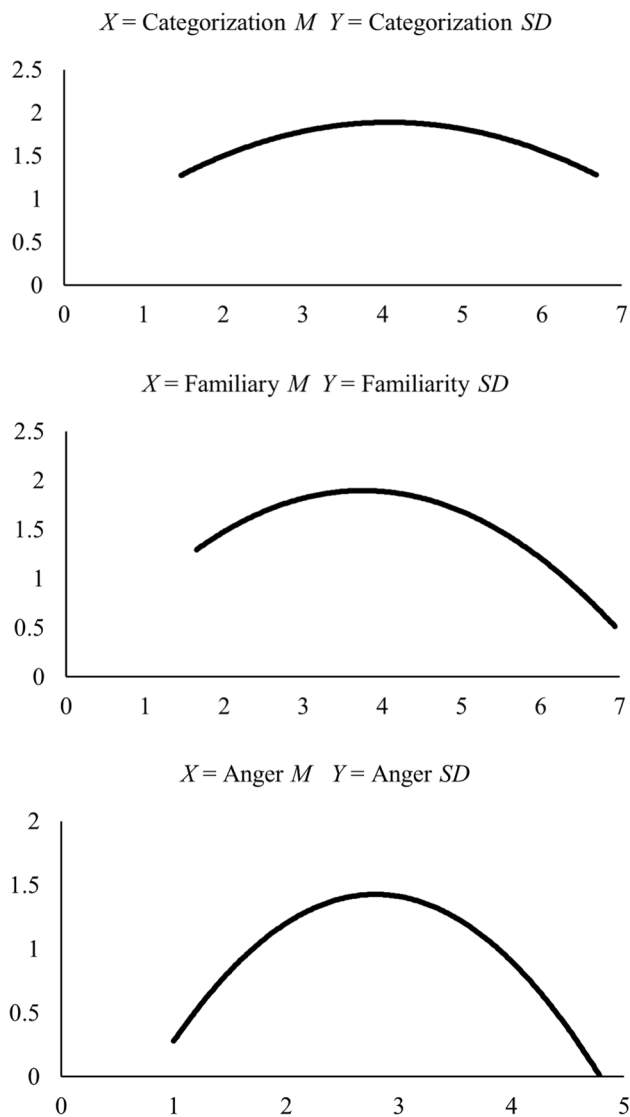
psychometrically reliable indexes and three predicted patterns have been detected with them – namely, that inter-attribute correlations of *M* intensity ratings vary inversely with the size of rating *SDs*, that rating *SDs* for individual attributes are inverted-U functions of *M* intensity ratings, and that recall is more accurate after encoding words with higher rating *SDs*. On the other hand, two fundamental questions remain about rating *SDs* as indexes of attribute ambiguity. The first is validity – whether attribute ambiguity is actually what rating *SDs* measure. That question is the focus of [Experiment 1](#). The other question is exactly how, at the level of specific retrieval processes, higher-*SD* items improve recall. That question is the focus of [Experiments 2–4](#).

## The validity question

The current validity situation is that rating *SDs* have high face validity as an ambiguity index, but no *data* have been reported that establish either concurrent validity or predictive validity. Concerning face validity, if the perceived intensity of an attribute (say, valence) is more ambiguous for one word than another (say, for *freedom* than for *talent*), people should disagree more in their intensity ratings of the first word than the second (Brainerd, 2018). Psychologically, the attribute features that are retrieved by the first word are less determinative than those that are retrieved by the second, leading to more variability in perceived intensity.

That is merely a conceptual argument, however, which must have empirical consequences if it is true – specifically, evidence of concurrent and predictive validity. With respect to concurrent validity, rating *SDs* ought to correlate with other variables that have high face validity as measures of attribute ambiguity. With respect to predictive validity, rating *SDs* ought to correlate with other variables that, according to theory, should be affected by ambiguity. Data on both forms of validity are reported in [Experiment 1](#).

Concurrent and predictive validity tests were conducted at the level of individual words, with data that were averaged across the performance of individual subjects. Concurrent validity was evaluated by having subjects (a) make conventional intensity ratings of six of the attributes in [Table 1](#) and (b) also make numerical ratings of their *certainty* that these intensity ratings were accurate. The concurrent validity tests evaluated whether the *SDs* of individual words' attribute intensity ratings correlated with the *M* certainty of the words' intensity ratings. Turning to predictive validity, we measured the latencies of subjects' attribute intensity ratings of words, as well as the numerical values of the ratings. Because theory posits that increased ambiguity triggers more controlled processing of attribute features, intensity ratings should slow as uncertainty increases. Therefore, the predictive



**Fig. 1** Best-fitting quadratic relations between *M* intensity ratings and rating *SDs* in word norms for categorization (Toglia & Battig, 1978), familiarity (Scott et al., 2019), and anger (Stadthagen-González et al., 2018)

validity tests evaluated whether the *SDs* of individual words' attribute intensity ratings correlated with the *M* latencies of the words' intensity ratings.

### The retrieval process question

Of the three predictions that have been tested for rating *SDs*, the most counterintuitive one is that recall is more accurate when rating *SDs* are higher than when they are lower. That is because *SDs* should have the opposite effect under the traditional assumption that measurement error increases as *SDs* increase, that intensity perceptions with higher *SDs* are inherently noisier than perceptions with lower *SDs*. For instance, Pollock (2018) and Neath and Surprenant (2020) tested the hypothesis that data showing that words with higher *M* concreteness intensity ratings yield better recall than words with lower *M* intensity ratings were *SD* artifacts – specifically, were due to high-concreteness words having lower rating *SDs* than low-concreteness words. Neither study confirmed that hypothesis. In other experiments in which *M* intensity and rating *SDs* were factorially manipulated for categorization, concreteness, meaningfulness, negative valence, and positive valence, recall was better when *SDs* were higher (Brainerd et al. 2021a, b).

The working explanation of the beneficial effects of higher *SDs* on recall is that the encoding of a word's semantic content becomes more controlled and thorough as an attribute's perceived intensity becomes more ambiguous. Brainerd et al. (2021b) noted that this hypothesis makes an obvious prediction that follows from prior research in which controlled processing has been encouraged by various manipulations (e.g., increased encoding time; Smith & Kimball, 2012). A well-known finding is that such manipulations enhance later recall, which means that they must enhance some of the retrieval operations that support recall. Which ones?

Here, we note, first, that a variety of evidence suggests that recall is supported by two distinct retrieval operations, direct verbatim access and reconstruction, plus a slave familiarity process (for a review, see Brainerd et al., 2009). To elaborate, subjects recall words by either (a) using direct access to recover literal verbatim traces of their presentation or (b) reconstructing them from gist traces of their semantic content. In the latter case, items that are recovered via reconstruction are subjected to familiarity checks before they are read out on recall tests. If higher rating *SDs* trigger more thorough encoding of semantic attributes, they ought to enhance recall by improving reconstructive retrieval because reconstruction is focused on semantic content. This prediction was evaluated in Experiments 2–4.

## Experiment 1

This was a norming experiment that was focused on the validity question. We used published norms for arousal, concreteness, familiarity, meaningfulness, negative valence, and positive valence to sample pools of words for these attributes. Half the words in each pool had received high *M* ratings on the target attribute and half had received low *M* ratings. Half the words in the high and low *M* rating groups had high-rating *SDs* and half had low-rating *SDs*. Within individual experimental conditions, subjects rated all the words in three of the six pools for the intensity of their respective attributes on a lower → higher numerical scale.

To test concurrent validity, our subjects also judged the certainty that their intensity ratings of each of the six attributes were accurate on a lower → higher numerical scale. Words' *M* certainty ratings should be inversely related to the *SDs* of the words' intensity ratings, if the latter are valid measures of attribute ambiguity. To evaluate predictive validity, we measured the latencies of our subjects' attribute intensity ratings and certainty ratings. The *M* latency of words' intensity ratings and of words' certainty ratings should both be positively related to the *SDs* of words' intensity ratings, if the latter are valid measures of attribute ambiguity.

## Method

### Subjects

This experiment was conducted over two academic semesters. The subjects were 85 undergraduates who participated to fulfill a course requirement. They received course credits for participation. Individual subjects were randomly assigned to 19 experimental conditions, in which they rated three randomly selected word pools that had been sampled from a larger set of six word pools. The individual words in each word pool were rated by an average of 43 subjects, which is comparable to the numbers of subjects who rated individual words in the Toglia and Battig (1978) norms and 30% larger than the numbers of subjects who rated individual words in the Scott et al. (2019) norms. The complete norming data for all attributes and measures appear in the Appendix.

### Materials

The materials consisted of six pools of 120 words, each of which was rated for the perceived intensity of a target attribute. The words in the first pool were selected from the Scott et al. (2019) arousal norms, and they represented a broad

range of intensities ( $M$ s ranged from 2.5 to 7.1) and ambiguities ( $SD$ s ranged from 1.6 to 3.2). The words in the second pool were selected from the Scott et al. (2019) concreteness norms, and they represented a broad range of intensities ( $M$ s ranged from 2.2 to 6.8) and ambiguities ( $SD$ s ranged from 0.7 to 2.3). The words in the third pool were selected from the Scott et al. (2019) familiarity norms, and they represented a broad range of intensities ( $M$ s ranged from 2.5 to 6.7) and ambiguities ( $SD$ s ranged from 0.8 to 2.4). The words in the fourth pool were selected from the Toglia and Battig (1978) meaningfulness norms, and they represented a broad range of intensities ( $M$ s ranged from 1.4 to 5.1) and ambiguities ( $SD$ s ranged from 1.5 to 2.3). The words in the fifth pool were selected from the Scott et al. (2019) valence norms, and they represented a broad range of negative intensities ( $M$ s ranged from 1.4 to 5) and ambiguities ( $SD$ s ranged from 0.8 to 2.2). The words in the sixth pool were selected from the Scott et al. (2019) valence norms, and they represented a broad range of positive intensities ( $M$ s ranged from 5 to 8.4) and ambiguities ( $SD$ s ranged from .9 to 2.4).

## Procedure

Except for the specific rated attributes, the procedure was the same for all 19 conditions. First, the subject received instructions about the nature of the experiment – that he or she would be rating three pools of words and that each pool would be rated for one of three semantic attributes. Next, the subject received instructions about how to make ratings for one of six attributes (arousal, concreteness, familiarity, meaningfulness, negative valence, or positive valence). The rating instructions for meaningfulness were drawn from the Toglia and Battig (1978) norms, and the rating instructions for the other five attributes were drawn from the Scott et al. (2019) norms.

Following the first set of instructions, the items in the first pool were rated for the target attribute. Each word was centered on the computer screen and printed in 36-pt black font. The subject then rated the words for one of the six attributes. All ratings were self-paced. Two ratings were made for each word. First, the strength of the target attribute was rated (e.g., how arousing the word was, how concrete it was) by clicking a number between 1 and 7 on a scale that appeared below the word (from 1 = extremely low to 7 = extremely high). After the subject chose a number, the strength rating was cleared from the screen. Second, the subject rated his or her certainty that the word's strength rating was accurate by checking a number between 1 and 7 on a lower → higher rating scale (from 1 = not at all certain to 7 = extremely certain). Following the certainty judgment, the screen cleared, and the next word was presented for strength and certainty ratings. The latencies as well as the numerical responses were recorded for both judgments.

After the strength and certainty judgments for the words in the first randomly selected pool, the instructions about how to make ratings for the second attribute were presented. The subject then made strength and certainty ratings of the 120 words in the second pool. After the subject had completed all the words in the second pool, the instructions about how to make ratings for the final attribute were presented, followed by strength and certainty ratings of the 120 words in the third pool.

The complete norms for each word in each of the six pools that were generated by this procedure appear in the Appendix. The norms for the words in the arousal and concreteness pools appear in Appendix Table 8. The norms for the words in the familiarity and meaningfulness pools appear in Appendix Table 9. The norms for the words in the positive and negative valence pools appear in Appendix Table 10.

## Results

In the first subsection below, as a manipulation check, we consider whether the rating  $SD$ s for these six attributes behaved in the manner that rating  $SD$ s in published norms are known to behave. The results that bear on the concurrent and predictive validity of attribute rating  $SD$ s are reported in the second subsection.

### Quadratic relations between $M$ ratings and rating $SD$ s

We saw that rating  $SD$ s in published norms are inverted-U functions of  $M$  intensities; that is, functions of the form  $SD = aM^2 + bM + c$ , whose first derivatives are negative when they are maximized (which occurs when  $a < 0$  and  $b > 0$ ). The results in Table 2 reveal this same pattern in the present experiment. First, note that  $a < 0$  and  $b > 0$  for all six attributes, so that all the functions are inverted Us. Second, note the values of the fit statistics in the  $r^2_Q$ ,  $r^2_L$ , and  $r^2_C$  columns. When rating  $SD$ s are regressed on  $M$  ratings, these are the proportions of variance accounted for by the general quadratic equation ( $r^2_Q$ ), by the model that is one step simpler ( $r^2_L$  for the general linear equation), and by the model that is one step more complex ( $r^2_C$  for the general cubic equation). To conclude that  $SD = f(M)$  is quadratic, it must be found that  $r^2_Q$  is significantly larger than  $r^2_L$ , but that  $r^2_Q$  does not differ significantly from  $r^2_C$ . This is exactly the pattern for the six attributes in Table 2. Over the six attributes, the mean values of  $r^2_Q$ ,  $r^2_L$ , and  $r^2_C$  were .53, .25, and .54, respectively.

As in our prior studies of  $M$ - $SD$  functions, we conducted three supplementary analyses to ensure that the results in Table 2 were not artifacts of certain boundary effects (see Brainerd et al. 2021a) that are possible with numerical rating scales – namely, situations in which the bulk of the ratings pile up at the high and low endpoints (ceiling and floor effects), where variance is lower because it can only move



**Table 2** Tests of the quadratic law for six semantic attributes

Attribute	Parameter estimates and fit						
	<i>a</i>	<i>b</i>	<i>c</i>	<i>Max A</i>	$r_Q^2$	$r_L^2$	$r_C^2$
Arousal	-.17	1.41	-.86	4.15	.60*	.39*	.62*
Concreteness	-.18	1.47	-1.33	4.08	.75*	.30*	.77*
Familiarity	-.14	1.15	-.24	4.11	.87*	.70*	.87*
Meaningfulness	-.07	.54	.69	3.88	.22*	.11*	.23*
Negative valence	-.09	.68	.61	3.78	.45*	.02	.45*
Positive valence	-.12	1	.28	4.17	.31*	.00	.31*

*Note.* The objective midpoint of all rating scales is 4. The subjective midpoint of each scale is the maximum of the first derivative of the quadratic function  $SD = aM^2 + bM + c$ , which is  $Max A = -b/2a$ . Second, concerning parameter values, estimates of *a*, *b*, and *c* appear in the first three columns of Table 1. The parameter *a* measures the steepness of an inverted U, with negative coefficients with larger absolute values producing steeper curves than negative coefficients with smaller absolute values. The parameter *b* measures the location of an inverted U along the *X* axis, with positive values tilting the curve down toward higher values of the *X* and negative values tilting it down toward lower values. The *c* parameter measures the lengths of the two arms of an inverted-U; that is, how far each arm extends down from the curve’s maximum toward the minimum value of *Y*

\**p* < .001

in one direction (back toward the mean). First, we generated the frequency distributions for the intensity ratings of the six attributes. Those distributions produced no evidence of either ceiling or floor effects for any of the attributes. Second, we refit all the linear, quadratic, and cubic functions in Table 2 using a data transformation of the raw *M*s and *SD*s that eliminates boundary effects (arcsine). The refits of the transformed data produced the same patterns as before – for each attribute, the quadratic fit still accounted for significantly more variance than the corresponding linear fit, and the cubic fit still did not account for significantly more variance than the quadratic fit. Finally, we refit the functions in Table 2 using only the data of words whose *M* intensities were well above the low boundary (*M*s ≥ 3) and well below the upper boundary (*M*s < 6). Those refits also produced the same patterns as in Table 2.

**Validity**

First, we report descriptive statistics for three variables that figured in the validity analyses: the *SD*s of the intensity ratings of the six attributes, the mean latencies of the intensity ratings, and the means of the certainty ratings of the intensity ratings. Those results appear in Table 3. Visual inspection reveals differences among the six attributes on each of the three variables, and one-way analyses of variance (ANOVAs) established that the differences were significant for each variable. For rating *SD*s,  $F(5, 595) = 23.25$ ,  $MSE = .07$ ,  $\eta_p^2 = .16$ ,  $p < .001$ , for *M* intensity rating latency,  $F(5, 595) = 132.27$ ,  $MSE = .09$ ,  $\eta_p^2 = .53$ ,  $p < .001$ , and for *M* certainty rating,  $F(5, 595) = 135.71$ ,  $MSE = .137$ ,  $\eta_p^2 = .53$ ,  $p < .001$ . Post hoc tests indicated that the magnitude orderings were: (a) concreteness < familiarity = positive valence = meaningfulness < negative valence = arousal for

rating *SD*s; (b) familiarity < arousal = positive valence < negative valence < meaningfulness = concreteness for *M* rating latencies; and (c) meaningfulness < negative valence = positive valence < arousal = concreteness < familiarity for *M* certainty ratings. It must be remembered that subjects rated different word pools for each attribute. Hence, the differences among the attributes in rating *SD*s, *M* intensity rating latencies, and *M* certainty ratings cannot be interpreted as inherent attribute differences in any of the three variables. Rather, the data in Table 3 should be viewed as descriptive results for each attribute that may suggest psychometric explanations of some of the validity findings.

We report two sets of validity results, one for concurrent validity and the other for predictive validity. The concurrent validity question asks whether the *SD*s of attribute intensity ratings of individual words are correlated with a direct, intuitive measure of the ambiguities of those ratings: the *M* certainty that subjects express about their intensity ratings of each word. The relevant attenuation-corrected validity coefficients appear in the first column of Table 4 for arousal, concreteness, familiarity, meaningfulness, negative valence, and positive

**Table 3** Means of attribute intensity rating standard deviations (*SD*s), attribute intensity rating latencies, and certainty ratings of attribute intensity ratings (numbers in parentheses are *SD*s)

Attribute	<i>M</i> Intensity rating <i>SD</i> latency	<i>M</i> rating latency	<i>M</i> certainty
Arousal	1.80 (.29)	2.09 (.22)	5.70 (.20)
Concreteness	1.49 (.35)	2.51 (.47)	5.74 (.54)
Familiarity	1.57 (.34)	1.60 (.23)	6.45 (.26)
Meaningfulness	1.65 (.18)	2.40 (.35)	5.33 (.33)
Negative valence	1.77 (.18)	2.39 (.27)	5.56 (.44)
Positive valence	1.62 (.19)	2.10 (.20)	5.62 (.27)

**Table 4** Attenuation-corrected concurrent and predictive validity correlations for attribute *M* certainty ratings, *M* attribute rating latencies, and attribute rating *SDs* (raw correlations in parentheses)

Attribute	<i>M</i> Certainty (concurrent)	<i>M</i> Latency (predictive)	
		Rating <i>SD</i>	<i>M</i> Certainty
Arousal	-.25*(-.20)	.23*(.10)	-.42*(-.14)
Concreteness	-.99**(-.64)	.88**(.55)	-.99**(-.72)
Familiarity	-.99**(-.79)	.99**(.72)	-.99**(-.74)
Meaningfulness	-.57**(-.30)		
Negative Valence	-.58**(-.22)	.69**(.22)	-.99**(-.54)
Positive Valence	-.56(-.30)	.29*(-.22)	-.44**(-.20)

*Note.* Validity coefficients were computed with the equation  $\rho_{XY} = r_{XY} \div [\sqrt{(r_{XX} \times r_{YY})}]$ , where  $r_{XY}$  is the raw sample correlation,  $r_{XX}$  is the estimated reliability of the *X* variable, and  $r_{YY}$  is the estimated reliability of the *Y* variable. For the variables in this table, the reliability estimates ranged from .49 to .90, except for meaningfulness rating latencies. As the reliability values for intensity rating *SDs*, we used the means of two reliability coefficients that had been computed for each attribute (a split-half coefficient and a parallel-forms coefficient). As the reliability values for *M* intensity rating latencies, we computed modified split-half reliabilities, using the same procedure that generated the split-half coefficients for *SDs*. All the reliability coefficients were significant, except for meaningfulness. Hence, predictive validity tests were not computed for that variable. As the reliability values for *M* certainty ratings, we computed modified split-half reliabilities, using the same procedure that generated the split-half coefficients for *SDs*. Any validity correlation whose absolute value is  $\geq .16$  is significant beyond the .05 level

\* $p < .003$

\*\* $p < .001$

valence. To begin, consider the coefficients for concreteness and familiarity. For these attributes, both concurrent validity coefficients are virtually perfect, and thus, there was extremely strong support for the hypothesis that the subjective certainty that their intensity ratings are correct declines as those ratings become more variable. For the six attributes as a group, all the concurrent validity coefficients were statistically reliable. The mean value of the six coefficients was .66, which is a moderately strong relation in the language of correlations. Thus, the *SDs* of intensity ratings passed concurrent validity tests because rating *SDs* increased reliably as the judged uncertainty of intensity ratings increased.

Turning to predictive validity, the question is whether the intensity rating *SDs* of individual words forecasts the words' *M* values for other variables that theory expects that they should predict. Here, theory makes a pair of predictions (Brainerd et al., 2021b). The first falls out of the working explanation of why recall is better for words with higher rating *SDs*, which is that such words trigger more thorough and detailed encoding. On the hypothesis that thorough encoding requires more time than superficial encoding, rating *SDs* ought to predict the *M* latencies of intensity ratings. The other prediction is concerned with

the certainty judgments that subjects made about their intensity ratings. If rating *SDs* and *M* certainty ratings are both measuring ambiguity, *M* certainty ratings should also predict the *M* latencies of intensity ratings. The results for the two forms of predictive validity appear in the second and third columns of Table 4.

Taking *SD*-latency first, consider the validity coefficients for concreteness and familiarity –two attributes for which concurrent validity was especially high. Strong positive correlations would be expected, and they were observed, with the average value being .94. For the other four attributes, all the validity coefficients were reliable and positive, with a mean value of .48.

Continuing to *M* certainty ratings, consider the validity coefficients for concreteness and familiarity first, as before. Strong negative correlations would be expected (greater certainty = shorter latencies), and they were observed, with both being nearly perfect. For the other four attributes, the general picture was one of stronger predictive validity correlations for *M* certainty than for rating *SDs*. In fact, all the predictive validity coefficients in the third column are larger than the corresponding ones in the second column, except for familiarity.

A cautionary comment about the predictive validity results is in order. As the coefficients in the third column of Table 4 are larger than those in the second column, it is tempting to entertain theoretical interpretations of that difference, such as positing that certainty judgments somehow tap attribute ambiguity more directly than rating *SDs*. Any such hypothesis would be speculative, however, because our experiments did not include procedures to measure the relative sensitivity of different indexes of attribute ambiguity.

## Experiments 2–4

In prior experimentation, increases in attribute rating *SDs* have produced improvements in recall for multiple attributes when their rating *SDs* and *M* intensity ratings were manipulated factorially. The exact reason is unclear. The working explanation is that increases in ambiguity trigger more thorough, detailed encoding of attribute features (Brainerd, et al., 2021a), which receives direct support from the finding (Table 4) that intensity rating latencies are longer when intensity levels are more uncertain (higher rating *SDs*). To improve recall, however, that encoding effect must translate into downstream effects on specific retrieval processes. Our aim in Experiments 2–4 was to identify those retrieval processes.

We mentioned earlier that because attribute ambiguity is concerned with semantic features, retrieval operations that process semantic content, reconstruction in particular, should be affected by ambiguity-driven enhancements in encoding. To evaluate that possibility, it is necessary to separate reconstructive retrieval from retrieval operations

that process literal verbatim content. Modeling techniques that accomplish this are available for both recognition (see Abadie et al., 2013; Abadie & Camos, 2019; Greene et al., *in press*; Greene & Naveh-Benjamin, 2020, 2021; Lampinen et al., 2005; Lampinen & Odegard, 2006; Niedzialkowska & Nieznański, *in press*; Nieznański & Obidzinski, 2019; Singer & Remillard, 2008; Singer & Spear, 2015) and recall (see Barnhardt et al., 2006; Bouwmeester & Verkoeijen, 2011; Bruer & Pozzulo, 2014; Marche et al., 2016; Palmer & Dodson, 2009; Storbeck & Clore, 2005; Tractenberg et al., 2015).

Returning to the working explanation of ambiguity-driven improvements in recall, the explanation obviously predicts that these improvements should be tied to enhanced reconstructive retrieval. This prediction is easily tested with the dual-retrieval model (Gomes et al., 2014; Tractenberg et al., 2015), which was developed to secure separate estimates of reconstruction, direct verbatim access, and familiarity judgment for recall. It supplies maximum likelihood estimates of the six retrieval parameters in Table 5 for free, cued, and serial recall data.

Two parameters measure successful reconstruction ( $R_1$  and  $R_2$ ), two measure successful verbatim access ( $D_1$  and  $D_2$ ), and two measure successful familiarity judgment ( $J_1$  and  $J_2$ ). In prior experiments with this model, classic semantic manipulations (e.g., deep vs. shallow encoding instructions, large vs. small numbers of category exemplars, subjects with normal vs. impaired meaning comprehension) have affected the reconstruction parameters, whereas classic item distinctiveness manipulations (e.g., full vs. divided attention, shorter vs. longer lists, familiar vs. unusual fonts) have affected the direct access parameters (for a review, see Gomes et al., 2014). Consequently, if semantic content is

encoded more thoroughly as attribute ambiguity increases, the values of one or both reconstruction parameters ought to be larger for words with higher rating *SDs* than for words with lower *SDs*. This effect might spill over to the direct access or familiarity judgment parameters, too, but the key prediction is that one or both reconstruction parameters ought to be affected if the working explanation is correct.

We evaluated this prediction in three experiments in which attribute ambiguity and attribute intensity were manipulated factorially. They followed the multi-trial methodology that has been previously used to identify memory effects of attribute ambiguity. This multi-trial procedure is required to implement the dual-retrieval model. As the model contains six parameters, experiments with three separate trials are the minimum designs that generate sufficient degrees of freedom to estimate all the parameters and compute goodness-of-fit tests (see Brainerd et al., 2009).

In **Experiment 2**, the target attribute was meaningfulness. Subjects learned to recall lists of words on which the items were high or low in normed values of *SDs* for meaningfulness intensity ratings and were high or low in normed values of the *Ms* of those ratings. We located two previously reported experiments of the same type that had produced robust ambiguity effects for two other attributes (categorization and concreteness) but had not been analyzed with the dual-retrieval model. We conducted a retrospective modeling analysis of those studies (**Experiments 3 and 4**) so that any conclusions about the retrieval effects of attribute ambiguity would cut across multiple attributes and be grounded in a large amount of data.

Our focus in **Experiments 2–4** was squarely on using the dual-retrieval model to explain how ambiguity enhances recall by pinpointing its retrieval locus. A secondary objective was to determine whether that locus was the same as the

**Table 5** Retrieval processes that are measured by the dual-retrieval model

Retrieval process/parameter	Definition
Reconstruction:	
$R_1$	For any item whose verbatim trace cannot be accessed on a recall test following the first study cycle, the probability that the item can be reconstructed on that recall test
$R_2$	For any item for whose verbatim trace cannot be accessed on a recall test following any study cycle and that could not be reconstructed on a recall test following the first study cycle, the probability that the item can be reconstructed following any subsequent recall test
Familiarity judgment:	
$J_1$	For any item that is reconstructed on a recall test that follows the first study cycle, the probability that the reconstruction is judged to be familiar enough to output
$J_2$	For any item that is reconstructed on a recall test that follows any study cycle after the first one, the probability that the reconstruction is judged to be familiar enough to output
Direct verbatim access:	
$D_1$	The probability that verbatim traces of an item's presentation on the study list can be accessed on a recall test that follows the first study cycle
$D_2$	The probability that verbatim traces of an item's presentation on a study list can be accessed on a recall test that follows the second or later study cycles, if it could not be accessed following the first study cycle



corresponding one for the effects of attribute intensity. It is sensible that it should be, for two reasons. First, as noted, it has been previously reported for attributes such as concreteness and imagery that variations in attribute intensity affect reconstructive retrieval. Second, for any specific attribute, variations in intensity and ambiguity involve variations in *the same* semantic features.

## Method

### Subjects

The subjects in [Experiment 2](#) were a sample of 72 undergraduates, the subjects in [Experiment 3](#) were an independent sample of 68 undergraduates, and the subjects in [Experiment 4](#) were an independent sample of 83 undergraduates. The experiments were conducted during four semesters over two academic years. The subjects in all experiments participated to fulfill course requirements. Power analyses that relied on data from past experiments with this paradigm (Brainerd, et al., 2021a) indicated that a sample size of 65 would provide a  $> .80$  probability of detecting all main effects and interactions at the .05 level. However, we tested all additional subjects who volunteered for individual experiments – hence, the somewhat different sample sizes.

### Materials

In [Experiment 2](#), the materials were word lists sampled from a pool of words that had been derived from the Toggia and Battig (1978) semantic word norms, in which *meaningfulness* was rated on a 1–7 scale for 2,852 words. The pool of words from which the items on the lists that were administered to individual subjects were sampled was constructed in such a way that their meaningfulness ambiguities varied from high to low and their meaningfulness intensities varied from high to low. Each subject studied and recalled two lists of 36 words (nine high ambiguity/high intensity, nine high ambiguity/low intensity, nine low ambiguity/high intensity, and nine low ambiguity/low intensity). Thus, each list that was administered to individual subjects had a 2 (meaningfulness ambiguity: high vs. low)  $\times$  2 (meaningfulness intensity: high vs. low) factorial structure. Over lists, items' mean intensities were  $M_I = 8.8$  (high) and 4.6 (low), and their mean ambiguities were  $M_A = 4.1$  (high) and 2.7 (low). Each list was composed of 36 words, and the differences between the high and low  $M_{A,S}$  and the high and low  $M_{I,S}$  over the lists were highly reliable (all  $ps < .0001$  by  $t$  tests). The differences in  $M_{A,S}$  and  $M_{I,S}$  were uncorrelated; that is, the high and low  $M_A$  conditions did not differ reliably in  $M_I$ , and conversely.

In [Experiment 3](#), the word lists that were administered to individual subjects were constructed in the same way, using the Toggia and Battig (1978) norms, except that the word pool

from which they were sampled was derived from subjects' ratings of *concreteness*. Thus, each subject studied and recalled two lists of 36 words, for which the list design was 2 (concreteness ambiguity: high vs. low)  $\times$  2 (concreteness intensity: high vs. low). Over lists, the mean intensities were  $M_I = 5.3$  (high) and 3.1 (low), and the mean ambiguities were  $M_A = 2.2$  (high) and 1.6 (low) on the 1–7 concreteness scale. The intensity and ambiguity differences were highly reliable (all  $ps < .001$  by  $t$  tests), and the differences were uncorrelated.

In [Experiment 4](#), the word lists that were administered to individual subjects were constructed in the same way, using the Toggia and Battig (1978) norms, except the word pool from which they were sampled was based on subjects' ratings of *categorization*. Thus, each subject studied and recalled two lists of 36 words, for which the list design was 2 (categorization ambiguity: high vs. low)  $\times$  2 (categorization intensity: high vs. low). Over lists, the mean intensities were  $M_I = 5.5$  (high) and 3.1 (low), and the mean ambiguities were  $M_A = 2.0$  (high) and 1.6 (low) on the 1–7 categorization *SD* scale. The intensity and ambiguity differences were highly reliable (all  $ps < .0001$  by  $t$  tests), and the differences were uncorrelated.

A final consideration about the materials is a point that Pollock (2018) raised in connection with word lists that have been used to manipulate *M* intensity in prior studies of the concreteness attribute. Pollock listed five linguistic variables that tended to be correlated with *M* intensity differences in those experiments and, hence, might have produced the observed memory effects: word frequency, age of acquisition, number of phonemes, word length, and number of syllables. Neath and Surprenant (2020, Experiment 1) controlled all these variables in a recall study in which concreteness rating *SDs* and *M* concreteness ratings were manipulated factorially. Recall was not affected by either rating *SDs* or *M* ratings. Consequently, we analyze the word pools for our three experiments for each of the above linguistic variables, using WordMine2 (Durda & Buchanan, 2006). We found that these five variables were not reliable predictors of differences in words ambiguity or intensity ratings. In each experiment, we tested the null hypotheses that this set of variables did not differ reliably for words with high versus low attribute ambiguity scores and did not differ reliably for words with high versus low attribute intensity scores. In each case, the core analysis was a 2 (attribute ambiguity: high vs. low ambiguity words)  $\times$  2 (attribute intensity: high vs. low intensity words)  $\times$  5 (linguistic variables: age of acquisition, frequency, length, number of phonemes, number of syllables) mixed ANOVA. The main effect for attribute ambiguity tested the first null hypothesis, and the main effect for attribute intensity tested the second. Neither of those main effects was significant in any of the three experiments. The statistical details of these analyses can be found in [Appendix Tables 11 and 12](#).

## Procedure

The free-recall procedure was the same for all three experiments. The subject first read instructions on the computer screen, which explained the multi-trial procedure. The instructions stated that the subject would participate in a series of trials in which a word list would be studied and recalled, and would then participate in a second series of trials in which a second word list would be studied and recalled. Next, the subject participated in the first phase of the experiment, which consisted of three study-buffer-test trials on List 1. This was followed by a second phase in which three study-buffer-test trials on List 2 were administered. During the first phase, the study cycle of each trial consisted of presenting the 36 words individually on the screen, with the subject being instructed to read the words silently as they were presented. Individual words were presented in random order at a 3-s rate, centered in 36-pt black font. After the last word had been presented, the subject performed a 45-s buffer task, which consisted of typing answers to simple arithmetic problems that appeared on the screen. The buffer task was followed by a 1-min free-recall test. The subject was instructed to type as many words as could be remembered from the list in any order. The subject was also told that accurate spelling was not required for responses to be counted as correct. Misspellings were rare, and when they occurred, it was obvious that they referred to one of the words on this list. After the first recall test was completed, the second trial of study-buffer-test began. The procedure for the second trial was identical to the procedure for the first trial. After the second recall test was completed, the third trial of study-buffer-test began, and it followed the same procedure as the first two trials.

Following the third recall test, there was a 1-min rest period before the second phase of the experiment. The second phase began with instructions that explained the upcoming procedure. The subject then participated in three trials of study-buffer-test on List 2. The procedure for these three trials was identical to the procedure for the three trials of the first phase.

## Results

Principal interest attaches to the dual-retrieval analyses, rather than the usual descriptive findings. However, we begin with a brief synopsis of such findings, simply to determine whether the effects of ambiguity and intensity on raw recall were robust across attributes. In all the results reported below, the dependent variable was correct free recall.

### Descriptive results

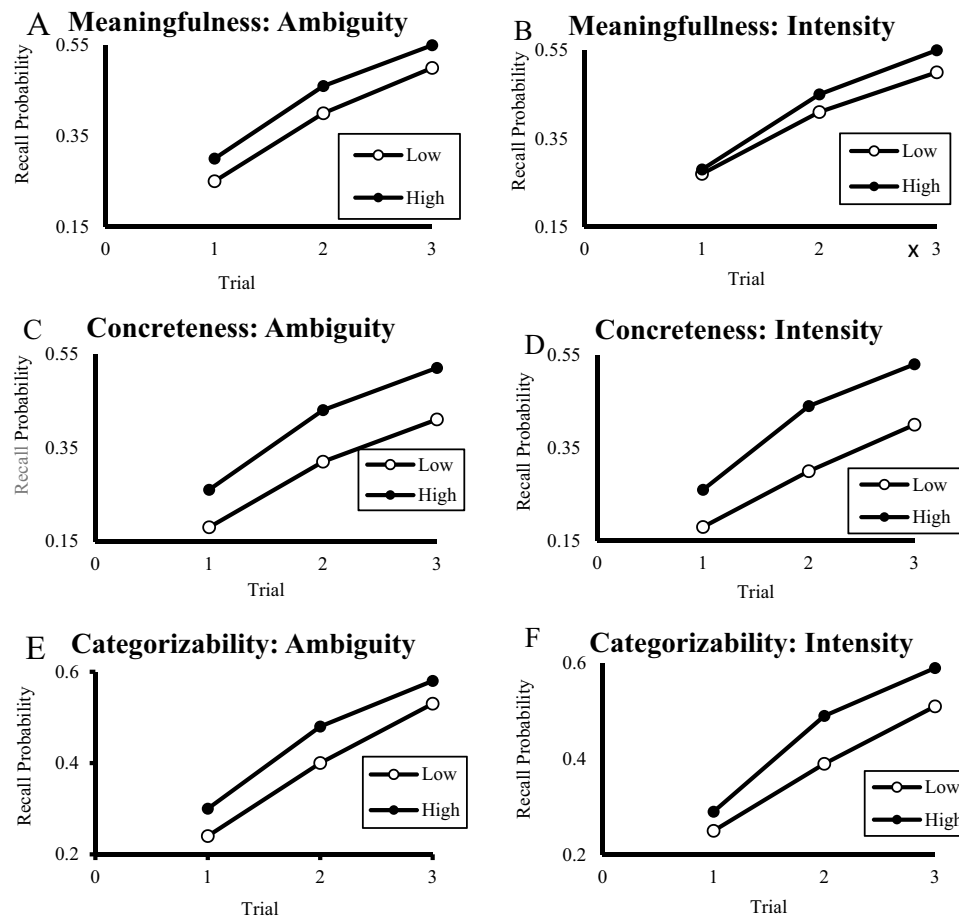
Figure 2 provides an overall picture of the recall effects of ambiguity and intensity on Trials 1–3 for meaningfulness,

concreteness, and categorization. The plotted points are for the pooled data of List 1 and List 2, as performance on the two lists did not differ appreciably. Panels A, C, and E on the left side of Fig. 2 depict the ambiguity effects for meaningfulness, concreteness, and categorization, respectively. It is obvious at a glance that recall was better when attribute ambiguity was high than when it was low, on all three trials and regardless of attribute. Panels B, D, and F on the right side depict the corresponding intensity effects for these attributes. Except for Trial 1 in Panel B, it can be seen that recall was better when attribute intensity was high than when it was low, on all trials for all attributes.

Turning to statistical tests of these patterns, we computed a 2 (attribute ambiguity: high vs. low)  $\times$  2 (attribute intensity: high vs. low)  $\times$  3 (trial: 1, 2, 3) ANOVA of the proportions of correct recall in each experiment, for the pooled data of Lists 1 and 2. The results were consistent with the patterns in Fig. 2. In Experiment 2, there were main effects for meaningfulness ambiguity,  $F(1, 71) = 24.21$ ,  $MSE = .03$ ,  $\eta_p^2 = .25$ ,  $p < .001$ , meaningfulness intensity,  $F(1, 71) = 8.98$ ,  $MSE = .03$ ,  $\eta_p^2 = .11$ ,  $p < .004$ , and trial,  $F(2, 142) = 188.67$ ,  $MSE = .02$ ,  $\eta_p^2 = .73$ ,  $p < .001$ . The ambiguity and intensity factors did not interact, but there was a small Intensity  $\times$  Trial interaction,  $F(2, 142) = 4.23$ ,  $MSE = .01$ ,  $\eta_p^2 = .06$ ,  $p < .01$ . Post hoc analysis revealed what is obvious in Fig. 2: The meaningfulness intensity effect was larger on Trials 2 and 3 than on Trial 1, and indeed, the intensity effect was not reliable on Trial 1. Turning to Experiment 3, there were main effects for concreteness ambiguity,  $F(1, 67) = 93.64$ ,  $MSE = .02$ ,  $\eta_p^2 = .58$ ,  $p < .001$ , concreteness intensity,  $F(1, 67) = 83.39$ ,  $MSE = .03$ ,  $\eta_p^2 = .55$ ,  $p < .0001$ , and trial,  $F(2, 134) = 575.39$ ,  $MSE = .01$ ,  $\eta_p^2 = .90$ ,  $p < .001$ . There was no Ambiguity  $\times$  Intensity interaction, and neither attribute factor interacted with the trial factor. Last, for Experiment 4, there was a main effect for categorization ambiguity,  $F(1, 82) = 44.15$ ,  $MSE = .02$ ,  $\eta_p^2 = .35$ ,  $p < .001$ , a main effect for categorization intensity,  $F(1, 82) = 66.66$ ,  $MSE = .02$ ,  $\eta_p^2 = .45$ ,  $p < .001$ , and a main effect for trial,  $F(2, 156) = 216.33$ ,  $MSE = .03$ ,  $\eta_p^2 = .73$ ,  $p < .001$ . There was also an Ambiguity  $\times$  Intensity interaction,  $F(1, 82) = 18.82$ ,  $MSE = .02$ ,  $\eta_p^2 = .19$ ,  $p < .001$ . Post hoc tests indicated that the ambiguity effect was larger when intensity was low than when it was high, but it was reliable at both intensity levels.

### Dual-retrieval model

The results of principal interest involved fitting the dual-retrieval model to the data of these experiments and estimating its parameters for the various conditions. Statistical procedures for this model are provided in Gomes et al. (2014). Briefly, the following steps are involved in fitting the model in Table 5 to the data of all the conditions of an experiment. The individual words on each list that the subjects learned



**Fig. 2** Memory effects of meaningfulness ambiguity (**Panel A**), meaningfulness intensity (**Panel B**), concreteness ambiguity (**Panel C**), concreteness intensity (**Panel D**), categorization ambiguity (**Panel E**), and categorization intensity (**Panel F**)

to recall generate a data space with eight distinct patterns of errors and successes:  $C_1C_2C_3$ ,  $C_1C_2E_3$ , ...,  $E_1E_2E_3$ , where  $C_i$  denotes a correct recall on the  $i$ th test, and  $E_i$  denotes an error. When probabilities are attached to these patterns, there are seven independent probabilities for each of the four conditions of each experiment (2 levels of attribute ambiguity  $\times$  2 of attribute intensity). This means that the model in Table 5 is fit to the data of each condition with one degree of freedom: There are seven free probabilities in the data space versus six probabilities (parameters) in the model.

The fit test runs as follows. The likelihood of the data is estimated with all probabilities free to vary, and then it is re-estimated with only the model's parameters free to vary. Twice the negative natural log of the ratio of these likelihoods is a  $G^2(1)$  statistic with a critical value of 3.84 to reject the null hypothesis that the model fits the target data. If this null hypothesis cannot be rejected for the conditions of an experiment, the model's parameters are then estimated and can be compared to determine if they differ reliably between conditions. These between-condition significance tests for parameters also involve computing a pair of likelihoods

– one in which all the model's parameters are completely free to vary in both target conditions and a second in which one of the model's parameters is constrained to be equal in the two conditions. The likelihood ratio statistics is  $G^2(1)$ , with a 3.84 critical value to reject the null hypothesis that the parameter has the same value in the two conditions.

First, we computed the model fit tests for the conditions of these experiments, using the 2 (attribute ambiguity)  $\times$  2 (attribute intensity) factorial structure pooled over the three trials as the four conditions of each experiment. The mean values of the four  $G^2(1)$  fit tests for each experiment were 2.02 (Experiment 2, meaningfulness), 3.38 (Experiment 3, concreteness), and .56 (Experiment 4, categorization). As a value of 3.84 is required to reject the null hypothesis of fit, this hypothesis could not be rejected in any of the experiments, and hence, we estimated the retrieval parameters for the four list conditions for each of the experiments. The results are shown in Table 6. There, the parameter values are organized by retrieval process, with the values for direct verbatim access blocked in columns 1 and 2, the values for

reconstruction blocked in columns 3 and 4, and the values for familiarity judgment blocked in columns 5 and 6.

The focal question is where, at the level of retrieval processes, the effects of attribute ambiguity are localized – reconstruction, direct verbatim access, or familiarity judgment. Recall that our working explanation predicts that it will be a reconstruction effect; that it will be easier to learn how to reconstruct list items from their semantic content when rating *SDs* are high than when they are low because attribute features are more thoroughly processed for high-*SD* items. Scanning down the paired columns in Table 6, a consistent finding was that ambiguity affected reconstructive retrieval – specifically, the  $R_1$  parameter, which measures the proportion of items that can be reconstructed after the first study cycle. Over the three attributes, that proportion was .37 when rating *SDs* were high and .27 when they were low, a highly reliable difference. Attribute ambiguity did not have a reliable effect on  $R_2$ , which measures the proportion of items that can be reconstructed after two or more study cycles.

These results yield a simple account of how attribute ambiguity enhances recall. Prior research on the dual-retrieval model indicates that words differ substantially in how easy it is to learn to reconstruct them from their semantic features on

recall tests. For some, a single study trial is all that is needed to learn a search algorithm that will reconstruct them, whereas others require multiple trials. The two reconstruction parameters are sensitive to this difference:  $R_1$  can be interpreted as the proportion of list items that are easier to learn how to reconstruct, and  $R_2$  can be interpreted as the proportion that are harder to learn to reconstruct (Brainerd et al., 2014). (Some items can never be reconstructed, and that proportion is  $1 - R_1 - (1 - R_1)R_2 - (1 - R_1)(1 - R_2)R_2$  in these experiments.) Thus, the primary effect of the ambiguity manipulation in all these experiments was to enlarge the subset of easier-to-reconstruct items, while not affecting the subset of harder-to-reconstruct items. Based on the grand means of  $R_1$  for these experiments, the easier-to-reconstruct subset was more than one-third larger when rating *SDs* were high, relative to when they were low.

Attribute ambiguity also produced some reliable effects on other retrieval processes, but they were not consistent across experiments. With respect to direct verbatim access, ambiguity affected one of the two *D* parameters for both concreteness and categorization but not the same one, and it did not affect either *D* parameter for meaningfulness. With respect to familiarity judgment, ambiguity affected one of the two *J* parameters for concreteness, but it did not affect either of them for meaningfulness or categorization.

**Table 6** Probabilities of direct verbatim retrieval (*D* parameters), reconstructive retrieval (*R* parameters), and familiarity judgment (*J* parameters)

Retrieval processes	Verbatim retrieval		Reconstruction		Familiarity judgment	
	$D_1$	$D_2$	$R_1$	$R_2$	$J_1$	$J_2$
Experiment 2						
Ambiguity effect:						
High ambiguity	.12	.13	.42*	.19*	.51	.56
Low ambiguity	.10	.12	.36*	.15*	.46	.53
Intensity effect:						
High intensity	.12	.16*	.44*	.13*	.44*	.52
Low intensity	.10	.09*	.37*	.20*	.53*	.57
Experiment 3						
Ambiguity effect:						
High ambiguity	.12*	.11	.32**	.24	.51*	.57
Low ambiguity	.06*	.06	.20**	.25	.64*	.58
Intensity effect:						
High intensity	.11	.09	.31**	.27	.57	.64*
Low intensity	.07	.08	.20**	.21	.58	.52*
Experiment 4						
Ambiguity effect:						
High ambiguity	.16	.20*	.38*	.09	.43	.51
Low ambiguity	.13	.15*	.27*	.14	.50	.44
Intensity effect:						
High intensity	.18	.21*	.35*	.12	.45	.55
Low intensity	.14	.16*	.26*	.17	.41	.54

\* $p < .05$

\*\* $p < .01$

Turning to attribute intensity, because differences in  $M$  intensity ratings are concerned with the same semantic features as differences in rating  $SD$ s, the theoretical expectation is that the retrieval loci of their effects should be similar. That assumption proved to be correct. Intensity consistently affected reconstructive retrieval – in particular, high intensity increased  $R_1$  but not  $R_2$ , relative to low intensity, for all three attributes. Theoretically, then, attribute intensity also increases the subset of easier-to-reconstruct items without affecting the subset of harder-to-reconstruct items. This similar retrieval effect for intensity could not have been due to correlations between the  $M$  intensity and rating  $SD$  manipulations because, as mentioned, they were uncorrelated in the lists that were administered in these experiments. Also similar to ambiguity, intensity did not produce consistent effects for direct verbatim access or familiarity judgment. With respect to the  $D$  parameters,  $D_2$  was significantly larger for high-intensity items in two of the experiments but not in the third. With respect to the  $J$  parameters, intensity had a reliable effect on  $J_1$  in Experiment 2 and a reliable effect on  $J_2$  in Experiment 3, but the effects ran in opposite directions.

Summing up, the dual-retrieval analyses converge on the conclusion that the ambiguity of these three semantic attributes affects reconstructive retrieval, and it does so in a specific way.

When one increases normed values of the ambiguity of attributes' perceived intensity, the total number of items that can be reconstructed increases. Moreover, this increase occurs because the size of the subset of easier-to-reconstruct items expands while the size of the subset of harder-to-reconstruct items remains invariant.

By way of qualification, it should be noted that all the ambiguity and intensity effects that were observed in [Experiments 2–4](#) were detected in mixed-list designs. In the literature, there is an established pattern in which various list manipulations are more likely to produce memory effects with mixed-list designs than with pure-list designs (see [Gomes et al., 2013](#), for a discussion of this point in connection with valence). Therefore, it remains to be seen whether the ambiguity and intensity effects that we reported will also be detected in pure-list experiments.

## General discussion

Rating norms for words' semantic attributes have been used to manipulate the types of meaning that subjects encode in a wide range of memory paradigms. Those manipulations have centered on varying attributes' average perceived intensities, as indexed by words'  $M$  intensity ratings (e.g., [Citron et al., 2014](#); [Mikels et al. 2005a](#); [Mikels et al. 2005b](#)). It is well established that such intensity manipulations produce robust memory effects for attributes such as arousal, concreteness,

familiarity, and valence. Recently, evidence has accumulated that a second property, the level of ambiguity that attaches to attributes' perceived intensity, produces separate memory effects. However, some fundamental questions remain about the normed quantity that is used to manipulate ambiguity (rating  $SD$ s). We investigated two of them: Is this quantity a valid measure of ambiguity? Which retrieval processes are affected when this quantity is varied?

The validity question was investigated for six attributes that have long histories in the memory literature (arousal, concreteness, familiarity, meaningfulness, negative valence, and positive valence). We normed separate word pools for these attributes in the traditional way, using lower  $\rightarrow$  higher intensity rating scales, and subjects also rated their degree of certainty that words' intensity ratings were accurate. In addition, we normed these word pools for the latencies of words' intensity ratings. Those norms are provided in the [Appendix](#).

We computed concurrent validity correlations between words' rating  $SD$ s and the  $M$  certainty of words' intensity ratings. To provide further evidence of validity, we computed predictive validity correlations between words' rating  $SD$ s and the  $M$  latencies of words' intensity ratings. We also computed predictive validity correlations between the  $M$  certainty of words' intensity ratings and the  $M$  latencies of words' intensity ratings. All those results converged on the conclusion that attribute rating  $SD$ s measure the ambiguity of attributes' perceived intensities.

Next, we investigated the retrieval question in recall experiments with three different semantic attributes. We used the data of experiments in which the rating  $SD$ s and  $M$  ratings of those attributes were manipulated factorially to pinpoint the retrieval loci of ambiguity's effects on recall and compare them to the retrieval loci of intensity's effects. The high-low  $SD$  manipulation had a specific retrieval effect across all attributes: The pool of items that subjects learned to reconstruct was roughly one-third larger when their rating  $SD$ s were high rather than low. That pattern was predicted by the working explanation of why ambiguity enhances recall, which posits that uncertainty about attribute intensity triggers more thorough encoding of words' semantic features. Importantly, the retrieval locus of the effects of the high-low intensity manipulation was the same as that of the high-low ambiguity manipulation, which also was anticipated on theoretical grounds. However, the two manipulations did not interact in two of the three experiments.

In closing, an important remaining question is concerned with conceptual differences between attribute ambiguity and more traditional notions of ambiguity. Customarily, we say that a word is ambiguous when (a) its meaning is unfamiliar, or (b) we know that it has multiple salient meanings. For example, most people are not aware that *insouciant* means indifferent and *moiety* means half, but most people know that *toast* has two salient meanings (bread, speech) and so does *bank* (money, river).



Both these forms of uncertainty are concerned with a word's semantic content as a whole, whereas attribute ambiguity is a componential form of uncertainty that targets a specific property (intensity) of a specific semantic attribute (e.g., concreteness, valence). Attribute ambiguity can be present at high levels notwithstanding that a word is familiar and that it does not have multiple salient meanings. For instance, *angel*, *gravity*, *unicorn*, and *vampire* are all familiar and none has multiple salient meanings, but the uncertainty of their perceived concreteness levels is very high (all rating *SDs* > 2 on a 7-point rating scale).

Although attribute ambiguity is conceptually distinct from these traditional forms of ambiguity, it might be strongly correlated with them. Indeed, a plausible hypothesis is that uncertainty about a word's meaning, either because it is unfamiliar or it has multiple meanings, would drive up uncertainty about the intensity of specific semantic attributes and, thus, produce strong correlations between their respective measures. This hypothesis leads to two obvious predictions that can be used to test it. The first is that words' overall familiarity levels will be well correlated with words' rating *SDs* for individual semantic attributes. The second is that relative to words that do not have multiple meanings, words that do (*toast*, *bank*) should display elevated rating *SDs* for individual semantic attributes. We tested both these predictions, and neither proved to be accurate.

The first prediction was evaluated by correlating attribute rating *SDs* with word familiarity measures, using published attribute rating norms. Two intuitive measures of word familiarity are available to conduct such tests: (a) familiarity ratings of words on lower → higher scales and (b) word frequencies in printed text and spoken language. The most extensive sources of evidence on both measures are the Toggia and Battig (1978) and Scott et al. (2019) norms, in which subjects rated familiarity and several other attributes for thousands of words. We added word frequencies to both norms, using the Wordmine2 database (Durda & Buchanan, 2006). We computed the predictive validity correlations between each of the familiarity measures and the rating *SDs* of six attributes in the Toggia-Battig norms and the rating *SDs* of eight other attributes in the Scott et al norms. The results appear in Table 7.

These validity correlations do not provide support for the notion that attribute ambiguity increases as words' meanings become more unfamiliar. If that were true, the coefficients would be large and negative; instead, they are mostly very small, accounting for an average of 1% of the variance. Further, one-third of them are positive, with the two strongest ones (for the *SDs* of meaningfulness and number of features in the Toggia-Battig norms) being positive. In short, the relations between words' judged familiarity and their frequency in the lexicon, on the one hand, and the rating *SDs* of several semantic attributes, on the other, are weak and inconsistent.

We conducted a similar validity analysis for the prediction that attribute ambiguity will be greater for words with multiple

salient meanings using data from the Scott et al. (2019) norms. The 5,553 words in those norms includes a subsample of 379 words that each have two salient meanings (e.g., *bank*, *figure*, *live*, *medium*, *toast*). We computed the rating *SDs* of the nine rated attributes separately for these multiple-meaning words versus the remaining 5,174 words. The prediction, of course, is that for all nine attributes, the mean rating *SDs* of the multiple-meaning words will be higher than the mean rating *SDs* of other words. To test that prediction, using rating *SDs* as the dependent variable, we computed a 2 (word ambiguity: multiple-meaning words vs. other words) × 9 (rated attribute) ANOVA. The key result is the main effect for word ambiguity, which should be significant if multiple meanings increase attribute ambiguity. It was not significant, however; the rating *SDs* of multiple-meaning words did not differ reliably from the rating *SDs* of other words.

To conclude, beyond the results in Experiments 1–4, it seems that the normed quantitative measure of attribute ambiguity is empirically distinct from more traditional conceptions of semantic ambiguity. The data that were just reported reveal that attribute ambiguity is unrelated to either the familiarity or multiple-meanings conception of ambiguity: There was no evidence of strong negative correlations between rating *SDs* and words' rated familiarity or words' objective frequency in linguistic usage, and there was no evidence that rating *SDs* are higher for words with multiple salient meanings.

**Table 7** Predictive validity correlations between two measures of word familiarity and rating *SDs* of semantic attributes in the Toggia and Battig (1978) and Scott et al. (2019) norms

Measure	Familiarity ratings	
	Familiarity ratings	Word frequencies
	Toggia and Battig (1978)	
Categorization	-.06	-.05
Concreteness	-.18	-.01
Imagery	-.12	-.01
Meaningfulness	.40	.08
Number of features	.38	.12
Pleasantness	.11	-.08
	Scott et al. (2019)	
Age of acquisition	-.23	-.13
Arousal	-.03	-.03
Dominance	.07	.00
Concreteness	-.11	.09
Gender	-.03	-.12
Imagery	-.24	.05
Size	-.10	-.01
Valence	-.04	.00

**Appendix**

**Table 8** Arousal (AROU) and Concreteness (CON) Intensity, Latency, and Certainty Norms

AROU		CON											
Word	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency	Word	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency
accordion	2.88	1.9	2.08	5.46	1.69	0.83	abode	4.37	1.94	2.77	4.95	1.91	0.95
alien	4.46	1.91	1.79	6.02	1.13	0.72	adult	5.53	1.55	3.45	5.63	1.43	1
angle	2.34	1.41	2.27	5.56	1.5	0.95	alert	3.63	1.9	2.78	5.4	1.56	1
arch	2.76	1.7	1.91	5.32	1.79	0.93	alien	3.93	1.93	2.8	5.12	1.65	0.94
berry	3.56	1.69	1.61	5.51	1.25	0.83	alter	4.23	2.07	3.16	5.81	1.47	0.89
birth	4.51	1.78	2.19	5.61	1.66	0.84	apex	3.67	1.69	2.54	4.7	2.21	0.93
blacksmith	2.71	1.74	1.92	5.59	1.56	0.81	art	4.12	1.94	2.89	5.81	1.16	0.89
bomb	4.73	2.09	1.75	5.88	1.44	0.84	baby	5.74	1.62	2.19	6.14	1.19	0.92
bored	1.95	1.4	1.92	6.1	1.02	0.71	ball	6.28	1.26	2.13	6.3	1.06	0.8
botany	2.9	1.99	2.18	5.66	1.3	0.84	barrel	5.98	1.55	1.8	6.26	1.16	0.84
bright	4.76	1.87	2.28	5.63	1.51	0.75	beach	6.19	1.22	2.2	6.14	1.26	0.91
brush	2.68	1.6	2.25	5.88	1.14	1	beef	6.02	1.49	2.27	6.49	0.86	0.79
cabin	3.2	1.91	1.86	5.51	1.58	0.87	belief	2.26	1.31	2.35	5.63	1.5	1
case	2.68	1.57	2.02	5.46	1.63	0.83	black	4.7	1.86	2.41	5.58	1.43	0.84
cat	3.9	2.15	2.12	5.66	1.57	0.8	blue	4	1.91	2.47	5.72	1.24	0.88
caterpillar	2.98	1.94	2.22	5.37	1.71	0.87	bomb	6.21	1.15	2.17	6.16	1.17	0.92
choose	3.34	1.93	1.84	5.8	1.31	0.78	bored	3.09	1.64	2.82	5.47	1.35	0.9
church	2.93	1.89	1.91	5.76	1.26	0.82	calm	2.81	1.24	3.11	5.33	1.19	1.05
cigar	2.61	1.51	1.94	5.71	1.36	0.81	candle	6.47	1.03	1.76	6.53	0.85	0.74
clarinet	2.68	1.74	2.39	5.59	1.41	0.93	cement	6.51	0.98	2.14	6.49	0.98	0.89
cream	3.12	1.81	2.1	5.56	1.34	0.87	chair	6.53	0.96	1.74	6.47	1.14	0.78
course	4.37	2.01	2.46	5.9	1.3	0.8	cheese	6.33	1.38	1.93	6.56	0.91	0.77
dawn	3.29	1.95	2.31	5.78	1.17	0.87	clover	6.12	1.4	2.13	6.4	0.93	0.8
dead	3.54	2.26	2	5.71	1.5	0.79	conference	4.86	1.75	2.72	5.56	1.42	1.02
dear	3.2	1.79	2.36	5.56	1.48	0.77	crow	6.42	1.16	2.26	6.4	1.18	0.79
death	3.71	2.36	2.11	5.66	1.76	0.83	dancer	5.77	1.32	2.26	6.05	1.13	0.91
deceit	3.98	1.93	2.37	5.51	1.57	0.82	death	3.74	2.09	3.47	5.19	1.72	1.03
despair	2.98	1.93	2.02	5.95	1.16	0.81	deceit	2.79	1.52	2.57	5.37	1.5	1.01
despise	4.2	1.83	2.23	5.41	1.55	0.99	die	3.42	1.99	2.91	5.23	1.59	1.02
die	3.8	2.3	1.92	5.61	1.76	0.79	door	6.42	1.14	1.92	6.44	1.08	0.78
disappointed	2.39	1.48	1.9	5.85	1.22	0.83	eel	6.3	1.47	1.66	6.51	0.91	0.88
dough	2.46	1.57	2.22	5.51	1.45	0.85	elegant	3.16	1.66	2.77	5.42	1.28	1

Table 8 (continued)

ARO	CON												
	Word	MRating	SDRating	MRatingLa- tency	MCertainty	SDCertainty	MCertaintyLa- tency	MRating	SDRating	MRatingLa- tency	MCertainty	SDCertainty	MCertaintyLa- tency
dusty	2.17	1.3	1.99	5.9	1.22	0.87	empty	3.07	1.5	2.62	5.33	1.32	1.02
easy	2.98	1.9	2.48	5.51	1.57	0.77	fairy	3.88	2.14	3.32	5.33	1.34	1.12
egg	2.46	1.42	1.92	5.83	1.34	0.82	fall	3.7	1.63	3.29	5.7	1.21	1.09
fear	4.37	2.06	2.11	6.1	1.16	0.83	foolish	2.95	1.33	2.31	5.44	1.39	0.91
flow	2.83	1.66	1.82	5.63	1.37	0.74	freedom	2.26	1.33	2.92	5.67	1.43	1.06
gain	3.56	1.64	2.2	5.49	1.38	0.83	friend	5.05	1.59	2.97	5.72	1.22	1.14
garlic	3.15	1.96	2.07	5.51	1.49	0.85	fuse	4.56	1.65	3.08	5.23	1.43	0.89
general	2.56	1.61	1.84	5.41	1.55	0.85	gain	3.44	1.28	3.42	5.23	1.32	0.85
gentleman	3.85	1.89	2.22	5.59	1.3	0.91	gender	3.02	1.75	3.08	5.23	1.49	1.2
gorilla	4.39	2.2	1.86	5.93	1.4	0.76	ghost	2.72	1.5	3.04	4.98	1.6	1.06
graph	2.32	1.51	1.76	5.61	1.63	0.9	gist	2.72	1.33	2.41	5.23	1.6	1.09
grief	3.39	2.1	1.84	5.51	1.73	0.73	half	4.07	1.71	2.83	5.65	1.4	1.02
gun	4.59	2.26	2.09	5.95	1.4	0.69	hard	3.72	1.91	2.77	5.67	1.32	1.03
hate	4.49	2.19	2.25	5.93	1.19	0.92	harmony	3.12	1.58	2.71	5.26	1.33	1.02
heel	2.66	1.7	1.89	5.49	1.53	0.88	henchman	4.91	1.8	2.49	5.4	1.65	0.82
hierarchy	3.2	1.72	2.13	5.41	1.52	0.88	home	4.77	1.76	3.21	5.84	1.19	0.93
infinite	4.68	1.97	2.44	5.76	1.43	0.83	interest	2.84	1.54	2.57	5.44	1.53	1.15
intense	5.71	1.36	2.06	6.22	1.01	0.75	lamb	6.47	1.03	2.65	6.35	1.11	0.8
jam	3.22	1.96	1.77	5.61	1.53	0.88	last	3.28	1.52	2.51	5.05	1.57	0.89
juggler	3.93	2	2.46	5.71	1.27	0.93	lawyer	6.14	1.17	2.1	6.19	1.28	0.79
kill	4.41	2.27	2.24	5.85	1.49	0.93	lemon	6.35	1.46	1.61	6.65	0.75	0.73
kitten	4.56	2.29	1.97	5.9	1.26	0.89	library	6.28	1.03	1.97	6.09	1.25	0.72
lawn	2.24	1.41	2.29	5.73	1.41	0.71	lion	6.37	1.38	1.63	6.3	1.44	0.78
lazy	1.93	1.4	2.12	5.73	1.61	0.76	live	3.09	1.66	2.96	5.47	1.58	1.07
lesson	2.49	1.57	2.2	5.71	1.25	0.9	load	3.44	1.39	2.9	5.07	1.55	1.13
load	2.8	1.62	2.16	5.61	1.48	0.86	long	3.56	1.68	2.63	5.49	1.35	0.95
locker	2.8	1.47	1.99	5.66	1.33	0.97	lynx	5.4	1.98	2.46	5.37	1.85	0.87
lovely	4.51	1.73	1.9	5.83	1.12	1	mallet	5.77	1.7	2.3	5.88	1.38	0.84
lure	3.85	1.74	2.29	5.54	1.45	0.78	mathematics	4.19	2.04	2.96	5.4	1.58	0.89
morgue	3	2.16	2.18	5.61	1.66	0.84	memory	2.84	1.4	2.59	5.35	1.36	1.01
moss	2.07	1.42	2.23	5.66	1.48	0.85	mild	2.86	1.21	2.67	5	1.53	1.03
movie	4.68	1.63	2.2	5.68	1.42	0.83	minute	4.33	2.03	2.78	5.4	1.64	0.89
murder	4.39	2.33	1.77	5.98	1.25	0.73	moth	6.33	1.48	1.88	6.6	0.85	0.69

Table 8 (continued)

ARO	CON													
	Word	MRating	SDRating	MRatingLa- tency	MCertainty	SDCertainty	MCertaintyLa- tency	Word	MRating	SDRating	MRatingLa- tency	MCertainty	SDCertainty	MCertaintyLa- tency
	nail	2.54	1.48	2.49	5.46	1.4	0.8	mother	5.53	1.52	2.48	5.67	1.38	0.92
	navel	2.88	1.73	2.19	5.59	1.61	0.82	none	2.42	1.48	2.72	5.14	1.52	1.04
	needle	3.22	1.68	2.04	5.59	1.3	0.84	nursery	5.72	1.32	2.41	6.02	1.26	0.89
	none	1.59	1.09	1.65	6.15	1.17	0.8	onion	6.4	1.31	1.93	6.42	1.01	0.85
	oblivious	2.56	1.38	2.27	5.63	1.26	0.83	owl	6.44	1.22	1.69	6.6	0.82	0.79
	organ	3.1	1.62	1.99	5.44	1.79	0.79	pain	3.28	1.59	2.62	5.26	1.66	0.94
	pain	4.51	2.1	2.1	5.85	1.3	0.84	paint	5.74	1.57	2.09	6.16	1.02	0.78
	peace	3.27	2.1	1.76	5.88	1.33	0.78	paper	6.3	1.12	2.09	6.44	1.12	0.98
	peaceful	3	2.28	2.03	5.95	1.14	0.82	pear	6.56	0.67	1.91	6.58	0.79	0.76
	philosophy	2.66	1.87	2.13	5.95	1.41	0.89	penicillin	6.16	1.48	1.84	6.4	0.98	0.78
	phrase	2.39	1.48	1.97	5.63	1.44	0.82	permanent	3.77	1.93	3.07	5.53	1.4	1.04
	pigeon	2.8	1.9	1.99	5.9	1.3	0.78	physics	4.35	1.78	2.48	5.35	1.51	1.02
	pliers	2.39	1.58	2.22	5.59	1.48	0.94	pick	3.4	1.59	3.11	5.09	1.38	0.99
	plum	3.05	1.63	1.95	5.63	1.28	0.76	pill	6.28	1.18	2.32	6.26	1.27	0.82
	pope	2.83	2	1.81	5.56	1.63	0.82	pork	6.42	1.01	2.1	6.51	0.86	0.76
	prince	4	1.82	2.25	5.59	1.47	0.84	prayer	3.65	1.8	2.75	5.26	1.4	1.17
	quack	3.61	2.18	1.88	5.51	1.6	0.73	real	3.21	1.81	3.27	5.28	1.62	1.2
	rabbi	2.49	1.72	2.08	5.66	1.49	0.82	rhapsody	3.21	1.6	3.09	4.74	1.77	1.15
	rancid	3.46	2.3	2.94	5.78	1.44	0.73	rocket	6.42	0.91	1.85	6.4	0.98	0.71
	ration	2.29	1.21	2.62	5.32	1.68	0.78	rough	3.84	1.57	2.7	5.42	1.3	1.04
	regulation	2.34	1.41	2.12	5.66	1.33	0.79	saint	3.84	1.8	2.61	5.3	1.44	1.07
	religion	2.68	1.78	2.28	5.63	1.55	0.76	salty	4.16	1.65	2.87	5.28	1.62	0.99
	river	3.44	1.72	2.2	5.46	1.45	0.8	sane	2.86	1.37	3.06	5.07	1.3	1.21
	role	2.63	1.5	1.83	5.71	1.47	0.8	scent	3.98	1.63	2.66	5.26	1.51	0.88
	roll	2.93	1.65	2.24	5.66	1.24	0.82	science	4	1.88	2.91	5.23	1.67	1.03
	salty	3.51	2.05	2.15	5.85	1.28	0.71	sense	3.37	1.56	2.41	5.35	1.33	1.08
	sea	4.44	1.98	2.38	5.83	1.3	0.82	skin	6.05	1.31	2.4	6.23	0.97	0.85
	sheer	3.02	1.65	2.56	5.51	1.29	0.69	smile	5.33	1.52	2.4	5.79	1.34	0.96
	shoot	4.37	2.14	2.15	5.8	1.35	0.78	smooth	3.81	1.61	2.46	5.4	1.29	1.1
	shriek	4.61	2.1	1.79	5.83	1.41	0.73	snake	6.47	1.08	2.14	6.37	1.33	0.85
	sing	4.8	1.54	1.92	5.85	1.17	0.91	soar	3.37	1.53	2.27	5.26	1.54	1.09
	skirt	3.27	1.7	2.15	5.24	1.59	0.83	sore	3.74	1.5	3	5.51	1.3	0.94
	skunk	3.37	1.84	2.14	5.63	1.24	0.83	space	3.3	1.5	3.68	5.26	1.66	0.97

Table 8 (continued)

ARO		CON											
Word	MRating	SDRating	MRatingLa- tency	MCertainty	SDCertainty	MCertaintyLa- tency	Word	MRating	SDRating	MRatingLa- tency	MCertainty	SDCertainty	MCertaintyLa- tency
sleep	2.9	2.2	2.11	5.88	1.23	0.8	submarine	6.72	0.63	2.11	6.6	1	0.82
smoke	3.2	1.78	2.13	5.71	1.45	0.86	suit	6.12	1.2	1.76	6.16	1.27	0.87
spear	3.78	1.88	2	5.51	1.43	0.75	sweet	3.58	1.74	2.77	5.16	1.53	0.94
spoon	2.78	1.82	1.76	5.83	1.24	0.73	swimming	5.4	1.5	2.68	5.81	1.45	0.96
stay	2.49	1.57	1.66	5.54	1.38	0.79	sword	6.4	0.95	2.13	6.35	1.02	0.94
stem	2.8	1.68	1.92	5.66	1.46	0.79	teeth	6.7	0.74	1.75	6.67	0.78	0.82
string	2.27	1.38	2.01	5.93	1.15	0.78	theory	2.98	1.71	2.39	5.3	1.7	0.97
suffocate	4.32	2.31	2.57	5.85	1.67	0.84	tickle	4.44	1.69	2.37	5.53	1.5	0.84
sunset	4.68	2.18	2.23	5.93	1.1	0.94	tiger	6.7	0.74	2.15	6.6	0.93	0.81
taste	3.9	1.96	2.33	5.78	1.19	0.92	tooth	6.67	0.75	1.92	6.63	0.93	0.87
threat	4.46	2.27	2.34	5.98	1.27	0.89	trace	3.56	1.31	3.45	4.91	1.56	0.93
throat	3.22	1.42	2.12	5.39	1.64	0.75	veal	5.74	1.59	1.99	5.79	1.77	0.81
toilet	2.54	1.6	2.17	5.54	1.61	0.93	verb	4.07	2.05	2.71	5.42	1.78	0.98
traitor	4.17	2	1.75	5.78	1.37	0.77	vinegar	6.44	1.05	1.65	6.47	0.85	0.83
treat	4.61	1.84	2.09	5.93	1.21	0.87	vision	3.84	1.59	2.13	4.93	1.68	0.96
triumph	5.17	1.84	2.15	6.12	0.98	0.8	wall	6.4	0.9	2.1	6	1.27	1.01
war	4.56	2.12	2.1	5.78	1.37	0.79	wallet	6.58	0.91	2.81	6.56	0.93	0.71
week	2.37	1.46	1.98	5.78	1.56	0.86	water	6.12	1.14	2.36	6.23	1.02	0.86
wheel	2.51	1.49	2.16	5.66	1.33	0.83	wise	2.51	1.45	2.36	5.7	1.32	0.98
whiff	2.98	1.65	1.85	5.39	1.56	0.78	witch	4.37	2.18	2.48	5.58	1.26	0.96
wish	4.61	2.01	1.94	5.8	1.27	0.81	wizard	4.49	2.16	2.96	5.67	1.32	0.99
zero	1.68	1.15	1.93	5.98	1.15	0.84	yellow	4.3	1.88	2.58	5.23	1.74	1.07



**Table 9** Familiarity (FAM) and Meaningfulness (MNG) Intensity, Latency, and Certainty Norms

Word	FAM					MNG							
	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty			
accordion	4.45	1.97	1.94	6.32	1.01	0.92	aisle	3.28	1.92	2.54	5.15	1.27	0.91
alter	5.14	1.61	1.82	6.18	0.99	0.84	ant	3	1.72	2.14	5.31	1.17	0.96
apathy	4.55	1.85	1.81	6.02	1.17	0.82	arm	4.03	1.69	2.78	5.1	1.33	0.83
aunt	6.43	1.19	1.51	6.73	0.69	0.67	art	5.51	1.5	2.92	5.74	1.31	0.88
bang	5.75	1.74	1.41	6.39	0.99	0.7	awkward	4.18	1.41	1.8	5.05	1.32	0.81
bar	6.18	1.5	1.32	6.7	0.63	0.74	balloon	3.23	1.6	2.59	4.95	1.3	1.01
beet	5.25	1.97	1.58	6.34	0.99	0.78	basketball	4.21	1.58	1.97	5.28	1.28	0.9
beg	5.82	1.56	1.36	6.3	1.13	0.71	beet	2.72	1.69	2.25	5.38	1.21	0.92
bishop	4.09	2.25	1.75	6.16	1.01	0.77	blunder	3.31	1.66	2.7	5.18	1.45	0.86
board	6.2	1.23	1.91	6.5	0.73	0.87	boat	4.21	1.4	1.93	5.33	1.15	0.96
boot	6.18	1.24	1.27	6.43	1.19	0.78	botany	3.15	1.66	2.33	5.08	1.35	0.91
botany	4	1.99	1.76	6.02	1.34	0.88	brush	3.38	1.43	2.48	5.33	1.18	0.82
boy	6.43	1.35	1.21	6.86	0.35	0.61	butterfly	3.72	1.67	2.37	5.36	1.11	0.89
café	6.18	1.28	1.62	6.73	0.66	0.69	carol	2.74	1.79	2.36	4.92	1.46	0.94
charlatan	2.8	1.82	1.85	5.98	1.36	0.85	cast	3.59	1.76	2.35	5.08	1.4	0.91
clean	6.36	1.24	1.59	6.73	0.59	0.86	choral	2.95	1.86	2.88	4.82	1.54	0.77
cloak	4.64	1.88	1.77	5.95	0.99	0.95	chute	2.67	1.71	1.82	5.03	1.48	0.94
clumsy	5.86	1.59	1.72	6.59	0.69	0.7	circular	3.69	1.82	2.45	4.87	1.59	0.8
coast	6.09	1.2	1.56	6.39	0.92	0.69	clove	2.79	1.63	2.4	5.08	1.42	0.96
coffee	6.41	1.32	1.59	6.68	0.8	0.7	clue	3.41	1.48	2.6	5	1.03	0.85
court	5.48	1.68	1.57	6.2	1.13	0.86	coarse	3.1	1.47	2.17	4.87	1.26	0.9
crucifix	3.84	1.93	1.8	5.84	1.26	0.94	coffee	4.69	1.69	1.76	5.59	1.14	0.87
bup	6.5	1.19	1.48	6.84	0.48	0.66	corner	3.31	1.45	2.66	4.95	1.28	0.88
dame	3.43	2.03	1.76	5.98	1.09	0.88	culture	5.82	1.14	1.94	5.9	1.23	0.86
deck	5.61	1.78	1.67	6.52	0.82	0.77	dance	4.56	1.47	2.08	5.36	0.93	0.78
dense	5.8	1.34	1.61	6.5	0.7	0.98	dandruff	2.31	1.62	2.22	5.18	1.37	0.97
dentist	6	1.45	1.44	6.52	0.93	0.76	decorate	3.92	1.55	2.86	4.85	1.42	0.83
desk	6.18	1.54	1.44	6.55	1.07	0.77	doctor	5.18	1.54	2.04	5.74	1.12	0.72
destroy	5.77	1.76	1.35	6.55	0.76	0.71	edge	3.44	1.55	2.22	5.21	1.15	0.9
doctor	6.36	1.2	1.38	6.73	0.59	0.7	eel	3.08	1.64	2.41	5.31	1.2	0.85
duke	4.57	1.96	1.75	6.07	1.17	0.77	evaluate	3.74	1.53	2.69	5.15	1.37	0.92
dungeon	4.57	2.18	1.88	6.09	1.14	0.75	ever	3.33	1.92	2.84	5.23	1.53	0.95
dye	5.41	1.88	1.8	6.61	0.81	0.75	fashion	4.87	1.54	2.61	5.74	1.16	0.82

Table 9 (continued)

FAM	MNG													
	Word	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency	Word	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency
	elephant	6.02	1.58	1.56	6.52	0.85	0.64	fate	4.77	1.63	2.78	5.51	1.32	0.77
	feet	6.45	1.25	1.32	6.77	0.48	0.68	feat	3.38	1.77	2.62	5.18	1.25	0.85
	frugal	5.14	1.79	1.66	6.25	0.97	0.72	first	4.44	1.76	3.09	5.33	1.44	0.89
	garbage	6.23	1.46	1.41	6.68	0.74	0.76	garden	4.51	1.41	2.32	5.54	1.02	0.91
	glass	6.3	1.41	1.38	6.77	0.57	0.68	gentleman	4.59	1.73	2.36	5.15	1.27	0.94
	good	6.57	1.17	1.41	6.84	0.57	0.64	gist	3.15	1.93	2.28	5.44	1.27	1.01
	grape	6.32	1.46	1.44	6.64	0.81	0.74	god	5.62	1.8	2.29	5.95	1.23	0.86
	hands	6.41	1.39	1.28	6.77	0.52	0.67	happiness	5.97	1.37	1.93	5.97	1.35	0.74
	heavy	6.41	1	1.52	6.61	0.87	0.69	happy	5.97	1.14	1.6	5.95	1.45	0.78
	hospital	6.41	1.11	1.33	6.73	0.59	0.73	hate	5.41	1.68	1.88	5.69	1.32	0.89
	hymn	4.14	2	1.77	6.16	0.96	0.88	hell	4.72	1.76	2.2	5.54	1.19	0.93
	impropriety	3.14	1.94	2.3	5.98	1.05	1.1	herb	3.41	1.52	2.26	5.28	1.26	0.87
	indifferent	5.64	1.5	1.84	6.32	0.88	0.91	honesty	5.46	1.5	2.46	5.72	1.41	0.89
	islander	5.09	2.02	1.84	6.27	1.19	0.78	incompetent	3.95	1.79	2.71	5	1.67	0.85
	journal	6.11	1.43	1.42	6.68	0.64	0.75	juggler	2.92	1.61	2.13	4.97	1.56	0.9
	judge	5.64	1.71	1.49	6.57	0.76	0.85	kind	5.38	1.62	1.73	5.92	1.26	0.75
	lady	6.34	1.46	1.29	6.61	0.89	0.66	lady	4.85	1.68	3.11	5.28	1.47	0.85
	lamp	6.36	1.28	1.34	6.66	0.83	0.8	lane	3.41	1.6	1.95	4.95	1.39	0.85
	lawn	5.84	1.7	1.58	6.59	0.82	0.73	lean	3.49	1.67	2.93	4.97	1.42	0.95
	learn	6.57	1.09	1.44	6.82	0.66	0.7	lift	3.69	1.58	2.73	4.92	1.31	0.76
	lice	5.39	1.82	1.63	6.18	1.04	0.78	love	6.28	1.21	2.58	6.33	1.01	0.8
	limousine	5.52	1.82	1.67	6.41	0.79	0.78	macaroni	3.41	1.79	2.78	5.41	1.29	0.88
	lips	6.27	1.19	1.28	6.68	0.67	0.7	marry	4.9	1.29	2.73	5.13	1.44	0.89
	man	6.57	1.11	1.31	6.77	0.71	0.66	mattress	3.87	1.54	2.59	5.1	1.21	0.91
	many	6.36	1.42	1.63	6.68	0.74	0.67	mischievous	3.97	1.48	2.25	5.18	1.32	0.86
	meeting	6.25	1.35	1.56	6.7	0.67	0.71	mosque	3.67	1.8	2.33	4.82	1.85	1.07
	memory	6.43	0.93	1.31	6.57	0.73	0.83	mother	5.38	1.55	2.06	5.77	1.29	0.84
	molecule	5.52	1.56	1.68	6.55	0.7	0.72	movie	4.51	1.59	2.13	5.69	0.95	0.82
	moon	6.34	1.49	1.61	6.8	0.51	0.68	mustard	2.85	1.58	2.75	5.23	1.4	0.97
	morgue	4.3	2.02	1.9	6.18	1.26	0.75	need	4.56	1.86	2.51	5.46	1.45	0.98
	mouth	6.36	1.33	1.51	6.77	0.48	0.66	none	3.49	1.9	2.34	4.95	1.52	0.83
	navel	3.73	2.02	2.03	5.93	1.07	0.86	novel	4.54	1.5	2.4	5.64	1.16	0.79
	new	6.41	1.26	1.38	6.66	0.71	0.71	obese	4.1	1.67	2.97	5.44	1.27	0.96

Table 9 (continued)

FAM	MNG													
	Word	MRating	SDRating	MRatingLatency	MCertaintyLatency	Word	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency		
	novel	5.61	1.57	1.85	6.43	1.07	0.77	oblivious	3.49	1.75	2.39	5.13	1.22	0.97
	oboe	3.98	2.15	1.79	6.2	1.05	0.84	oboe	2.79	1.76	2.73	5.03	1.71	0.86
	organ	5.61	1.59	1.57	6.32	1.07	0.74	older	4.51	1.71	2.33	5.33	1.3	0.86
	oyster	5.34	1.75	1.54	6.18	0.97	0.9	onion	3.21	1.34	2.39	5.28	1.26	0.96
	page	6.14	1.46	1.45	6.68	0.67	0.69	pause	3.69	1.45	2.53	5.23	1.25	0.93
	palace	5.32	1.84	1.55	6.27	0.95	0.75	peaceful	5.38	1.44	2.4	5.72	1.5	0.84
	paradox	5.11	1.75	1.86	6.27	0.79	0.74	pearl	3.31	1.62	2.55	5.18	1.19	0.88
	parish	3.59	2.02	2.05	5.89	1.24	0.94	penicillin	3.44	1.83	3.27	5.28	1.36	0.97
	park	6.32	1.34	1.4	6.7	0.59	0.67	pickle	3.26	1.79	2.44	5.41	1.14	0.99
	path	6	1.66	1.6	6.48	0.95	0.74	pigeon	2.97	1.66	2.12	5.36	1.33	0.79
	peal	4.2	2.27	2.02	6.27	1.06	0.93	plea	3.41	1.65	2.56	5.23	1.2	0.92
	pelt	4	2.09	1.75	5.98	1.25	0.77	potato	3.95	1.89	2.4	5.31	1.36	1.02
	phrase	5.82	1.4	1.53	6.7	0.55	0.71	projector	3.31	1.75	2.01	5.13	1.28	1.03
	pint	5.25	1.78	1.76	6.3	1.02	0.85	pump	3.36	1.71	2.37	4.9	1.52	0.89
	point	6.34	1.03	1.38	6.48	1.05	0.72	quack	3.08	1.61	2.18	4.92	1.38	1.01
	potato	6.55	0.85	1.35	6.66	0.71	0.73	queen	4.67	1.9	2.18	5.33	1.49	0.86
	prayer	5.45	1.72	1.66	6.32	1.09	0.83	rabbi	3.54	1.94	2.06	5.08	1.58	0.92
	pulpit	2.59	1.66	1.95	6.02	1.28	0.89	rape	4.87	2.05	3	5.46	1.65	0.99
	punch	5.8	1.73	1.79	6.52	0.82	0.87	red	3.87	1.92	2.18	5.38	1.31	0.8
	quietly	6.11	1.4	1.92	6.45	1.17	0.81	reel	2.77	1.53	2.06	5.15	1.18	1
	rabbi	3.98	2.12	1.92	6.02	1.11	0.9	rest	4.74	1.57	2.17	5.67	1.11	0.84
	rattlesnake	4.82	2	1.97	6.27	0.97	0.76	rich	5.33	1.47	2.12	5.85	1.11	0.78
	red	6.36	1.3	1.39	6.82	0.39	0.64	role	3.82	1.8	2.31	5.46	1.14	0.86
	refrigerator	6.3	1.29	1.29	6.59	0.9	0.69	roll	3.38	1.52	2.74	4.95	1.47	0.8
	region	5.84	1.33	1.64	6.43	0.85	0.8	sheep	3.08	1.48	1.89	5.38	1.11	0.93
	rough	6.16	1.36	1.43	6.64	0.78	0.67	sin	4.64	1.77	3.15	5.23	1.42	0.91
	sapphire	4.66	2	1.76	6.25	0.92	0.85	sluggish	3.64	1.51	2.04	5.23	1.13	0.89
	school	6.59	1.15	1.23	6.73	0.66	0.71	smear	2.82	1.5	2.69	4.95	1.64	0.92
	sheepskin	4.11	2.07	2.04	6.05	1.12	0.82	song	4.97	1.53	2.49	5.38	1.18	0.98
	shriek	5.09	2	1.67	6.41	0.76	0.81	sour	4.23	1.51	2.98	5.31	1.08	0.85
	Slush	4.84	1.99	1.75	6.14	1.17	0.78	steam	3.41	1.57	2.83	4.85	1.35	0.82
	smear	4.89	1.97	1.74	6.32	0.96	0.7	stem	3.44	1.65	2.46	5.08	1.46	1.01
	Soup	6.52	1.05	1.33	6.61	0.89	0.71	store	4.56	1.92	2.93	5.51	1.07	0.92

Table 9 (continued)

FAM	MNG													
	Word	MRating	SDRating	MRatingLa- tency	MCertainty	SDCertainty	MCertain- tyLa- tency	Word	MRating	SDRating	MRatingLa- tency	MCertainty	SDCertainty	MCertain- tyLa- tency
	speak	6.52	1.13	1.42	6.57	0.87	0.75	sweet	5.26	1.58	2.71	5.79	1.22	0.86
	store	6.32	1.36	1.27	6.59	0.97	0.71	synagogue	3.28	1.83	2.58	5.15	1.46	1.07
	suit	6.14	1.17	1.41	6.52	0.79	0.67	team	5.31	1.47	2.7	5.92	1.04	0.72
	sulphur	4.61	2.05	1.95	6.2	1.09	0.8	thought	5.13	1.94	2.39	5.87	1.2	0.8
	theoretical	4.98	1.76	1.98	6.3	0.88	0.85	tiger	3.9	1.73	2.18	5.23	1.27	0.82
	top	6.34	1.33	1.41	6.73	0.73	0.72	till	2.77	1.68	2.09	4.82	1.5	0.86
	tourist	5.89	1.48	1.57	6.57	0.76	0.84	time	5.62	1.5	2.74	6.08	1.01	0.86
	town	6.39	1.2	1.54	6.77	0.48	0.8	tomb	3.51	1.76	2.04	5.21	1.44	0.85
	train	6.34	1.24	1.41	6.75	0.53	0.74	typewriter	2.85	1.69	2.01	5.38	1.09	0.94
	trash	6.45	1.07	1.57	6.7	0.55	0.75	tyranny	4.15	1.74	2.25	5.05	1.43	0.87
	trout	4.59	1.81	1.75	6.34	1.01	0.83	university	5.69	1.2	2.25	5.87	0.95	0.91
	truth	6.14	1.44	1.46	6.68	0.71	0.75	up	4.1	1.96	2.56	5.21	1.44	0.96
	unite	5.68	1.6	1.7	6.27	1.23	0.85	wealth	5.38	1.48	1.88	5.72	1.19	0.79
	vet	6.14	1.21	1.42	6.61	0.72	0.65	white	4.46	1.83	2.67	5.46	1.25	0.78
	walrus	5.25	1.89	1.69	6.39	1.04	0.72	woman	5.59	1.67	1.88	5.87	1.28	0.86
	watch	6.23	1.33	1.41	6.73	0.62	0.7	womb	4	1.69	2.27	5.13	1.3	0.79
	whalebone	3	1.99	1.86	5.95	1.2	0.89	work	5.15	1.44	1.92	5.92	1.04	0.82
	whistle	5.89	1.59	1.38	6.57	0.73	0.82	worth	4.72	1.75	2.83	5.67	1.2	0.87
	wife	6.34	1.35	1.44	6.64	0.87	0.72	yellow	4.38	1.73	2	5.33	1.18	0.95
	witch	5.55	1.8	1.56	6.41	1.06	0.84	yolk	2.79	1.63	2.81	5.46	1.23	0.89
	worse	6.27	1.28	1.64	6.68	0.64	0.74	youth	5.1	1.29	2.76	5.36	1.39	0.81

Table 10 Positive Valence (POS) and Negative Valence (NEG) Intensity, Latency, and Certainty Norms

POS	NEG										
	MRating	SDRating	MRatingLatency	MCertaintyLatency	Word	MRating	SDRating	MRatingLatency	MCertaintyLatency	SDCertainty	MCertaintyLatency
alcohol	4.09	1.91	2.43	5.72	abdication	3.83	1.54	2.98	4.19	1.95	0.89
all	3.91	1.99	2.08	5.33	abduction	6.07	1.31	2.42	6.14	1.28	0.85
alter	2.91	1.52	2.32	5.24	ache	5.45	1.5	2.09	5.98	1.18	1.02
ambassador	3.22	1.4	2.22	5.11	afraid	5.71	1.45	1.75	6.29	1.11	0.76
appear	3.26	1.56	2.18	5.3	agony	5.81	1.6	1.74	6.36	1.19	0.81
artist	5.2	1.33	2.09	5.7	apathy	4.86	1.82	2.61	5.05	1.65	0.83
baby	5.11	1.78	1.96	5.83	away	3.71	2.06	2.84	5.62	1.38	0.8
ball	4.26	1.48	2.18	5.41	bad	5.67	1.6	1.94	6.26	0.96	0.75
balloon	4.91	1.56	2.17	5.93	bandage	3.79	1.82	2.43	5.48	1.17	0.86
basket	3.67	1.63	2.08	5.39	bang	3.6	1.82	2.39	5.02	1.84	0.89
beer	3.67	1.78	2.04	5.65	beak	2.02	1.66	2.26	5.67	1.71	0.99
bell	3.52	1.6	2.05	5.35	bland	4.26	1.7	2.26	5.36	1.32	0.84
beneficial	5.57	1.26	2.08	6.09	blow	2.76	1.74	2.54	5.38	1.61	0.85
berry	5.3	1.36	1.66	5.83	bomb	6.02	1.51	1.83	6.26	1.08	0.74
best	5.5	1.7	2.03	6.28	bored	4.9	1.85	2.38	6.02	1.2	0.73
bishop	2.39	1.31	1.94	5.41	boulder	2.74	1.81	2.55	5.43	1.5	0.79
bra	3.35	1.78	2.03	5.57	burner	3.17	1.81	2.61	4.98	1.72	0.87
choir	4.2	1.67	1.74	5.7	carp	2.64	1.64	2.43	4.52	1.92	0.67
circus	4.07	1.98	1.97	5.59	ceiling	2	1.56	2.17	5.48	1.74	1.13
coast	5	1.63	2.1	5.98	cement	2.88	1.93	1.94	5.62	1.46	0.86
coke	3.96	1.86	2.46	5.61	charlatan	3.48	1.99	2.82	4.21	1.94	0.84
concrete	2.65	1.51	1.95	5.41	church	2.83	1.83	2.83	5.14	1.8	0.91
cottage	4.57	1.67	2.1	5.59	cigar	3.9	2.16	2.09	5.71	1.33	0.91
count	2.72	1.5	2.1	5.43	clown	3.81	1.95	2.43	5.57	1.45	0.92
custom	3.87	1.67	2.26	5.35	command	4.1	1.86	2.14	5.14	1.68	0.82
dancer	4.87	1.64	1.75	5.7	condemn	5.05	1.96	2.09	6.1	1.03	0.85
democracy	4.37	1.6	2.46	5.28	convent	2.71	1.67	2.63	4.64	1.74	0.79
diamond	5.15	1.56	1.97	5.87	crowd	3.43	1.86	2.36	5.33	1.63	0.83
dog	5.85	1.73	1.5	6.37	cry	5.6	1.67	1.94	6.19	1.4	0.77
dusk	4.35	1.35	2.22	5.33	decay	5.05	1.79	2.04	5.79	1.46	0.81
eagle	3.89	1.7	1.92	5.43	deceit	5.57	1.71	2.23	6.07	1.3	0.96
empire	3.33	1.59	2.09	5.26	demolish	5.12	1.56	2.31	5.48	1.49	0.83
extravagant	5.02	1.78	2.07	5.8	despise	5.69	1.65	1.86	6.24	1.08	0.83



Table 10 (continued)

POS	NEG												
	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency	Word	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency
farm	4.22	1.82	2.28	5.48	1.35	0.88	dim	3.33	1.73	2.14	5.43	1.43	1.02
fashion	4.98	1.77	1.87	5.96	1.11	0.83	ditch	4.43	2	2.38	5.55	1.38	0.91
file	2.67	1.4	1.84	5.67	1.33	0.89	empty	5.05	1.68	2.05	5.81	1.38	0.91
firm	2.8	1.36	2.14	5.43	1.28	0.96	fail	5.83	1.53	2.34	6.29	1.04	0.79
flicker	2.93	1.57	2.15	5.43	1.28	0.86	fear	5.81	1.53	1.86	6.33	0.87	0.8
fly	4.59	1.76	2.02	5.59	1.39	0.78	fire	4.33	1.91	2.26	5.57	1.36	0.9
food	6.2	1.07	1.92	6.41	0.93	0.83	fox	2.21	1.57	2.08	5.86	1.34	1.06
garment	3.74	1.5	2.17	5.28	1.31	0.95	grave	5.57	1.31	2.32	6.07	1.18	0.75
gem	4.93	1.64	2.08	5.61	1.26	0.92	gravel	2.93	1.9	2.28	5.48	1.52	0.89
god	3.83	1.97	2.25	5.35	1.62	0.99	grind	3.6	1.87	2.43	5.48	1.6	0.83
good	6.09	1.26	1.87	6.26	0.95	0.69	hate	5.98	1.55	2.06	6.36	1.21	0.87
grasshopper	2.76	1.65	2.06	5.72	1.28	0.92	heavy	3.93	1.9	2.61	5.29	1.31	0.96
green	4.48	1.62	2.05	5.83	1.2	0.87	highway	2.79	1.55	2.27	5.19	1.52	0.87
gym	4.09	1.71	2.05	5.54	1.21	0.98	hymn	2.4	1.8	2.6	5.67	1.6	0.93
hands	4.26	1.56	2.18	5.63	1.16	0.82	inferior	5.69	1.62	2.01	6.29	1.13	0.7
help	4.26	1.76	2.29	5.63	1.12	0.91	insect	3.79	1.96	2.32	5.81	1.23	0.88
high	4.2	1.96	2.27	5.54	1.33	0.79	insolent	4.43	1.65	2.22	5.02	1.62	0.88
honeymoon	5.48	1.75	2.27	5.98	1.2	0.98	iron	2.48	1.77	1.99	5.38	1.61	0.91
item	3	1.4	2.36	5.46	1.21	0.84	jellyfish	2.71	1.97	2.18	5.76	1.3	0.93
kind	6.07	1.29	1.9	6.2	1.2	0.8	judge	4.1	1.96	2.38	5.74	1.29	0.83
king	3.63	1.62	1.81	5.5	1.3	0.88	lark	2.36	1.61	2.36	4.86	2.03	0.85
kingdom	3.89	1.99	2.19	5.65	1.22	0.89	liquor	2.95	1.78	2.5	5.74	1.56	0.96
know	4.24	1.62	1.99	5.54	1.03	0.98	little	2.95	1.94	2.43	5.43	1.68	0.83
lantern	4.24	1.45	2.11	5.61	1.11	0.82	lure	4.31	1.96	2.32	5.19	1.6	0.71
lead	3.57	1.78	2.17	5.48	1.17	0.9	metal	2.6	1.85	2.27	5.86	1.14	0.87
leaf	4.26	1.57	2.07	5.72	1.11	1.02	mimic	3.36	1.85	2.06	5.36	1.43	0.83
letters	3.78	1.69	2.54	5.33	1.4	0.91	miner	2.76	1.87	2.98	5.31	1.73	0.85
lobster	4.13	2.09	1.83	5.87	1.15	0.84	minor	3.12	1.77	2.86	5.38	1.46	0.91
marry	5.2	1.54	2.15	5.85	1.25	0.77	mischievous	3.86	1.79	2.7	5.24	1.61	0.91
mirror	3.72	1.6	2.28	5.28	1.46	1	mosquito	5.14	1.91	1.97	6.05	1.15	0.75
moment	4.54	1.8	2.18	5.3	1.52	0.92	nail	2.74	1.99	2.35	4.95	1.75	0.9
novel	4.59	1.68	2.07	5.67	1.23	0.74	naval	2.4	1.43	2.58	5.12	1.7	0.87
obedient	2.87	1.6	2.33	5.74	0.95	0.99	need	2.93	1.74	2.16	5.4	1.38	1.06

Table 10 (continued)

POS	NEG												
	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency	Word	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency
older	3.2	1.31	2.18	5.52	1.21	0.99	needle	3.95	1.82	2.04	5.43	1.45	1.05
orange	4.3	1.67	2.25	5.67	1.19	0.86	news	3.31	1.77	2.36	5.31	1.52	0.8
owner	3.09	1.66	2.19	5.3	1.46	1	none	3.29	1.92	2.71	5.17	1.7	0.78
pacifism	3.5	1.85	2.5	5.17	1.62	0.89	nose	2.43	2	2.14	5.83	1.48	1.07
paint	4.59	1.72	1.9	5.8	1.02	0.84	ounce	1.86	1.47	2.27	5.74	1.5	0.96
palace	4.89	1.65	2.24	5.8	1.24	0.85	parish	3.86	2.29	2.51	5.02	2.03	0.88
paper	3.11	1.52	2.05	5.5	1.31	0.84	patch	2.29	1.57	2.19	5.14	1.73	1.02
park	5.28	1.36	2.16	5.87	1.05	0.89	pause	2.81	1.86	2.05	5.36	1.57	0.88
pea	3.37	1.6	1.96	5.74	1.1	0.93	peel	2.6	1.87	2.44	5.45	1.76	0.87
pick	2.96	1.15	2.06	5.39	1.34	0.95	perjury	4.95	1.53	2.24	5.24	1.54	0.85
pineapple	5.26	1.5	2.05	5.96	1.09	0.93	phantom	4	1.83	2.33	5.36	1.59	0.78
plum	4.09	1.47	2.08	5.59	1.02	0.93	phase	2.5	1.73	2.15	5.29	1.63	0.97
pocket	3.78	1.47	2.01	5.48	1.26	0.98	punch	4.57	1.88	2.45	5.86	1.22	0.94
polite	5.17	1.29	2.07	5.61	1.29	0.88	punish	5.6	1.61	2.07	6.12	1.09	0.77
power	4.65	1.74	2.28	5.72	1.2	0.77	quack	2.05	1.74	2.2	5.81	1.4	0.93
purse	4.13	1.63	2.04	5.48	1.21	0.9	rail	2.5	1.63	2.45	5.55	1.29	0.89
red	3.63	1.7	2.45	5.43	1.39	0.89	rain	3.26	2.05	2.48	5.79	1.35	0.95
resort	5.48	1.81	2.02	6.04	1.03	0.81	rear	2.62	1.86	2.45	5.38	1.7	0.88
restore	4.33	1.62	2.13	5.43	1.28	0.91	regulation	3.93	1.79	2.23	5.29	1.45	0.88
ride	4.54	1.6	2.08	5.74	1.12	0.95	rough	4.29	1.89	2.45	5.26	1.64	0.83
river	4.91	1.46	2.14	5.43	1.54	0.84	sad	6.12	1.52	2.02	6.52	0.94	0.76
scholar	4.65	1.43	2.45	5.7	1.28	0.85	shrimp	2.17	1.75	2.01	5.5	1.67	0.9
school	3.93	1.58	2.14	5.54	1.35	1.02	skyscraper	1.95	1.4	2.28	5.55	1.55	0.94
seek	3.5	1.63	2.16	5.52	1.22	1	slime	2.9	1.76	2.49	5.29	1.67	0.92
shape	3.26	1.51	2.13	5.04	1.55	0.87	sluggish	4.93	1.6	2.05	5.6	1.43	0.72
show	4.67	1.65	1.85	5.63	1.14	0.82	smoke	4.12	1.84	2.65	5.64	1.27	0.9
sister	4.76	1.82	2.23	5.87	1.11	0.86	snake	4.43	2.17	1.81	5.83	1.54	0.83
soar	4.76	1.85	1.9	5.87	1.19	0.87	sore	4.55	1.99	2.14	5.67	1.48	0.93
soft	5.41	1.44	2.04	5.85	1.35	0.75	spade	2.43	1.65	1.95	5.17	1.67	0.95
song	5.83	1.4	1.74	6.09	1.05	0.91	spray	2.52	1.86	2.27	5.19	1.6	0.94
steak	4.54	1.91	1.95	5.87	1.24	0.84	stick	2.38	1.67	1.81	5.57	1.4	0.81
stem	3.07	1.78	2.24	5.28	1.33	0.84	stomach	3.05	1.86	2.63	5.48	1.61	0.91
tangerine	4.91	1.6	2.08	5.98	1.04	0.94	strand	2.57	1.85	2.38	5.29	1.74	0.84

Table 10 (continued)

POS	NEG													
	Word	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency	Word	MRating	SDRating	MRatingLatency	MCertainty	SDCertainty	MCertaintyLatency
	taste	5.13	1.44	2.04	5.65	1.4	0.84	straw	2.14	1.57	1.98	5.79	1.51	0.82
	telephone	3.96	1.52	2.49	5.28	1.39	0.86	stroke	5.14	1.96	2.51	6.07	1.2	0.79
	temple	3.35	1.51	2.18	5.59	1.2	0.85	stumble	4.43	1.63	2.17	5.69	1.24	0.89
	thinner	3.59	1.8	2.5	5.52	1.3	0.98	sweat	3.57	2.03	2.28	5.43	1.65	0.82
	thoughtful	5.54	1.46	1.81	6.04	1.13	0.99	sweep	2.76	1.85	2.48	5.05	1.97	0.95
	thrifty	4.17	1.8	2.3	5.7	1.07	0.87	take	3.12	1.77	2.25	5.6	1.53	0.85
	urban	4.17	1.47	2.24	5.46	1.11	0.87	test	4.69	1.77	2.76	6.02	1.22	0.87
	vase	3.43	1.71	1.97	5.54	1.21	0.89	thermometer	2.24	1.74	2.51	5.64	1.46	1.02
	vet	3.37	1.65	2.07	5.35	1.62	0.86	thin	3.17	1.79	2.71	5.07	1.88	0.81
	victory	5.85	1.55	1.79	6.11	1.12	0.93	throat	2.48	1.8	2.08	5.33	1.59	0.94
	village	4.07	1.76	1.72	5.37	1.4	0.83	tomb	4.93	1.83	2.48	5.83	1.17	0.93
	violin	4.07	1.65	1.86	5.72	1.07	0.83	tool	2.31	1.72	2.27	5.5	1.6	0.93
	vow	4.37	1.88	2	5.48	1.35	0.74	tough	3.83	1.86	2.26	5.4	1.47	0.97
	walnut	3.04	1.52	2	5.59	1.17	0.82	trance	3.33	1.66	2.5	5.07	1.54	0.93
	welfare	4.26	1.79	2.56	5.5	1.41	1.08	triangle	1.93	1.6	2.31	5.93	1.57	0.97
	white	3.13	1.36	2.74	5.33	1.28	0.97	war	5.98	1.63	1.83	6.4	0.91	0.8
	wolf	3.48	1.66	2.18	5.54	1.17	0.9	warn	1.93	1.74	2.3	6.24	0.98	0.95
	world	4.61	1.67	2	5.7	1.36	0.78	whiff	2.6	1.61	2.31	5.31	1.62	1.02
	yolk	2.78	1.5	2.07	5.72	1.2	0.81	wind	2.64	1.86	2.09	5.62	1.34	0.95
	zoo	4.33	1.98	2.15	5.8	1.31	0.85	wither	4.74	2.08	2.22	5.67	1.57	0.72
	zoology	3.61	1.83	2.16	5.33	1.59	0.88	zero	2.69	1.87	2.06	5.55	1.53	0.95

**Table 11** Analysis of variance results for relations between five linguistic variables and *M* intensity ratings and rating *SDs* in Experiments 2, 3, and 4

	Statistic			
	<i>F</i> (ambiguity)	<i>p</i> (ambiguity)	<i>F</i> (intensity)	<i>p</i> (intensity)
Experiment 2	2.59	.11	2.57	.11
Experiment 3	1.46	.23	.88	.35
Experiment 4	1.85	.18	3.81	.06

*Note:* In each experiment, a 2 (attribute ambiguity: high vs. low ambiguity words)  $\times$  2 (attribute intensity: high vs. low intensity words)  $\times$  5 (linguistic variables: age of acquisition, frequency, length, number of phonemes, number of syllables) mixed ANOVA with repeated measures on the third factor was computed. The purpose of each ANOVA was to determine whether this set of linguistic variables differed reliability as a function of attribute ambiguity or intensity. *F*(ambiguity) is the omnibus test of whether the linguistic variables differed as a function of attribute intensity, and *p*(ambiguity) is its significance level. *F*(intensity) is the omnibus test of whether the linguistic variables differed as a function of attribute intensity, and *p*(intensity) is its significance level

**Table 12** Values of five linguistic variables and values of attribute ambiguity (*SD*) and attribute intensity (*M*) for three semantic attributes in Experiments 2–4

Word	Meaningfulness_ <i>M</i>	Meaningfulness_ <i>SD</i>	<i>M</i> meaning	<i>SD</i> meaning	AoA	Frequency	Length	Number of Phonemes	Number of Syllables
Experiment 2									
accord	3.07	1.61	low	low	10.51346	14.29	6	5	2
Ah	3.22	2.18	low	high		152.16	2	1	1
Ail	3.02	1.51	low	low	9.71	0.76	3	2	1
battle	4.95	1.62	high	low	6.697436	89.06	6	4	2
because	3.1	1.97	low	high	5.037854	790.22	7	5	2
bible	4.89	1.98	high	high	4.95	32.74	5	4	2
birth	5.05	1.96	high	high	5.807605	55.51	5	3	4
black	4.93	1.97	high	high	3.734473	307.09	5	4	2
both	3.2	1.56	low	low	4.453318	573.15	4	3	2
brain	4.94	1.91	high	high	6.180865	72.26	5	4	2
cancer	4.92	1.82	high	high	8.876412	19.41	6	5	1
carat	3.13	1.49	low	low	13.25	0.78	5	5	2
caucus	3.11	1.77	low	high	14.88	1.11	6	5	1
chute	3.17	1.52	low	low	9.005169	1.68	5	3	1
clean	4.93	1.49	high	low	4.219206	72.42	5	4	1
clove	3.15	1.53	low	low	10.11144	1.7	5	4	1
coast	4.93	1.8	high	high	6.568829	53.09	5	4	1
custard	3.13	1.56	low	low	8.481067	1.67	7	6	2
cymbal	3.08	1.91	low	high	9.94	0.31	6	5	2
dollar	4.92	1.97	high	high	4.859988	33	6	4	2
doorman	3.19	1.73	low	high	9.32	0.47	7	6	2
earn	4.92	1.39	high	low	7.559425	19.7	4	2	1
estate	4.95	1.37	high	low	10.07954	42.08	6	5	2
ever	3.17	2.03	low	high	5.648831	540.94	4	3	2
fetish	3.17	1.8	low	high	12.40721	1.95	6	5	2

**Table 12** (continued)

Word	Meaningfulness_M	Meaningfulness_SD	Mmeaning	SDmeaning	AoA	Frequency	Length	Number of Phonemes	Number of Syllables
fore	3.08	1.85	low	high	11.69862	11.73	4	3	1
formaldehyde	3.08	1.99	low	high	11.39	0.5	12	11	4
gable	3.05	2.03	low	high	10.293	2.79	5	4	2
germ	4.85	1.51	high	low	6.422496	5.25	4	3	1
gone	3.19	1.79	low	high	3.365325	292.14	4	3	1
guest	5.02	1.41	high	low	6.427262	31.34	5	4	1
gullet	3.15	1.96	low	high	13.35	0.58	6	5	2
hoarse	3.21	1.53	low	low	9.019742	8.93	6	4	1
hope	4.93	1.96	high	high	5.928156	230.71	4	3	1
hurricane	4.93	1.9	high	high	7.35	5.62	9	7	3
husband	4.92	1.99	high	high	5.932859	144.85	7	7	2
idea	4.95	1.87	high	high	6.272087	2.36	4	4	2
insult	4.9	1.47	high	low	8.673219	13.92	6	6	2
iota	3.16	1.99	low	high	12.47	0.8	4	4	3
labyrinth	3.18	1.94	low	high	10.53	3.2	9	8	3
left	3.15	1.9	low	high	5.594797	572.6	4	4	1
lint	3.1	1.36	low	low	7.92643	0.69	4	4	1
literal	3.06	1.41	low	low	10.11	5.89	7	7	3
long	4.92	1.66	high	low	5.052004	896.36	4	3	1
lute	3.16	1.95	low	high	10.77	3.25	4	3	1
mince	3.17	1.37	low	low	10.95529	2.32	5	4	1
nervous	4.9	1.6	high	low	7.436119	42.73	7	5	2
other	3.07	1.94	low	high	5.751221	1675.63	5	3	2
paradox	3.15	1.63	low	low	13.78	5.03	7	8	3
pear	4.94	1.87	high	high	4.555914	3.84	4	3	1
piano	4.87	1.92	high	high	5.34485	23.9	5	5	3
sable	3.25	1.57	low	low	11.8549	2.19	5	4	2
shoe	4.85	1.96	high	high	3.128213	13.88	4	2	1
skiing	4.92	1.97	high	high	7.89	3.17	6	5	2
skyscraper	4.9	1.56	high	low	8.26	0.41	10	9	3
solemnity	3.14	1.59	low	low	12.33	5.44	9	9	4
soon	3.07	1.57	low	low	4.520285	327.06	4	3	1
star	4.9	1.66	high	low	4.1549	61.17	4	4	1
states	4.9	1.65	high	low		213.46	6	5	1
stout	4.92	1.6	high	low	10.79503	20.19	5	4	1
sweet	4.95	1.54	high	low	3.949022	108.9	5	4	1
symbol	4.9	1.84	high	high	7.708955	19.09	6	5	2
tear	4.98	1.6	high	low	5.200738	23.2	4	3	1
tennis	4.92	1.86	high	high	7.025303	18.69	6	5	2
tune	4.87	1.47	high	low	6.658415	17.78	4	3	1
ugly	4.9	1.55	high	low	4.808413	27.4	4	4	2
universal	4.87	1.98	high	high	10.21	38.1	9	8	4
vane	3.1	1.69	low	low	10.58	4.75	4	3	1
vertical	3.14	1.55	low	low	9.71	14.23	8	6	3
very	3.07	1.88	low	high	4.720517	1511.55	4	4	2
walking	4.92	1.46	high	low	3.982672	75.53	7	5	2
wrung	3.15	1.49	low	low		5.19	5	3	1
Experiment 3									
abhor	3.16	2.21	low	high	13.76	1.21	5	5	2
ambition	3.25	2.16	low	high	10.41335	22.62	8	6	3
berry	5.56	1.7	high	low	5.8701	7.58	5	4	2
bland	3.34	1.7	low	low	9.113564	5.58	5	5	1

Table 12 (continued)

Word	Meaningfulness_M	Meaningfulness_SD	Mmeaning	SDmeaning	AoA	Frequency	Length	Number of Phonemes	Number of Syllables
blunder	3.18	1.69	low	low	11.05615	4.61	7	6	2
bred	3.04	1.6	low	low	8.807372	15.92	4	4	1
busy	3.25	1.63	low	low	5.50319	59.73	4	4	2
cell	5.09	2.06	high	high	9.626948	37.3	4	3	1
considerable	2.98	2.12	low	high	10.42	92.36	12	10	5
cymbal	5.09	2.18	high	high	9.94	0.31	6	5	2
dandruff	5.42	2.14	high	high	10.16735	0.34	8	7	2
destroyer	5.09	2.2	high	high	7.74	2.46	9	7	2
dinner	5.38	2.19	high	high	3.9439	114.8	6	4	2
drink	5.09	1.57	high	low	3.570535	93.5	5	5	1
drizzle	5.54	1.51	high	low	7.56	1.89	7	5	2
drop	3.16	1.69	low	low	4.385436	61.13	4	4	1
easy	2.83	1.53	low	low	5.120017	144.16	4	3	2
eternity	2.98	2.26	low	high	8.6	11.76	8	7	4
ever	2.95	2.22	low	high	5.648831	540.94	4	3	2
excuse	3.13	2.14	low	high	6.63537	44.28	6	7	2
fishing	5.05	1.69	high	low	4.72	37.65	7	5	2
fork	5.58	2.17	high	high	4.040786	10.64	4	4	1
formaldehyde	5.34	2.27	high	high	11.39	0.5	12	11	4
freedom	3.03	2.11	low	high	7.687589	62.16	7	6	2
frugal	3	1.67	low	low	11.37	1.75	6	6	2
gasket	5.21	2.18	high	high	10.9	0.25	6	6	2
ginger	5.18	1.7	high	low	7.732188	6.7	6	5	2
graph	5.49	1.62	high	low	7.986992	5.51	5	4	1
grown	3	1.67	low	low	5.098	71.16	5	4	1
guest	5.15	1.73	high	low	6.427262	31.34	5	4	1
gullet	5.33	2.11	high	high	13.35	0.58	6	5	2
higher	3.2	1.71	low	low	4.525095	127.71	6	3	1
innate	3.04	2.15	low	high	12.83718	5.18	6	4	2
lad	5.04	2.18	high	high	7.568821	39.6	3	3	1
lady	5.58	1.49	high	low	4.344326	2272.81	4	4	2
lard	5.13	2.17	high	high	9.808078	1.54	4	4	1
lenient	2.81	1.54	low	low	11.67	1.84	7	7	2
light	5.46	1.62	high	low	4.670701	2377.06	5	3	1
lunch	5.46	1.7	high	low	4.459989	40.18	5	4	1
method	3.07	1.51	low	low	9.865472	85.8	6	5	2
must	2.93	2.21	low	high	5.034231	1009.42	4	4	1
No	3.14	2.35	low	high	1.813204	2845.05	2	2	1
oblivious	3.02	2.2	low	high	12.56	4.56	9	8	3
once	3.11	2.23	low	high	4.972281	597.48	4	4	1
oregano	5.39	2.26	high	high	10.88	0.32	7	7	4
pelt	5.31	2.16	high	high	10.836	0.88	4	4	1
pew	5.38	2.21	high	high	8.56	4.22	3	3	1
punch	5.48	1.62	high	low	5.7677	11.93	5	4	1
pup	5.4	2.12	high	high	5.075359	2.62	3	3	1
races	5.05	1.67	high	low	5.262667	29.38	5	5	2
raise	3.36	1.69	low	low	6.530394	49.49	5	3	1
reluctant	2.59	1.53	low	low	11.11	12.72	9	9	3
rule	2.82	1.5	low	low	5.639677	86.08	4	3	1
scorching	4.72	1.66	high	low		2.45	9	7	2
sent	2.88	1.5	low	low	5.846357	213.72	4	4	1
sheets	5.46	1.67	high	low	5.329067	20.31	6	4	1

**Table 12** (continued)

Word	Meaningfulness_M	Meaningfulness_SD	Mmeaning	SDmeaning	AoA	Frequency	Length	Number of Phonemes	Number of Syllables
shower	5.88	1.44	high	low	6.33164	17.82	6	3	1
slice	4.23	1.59	high	low	5.562588	7.89	5	4	1
slide	5	1.6	high	low	4.773662	13.56	5	4	1
sword	5.32	2.17	high	high	6.048241	50.8	5	4	1
sympathetic	3.16	2.21	low	high	9.39	46.57	11	10	4
tame	3.31	1.6	low	low	8.147734	7.25	4	3	1
theory	3.24	2.12	low	high	10.02059	103.89	6	5	2
they	3.12	2.11	low	high	4.073859	4280.32	4	2	1
thimble	5.25	2.19	high	high	8.05	1.21	7	5	2
thought	3.09	2.15	low	high	5.098973	758.37	7	3	1
trouble	3.23	1.69	low	low	5.417133	136.74	7	5	2
truth	2.72	2.13	low	high	5.14693	176.71	5	4	1
upon	3.07	2.11	low	high	5.664352	1007.36	4	4	2
veal	5.24	2.21	high	high	10.9302	1.8	4	3	1
warn	3.11	1.59	low	low	6.376662	14.28	4	4	1
wind	5.48	1.53	high	low	4.319437	139.62	4	4	1
Experiment 4									
accordion	5.43	2.1	high	high	8.61	0.79	9	8	3
aim	3.09	1.71	low	high	6.648794	42.54	3	2	1
aluminum	5.5	1.96	high	high	7.26	0.9	8	8	4
ambition	3.07	1.55	low	low	10.41335	22.62	8	6	3
ampere	3.05	1.98	low	high		0.27	6	5	2
atom	5.48	1.95	high	high	11.51374	6.85	4	4	2
bacteria	5.48	1.95	high	high	9.28	6.15	8	8	3
ball	5.48	1.49	high	low	3.044339	64.47	4	3	1
balloon	5.48	1.92	high	high	3.66557	13.65	7	5	2
beneficial	3.13	1.58	low	low	10.84	9.04	10	8	4
bier	3.1	1.89	low	high		1.56	4	3	1
biology	5.52	1.57	high	low	9.915135	6.7	7	7	4
broader	3.13	1.58	low	low	8.741007	9.38	7	5	2
cardinal	5.55	1.88	high	high	8	6.15	8	7	3
cheeks	5.48	1.58	high	low	5.036467	40.69	6	4	1
condemn	3.08	1.59	low	low	11.22629	6.17	7	6	2
dinner	5.53	2.05	high	high	3.9439	114.8	6	4	2
distinct	3.07	1.59	low	low	10.49228	38.22	8	8	2
emancipation	3.05	1.95	low	high	12.15	0.41	12	10	5
farthing	3.18	2.07	low	high	12.67	2.06	8	6	2
few	3.03	1.51	low	low	5.642993	508.28	3	3	1
flair	3.12	1.88	low	high	10.2317	1.59	5	4	1
garbage	5.52	1.65	high	low	5.26037	2.87	7	6	2
gardenia	5.6	2.09	high	high	11.94	0.33	8	8	3
gravel	5.45	1.57	high	low	7.277124	12.32	6	5	2
helmet	5.52	1.53	high	low	5.346782	7.53	6	6	2
hexagon	5.5	1.94	high	high	7.95	0.4	7	8	3
humble	3.08	1.69	low	low	9.242345	6.17	6	5	2
husband	5.55	1.55	high	low	5.932859	144.85	7	7	2
infinite	3.17	2.17	low	high	9.6	23.55	8	7	3
jangle	3.15	1.34	low	low	8	0.65	6	5	2
lantern	5.52	1.62	high	low	8.55	12.48	7	6	2
letter	5.53	1.58	high	low	5.340541	221.36	6	4	2
mattress	5.53	1.59	high	low	5.850752	4.59	8	6	2
mirror	5.53	1.83	high	high	4.998657	38.42	6	3	1

Table 12 (continued)

Word	Meaningfulness_M	Meaningfulness_SD	Mmeaning	SDmeaning	AoA	Frequency	Length	Number of Phonemes	Number of Syllables
misuse	3.13	1.49	low	low	9.68	2.36	6	6	2
need	3.1	1.97	low	high	4.629734	386.89	4	3	1
once	3.05	1.94	low	high	4.972281	497.48	4	4	1
orchestra	5.58	1.54	high	low	9.44	7.53	9	8	3
out	3.07	1.99	low	high	4.028419	350.65	3	2	1
over	3.13	1.85	low	high	5.254416	1414.08	4	3	2
pardon	3.08	1.58	low	low	7.964002	35.64	6	5	2
pause	3.15	1.57	low	low	7.067198	40.93	5	3	1
penicillin	5.47	2.07	high	high	9.56	0.84	10	9	4
plait	3.08	2.04	low	high		1	5	4	1
plate	5.48	1.55	high	low	4.313011	41.67	5	4	1
pliers	5.48	1.9	high	high	8.297345	0.43	6	5	1
pope	5.58	1.81	high	high	8.83	22.74	4	3	1
portray	3.1	1.65	low	low	13.69	1.99	7	6	2
propeller	5.47	1.56	high	low	7	3.39	9	7	3
real	3.17	1.99	low	high	5.33285	221.7	4	4	1
refresh	3.12	1.35	low	low	10.83	2.22	7	6	2
requirement	3.07	1.93	low	high	8.44	15.39	11	10	3
responsive	3.05	1.43	low	low	10.58	5.03	10	9	3
sandal	5.62	1.98	high	high	5.21	1	6	5	2
sergeant	5.5	1.91	high	high	9.378242	24.22	8	7	2
sine	3.17	2.2	low	high	13.69	2.06	4	3	1
soap	5.48	1.62	high	low	3.892256	5.87	4	3	1
spatula	5.52	2.03	high	high	7.32	0.25	7	7	3
squib	3.1	1.95	low	high		0.3	5	5	1
steel	5.53	1.57	high	low	8.024475	42.38	5	4	1
taper	3.17	1.6	low	low	11.04445	2.1	5	4	2
teacher	5.5	1.86	high	high	4.411538	55.31	7	4	2
theoretical	3.12	1.9	low	high	13.44	16.17	11	9	4
trouble	3.15	1.55	low	low	5.417133	136.74	7	5	2
two	5.48	2.09	high	high	2.810897	1461.3	3	2	1
valor	3.12	1.68	low	low	12.44766	35.64	5	4	2
vase	5.46	1.5	high	low	6.815105	5.87	4	3	1
verb	5.55	1.56	high	low	8.154128	6.09	4	3	1
view	3.1	1.68	low	low	6.775729	211.4	4	3	1
We	3.05	2.05	low	high	3.404063	361.25	2	2	1
zipper	5.52	1.36	high	low	4.910885	0.85	6	4	2

## References

- Abadie, M., & Camos, V. (2019). False memory at short and long term. *Journal of Experimental Psychology: General*, *148*(8), 1312–1334.
- Abadie, M., Waroquier, L., & Terrier, P. (2013). Gist memory and the unconscious thought effect. *Psychological Science*, *24*(7), 1253–1259.
- Barnhardt, T. M., Choi, H., Gerkens, D. R., & Smith, S. M. (2006). Output position and word relatedness effects in a DRM paradigm: Support for a dual-retrieval process theory of free recall and false memories. *Journal of Memory and Language*, *55*(2), 213–231.
- Bookbinder, S. H., & Brainerd, C. J. (2016). Emotion and false memory: The context-content paradox. *Psychological Bulletin*, *142*(12), 315–351.
- Bouwmeester, S., & Verkoeijen, P. P. J. L. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language*, *65*(1), 32–41.
- Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings* (tech. Rep. No. C-1). Gainesville, FL: The Center for Research in psychophysiology, University of Florida.
- Brainerd, C. J. (2018). The emotional ambiguity hypothesis: A large-scale test. *Psychological Science*, *25*(10), 1706–1715.
- Brainerd, C. J., Chang, M., & Bialer, D. M. (2021a). Emotional ambiguity and memory. *Journal of Experimental Psychology: General*, *150*(8), 1476–1499.
- Brainerd, C. J., Chang, M., Bialer, D. M., & Toglia, M. P. (2021b). Semantic ambiguity and memory. *Journal of Memory and Language*. Article Number, 104286.



- Brainerd, C. J., Nakamura, K., & Reyna, V. F. (2017). Overdistribution: Categorical judgments produce them, confidence ratings reduce them. *Journal of Experimental Psychology: General*, *146*(1), 20–40.
- Brainerd, C. J., Reyna, V. F., Gomes, C. F. A., Kenney, A. E., Gross, C. J., Taub, E. S., & Spreng, R. N. (2014). Dual-retrieval models and neurocognitive impairment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 41–65.
- Brainerd, C. J., Reyna, V. F., & Howe, M. L. (2009). Trichotomous processes in early memory development, aging, and neurocognitive impairment: A unified theory. *Psychological Review*, *116*(4), 783–832.
- Bruer, K. C., & Pozzulo, J. D. (2014). Familiarity and recall memory for environments: A comparison of children and adults. *Journal of Applied Developmental Psychology*, *35*(4), 318–325.
- Bugg, J. M. (2014). Conflict-triggered top-down control: Default mode, last resort, or no such thing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 567–587.
- Citron, F. M., Weekes, B. S., & Ferstl, E. C. (2014). How are affective word ratings related to lexicosemantic properties? Evidence from the Sussex affective word list. *Applied Psycholinguistics*, *35*(2), 313–331.
- Durda, K., & Buchanan, L. (2006). WordMine2 [online] Available from: <http://web2.uwindsor.ca/wordmine>
- Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 717–726.
- Engelthaler, T., & Hills, T. T. (2018). Humor norms for 4,997 English words. *Behavior Research Methods*, *50*(3), 1116–1124.
- Ferré, P., Guasch, M., Martínez-García, N., Fraga, I., & Hinojosa, J. A. (2017). Moved by words: Affective ratings for a set of 2,266 Spanish words in five discrete emotion categories. *Behavior Research Methods*, *49*(3), 1082–1094.
- Gomes, C. F. A., Brainerd, C. J., Nakamura, K., & Reyna, V. F. (2014). Markovian interpretations of dual retrieval processes. *Journal of Mathematical Psychology*, *59*, 50–64.
- Gomes, C. F. A., Brainerd, C. J., & Stein, L. M. (2013). Effects of emotional valence and arousal on recollective and nonrecollective recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 663–677.
- Greene, N. R., Chism, S., & Naveh-Benjamin, M. (in press). *Levels of specificity in episodic memory: Insights from response accuracy and subjective confidence ratings in older adults and in younger adults under full or divided attention*. General.
- Greene, N. R., & Naveh-Benjamin, M. (2020). A specificity principle of memory: Evidence from aging and associative memory. *Psychological Science*, *31*(3), 316–331.
- Greene, N. R., & Naveh-Benjamin, M. (2021). *The effects of divided attention at encoding on specific and gist-based associative episodic memory*. Advance online publication.
- Henriques, J. B., & Davidson, R. J. (2000). Decreased responsiveness to reward in depression. *Cognition and Emotion*, *14*(5), 711–724.
- Hinojosa, J. A., Martínez-García, N., Villalba-García, C., Fernández-Folgueiras, U., Sánchez-Carmona, A., Pozo, M. A., & Montoro, P. R. (2016). Affective norms of 875 Spanish words for five discrete categories and two emotional dimensions. *Behavior Research Methods*, *48*(1), 272–284.
- Juhász, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131*(5), 684–712.
- Juhász, B. J., Lai, Y.-H., & Woodcock, M. L. (2015). A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods*, *47*(4), 1004–1019.
- Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, *1*(2), 99–113.
- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 804–823.
- Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*(4), 917–940.
- Lampinen, J. M., & Odegard, T. N. (2006). Memory editing mechanisms. *Memory*, *14*(6), 649–654.
- Lampinen, J. M., Odegard, T. N., Blackshear, E., & Togliani, M. P. (2005). Phantom ROC. In D. T. Rosen (Ed.), *Trends in experimental psychology research* (pp. 235–267). NOVA Science Publishers.
- Marche, T. A., Briere, J. L., & von Baeyer, C. L. (2016). Children's forgetting of pain-related memories. *Journal of Pediatric Psychology*, *41*(2), 220–231.
- Marschark, M., Richman, C. L., Yuille, J. C., & Hunt, R. R. (1987). The role of imagery in memory: On shared and distinctive information. *Psychological Bulletin*, *102*(2), 28–41.
- Mattek, A. M., Wolford, G. L., & Whalen, P. J. (2017). A mathematical model captures the structure of subjective affect. *Perspectives on Psychological Science*, *12*(3), 508–526.
- Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005a). Emotional category data on images from the international affective picture system. *Behavioral Research Methods*, *37*(4), 626–630.
- Mikels, J. A., Larkin, G. R., Reuter-Lorenz, P. A., & Carstensen, L. L. (2005b). Divergent trajectories in the aging mind: Changes in working memory for affective versus visual information with age. *Psychology and Aging*, *20*(4), 542–553.
- Neath, I., & Surprenant, A. M. (2020). Concreteness and disagreement: Comment on Pollock (2018). *Memory & Cognition*, *48*(4), 683–690.
- Niedzialkowska, D., & Nieznański, M. (in press). Recollection of "true" feedback is better than "false" feedback independently of a priori beliefs: An investigation from the perspective of dual-recollection theory. *Memory*.
- Nieznański, M., & Obidzinski, M. (2019). Verbatim and gist memory and individual differences in inhibition, sustained attention, and working memory capacity. *Journal of Cognitive Psychology*, *31*(1), 16–33.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, *76*(1, Pt.2), 1–25.
- Palmer, J. E., & Dodson, C. S. (2009). Investigating the mechanisms fueling reduced false recall of emotional material. *Cognition & Emotion*, *23*(2), 238–259.
- Pexman, P. M., Muraki, E., Sidhu, D. M., Siakaluk, P. D., Siakaluk, P. D., & Yap, M. J. (2019). Quantifying sensorimotor experience: Body-object interaction ratings for more than 9,000 English words. *Behavior Research Methods*, *51*(2), 453–466.
- Pollock, L. (2018). Statistical and methodological problems with concreteness and other semantic variables: A list memory experiment case study. *Behavior Research Methods*, *50*(3), 1198–1216.
- Schock, J., Cortese, M. J., Khanna, M. M., & Toppi, S. (2012). Age of acquisition estimates for 3,000 disyllabic words. *Behavior Research Methods*, *44*(4), 971–977.
- Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, *51*(3), 1258–1270.
- Singer, M., & Remillard, G. (2008). Veridical and false memory for text: A multiprocess analysis. *Journal of Memory and Language*, *59*(1), 18–35.

- Singer, M., & Spear, J. (2015). Phantom recollection of bridging and elaborative inferences. *Discourse Processes*, 52(5–6), 356–375.
- Smith, T. A., & Kimball, D. R. (2012). Revisiting the rise and fall of false recall: Presentation rate effects depend on retention interval. *Memory*, 20(6), 535–553.
- Stadthagen-González, H., Ferré, P., Pérez-Sánchez, P. A., Imbault, C., & Hinojosa, J. A. (2018). Norms for 10,491 Spanish words for five discrete emotions: Happiness, disgust, anger, fear, and sadness. *Behavior Research Methods*, 50(5), 1943–1952.
- Stevenson, R. A., Mikels, J. A., & James, T. W. (2007). Characterization of the affective norms for English words by discrete emotional categories. *Behavior Research Methods*, 39(4), 1020–1024.
- Storbeck, J., & Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory: Mood and the false memory effect. *Psychological Science*, 16(10), 785–791.
- Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Erlbaum.
- Tractenberg, S. G., Viola, T. W., Gomes, C. F. A., Wearick-Silva, L. E., Kristensen, C. H., Stein, L. M., & Grassi-Oliveira, R. (2015). Dual-memory processes in crack cocaine dependents: The effects of childhood neglect on recall. *Memory*, 23(7), 955–971.
- Underwood, B. J., & Schulz, R. W. (1960). *Meaningfulness and verbal learning*. Chicago: Lippincott.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.