Psychonomic Bulletin & Review (2023) 30:2315–2327
https://doi.org/10.3758/s13423-023-02291-5

BRIEF REPORT

The recognition effects of attribute ambiguity

Minyu Chang[1] · C. J. Brainerd[2]

Accepted: 2 April 2023 / Published online: 2 May 2023

Abstract

When examining memory effects of semantic attributes, it is common practice to manipulate normed mean (*M*) ratings of the attributes (i.e., attribute intensity) in learning materials. Meanwhile, the standard deviations (*SD*s) of attribute ratings (i.e., attribute ambiguity) are usually treated as indexes of measurement error. However, some recent research found that recall accuracy varied as a function of both the intensity and ambiguity of semantic attributes such as valence, categorization, concreteness, and meaningfulness. These findings challenged the traditional interpretation of attribute rating *SD*s as noise indexes. In the current study, we examined the recognition effects of ambiguity, intensity, and Ambiguity × Intensity interactions for 21 attributes using mega study data for over 5,000 words. Our results showed that attribute ambiguity had reliable recognition effects beyond those of attribute intensity, and that it sometimes explained more unique variance in recognition than attribute intensity. Thus, we concluded that attribute ambiguity is a distinct psychological dimension of semantic attributes, which is processed separately from attribute intensity during encoding. Two theoretical hypotheses had been proposed for the memory effects of attribute ambiguity. We discuss the implications of our findings for the two theoretical hypotheses about how attribute ambiguity influences episodic memory.

Keywords Semantic attributes · Attribute ambiguity · Attribute intensity · Recognition memory · Megastudy data

Introduction

In the memory literature, it has often been found that accuracy is affected by the intensities of certain semantic attributes, such as valence and arousal (Adelman & Estes, 2013; Chang et al., 2021; Cortese & Khanna, 2022), concreteness (Fliessbach et al., 2006; Hamilton & Rajaram, 2001), meaningfulness (Rae, 1979), and many others. Attribute intensity is typically manipulated by varying mean (*M*) attribute ratings from rating norms (e.g., Brysbaert et al., 2014; Pexman et al., 2019; Scott et al., 2019; Toglia & Battig, 1978; Warriner et al., 2013). Recently, a second property of attribute ratings, their ambiguity, has also been found to affect memory accuracy – where ambiguity is manipulated by varying *SD*s of attribute ratings (Brainerd et al., 2020, 2021, 2022). Further, the effects of ambiguity on memory have been dissociated from intensity's effects.

These recent studies of attribute ambiguity are theoretically significant because they demonstrate that people process two distinct forms of information about semantic attributes as they encode items in memory experiments. However, the generality of the studies' findings is limited in two important respects: They are confined to recall, and only four attributes have been investigated to date (valence, categorization, concreteness, and meaningfulness). We address both of those limitations in the present study. More explicitly, we investigated the effects of attribute ambiguity on recognition, and we investigated those effects for 21 different attributes. To do so, we relied on data from two mega recognition studies (Cortese et al., 2010, 2015), combined with ambiguity ratings from published attribute rating norms. Before reporting our findings, we provide a sketch of recent research on attribute ambiguity.

Attribute ambiguity: A brief overview

To study the memory effects of semantic content, it is common to manipulate semantic attributes in study materials and

✉ Minyu Chang
minyu.chang@mcgill.ca

1 Department of Psychology, McGill University, 2001 Avenue McGill College, Montréal, QC H3A 1G1, Canada

2 Department of Psychology and Human Neuroscience Institute, Cornell University, Ithaca, NY, USA

test how memory accuracy is affected. In that connection, researchers often rely on data from attribute-norming studies, in which large numbers of participants rate words (or pictures) for the perceived intensity of one or more attributes. Then, the attributes' memory effects are studied by manipulating the normed intensity of to-be-remembered items, which is indexed by their *M* ratings.

In addition to items' *M* ratings, semantic word norms provide their rating *SD*s, which are customarily interpreted as indexes of measurement error; that is, ratings with higher *SD*s are viewed as less reliable than those with lower *SD*s (Brainerd et al., 2020; Pollock, 2018; Toglia & Battig, 1978). However, recent findings support an alternative interpretation that rating *SD*s are indexes of a psychological property – namely, the level of ambiguity that attaches to an attribute's perceived intensity. This line of investigation began with Mattek et al.'s (2017) study of valence ambiguity, which showed that the judged ambiguity of emotional pictures' valence determines how strongly their valence intensity ratings correlate with their arousal intensity ratings. Specifically, correlations were high when the valence ambiguity of emotional pictures was judged to be low (i.e., clearly positive or clearly negative), but correlations were low or insignificant when valence ambiguity was judged to be high (i.e., neither clearly positive nor clearly negative, such as surprised facial expressions).

Brainerd (2018) extended Mattek et al.'s work with a large-scale analysis of multiple emotional word and picture norms, which used rating *SD*s to measure the ambiguity of items' perceived valence. Consistent with Mattek et al.'s (2017) results, he found that the size of valence-arousal intensity correlations depended on valence ambiguity. Across all rating norms, these correlations decrease linearly as ambiguity increases. Later, Brainerd et al. (2020) found that when valence ambiguity and intensity were factorially manipulated in word lists, recall was better both when ambiguity was high and when intensity was high.

Brainerd et al. (2021) then conducted some follow-up experiments in which ambiguity and intensity were factorially manipulated for three additional attributes (concreteness, categorization, and meaningfulness). As with valence, recall was better when ambiguity was high and when intensity was high. Moreover, Brainerd et al. (2021) expanded Mattek et al.'s (2017) and Brainerd's (2018) finding that correlations between valence-arousal intensity ratings are controlled by the ambiguity of valence ratings. For intensity ratings of categorization, concreteness, imageability, familiarity, meaningfulness, number of features, and pleasantness, bivariate correlations between intensity ratings of pairs of these attributes varied inversely with attribute ambiguity. Last, Brainerd et al. reported that ambiguity displays an inverted-U relation with intensity across multiple attributes, suggesting that items that fall on the extreme ends of the intensity rating scale tend to have lower ambiguity for the given attribute whereas those that fall on the intermediate range tend to have higher ambiguity.

Although increasing the ambiguity of four different attributes improved recall in the above studies, there are two others in which a similar ambiguity manipulation of one attribute (concreteness) failed to affect recall (Neath & Surprenant, 2020; Pollock, 2018). There are some salient methodological differences between these latter studies and those discussed above that may be responsible for the discrepant results. For example, Neath and Surprenant administered serial recall after the presentation of each word list, whereas the above studies used free recall after multiple word lists had been presented. However, it would be speculative to suggest that any one of those differences is likely to be responsible for the different results.

To sum up, some recent findings indicate that the *SD*s of semantic attribute ratings are more than noise indexes, and instead, that they measure the psychologically important property of ambiguity in perceived attribute intensity. This interpretation derives support from empirical effects – specifically, that ambiguity seems to control the strength of inter-attribute intensity correlations and that it affects the accuracy of memory. However, the second effect, which is a more critical one for memory research, is limited in two ways. First, the number of attributes for which ambiguity's memory effects have been measured is not large, and second, those effects have only been measured for recall. We target both of those limitations here.

The current research

This study was aimed at measuring ambiguity effects for a large number of semantic attributes using data from mega recognition studies. In particular, we merged the recognition data from two mega studies (Cortese et al., 2010, 2015) with the ambiguity (*SD*) and intensity (*M*) ratings of 21 semantic attributes from large-scale norming studies. Then, we used hierarchical linear regressions to separate the recognition effects of ambiguity, intensity, and their interaction for each attribute, controlling for a variety of lexical and semantic variables that are known to affect recognition.

We implemented this approach because it has some clear advantages over prior studies, in which ambiguity was manipulated for one attribute at a time in individual experiments. First and most obviously, it allowed us to investigate, in a single study, the ambiguity effects of a very large number of semantic attributes, the intensity effects of those same attributes, and whether ambiguity and intensity effects interact in a consistent way. Results of that sort would support some broad conclusions about the questions that motivated this research. The second advantage is that mega studies use much larger numbers of items and, thus, provide higher levels of statistical power per participant (Adelman & Estes, 2013). Whereas prior experiments collected memory data for tens or hundreds of words, Cortese et al. (2010, 2015) provide recognition data for over 5,000 words. Third, word stimuli in mega studies cover the full range of semantic attribute ratings, rather than just extreme cases. In previous experiments, researchers usually sampled items with extremely high- or low-rated intensities (e.g., words with very high *M* concreteness vs. words with very low *M* concreteness) to maximize the probability of detecting reliable effects. However, that could both spuriously inflate the effects of attribute intensity (e.g., if participants notice that words differ considerably in *M* concreteness, they may be over-sensitive to its effects) and spuriously deflate the effects of attribute ambiguity (e.g., words at the extreme ends of the *M* concreteness spectrum are low in *SD* concreteness). In contrast, the words used in Cortese et al.'s (2010, 2015) mega studies varied broadly in both the *M*s and *SD*s of multiple semantic attributes (see Appendix Table 6). This rules out item selection biases from extreme values.

Method

Materials

The outcome variables in our regression models are recognition memory measures (hit, false alarm, *d'*, *C*, accuracy) from Cortese et al. (2010, 2015). We combined the datasets from Cortese et al. (2010) and Cortese et al. (2015), which contain recognition memory data for 2,578 monosyllabic words and for 2,879 disyllabic words, respectively. For both studies, the participants were undergraduate students (*N*s = 117 and 120). Each participant studied 30 lists of 50 words, and for each list, they completed an item recognition test of 100 words (50 studied words plus 50 new words).

The predictor variables are attribute ambiguity (*SD*) and intensity (*M*) of 21 attributes. In prior studies, the separate effects of attribute ambiguity and intensity on recall were studied for four of these attributes (valence, concreteness, categorization, and meaningfulness; see Brainerd et al., 2020, 2021; Neath & Surprenant, 2020; Pollock, 2018). Here, we examined their effects on recognition with a much larger pool of words and with a set of common covariates controlled. Moreover, we expanded such analyses to include 17 other attributes (six traditional semantic attributes: arousal, dominance, body-object interaction (BOI), imageability, semantic size, familiarity; plus 11 sensorimotor attributes: six sensory modalities (auditory, gustatory, haptic, interoceptive, olfactory, and visual), and five action effectors (leg/foot, arm/hand, head, mouth, torso)). Table 1 displays the norming studies from which we obtained the ratings for these semantic attributes and how these attributes were rated.

Besides the predictor variables that are of principal interest, we included a series of lexical and semantic covariates in the regression models as the prior studies did (Adelman & Estes, 2013; Cortese et al., 2010, 2015). These covariates were word frequency (Brysbaert & New, 2009), word length, orthographic Levenshtein distance (OLD; Balota et al., 2007), phonological Levenshtein distance (PLD; Balota et al., 2007), and age of acquisition (AOA; Cortese & Khanna, 2008; Schock et al., 2012).

Statistical analyses

To examine whether attribute ambiguity and intensity predict recognition memory and to what extent they independently explain variance in recognition memory, we conducted hierarchical linear regressions using the lme4 package (Bates et al., 2015) in R (R Core Team, 2019). First, we ran a linear regression model with only attribute ambiguity or only attribute intensity as a predictor. Second, we included both attribute ambiguity and intensity as predictors.

For each semantic attribute, we ran multiple sets of hierarchical linear regressions. In each set of models, we used one of the five recognition memory measures (hit, false alarm, *d'*, *C*, accuracy) as the outcome variable, and we used attribute ambiguity, attribute intensity, or both as predictors. In all the models, we included all the aforementioned covariates. We only included words with complete data for attribute ambiguity, attribute intensity, and all covariates in our analyses. Additionally, we ran the models both including and excluding an interaction term between attribute ambiguity and attribute intensity, which is meant to capture the unique effect of Ambiguity × Intensity interaction. In all regression models excluding the interaction terms, the variance inflation factor (VIF) is below 5 for all predictors, suggesting that there is no multicollinearity problem.

We relied on the *p* values for regression coefficients to assess whether attribute ambiguity and intensity predict recognition memory. In addition, we used ΔR^2 to measure the two variables' unique contributions to model fits. The regression coefficients for main effects were obtained from the models without an interaction term. The unique variance accounted for by attribute ambiguity and attribute intensity, respectively, was calculated by comparing models that included either attribute intensity or ambiguity with

Table 1 Norming studies and rating instructions for the 21 semantic attributes

Attributes	Norming studies	Rating instructions	Scales
Concreteness	Brysbaert et al. (2014)	To what extent the word refers to things or actions that can be directly experienced through the five senses?	From 1 (abstract) to 5 (concrete)
Valence	Warriner et al. (2013)	How do you feel while reading the word?	From 1 (sad/calm/controlled) to 9 (happy/excited/in control)
Arousal	Warriner et al. (2013)		
Dominance	Warriner et al. (2013)		
Body-object interaction	Pexman et al. (2019)	How easily can a human body physically interact with the word's referent?	From 1 (not at all) to 7 (easily)
Imageability	Scott et al. (2019)	How easily can you image or picture the word's referent?	From 1 (very unimageable) to 7 (very imageable)
Familiarity	Scott et al. (2019)	How frequently do you experience the word and how easily do you recognize it?	From 1 (very unfamiliar) to 7 (very familiar)
Semantic size	Scott et al. (2019)	How large is the word's physical or conceptual referent?	From 1 (very small) to 7 (very big)
Categorization	Toglia and Battig (1978)	How easily does a word fit into some large categories?	From 1 (not at all) to 7 (easily)
Meaningfulness	Toglia and Battig (1978)	How easily can the word retrieve other words as associates to them?	
Auditory modality	Lynott et al. (2020)	To what extent do you experience the word by hearing/tasting/feeling through touch/sensation inside your body/smelling/seeing?	From 0 (not experienced at all with that sense) to 5 (experienced greatly with that sense)
Gustatory modality	Lynott et al. (2020)		
Haptic modality	Lynott et al. (2020)		
Interoceptive modality	Lynott et al. (2020)		
Olfactory modality	Lynott et al. (2020)		
Visual modality	Lynott et al. (2020)		
Leg/foot action	Lynott et al. (2020)	To what extent do you experience the word by performing an action with the foot/leg, hand/arm, head excluding mouth, mouth/throat, and torso?	From 0 (not experienced at all with that action) to 5 (experienced greatly with that action)
Arm/hand action	Lynott et al. (2020)		
Head action	Lynott et al. (2020)		
Mouth action	Lynott et al. (2020)		
Torso action	Lynott et al. (2020)		

models that included both. The unique variance explained by the Ambiguity × Intensity interaction was calculated by comparing models that included attribute ambiguity, attribute intensity, and their interaction with models that only included attribute ambiguity and intensity.

Results

We focus on two types of findings. First, we consider the regression coefficients for attribute ambiguity, attribute intensity, and their interaction. Second, we examine the unique variance that these variables account for. As a preview, two major patterns emerged: (a) for the majority of the attributes, ambiguity and intensity were distinct predictors of recognition memory; and (b) ambiguity sometimes accounted for more unique variance (i.e., was a better predictor of recognition memory) than intensity. The more detailed regression results can be found in the Online Supplementary Materials (OSM).

Regression coefficients

Table 2 displays the regression coefficients of attribute ambiguity and intensity when both were included in a linear regression model. As can be seen there, in the majority of the cases, attribute ambiguity and intensity both predicted recognition memory, rather than only intensity predicting recognition. For example, the regression coefficients for attribute ambiguity were significant with hits, d', and accuracy for 14 of the 21 attributes. They were also significant with false alarms and C for 13 and nine attributes, respectively. As a comparison, for hits, d', and accuracy, the regression coefficients for attribute intensity were significant for 14, 16, and 17 attributes, respectively, and for false alarms and C, they were significant for 14 and 10 attributes, respectively.

Table 2 Regression coefficients for attribute intensity and attribute ambiguity

Attributes	*N*	Hit		FA		*d'*		*C*		Acc	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Concreteness	5,303	0.028***	-0.007*	-0.011***	0.018***	0.130***	-0.085***	-0.023***	-0.018	0.020***	-0.013***
Valence	4,837	-0.003***	0.038***	-0.005***	-0.019***	0.008	0.198***	0.015***	-0.023*	0.001	0.029***
Arousal	4,837	0.010***	0.016***	0.005**	-0.015***	0.014	0.105***	-0.026***	0.005	0.002*	0.015***
Dominance	4,837	-0.010***	0.016***	0.002	-0.001	-0.039***	0.054**	0.012***	-0.022*	-0.006***	0.009**
BOI	4,377	0.015***	0.019***	-0.011***	0.015***	0.088***	0.002	-0.004	-0.056***	0.013***	0.002
Imageability	2,606	0.028***	0.002	-0.017***	0.007	0.147***	-0.029	-0.013**	-0.019	0.022***	-0.002
Familiarity	2,606	0.004	0.056***	0.001	-0.009	0.008	0.197***	-0.012	-0.082***	0.001	0.032***
Size	2,606	-0.003	0.029***	0.008***	-0.035***	-0.042***	0.212***	-0.009	0.018	-0.006***	0.032***
Categorization	1,802	0.033***	0.013	-0.024***	-0.004	0.191***	0.059	-0.004	-0.013	0.028***	0.009
Meaningful	1,802	0.026***	0.049***	0.001	-0.094***	0.067***	0.546***	-0.040***	0.111***	0.012***	0.072***
Auditory	5,293	-0.001	-0.008**	-0.001	0.012***	-0.003	-0.071***	0.004	-0.012	0.000	-0.010***
Gustatory	5,293	0.012***	-0.008**	-0.011***	0.006*	0.081***	-0.044***	0.002	-0.001	0.012***	-0.007***
Haptic	5,293	0.008***	0.007*	-0.004*	0.007*	0.038***	-0.001	-0.006	-0.023**	0.006***	0.000
Interoceptive	5,293	0.002	-0.008**	0.007***	0.007*	-0.023*	-0.047**	-0.016***	-0.001	-0.003*	-0.007***
Olfactory	5,293	0.016***	0.000	-0.011***	0.000	0.090***	0.005	-0.004	0.000	0.014***	0.000
Visual	5,293	0.014***	-0.001	-0.004*	0.008	0.059***	-0.028	-0.015***	-0.014	0.009***	-0.004
Leg/foot	5,293	-0.002	0.005	0.005*	0.010**	-0.028**	-0.021	-0.007	-0.025**	-0.004*	-0.003
Arm/hand	5,293	0.000	0.011**	0.010***	-0.006	-0.037***	0.059***	-0.018***	-0.005	-0.005***	0.009***
Head	5,293	-0.001	0.000	0.002	-0.022***	-0.013	0.092***	-0.001	0.042***	-0.002	0.011**
Mouth	5,293	0.003*	-0.008**	-0.002	0.006	0.018*	-0.045**	-0.001	0.001	0.002*	-0.007**
Torso	5,293	0.011***	0.001	-0.001	0.013***	0.036**	-0.042*	-0.018**	-0.023*	0.006**	-0.006*

N = number of words included in the analyses, *M* = attribute intensity, *SD* = attribute ambiguity, *BOI* = body-object interaction

* $p < .05$, ** $p < .01$, *** $p < .001$. All coefficients are from the models that include attribute intensity, attribute ambiguity, and covariates, without interaction terms

Importantly, ambiguity displayed greater predictive power than intensity for some attributes. For instance, the ambiguity of familiarity, auditory modality, interoceptive modality, and arm/hand action predicted hits, whereas the intensity of those attributes did not. The opposite was sometimes true, too: The intensity of imageability, categorization, olfactory modality, and visual modality predicted hits whereas the ambiguity of those attributes did not. These patterns illustrate (a) that attribute ambiguity and intensity have separate influences on recognition memory, (b) that those effects are dissociable for certain attributes, and (c) that the recognition effects of attribute ambiguity are more robust than that of attribute intensity for certain attributes.

Further, it is noteworthy that the direction of the effects of ambiguity varies across attributes (which is also true for the effects of intensity). Here, we consider only regression coefficients that are statistically significant. Taking *d'* and accuracy as examples, we can see that ambiguity improves recognition for about half of the attributes but impairs recognition for the other half. These recognition results replicate patterns that were reported for recall in prior experiments on the ambiguity of valence, categorization, and meaningfulness (Brainerd et al., 2020, 2021), where ambiguity enhanced the accuracy of both recall and recognition. However, unlike these prior studies, we found the opposite pattern for concreteness, as ambiguity impaired recognition.

When it comes to hits and false alarms, ambiguity primarily displays mirror effects, wherein it simultaneously increases hits and decreases false alarms or vice versa. However, there are two exceptions: body-object interaction and haptic modality, as ambiguity of these two attributes increases both hits and false alarms. Additionally, the effects of ambiguity are very consistent for *C*, in that ambiguity decreases *C* across all attributes except for head action. In short, increasing attribute ambiguity overwhelmingly reduces response bias.

Beyond the aforementioned patterns, Table 3 shows that the Ambiguity × Intensity interaction predicted recognition memory for multiple attributes (e.g., body-object interaction, imageability, familiarity, semantic size, auditory modality, torso action, etc.). In particular, for those attributes, the effects of intensity were amplified or reduced by higher levels of ambiguity. This again highlights the fact that attribute ambiguity and intensity are psychologically distinct.

Table 3 Regression coefficients for Ambiguity × Intensity interaction

Attributes	*N*	Hit	FA	*d'*	*C*	Acc
Concreteness	5,303	-0.006	0.006	-0.039*	-0.006	-0.006*
Valence	4,837	0.004	-0.004	0.038*	0.003	0.004
Arousal	4,837	0.000	-0.008	0.030	0.016	0.004
Dominance	4,837	0.006	-0.001	0.028	-0.006	0.004
BOI	4,377	0.001	-0.008**	0.038**	0.013*	0.005**
Imageability	2,606	-0.005	0.023***	-0.115***	-0.041***	-0.014***
Familiarity	2,606	-0.017**	0.055***	-0.274***	-0.072***	-0.036***
Size	2,606	0.012	-0.022**	0.120***	0.016	0.017***
Categorization	1,802	-0.014	0.008	-0.059	0.013	-0.011*
Meaningful	1,802	0.005	0.040*	-0.135	-0.076	-0.017
Auditory	5,293	0.010***	-0.006*	0.060***	-0.003	0.008***
Gustatory	5,293	0.007**	0.002	0.016	-0.016*	0.003
Haptic	5,293	0.003	-0.006*	0.037**	0.008	0.005*
Interoceptive	5,293	0.006	-0.005	0.040**	0.000	0.006**
Olfactory	5,293	0.002	0.001	0.006	-0.003	0.000
Visual	5,293	0.006	0.007*	-0.011	-0.025**	-0.001
Leg/foot	5,293	0.008**	-0.006	0.051***	0.001	0.007***
Arm/hand	5,293	0.001	-0.004	0.019	0.006	0.002
Head	5,293	0.004	0.009*	-0.024	-0.022	-0.003
Mouth	5,293	0.003	-0.001	0.016	-0.002	0.002
Torso	5,293	0.009*	-0.013**	0.081***	0.012	0.011***

N = number of words included in the analyses, *M* = attribute intensity, *SD* = attribute ambiguity, *BOI* = body-object interaction

* $p < .05$, ** $p < .01$, *** $p < .001$. All coefficients are from the models that include attribute intensity, attribute ambiguity, covariates, and Ambiguity × Intensity interaction

Unique variance explained

Next, we turn to the unique variance explained by attribute ambiguity and intensity. The relevant findings are displayed in Table 4. Considering that ambiguity does not figure in traditional research on semantic attributes, it is quite surprising that across all five measures of recognition memory, ambiguity accounts for more unique variance than intensity for multiple attributes. For example, for false alarms, ambiguity had a predictive advantage over intensity for 11 of the 21 attributes. Thus, in addition to the finding that attribute ambiguity and intensity are distinct psychological dimensions, when it comes to the memory effects of semantic attributes, ambiguity effects are sometimes greater than intensity effects.

Further, for *d'* and accuracy, which are typically of primary interest to memory researchers, there is a very consistent pattern that ambiguity of valence, arousal, familiarity, semantic size, meaningfulness, auditory modality, interoceptive modality, head action, and mouth action accounts for more variance than the intensity of these attributes. Additionally, the relative predictive power of ambiguity and intensity vary across attributes: Ambiguity is more predictive than intensity for some attributes, while intensity is more predictive than ambiguity for others. Evaluation explanations of that pattern would seem to be a key target for future research.

Discussion

In the current study, we sought to greatly expand the data on how attribute ambiguity influences episodic memory by how it affects recognition for 21 attributes, using data from two mega recognition studies (Cortese et al., 2010, 2015). Two critical patterns emerged. First, the recognition effects of attribute ambiguity were distinct from those of attribute intensity. Second, attribute ambiguity was sometimes a better predictor of recognition than attribute intensity.

Regarding the first pattern, averaging over the five recognition measures (hit, false alarm, *d'*, *C*, accuracy), attribute ambiguity significantly predicted recognition for 61% of the attributes *when attribute intensity was controlled*. It is worth noting that the effects of attribute ambiguity were not only independent but also dissociable from those of attribute intensity: Ambiguity effects were sometimes significant

Table 4 Unique variance explained by attribute intensity and attribute ambiguity when covariates are included in the models

Attributes	*N*	ΔR^2 in %									
		Hit		FA		*d'*		*C*		Acc	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Concreteness	5,303	**15.215**	0.161	**6.591**	2.626	**18.893**	1.216	**3.356**	0.293	**19.192**	1.205
Valence	4,837	0.801	**8.410**	3.500	**4.923**	0.200	**11.455**	**4.763**	0.953	0.079	**11.320**
Arousal	4,837	**3.060**	1.021	1.828	**2.175**	0.321	**2.405**	**6.143**	0.028	0.445	**2.431**
Dominance	4,837	**3.539**	1.371	**0.397**	0.012	**2.902**	0.768	**1.702**	0.784	**3.197**	0.938
BOI	4,377	**13.822**	1.580	**20.425**	3.083	**27.706**	0.001	0.274	**4.304**	**27.926**	0.031
Imageability	2,606	**20.039**	0.008	**13.734**	0.223	**28.417**	0.119	**1.257**	0.293	**29.354**	0.037
Familiarity	2,606	0.157	**6.472**	0.032	**0.303**	0.037	**5.406**	0.266	**2.930**	0.053	**6.499**
Size	2,606	0.419	**1.971**	3.693	**4.801**	3.953	**6.710**	**0.543**	0.164	3.254	**6.841**
Categorization	1,802	**17.529**	0.147	**38.176**	0.053	**50.850**	0.240	**0.074**	0.040	**44.839**	0.213
Meaningful	1,802	**5.657**	1.436	0.080	**24.430**	4.576	**21.754**	**3.279**	1.801	5.841	**14.310**
Auditory	5,293	0.062	**0.355**	0.057	**2.433**	0.021	**1.983**	0.158	**0.225**	0.008	**1.806**
Gustatory	5,293	**2.825**	0.530	**5.940**	0.838	**7.988**	1.118	**0.023**	0.004	**7.380**	1.217
Haptic	5,293	**1.359**	0.291	0.805	**0.830**	**2.195**	0.000	0.233	**0.883**	**2.226**	0.000
Interoceptive	5,293	0.051	**0.393**	**2.020**	0.690	0.575	**0.868**	**1.103**	0.002	0.334	**0.979**
Olfactory	5,293	**2.710**	0.000	**3.661**	0.004	**5.616**	0.009	**0.059**	0.000	**5.683**	0.002
Visual	5,293	**4.876**	0.001	**0.997**	0.557	**5.666**	0.180	**1.456**	0.178	**5.869**	0.169
Leg/foot	5,293	0.048	**0.142**	0.971	**1.590**	**0.739**	0.171	0.195	**0.976**	**0.537**	0.136
Arm/hand	5,293	0.000	**0.560**	**6.233**	0.471	**2.291**	1.076	**2.145**	0.032	**1.846**	1.049
Head	5,293	**0.019**	0.001	0.249	**2.953**	0.243	**1.353**	0.012	**1.092**	0.150	**0.885**
Mouth	5,293	0.231	**0.365**	0.283	**0.556**	0.575	**0.891**	**0.003**	0.000	0.492	**0.850**
Torso	5,293	**0.949**	0.002	0.011	**1.841**	**0.776**	0.488	**0.687**	0.532	**0.841**	0.484

N = number of words included in the analyses, *M* = attribute intensity, *SD* = attribute ambiguity, *BOI* = body-object interaction. ΔR^2 was calculated by comparing models that included either intensity or ambiguity variable with models that included both. Covariates were included in all the models. The larger ΔR^2 (in %) in the *M* vs. *SD* comparison is shown in bold font for each attribute

when intensity effects were not, and vice versa. Moreover, attribute ambiguity had a consistent effect on *C*, such that higher ambiguity reduced response bias. This clearly runs counter to the traditional interpretation of *SD*s of attribute intensity ratings – namely, that they are indexes of measurement error (e.g., Toglia & Battig, 1978). In sum, our results demonstrate that attribute ambiguity is not merely noise in perceived attribute intensity, but, rather, is a distinct dimension that displays robust recognition effects for certain attributes.

Additionally, Ambiguity × Intensity interactions were significant predictors of recognition for 40% of the attributes. Here, the Ambiguity × Intensity interaction for familiarity suggests an intriguing explanation of the classic word frequency-mirror effect, in which low-frequency words are easier to recognize than high-frequency ones (e.g., Glanzer & Adams, 1990; Glanzer et al., 1999). Specifically, our results indicate that the recognition effect of familiarity is mainly tied to familiarity ambiguity. First, for the hit, *d'*, and accuracy measures, the regression coefficients for familiarity ambiguity in Table 2 are significant but those for familiarity intensity are not. Second, for the same measures, the Ambiguity × Intensity interaction for familiarity is significant, indicating that unfamiliar words only produced better recognition when ambiguity was low (see Table 3). Considering the close relation between frequency and familiarity (Karlsen & Snodgrass, 2004; Tanaka-Ishii & Terada, 2011), it is possible that the frequency/familiarity effects are primarily under the control of ambiguity rather than intensity.

Regarding the second pattern, attribute ambiguity accounted for more unique variance in recognition than attribute intensity for 41% of the attributes, averaging over the five recognition measures. In addition to producing independent effects, then, attribute ambiguity sometimes also has stronger recognition effects than attribute intensity. This is another finding that echoes the theme that attribute ambiguity is a distinct component of semantic attributes, which requires different process explanations than the effects of intensity.

Previous studies proposed different theoretical hypotheses about the reasons for ambiguity effects. For instance, Pollock (2018) proposed a *disagreement hypothesis*, in which items with high disagreement in their attribute ratings are harder to

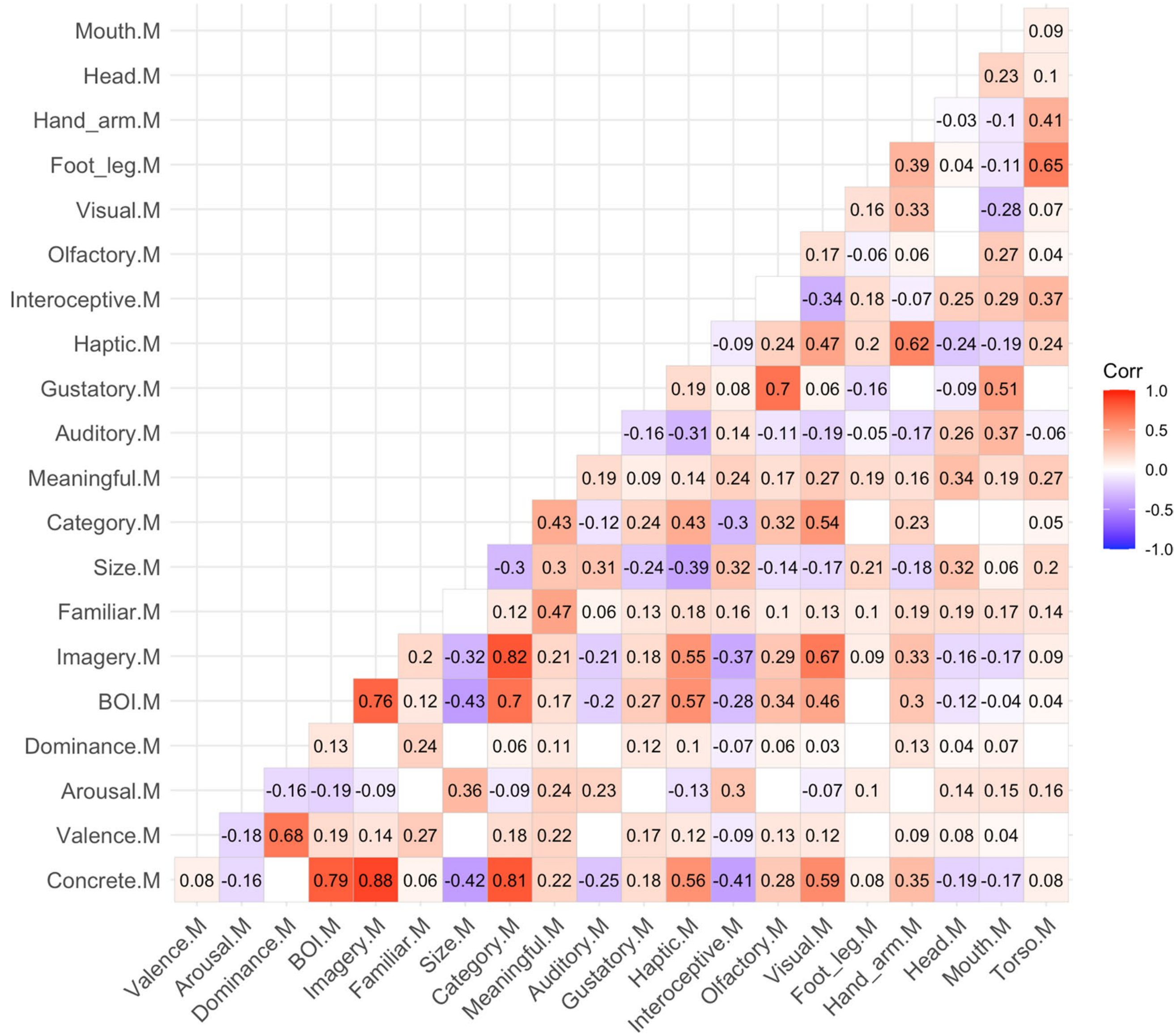

Fig. 1 Correlations between the intensity of attributes. The magnitude and direction for the correlation coefficients are indicated by the color and numerical values. Correlations with $p > .05$ were not shown.

process and are thus less memorable than items with low disagreement. In contrast, Brainerd et al. (2020, 2021) proposed a *categorical/quantitative hypothesis*. According to this hypothesis, when attribute ambiguity is low, people engage in superficial categorical processing of attribute intensity (e.g., whether an item is definitely high or low with respect to a certain attribute). However, when attribute ambiguity is high, people switch to a more thorough, quantitative processing of attribute intensity to resolve the uncertainty, and such deeper semantic processing redounds to the benefit of learning. In brief, the former hypothesis predicts negative memory effects of ambiguity, whereas the latter predicts positive memory effects.

In that connection, Brainerd et al. (2020, 2021) found positive recall effects of attribute ambiguity for four attributes, whereas Pollock (2018) and Neath and Surprenant (2020) both found that concreteness ambiguity did not affect recall. Obviously, the current study aligns more with the former experiments, inasmuch as attribute ambiguity had reliable effects on memory accuracy. At the level of individual attributes, however, the finding that increasing ambiguity improved recognition for some attributes is consistent with categorical/quantitative hypothesis, but the finding that ambiguity impaired recognition for other attributes is consistent with the disagreement hypothesis. It is also interesting to see that

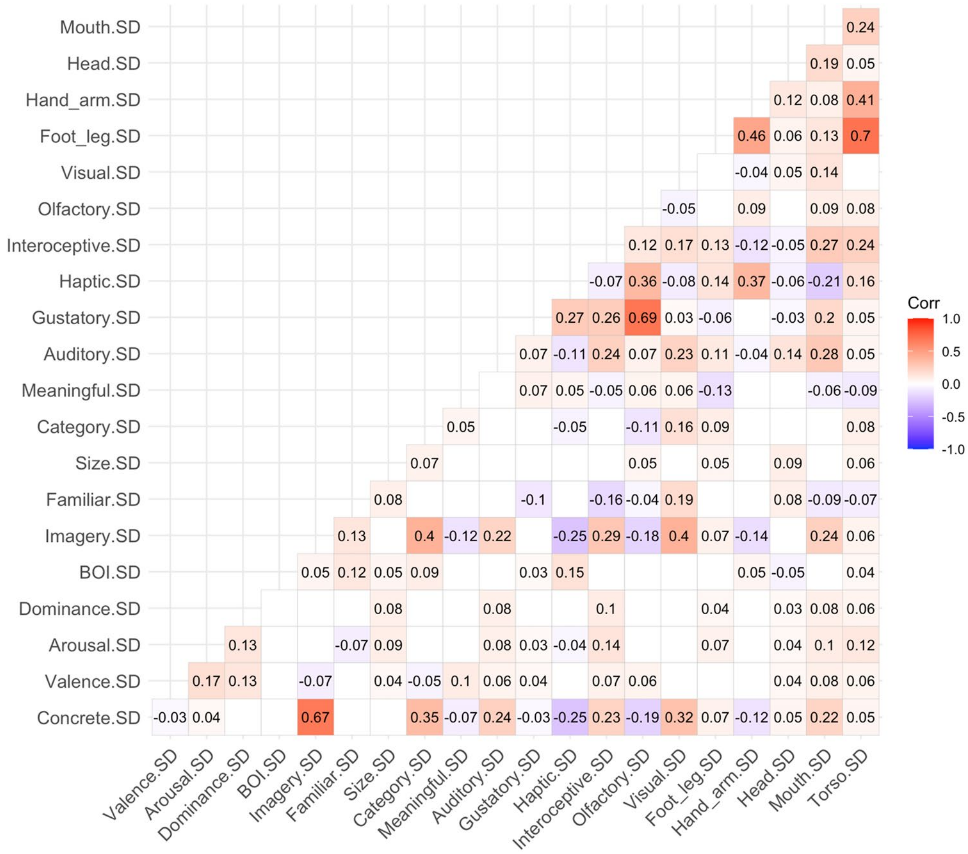

Fig. 2 Correlations between the ambiguity of attributes. The magnitude and direction for the correlation coefficients are indicated by the color and numerical values. Correlations with $p > .05$ were not shown.

ambiguity accounted for more variance in recognition than intensity for some attributes but not the others.

We offer two hypotheses about why ambiguity effects vary substantially across attributes. One is theoretical and the other is psychometric. First, an inspection of our findings revealed that attributes that have negative ambiguity effects on recognition were predominantly those grounded in certain sensorimotor experience (e.g., concreteness, auditory perception, gustatory perception, etc.), whereas attributes that have positive ambiguity effects tend to rely more on elaborative (e.g., affective, linguistic, semantic) processing (e.g., valence, familiarity, meaningfulness, etc.). One possibility is that higher *SD*s in the former attributes are associated with difficulty in generating a vivid representation of a word's referents. Here, *SD*s in those attributes signal ambiguity in how a word's referents can be experienced with senses or body actions. Thus, words with higher *SD*s are more likely to represent concepts that are less grounded in sensorimotor experiences, which in turn make it harder to generate supporting imagery. Note that imageability is the strongest positive predictor of recognition performance in the mega recognition dataset (Cortese et al., 2010, 2015). Thus, if *SD*s of some attributes are negatively associated with imageability, it follows that they will also be negatively associated with recognition. That is what the data show: Among the six attributes that produced negative ambiguity effects, *SD*s for five of them are negatively correlated with *M* imageability ($M_r = -.27$).[1] On the other hand, higher *SD*s in the latter type of attributes are not associated with difficulty of generating imagery. Among the eight attributes that produced positive ambiguity effects, the correlations between their *SD*s and *M* imageability are generally negligible ($M_r =$ .06). Instead, higher ambiguity in those attributes is likely to be associated with retrieval of broader ranges of information. For example, the word *liquor* has the fourth highest valence *SD* in our dataset, which can trigger associations ranging from positive information such as euphoria of celebratory events to negative information such as the disastrous consequences of driving under the influence. Thus, higher ambiguity in those attributes may provoke deeper and richer processing of the content of items, which redounds to recognition performance.

[1] We used the imageability ratings collected by Cortese et al. (2010, 2015) for the correlational analysis here, because that was collected for all words in the mega recognition dataset. To clarify, because imageability *SD*s was not provided by Cortese et al. (2010, 2015), we used imageability *M*s and *SD*s from the Scott et al. (2019) norms in the main analyses (see Table 1).

Our second hypothesis, which is psychometric, is that inter-attribute differences in the reliability of the ambiguity measure are contributing to the differences in ambiguity's recognition effects for different attributes. Suppose that attribute ambiguity tends to account for a larger proportion of variance in recognition when the reliability of this measure is at least moderately high, but it tends to explain a smaller proportion of variance in recognition when its reliability is low or non-significant. That would explain our results, but only if (a) the reliability of the ambiguity measures varies widely, and (b) is correlated with its recognition effects.

To test this hypothesis, we obtained the trial-level raw data of for seven attributes in the Scott et al. (2019) norms (valence, arousal, concreteness, dominance, familiarity, imageability, and semantic size) [2] and 11 attributes of the Lynott et al. (2020) norms (auditory, gustatory, haptic, interoceptive, olfactory, and visual perception, foot/leg, hand/arm, head, mouth, torso action). Then, we calculated the average Spearman-Brown corrected split-half reliability for attribute *SD*s across 100 random splits of the participants. For the Scott et al. (2019) norms, the split-half reliability of attribute ambiguity ranges between .31 and .70 (M = .47, SD = .14). For the Lynott et al. (2020) norms, the split-half reliability of attribute ambiguity ranges from .20 to .67 (M = .57, SD = .15). We found that the variance in recognition explained by attribute ambiguity alone is positively correlated with the reliability of the given attribute, *r*s = .52 and .44 for the Scott et al. norms and the Lynott et al. norms, respectively. Thus, in line with the second hypothesis, differences in reliability predict differences in ambiguity-recognition relations. The obvious implication is that an important target for future research is to identify an alternative measure of ambiguity whose reliabilities are uniformly high for all attributes. For instance, one possibility is to ask participants to rate the level of certainty that they attach to their intensity ratings of individual attributes (for an illustrative procedure, see Brainerd et al., 2022).

Note that the current finding of discrepant ambiguity effects across attributes is consistent with the earlier observation about attribute-specificity of ambiguity. In that connection, Brainerd et al. (2021) proposed that there may be a univariate psychological factor that triggers differences in rating ambiguity across different attributes. If that is the case, one would expect the ambiguities of different attributes would be strongly correlated with each other. However, they are not. Brainerd et al. (2021) used several attributes that had passed discriminative validity tests from Scott et al. (2019) norms and Toglia and Battig (1978) norms, and found that their ambiguities were either very weakly correlated or uncorrelated. We observed a similar pattern in the current dataset. Below, we present the inter-attribute correlations for intensity (Fig. 1) and ambiguity (Fig. 2) across all attributes included in our analyses. A comparison between Figs. 1 and 2 reveals two main patterns. First, inter-attribute ambiguity correlations were typically much weaker than inter-attribute intensity correlations. This argues against a univariate factor that controls rating ambiguity across different attributes. Second, the pattern of inter-attribute correlations for ambiguity to some extent tracks the pattern for intensity. For instance, attributes the are highly correlated in intensity (e.g., concreteness and imageability, foot/lag action and torso action) also tend to have a higher correlation in their ambiguity. This consistency is expected as *M*s are always involved in the computation of *SD*s.

On a related note, it is important to distinguish the *attribute ambiguity* discussed in this paper from the similar construct of *semantic ambiguity* or *semantic diversity* in the psycholinguistic literature (e.g., Hoffman et al., 2013; Johns,

Table 5 Correlations between attribute ambiguity and semantic ambiguity or semantic diversity

Variable	SemD	SD	SD-AP	UCD-SD
Concrete.SD	0.24	-0.01	0.02	0.05
Valence.SD	-0.10	-0.01	-0.06	-0.03
Arousal.SD	-0.01	0.06	0.04	0.09
Dominance.SD	0.03	0.06	0.07	0.08
BOI.SD	-0.17	-0.08	-0.11	-0.11
Imageability.SD	0.27	-0.02	0.02	0.03
Familiar.SD	-0.31	-0.55	-0.55	-0.62
Size.SD	-0.09	0.03	-0.04	-0.01
Category.SD	0.10	0.04	-0.01	0.00
Meaningful.SD	-0.04	-0.06	-0.12	-0.06
Auditory.SD	0.20	0.08	0.07	0.15
Gustatory.SD	0.03	0.03	0.04	0.07
Haptic.SD	-0.11	0.00	-0.02	-0.04
Interoceptive.SD	0.23	0.13	0.19	0.19
Olfactory.SD	-0.09	0.02	-0.02	0.00
Visual.SD	0.07	-0.15	-0.15	-0.11
Foot_leg.SD	0.12	0.15	0.15	0.11
Hand_arm.SD	-0.03	0.01	0.01	0.02
Head.SD	0.01	-0.01	-0.03	0.00
Mouth.SD	0.13	0.03	0.05	0.11
Torso.SD	0.06	0.11	0.12	0.08

SemD refers to Hoffman et al.'s (2013) measure of semantic ambiguity. SD refers to Jones et al.'s (2012) measure of semantic diversity. SD-AP refers to Johns et al.'s (2020) measure of author-prevalence semantic diversity. UCD-SD refers to Johns et al.'s (2021) user contextual diversity modified by the semantic distinctiveness model. *BOI* = body-object interaction

[2] To clarify, we used valence, arousal, dominance, and concreteness ratings from other norms in our main analyses (see Table 1), but we do not have access to the trial-level data of those norms, so we used the Scott et al. (2019) norms for this follow-up analysis.

2021; Johns et al., 2020; Jones et al., 2012). Whereas the former is a multivariate construct tapping interindividual variability in the perceived strength of specific semantic attributes, the latter is a univariate construct focusing on to what extent a word is used in distinct contexts, with the underlying assumption being that a word's meaning varies with the contexts it appears in. Attribute ambiguity is aligned with the componential approach to semantics, whereas semantic ambiguity/diversity is aligned with the distributional approach to semantics (Brainerd et al., 2023). In brief, the former approach assumes semantic attributes as critical units that capture salient aspects of meaning and derive such attributes from theories of memory representation – such as concreteness and imageability from dual-coding theory (Paivio, 1970) and body-object interaction and sensorimotor attributes based on theories of grounded cognition (Barsalou, 2008). In contrast, the distributional approach assumes that word meaning is acquired via the statistical redundancies in linguistic environments; that is, from how words co-occur in natural language (Günther et al., 2019; Kumar, 2021). Thus, the surrounding context of a word plays a key role in the word's meaning in this approach. A full review of the two approaches is beyond the scope of this paper. For our purposes, the critical point of distinction is simply that attribute ambiguity is conceptually distinct from semantic ambiguity/diversity. Empirically, we also found that attribute ambiguity is only weakly correlated with four common semantic ambiguity/diversity measures (see Table 5): The average absolute value of correlation coefficients was only .09. Thus, attribute ambiguity is both conceptually and empirically different from the other concepts of semantic ambiguity/diversity.

Appendix

Table 6 Descriptive Statistics for Word Stimuli Used in Cortese et al. (2010, 2015)

Attributes	*N*	Intensity (*M*)				Ambiguity (*SD*)			
		Mean	Variance	Min	Max	Mean	Variance	Min	Max
Concreteness	5303	3.542	1.072	1.070	5.000	1.087	0.141	0.000	1.850
Valence	4837	5.155	1.502	1.400	8.470	1.674	0.124	0.450	2.870
Arousal	4837	4.112	0.795	1.600	7.740	2.290	0.104	0.880	3.300
Dominance	4837	5.265	0.797	2.420	7.860	2.149	0.107	0.960	3.290
BOI	4377	3.372	2.074	1.120	6.880	1.717	0.147	0.332	2.732
Imagery	2606	4.858	1.828	1.875	6.941	1.397	0.171	0.235	2.402
Familiarity	2606	5.375	0.757	1.647	6.939	1.417	0.139	0.239	2.361
Size	2606	3.953	1.028	1.375	6.613	1.363	0.052	0.613	2.264
Categorization	1802	4.548	1.125	1.820	6.680	1.750	0.042	0.860	2.520
Meaningful	1802	4.275	0.424	2.100	6.160	1.740	0.026	1.230	2.400
Auditory	5293	1.508	1.158	0.000	5.000	1.463	0.237	0.000	2.426
Gustatory	5293	0.429	0.786	0.000	5.000	0.620	0.358	0.000	2.462
Haptic	5293	1.384	1.133	0.000	4.882	1.431	0.281	0.000	2.437
Interoceptive	5293	1.008	0.836	0.000	4.778	1.289	0.314	0.000	2.429
Olfactory	5293	0.523	0.609	0.000	5.000	0.784	0.363	0.000	2.422
Visual	5293	3.114	0.878	0.056	5.000	1.670	0.122	0.000	2.429
Leg/foot	5293	0.947	0.770	0.000	4.955	1.252	0.333	0.000	2.396
Arm/hand	5293	1.677	0.974	0.000	4.941	1.650	0.165	0.000	2.557
Head	5293	2.265	0.575	0.000	5.000	1.927	0.059	0.000	2.477
Mouth	5293	1.339	1.106	0.000	5.000	1.490	0.278	0.000	2.462
Torso	5293	0.914	0.553	0.000	4.391	1.298	0.263	0.000	2.635

N = number of words included in the analyses. *M* = attribute intensity. *SD* = attribute ambiguity. BOI = body-object interaction. The norming studies for the semantic attributes are described in Table 1

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-023-02291-5.

Acknowledgment We thank Xinya Liu for her assistance in calculating the reliability of attribute *SD*s.

Author contributions Conceptualization: Minyu Chang, C. J. Brainerd. Methodology: Minyu Chang. Formal analysis and investigation: Minyu Chang. Writing – original draft preparation: Minyu Chang. Writing – review and editing: Minyu Chang, C. J. Brainerd.

Data availability All data analyzed during this study have been published publicly (the acquisition methods are described in the Online Supplementary Materials).

Code availability Available upon request to the first author.

Declarations

Conflicts of interest/Competing interests Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication We consent to the publication of this article.

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition, 129*(3), 530–535.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445–459.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology, 59*(1), 617–645.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1–48.

Brainerd, C. J. (2018). The emotional-ambiguity Hypothesis: A large-scale test. *Psychological Science, 29*(10), 1706–1715.

Brainerd, C. J., Chang, M., & Bialer, D. M. (2020). Emotional ambiguity and memory. *Journal of Experimental Psychology: General.* https://doi.org/10.1037/xge0001011

Brainerd, C. J., Chang, M., Bialer, D. M., & Liu, X. (2022). How does attribute ambiguity improve memory? *Memory & Cognition.* https://doi.org/10.3758/s13421-022-01343-w

Brainerd, C. J., Chang, M., Bialer, D. M., Toglia, M., & P. (2021). Semantic ambiguity and memory. *Journal of Memory and Language, 121*, 104286.

Brainerd, C. J., Liu, X., Bialer, D. M., & Chang, M. (2023). The big three: Accuracy, organization, and retrieval effects of latent semantic attributes. *Journal of Experimental Psychology: General.* https://doi.org/10.1037/xge0001365

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904–911.

Chang, M., Brainerd, C. J., Toglia, M. P., & Schmidt, S. R. (2021). Norms for emotion-false memory lists. *Behavior Research Methods, 53*(1), 96–112.

Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods, 40*(3), 791–794.

Cortese, M. J., & Khanna, M. M. (2022). Relating emotional variables to recognition memory performance: A large-scale re-analysis of megastudy data. *Memory, 30*(7), 915–922.

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory, 18*(6), 595–609.

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology, 68*(8), 1489–1501.

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage, 32*(3), 1413–1421.

Glanzer, M., & Adams, J. (1990). The mirror effect in recognition Memory: Data and theory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 16*, 5–16.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 500–513.

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

Hamilton, M., & Rajaram, S. (2001). The Concreteness Effect in Implicit and Explicit Memory Tests. *Journal of Memory and Language, 44*(1), 96–117.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718–730.

Johns, B. T. (2021). Disentangling contextual diversity: Communicative need as a lexical organizer. *Psychological Review, 34*. https://doi.org/10.1037/rev0000265

Johns, B. T., Dye, M., & Jones, M. N. (2020). Estimating the prevalence and diversity of words in written language. *Quarterly Journal of Experimental Psychology, 73*(6), 841–855.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology, 66*(2), 115–124.

Karlsen, P. J., & Snodgrass, J. G. (2004). The word-frequency paradox for recall/recognition occurs for pictures. *Psychological Research, 68*(4), 271–276.

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review, 28*(1), 40–80.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods, 52*(3), 1271–1291.

Mattek, A. M., Wolford, G., & Whalen, P. J. (2017). A mathematical model captures the structure of subjective affect. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 12*(3), 508–526.

Neath, I., & Surprenant, A. M. (2020). Concreteness and disagreement: Comment on Pollock (2018). *Memory & Cognition, 48*(4), 683–690.

Paivio, A. (1970). On the functional significance of imagery. *Psychological Bulletin, 73*(6), 385–392.

Pexman, P. M., Muraki, E., Sidhu, D. M., 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.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/. Accessed 18 Oct 2022

Rae, G. (1979). Effect of Word Meaningfulness on Primary and Secondary Memory. *The Journal of General Psychology, 101*(2), 175–181.

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.

Tanaka-Ishii, K., & Terada, H. (2011). Word familiarity and frequency. *Studia Linguistica, 65*(1), 96–116.

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

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207.

Open practices statement All data analyzed in this study have already been published publicly. This study was not preregistered.