The Glasgow Norms: Ratings of 5,500 words on 9 scales

Graham G. Scott[1]*, Anne Keitel[2], Marc Becirspahic[3],

Patrick J. O'Donnell[3], and Sara C. Sereno[2,3]*

[1]School of Media, Culture and Society, University of the West of Scotland

[2]Institute of Neuroscience and Psychology, University of Glasgow

[3]School of Psychology, University of Glasgow

*_Corresponding authors_:

Dr. Graham G. Scott
School of Media, Culture and Society – Psychology
University of the West of Scotland
Paisley PA1 2BE
United Kingdom
graham.scott@uws.ac.uk
t: +44 (0)141 848-3830

Dr. Sara C. Sereno
Institute of Neuroscience and Psychology
School of Psychology
58 Hillhead Street
University of Glasgow
Glasgow G12 8QB
United Kingdom
sara.sereno@glasgow.ac.uk
t: +44 (0)141 330-5089

Running head: The Glasgow Norms

Abstract

The Glasgow Norms are a set of normative ratings for 5,553 English words on 9 psycholinguistic dimensions: arousal, valence, dominance, concreteness, imageability, familiarity, age of acquisition, semantic size, and gender association. The Glasgow Norms are unique in several respects. First, the corpus, itself, is relatively large while simultaneously providing norms across a substantial number of lexical dimensions. Second, for any given subset of words, the same participants provided ratings across all 9 dimensions (32 participants/word, on average). Third, two novel dimensions of semantic size and gender association are included. Finally, the corpus contains a set of 379 ambiguous words that are presented alone (e.g., *toast*) or with information that selects an alternative sense (e.g., *toast (bread)*, *toast (speech)*). Relationships between the dimensions of the Glasgow Norms were initially investigated by assessing their correlations. In addition, a principal component analysis revealed four main factors accounting for 82% of the variance ("visualization," "emotion," "salience," and "exposure"). The validity of the Glasgow Norms was established via comparisons of our ratings to 14 different sets of current psycholinguistic norms. Alternative senses of ambiguous words (i.e., disambiguated forms), when discordant on a given dimension, seemingly led to appropriately distinct ratings. Informal comparisons between ratings of ambiguous words and their alternative senses showed different patterns that likely depended on several factors (the number of senses, their relative strengths, and the rating scales, themselves). Overall, the Glasgow Norms provide a valuable resource, in particular, for researchers investigating the role of word recognition in language comprehension.

Keywords: psycholinguistic norms, arousal, valence, dominance, concreteness, imageability, familiarity, age of acquisition, semantic size, gender association

The Glasgow Norms provide a set of normative ratings for 5,553 English words on 9 psycholinguistic dimensions. Each word was rated on the dimensions of arousal, valence, dominance, concreteness, imageability, familiarity, age of acquisition, semantic size, and gender association. The aim was to develop a substantial set of standardized, freely available, psycholinguistic materials. The norms provide researchers with a considerable collection of materials that are not only reliably evaluated on specific dimensions of interest, but also on other potentially confounding dimensions. Accordingly, the norms allow for the creation and analysis of carefully controlled stimuli, facilitating continued investigations into these lexical dimensions as well as their interactions.

In comparison to previous word norms, the Glasgow Norms offer several significant, novel features. First, a relatively large number of lexical dimensions (9) was examined. Other norms typically assess only one to three dimensions. Second, the same participant provided ratings across all 9 dimensions for any given word. Currently, researchers interested in investigating more than a few lexical dimensions need to access different sets of norms which are tested on different populations of participants. Additionally, as different norms test non-overlapping word sets, it is often difficult to obtain ratings on all stimuli on all dimensions of interest. Third, two of the dimensions, semantic size and gender association, have not been investigated to date via an extensive set of norms. Finally, many words in the English lexicon are ambiguous, having more than one meaning (e.g., *bank*, having a 'money' or 'river' sense). The Glasgow Norms include ambiguous words presented in different forms (to different participants) – as isolated words (e.g., *bank*), and as words presented with disambiguating information (e.g., *bank (money)* or *bank (river)*). These key aspects of our approach make the Glasgow Norms a unique and valuable methodological contribution.

There are currently several sets of psycholinguistic norms that report ratings of words on particular psycholinguistic dimensions. Typically, such norms comprise ratings of either 1,000 or so words on a few dimensions or more than 10,000 words on a single dimension. **Table 1** summarizes such norms, limited to those based on more than 500 words. For each set of norms, information is provided about the lexical dimensions examined, the number of words used, and the number of participants tested.

Dimension(s)	*N* Items	Participants/Item	Source
AROU, VAL, DOM	1,034	50 on average	Bradley & Lang (1999)
AROU, VAL, DOM	13,915	18-30 for most items	Warriner, Kuperman, & Brysbaert (2013)
CNC	37,058	at least 25	Brysbaert, Warriner, & Kuperman (2014)
IMAG (monosyllabic words)	3,000	31	Cortese & Fugett (2004)
IMAG (disyllabic words)	3,000	35	Schock, Cortese, & Khanna (2012a)
CNC, IMAG	925	28 (CNC), 30 (IMAG)	Paivio, Yuille, & Madigan (1968)[a]
CNC, IMAG	1,080	50	Friendly, Franklin, Hoffman, & Rubin (1982)
CNC, IMAG, FAM	2,854	54-65	Toglia & Battig (1978)[b]
CNC, IMAG, FAM, AOA	1,944	35 (CNC), 37 (IMAG), 36 (FAM, AOA)	Gilhooly & Logie (1980)[c]
IMAG, FAM	2,311	16 (IMAG), 47-49 (FAM)	Clark & Paivio (2004)[d]
IMAG, AOA	2,694	78 (IMAG), 45 (AOA)	Bird, Franklin, & Howard (2001)
IMAG, FAM, AOA	1,526	20	Stadthagen-Gonzalez & Davis (2006)
AOA (monosyllabic words)	3,000	32	Cortese & Khanna (2008)
AOA (disyllabic words)	3,000	32	Schock, Cortese, Khanna, & Toppi (2012b)
AOA	30,124	18-22 for most items	Kuperman, Stadthagen-Gonzalez, & Brysbaert (2012)
GEND	600	356	Crawford, Leynes, Mayhorn, & Bink (2004)

Table 1. Inventory of English Word Norms. For each word norm, the relevant semantic dimension(s), number of words tested, number of participants per item, and citation are specified. Selected word norms comprise those having more than 500 lexical items. AROU = arousal, VAL = valence; DOM = dominance; CNC = concreteness; IMAG = imageability ("imagery" in earlier norms); FAM = familiarity; AOA = age of acquisition; GEND = gender association. [a]Paivio et al. (1968) also measured meaningfulness. [b]Toglia and Battig (1978) also measured meaningfulness, pleasantness, categorizability, and number of attributes or features. [c]Gilhooly and Logie (1980) also measured ambiguity. [d]Clark and Paivio (2004) also measured an additional 13 dimensions, but only on the original set of 925 items from Paivio et al. (1968).

It is beyond the scope of the current investigation to catalogue norms comprising fewer than 500 lexical items. Nevertheless, over the past several decades, such norms have proved valuable and have been extensively used (e.g., Morrison, Chappell, & Ellis' (1997) age of acquisition norms). Oftentimes, however, researchers need to use multiple sets of smaller norms to adequately describe the characteristics of their experimental stimuli (e.g., Scott, O'Donnell, & Sereno, 2012; Sereno, O'Donnell, & Sereno, 2009; Sereno, Scott, Yao, Thaden, & O'Donnell, 2015). Alternatively, researchers have frequently gathered local ratings on their stimuli to insure the validity of the lexical dimension(s) of interest (e.g., Altarriba, Bauer, & Benvenuto, 1999; Juhasz & Rayner, 2003; Kousta, Vinson, & Vigliocco, 2009; Sereno et al., 2009; Yao et al., 2013, 2017). In other cases, the dimension of interest, although pertinent to the study, is one that is either not widely employed or well-established. For example, researchers have evaluated words on the basis of "context availability" (Schwanenflugel, Harnishfeger, & Stowe, 1988), "danger" and "usefulness" (Wurm, 2007), "offensiveness" and

"tabooness" (Janschewitz, 2008), or "body-object interactivity" (Siakaluk, Pexman, Aguilera, Owen, & Sears, 2008).

The Glasgow Norms provide ratings of 5,553 words on 9 dimensions: arousal (AROU), valence (VAL), dominance (DOM), concreteness (CNC), imageability (IMAG), familiarity (FAM), age of acquisition (AOA), semantic size (SIZE), and gender association (GEND). The first three dimensions – AROU, VAL, and DOM – are typically used to characterize a word's emotional impact. AROU is a measure of internal activation (excitement, calmness), VAL is a measure of value or worth (positive, negative), and DOM indicates the degree of control one feels (dominant, controlled). Similar to existing emotion norms (e.g., Bradley & Lang, 1999; Warriner, Kuperman, & Brysbaert, 2013), these are measured on 9-point scales. In the psycholinguistic literature, emotion is generally represented within a two-dimensional framework (e.g., Osgood, Suci, & Tannenbaum, 1957; Russell, 1980), with greater emotionality associated with higher arousal and extreme valence. In behavioral terms, positive and negative emotion words tend to be recognized faster than comparable neutral words (e.g., Scott et al., 2009, 2012, 2014; Sereno et al., 2015; Yao et al., 2017).

All remaining dimensions of the Glasgow Norms are based on 7-point rating scales, a practice consistent with most existing norms. CNC represents the degree to which something can be experienced by our senses (concrete, abstract). Concrete words are typically recognized faster than abstract words (e.g., Juhasz & Rayner, 2003; Schwanenflugel et al., 1988; Whaley, 1978; Yao et al., 2013, 2017). Kousta, Vigliocco, Vinson, Andrews, and Del Campo (2011), however, proposed that abstract words tend to be more emotionally valenced than concrete words which gives rise to a residual processing advantage of abstract over concrete words, critically, once opposing effects of context availability and imageability are controlled. IMAG represents the degree of effort involved in generating a mental image of something (imageable, unimageable). In general, imageable words are facilitated in processing compared to less imageable words (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Cortese & Schock, 2013; Yao et al., 2017). CNC and IMAG, although highly correlated (see, e.g., Paivio, Yuille, & Madigan, 1968), are nevertheless considered to capture distinct semantic aspects of a word (Kousta et al., 2011; Richardson, 1976).

The measures of FAM and AOA are related in different subjective ways to the objective measure of word frequency, in which the relative number of occurrences of individual words within a substantial corpus (more often written than spoken) are calculated (e.g., the British National Corpus, 2007; Davies, 2004). FAM is a measure of a word's subjective experience (familiar, unfamiliar), and can be partially contrasted with subjective frequency estimates (Balota, Pilotti, & Cortese, 2001), which are considered to be less dependent on other meaning-level variables. Words that are more familiar are recognized faster than those that are less familiar (e.g., Connine, Mullennix, Shernoff, & Yelen, 1990). AOA is a measure of the age at which a word was initially acquired. Although there are alternative

ways of measuring AOA (e.g., Juhasz, 2005; Morrison et al., 1997), it is often assessed by adults providing an estimate of when they first learned a word, in spoken or written form, on a 7-point scale (a series of 2-year periods from 0-12 years and a final 13+ period). Zevin and Seidenberg (2002, 2004) suggested that our developmental experience with words may be better captured by measures of their cumulative frequency (summed lifetime usage) and frequency trajectory (how usage changes over time). Behaviorally, words acquired earlier in life are recognized faster than those acquired later (e.g., Cortese & Khanna, 2007; Johnston & Barry, 2006; Juhasz & Rayner, 2006; Sereno & O'Donnell, 2009).

The final dimensions of SIZE and GEND have only been the subject of more recent psycholinguistic investigations (e.g., Sereno & O'Donnell, 2009; Sereno et al., 2009; Yao et al., 2013). SIZE is a measure of magnitude (big, small) expressed in either concrete or abstract terms. That is, words can refer to objects or concepts that are considered bigger (e.g., *castle*, *wealth*) or smaller (e.g., *pocket*, *unique*). It has been demonstrated that words referring to bigger things are recognized faster than those referring to smaller ones (e.g., Sereno et al., 2009; Yao et al., 2013). GEND is a measure of the degree to which words are considered to be associated with male or female behavior (masculine, feminine). Recent norms have specifically examined gender perception of role nouns across languages (Garnham, Doehren, & Gygax, 2016; Misersky et al., 2014). Although reading studies have investigated gender role stereotypes (e.g., *electrician*, *secretary*) and their gendered mis/matching pronouns (e.g., Duffy & Keir, 2004; Kreiner, Sturt, & Garrod, 2008), there has been little if any research into gender associations to a much broader spectrum of content words. Sereno and O'Donnell (2009) investigated words rated as either male- or female-oriented (e.g., *frog*, *cigar*, *guitar* or *duck*, *flute*, *heaven*, respectively) in a lexical decision task (AOA was also manipulated). They found that while female participants demonstrated an advantage to same-gendered words (e.g., responses were faster to *tights* than *pliers*), male participants showed no such comparable bias (i.e., *pliers* was no faster than *tights*).

The current corpus of 5,553 words includes a range of content words (nouns, verbs, adjectives, adverbs) as well as 379 semantically ambiguous words (homographs) whose alternative meanings were additionally rated. To our knowledge, only three existing norms have explicitly included ambiguous words. Gilhooly and Logie (1980) had participants rate whether or not their words had multiple meanings, resulting in a set of ambiguous words (N = 649) that were then further rated for the relative dominance of alternative senses. Clark and Paivio (2004), in their extension of the Paivio et al. (1968; N = 925) norms, included number of meanings as an additional measure. Neither of these norms, however, collected ratings on the alternative meanings, themselves. Bird, Franklin, and Howard (2001) did acquire IMAG and AOA ratings on a subset of their items (N = 110) of noun-verb homographs (disambiguated by preceding the ambiguous word with *a* or *to*, respectively). While most ambiguous words are "biased," having a strongly dominant meaning and one or more subordinate meanings, some

are "balanced," having two more salient meanings with other possible subordinate senses (Sereno, O'Donnell, & Rayner, 2006). The ratings associated with ambiguous words from previous norming studies which have not explicitly disambiguated such words probably reflect participants' interpretation of the dominant meaning, although this is not a certainty. The ambiguous words identified in the Glasgow Norms were presented alone (e.g., *ball*), or in disambiguated form (e.g., *ball (sphere)* or *ball (dance)*), critically, to different participants.

The Glasgow Norms were collected by presenting our list of 5,553 words to participants in lists of either 101 or 150 words. For each list, participants rated words separately on all 9 dimensions described above. The relations among dimensions are explored and comparisons to other norms are provided.

Method

Participants

A profile detailing the number, age, and gender of participants is presented in **Table 2**. A total of 829 individuals ("unique participants") took part in the rating studies, with some completing more than one word list. When participants were tallied on the basis of completing a single list ("all participants"), regardless of whether they had completed other lists, the total came to 1,368. Overall, ages ranged from 16 to 73 years and there was slightly more than twice the number of females than males who took part. Participants were native English speakers from the University of Glasgow community. They were either paid at a rate of £6/hr or were given course credit for their participation. The study conformed to British Psychological Society ethical guidelines and protocols.

	Unique Participants			All Participants	
	***N* (%)**	**Age (SD)**	**# Lists (SD)**	***N* (%)**	**Age**
Female	599 (72)	21.5 (7.6)	1.6 (1.3)	960 (70)	22.6
Male	230 (28)	22.3 (6.9)	1.8 (1.6)	409 (30)	23.5
All	829 (100)	21.7 (7.4)	1.7 (1.4)	1368 (100)	22.8

Table 2. Age and gender profile of participants. The number of participants, age, and/or average number of lists completed are provided by grouping and gender. "Unique participants" comprise individuals, some of whom provided ratings for more than one list of words; "all participants" represent the total number of participants responding to all lists, and does not take into account whether any given participant took part in more than one list. The majority of participants, 69%, completed a single list of words. The remaining percentages of individual participants completing more than one list are as follows: 17% did 2 lists, 3% did 3 lists, 8% did 4 lists, and 3% did 8 lists.

Materials

A corpus of 5,553 words was assembled from an initial set of 808 words and a subsequent, larger set of 4,800 words (with 55 words included in both lists). The data acquired from these two sets were merged into a single corpus for subsequent analyses. Words ranged in length from 2 to 16 letters, with an average length of 6.10 letters (*SD* = 1.99).

The corpus included 379 ambiguous words. Each was presented in isolation or with disambiguating information following the word in parentheses (e.g., *solution*, or *solution (answer)* and *solution (chemical)*, respectively). The average number of disambiguated forms presented was 2.30 (*SD* = .58). The number of words having two, three, four, and five alternative meanings was 289, 69, 19, and 2, respectively. Thus, there was a total of 871 items in the corpus that were presented with disambiguation.

Procedure

Participants were recruited opportunistically via an experiment advertisement link on the home page of Psychology at the University of Glasgow. The experiment was run online via an in-house, experimental platform (*http://experiments.psy.gla.ac.uk*). Each participant rated a list of either 101 (8 possible lists of the 808 word set) or 150 words (32 lists of 4,800 word set). Lists of 101 or 150 words were generated by taking every 8^{th} or 32^{nd} item from alphabetized versions of either the 808 or 4,800 sets, respectively. This way, each list was representative of the set in terms of its distribution of word-initial letters and no list contained more than one instance of any given ambiguous word in any of its forms.

The general instructions for the experiment as well as the specific instructions for each of the 9 different rating tasks are presented in the **Appendix**. The same participant provided ratings across all 9 dimensions for any given word. Participants rated all words of a list on one scale, then all words on the next scale, and so on. The order of words within each scale was randomized as was the order of scales across participants. The approximate time to complete the experiment was 40 or 60 minutes for 101- or 150-item lists, respectively.

Results

Data were eliminated from further analyses if the response time (RT) to any word on any scale was less than 750 ms or if participants reported not knowing a word. For RT, examination of the trial-by-trial data revealed the presence of episodes of rapid responding by some participants, typically repeating a given rating value. In such cases, RTs were often less than 400 ms. In two recent, large-scale lexical decision experiments performed locally, average RTs to words were just under 600 ms (Sereno et al. (2015) used 240 words with 144 participants; Yao et al. (2017) used 270 words with 127 participants). Given that the judgements required in our rating tasks were, minimally, less definite and,

generally, more demanding than making lexical decisions, a lower RT cutoff of 750 ms was implemented. Of the total number of responses recorded (N = 1,712,607), the RT distribution was as follows: 5.30% were shorter than 750 ms, 75.51% were 750-3000 ms (with 36.03% 1500-2000 ms), 12.98% were 3000-5000 ms, and 6.27% were longer than 5000 ms. In terms of word knowledge, for all scales except for FAM, if participants did not know the meaning of a word, they could select the "unfamiliar word" button instead of rating it (see instructions in the **Appendix**). This option accounted for 0.33% of all responses. On average, there were 32.55 responses per word (SD = 3.89). A detailed profile of the number of responses across all 9 rating scales is presented in **Table 3**.

N		AROU	VAL	DOM	CNC	IMAG	FAM	AOA	SIZE	GEND
5,553	***M***	32.71	32.98	32.60	32.71	32.60	30.58	33.70	32.78	32.33
	SD	3.74	3.76	3.78	3.85	3.81	3.71	3.72	3.84	4.03
	range	13-69	14-70	14-69	10-70	14-69	16-66	17-70	12-69	13-69
55	***M***	64.62	64.67	63.95	63.29	62.80	56.73	66.11	65.11	62.24
	SD	1.59	1.95	2.47	3.28	3.51	3.95	1.71	1.92	3.31
	range	60-69	61-70	58-69	54-70	54-69	48-66	63-70	59-69	56-69
5,498	***M***	32.39	32.66	32.28	32.41	32.30	30.31	33.38	32.45	32.03
	SD	1.96	2.01	2.11	2.33	2.30	2.61	1.82	2.08	2.70
	range	13-36	14-36	14-36	10-36	14-36	16-35	17-36	12-36	13-36

Table 3. Profile of the number of responses across dimensions. Profile of the number of responses to the overall corpus (N = 5,553), the subset of words (N = 55) repeated across the 808- and 4,800-word lists, and the majority of words (N = 5,498) presented in only one of the two lists.

The descriptive statistics for the 9 rated dimensions are presented in **Table 4**. The Glasgow Norms are available as part of the **Supplementary Materials** to this article and are provided in .csv format. The Glasgow Norms present an alphabetized list of 5,553 words. The columns, from left to right, are as follows: Word, Length (which excludes possible disambiguating information), and, for each of the 9 dimensions, the mean rating (M), standard deviation (SD), and number of ratings (N) for each word.

Dimension	Scale Range	*M*	*SD*
AROU	1–9	4.65	1.10
VAL	1–9	5.10	1.56
DOM	1–9	5.06	0.92
CNC	1–7	4.63	1.42
IMAG	1–7	4.77	1.36
FAM	1–7	5.19	0.94
AOA	1–7	4.13	1.25
SIZE	1–7	4.09	1.03
GEND	1–7	4.13	0.93

Table 4. Descriptive statistics of the nine dimensions of the Glasgow norms. Scale ranges and mean (*M*) and standard deviation (*SD*) values for the nine psycholinguistic dimensions of the Glasgow Norms. AROU = arousal; VAL = valence; DOM = dominance; CNC = concreteness; IMAG = imageability; FAM = familiarity; AOA = age of acquisition; SIZE = semantic size; GEND = gender association.

Relations between the nine dimensions of the Glasgow Norms

To provide an initial overview of the relations between all 9 Glasgow Norms scales, we performed Spearman correlations and these are presented in **Table 5**. As Spearman correlations are rank-based, this method takes into account linear and non-linear relations between dimensions. We used the Bonferroni method to correct *p*-values for multiple tests and applied a significance threshold of $p = .01$. Due to the large number of items ($N = 5{,}553$), almost all correlations were significant. However, considering only large effects (i.e., with *r*s > .5; Cohen, 1988), the following correlations were particularly strong: CNC × IMAG ($r = .91$; the more concrete a word is, the easier it is to imagine); VAL × DOM ($r = .68$; the more positive a word is, the more it provokes feelings of dominance); FAM × AOA ($r = -.67$; the more familiar a word is, the earlier that word was learned); and SIZE × AROU ($r = .51$; the bigger the object or concept is to which a word refers, the more arousing it is).

	AROU	VAL	DOM	CNC	IMAG	FAM	AOA	SIZE	GEND
AROU	**–**								
VAL	.35	**–**							
DOM	.34	**.68**	**–**						
CNC	-.26	(.05)	(.05)	**–**					
IMAG	-.10	.09	.08	**.91**	**–**				
FAM	.18	.29	.23	.10	.21	**–**			
AOA	(.00)	-.19	-.14	-.38	-.49	**-.67**	**–**		
SIZE	**.51**	.12	.09	-.41	-.33	(.05)	.22	**–**	
GEND	-.10	-.42	-.09	.15	.08	-.21	.15	.15	**–**

Table 5. Correlations between dimensions of the Glasgow norms. Spearman coefficients for all combinations of scales. All correlations are significant ($p < .01$; Bonferroni corrected) except those listed in parentheses. Tests printed in bold are those considered large, with *r*s > .50 (Cohen, 1988).

AROU = arousal; VAL = valence; DOM = dominance; CNC = concreteness; IMAG = imageability; FAM = familiarity; AOA = age of acquisition; SIZE = semantic size; and GEND = gender association.

For a more detailed analysis of relations between scales, we fit linear and quadratic models to the data, using the MATLAB function *fitlm* (The MathWorks, Inc.). To account for outliers, the fits were computed using a robust least-squares method (bisquare weighting function). Reported R^2s were adjusted for the number of coefficients. The results for all combinations of dimensions using linear and quadratic fits are included in the **Supplementary Materials** to this article as **Table S1** and **Table S2**, respectively.

We will highlight effects of SIZE and GEND, as these two dimensions are relatively new and less well understood. **Figure 1** shows quadratic fits for all combinations with either SIZE or GEND that account for more than 18% of variance (see **Table S2**): SIZE × AROU (R^2 = .27); SIZE × CNC (R^2 = .19); VAL × SIZE (R^2 = .19); and VAL × GEND (R^2 = .18). For three of these, linear fits accounted for comparable (but numerically slightly less) amounts of variance (see **Table S1**) and are straightforward to interpret: SIZE × AROU (the semantically bigger a word is, the more arousing it is); SIZE × CNC (the semantically bigger a word is, the less concrete it is); and VAL × GEND (the more positive a word is, the more feminine it is). In contrast, VAL × SIZE was better explained by a quadratic (R^2 = .19) than linear (R^2 < .01) fit (the more extremely valenced – negative *or* positive – a word is, the semantically bigger it is).

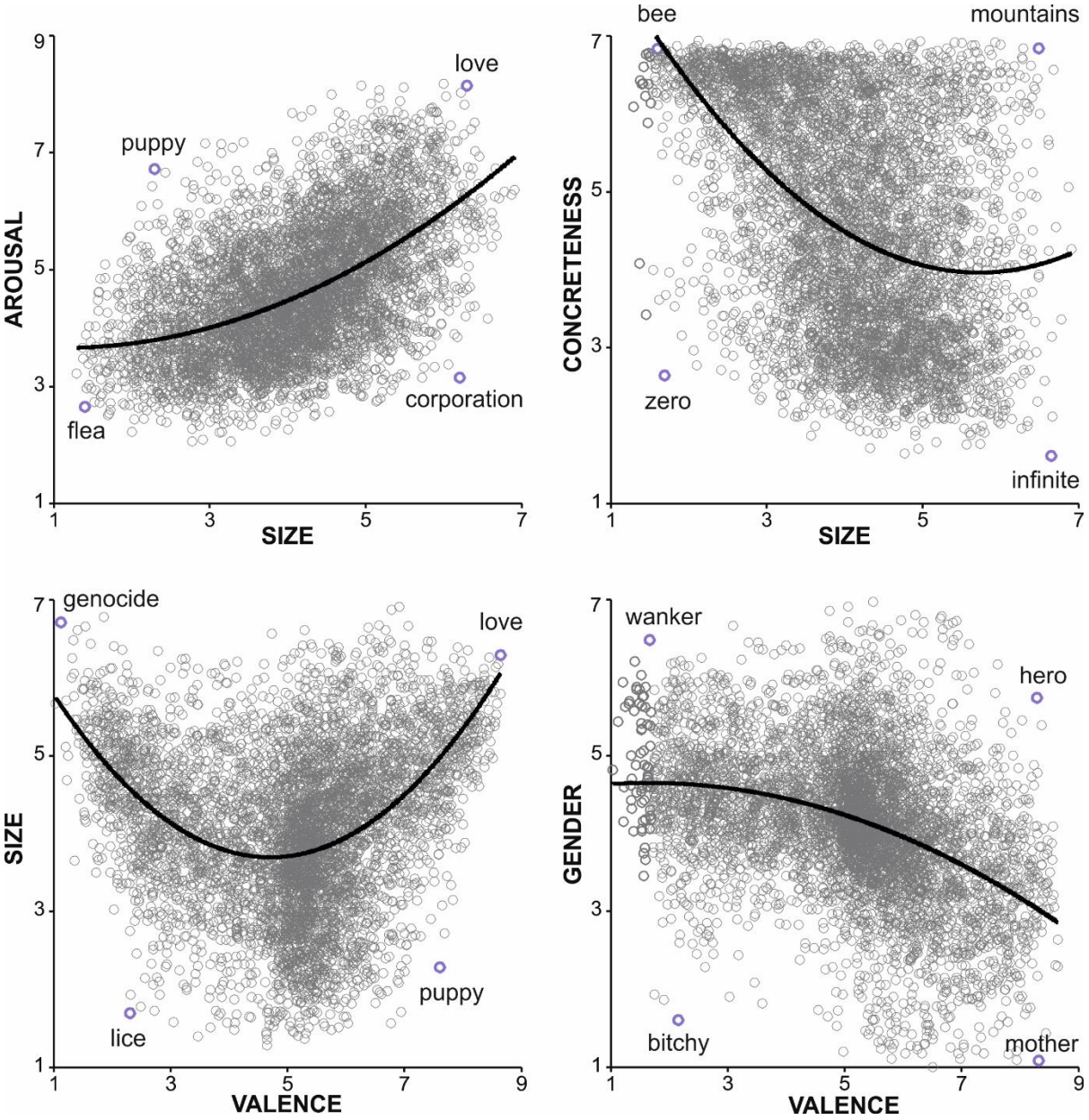

Figure 1. Quadratic relations between semantic size or gender association and other dimensions. Quadratic fits with SIZE or GEND that explain more than 18% of variance. R^2s, *F*-values, and significance for linear and quadratic fits for all combinations of dimensions with can be found in **Tables S1** and **S2** of the **Supplementary Materials**.

Factor analysis of dimensions

To summarize and interpret the correlation results and to explore the relative alignment of the newer scales of SIZE and GEND, we performed a factor analysis. Data for all 9 scales were submitted to a principal component analysis (PCA) with an oblique rotation (direct oblimin, Harman, 1976; Jennrich & Sampson, 1966). Note that using an orthogonal rotation (e.g., varimax, quartimax, or equamax) yielded comparable results. Factors with eigenvalues greater than one were included in the factor solution (Kaiser, 1960).

The factor analysis is presented in **Table 6** and yielded a solution with four factors. The first factor "visualization" accounted for 30% of the variance in the data and included the scales CNC and IMAG. The second factor "emotion" accounted for an additional 26% of the variance and included VAL and DOM. The third and fourth factors each accounted for 13% of the variance: a "salience" factor, including SIZE, GEND, and AROU, and an "exposure" factor, including FAM and AOA. Together, the four factors explained 82% of the common variance. The communality for each scale was above .6, indicating that the amount of variance accounted for by the retained factors was sufficient. In other words, the scales' variance was useful in delineating the extracted factors.

	Factor 1 "visualization"	**Factor 2 "emotion"**	**Factor 3 "salience"**	**Factor 4 "exposure"**
CNC	**0.935**	0.098	-0.093	-0.053
IMA	**0.887**	0.116	-0.020	-0.219
VAL	0.022	**0.920**	-0.156	-0.018
DOM	0.171	**0.913**	0.105	0.103
SIZE	-0.336	0.081	**0.759**	-0.071
GEND	0.453	-0.287	**0.675**	0.239
AROU	-0.236	0.437	**0.533**	-0.207
FAM	-0.028	-0.036	0.086	**-0.935**
AOA	-0.284	0.067	0.084	**0.845**
%Var	29.724	25.855	13.190	13.003
%CumVar	29.724	55.579	68.768	81.771

Table 6. Factor loadings and for all dimensions. Reported are loadings of an oblique rotation matrix (direct oblimin) on four factors. Loadings greater than .5 are highlighted in bold. Explained common variance (%Var) is given for individual factors as well as the cumulative variance (%CumVar).

It is noteworthy that most scales loaded relatively high (i.e., above an absolute value of .5) on one factor. However, AROU (loading highest on factor "salience") also loaded on the factor "emotion," and GEND (also loading highest on factor "salience") additionally loaded on the factor "visualization," indicating that these variables cannot be explained in terms of a single factor.

Correlations with other psycholinguistic norms

To confirm the validity of our ratings, we correlated the dimensions of the Glasgow Norms with 14 of the 16 different sets of English norms listed in **Table 1**. Norms that were excluded were those that were not easily accessible. Between one and seven norms were available for all dimensions except SIZE. As linear relations between norms were expected, we performed Pearson correlations for all shared words. These correlations are presented in **Table 7**. All correlations were highly significant ($ps < .0001$, Bonferroni corrected) and the vast majority showed a Pearson coefficient of great than .5,

indicating a large effect (Cohen, 1988) and, therefore, sufficient validity. The highest and most consistent correlations were achieved by VAL, CNC, AOA, and GEND (with nearly all *r*s > .9).

Norms	N_{source}	$N_{overlap}$	AROU	VAL	DOM	CNC	IMAG	FAM	AOA	GEND
1	1,034	951	**.66**	**.95**	**.82**					
2	13,915	4,073	**.62**	**.93**	**.69**					
3	37,058	4,445				**.93**				
4	3,000	1,363					**.88**			
5	3,000	1,308					**.89**			
6	925	789				**.93**	**.92**			
7	1,944	902				**.93**	**.89**	**.82**	**.92**	
8	2,311	1,390					.42	**.82**		
9	2,694	994					**.80**		**.86**	
10	1,526	1,370					**.91**	**.82**	**.94**	
11	3,000	1,363							**.91**	
12	3,000	1,308							**.91**	
13	30,124	4,283							**.89**	
14	600	336								**.96**

Table 7. Correlations between the Glasgow norms and other English word norms. Pearson coefficients for 14 sets of norms reporting scales corresponding to the Glasgow Norms (note that no norms were available for semantic size). For each of the norms, the number of total items (N_{source}) and the number of identical items within the Glasgow Norms ($N_{overlap}$) which were used for the correlations are indicated. All correlations were highly significant (*p*s < .0001, Bonferroni corrected). Correlations with a large effect (*r* > .5, see Cohen, 1988) are printed in bold. References for the 14 norms are as follows: 1 = Bradley and Lang (1999); 2 = Warriner et al. (2013); 3 =Brysbaert et al. (2014); 4 = Cortese and Fugett (2004); 5 = Schock et al. (2012a); 6 = Paivio et al. (1968); 7 = Gihooly and Logie (1980); 8 = Clark and Paivio (2004); 9 = Bird et al. (2001); 10 = Stadthagen-Gonzalez and Davis (2006); 11 = Cortese and Khanna (2008); 12 = Schock et al. (2012b); 13 = Kuperman et al. (2012); and 14 = Crawford et al. (2004).

Ambiguous words

We did not perform any formal analyses on the ratings of ambiguous words in the corpus, whether they occurred in isolation (e.g., *pen*) or in disambiguated form (e.g., *pen (ink)*, *pen (cage)*). Informal examination of the ratings, however, indicated certain patterns. First, when disambiguating information was provided, the alternative senses of ambiguous words received distinct ratings where relevant to the dimension in question. **Figure 2A** illustrates the ratings that alternative senses of several ambiguous words received across the 9 dimensions, in particular, where the alternative senses were expected to lead to disparate judgments. The second aspect of ambiguous word ratings concerned the relationship between a word's ambiguous and disambiguated forms. While ambiguous words typically have a dominant sense and one or more subordinate senses (Sereno et al., 2006), the relative strengths of these alternative senses can vary substantially, not only across items, but across individuals. It is

also possible that the dimensions, themselves, may have served as 'contexts' for ambiguous words presented in isolation (i.e., given the scale AROU, a participant may have rated *plot* in its 'story' sense, but later, given the scale CNC, rated *plot* in its 'land' sense, because these were the most accessible meanings). **Figure 2B** shows the ratings across all dimensions of the ambiguous word *shell* and two of its alternative senses ('sea' and 'military'). Although ratings for *shell* tended to be closer to its dominant 'sea' sense than its subordinate 'military' sense, they did not always overlap as might be expected. Moreover, we observed several different patterns of ratings across ambiguous items. Oftentimes, it did not appear that ambiguous words were rated according to only one of their senses. Without an independent measure of the dominance relationships among alternative senses of ambiguous words in our corpus, however, we are at present unable to characterize this data. Anecdotally, factors such as number of meanings, their relative strength, and the rating scale, itself, seem to play a role.

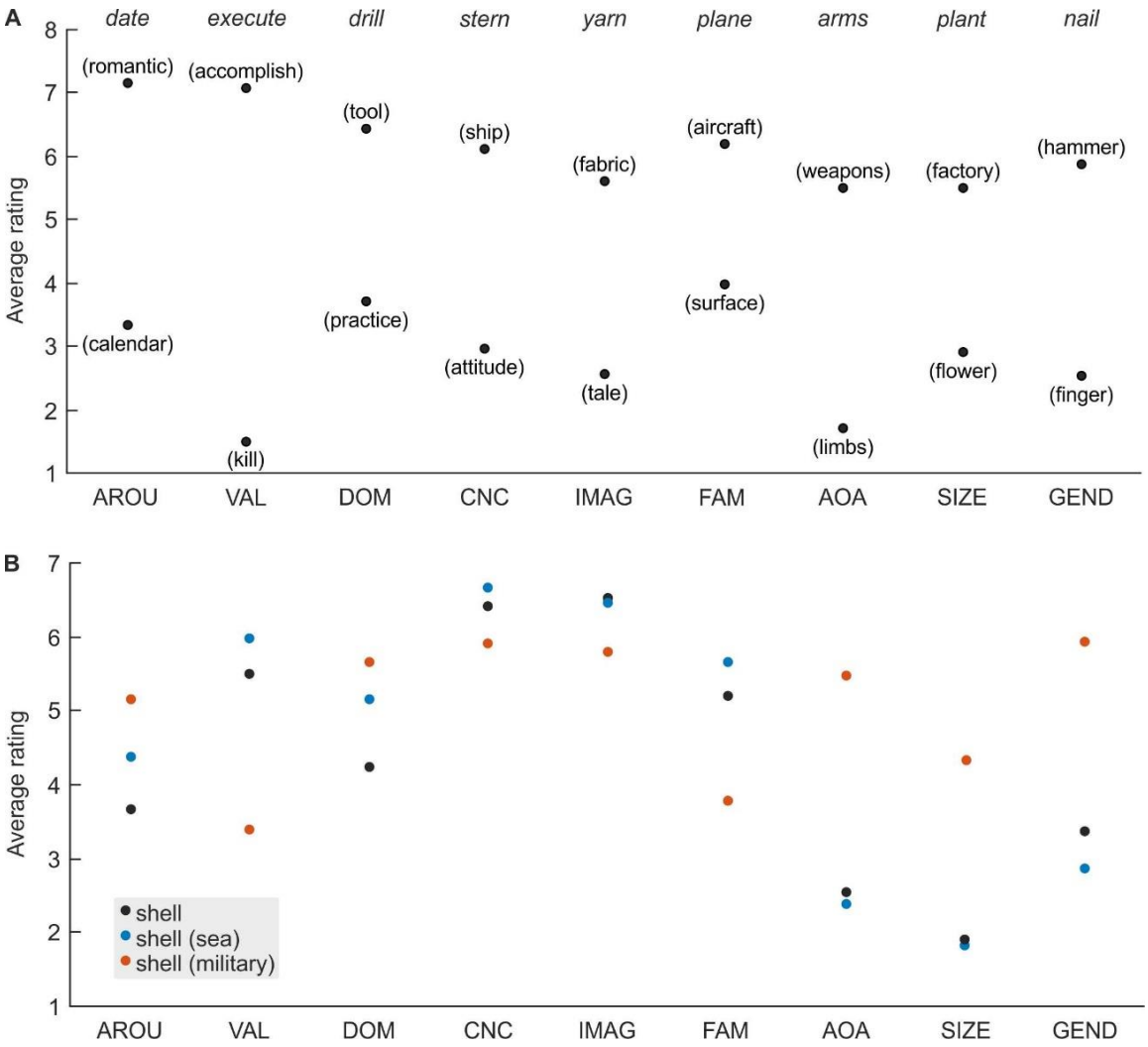

Figure 2. Example ratings for ambiguous words. A) Ratings of alternative senses of ambiguous words across dimensions. For each dimension of the Glasgow Norms, an example ambiguous word is listed across the top. The ratings corresponding to the alternative senses (defined in parentheses) are

indicated beneath each ambiguous word. **B)** Ratings of the ambiguous word *shell* and its alternative senses ('sea' and 'military') across dimensions of the Glasgow Norms. AROU = arousal; VAL = valence; DOM = dominance; CNC = concreteness; IMAG = imageability; FAM = familiarity; AOA = age of acquisition; SIZE = semantic size; and GEND = gender association.

Discussion

The Glasgow Norms examined 9 semantic dimensions of words in a corpus of 5,553 words, with an average of 32 participants contributing ratings to each word on each scale. Seven of the dimensions (AROU, VAL, DOM, CNC, IMAG, FAM, and AOA) are well-established and have been extensively investigated, while two dimensions (SIZE and GEND) are relatively novel and have not been investigated in a comprehensive way. In comparison to past norms, the Glasgow Norms provide an internally consistent set of ratings, not only across a sizeable corpus, but also across a considerable set of psycholinguistic dimensions. Moreover, the Glasgow Norms provide ratings for a significant number of ambiguous words (N = 379), presented in isolation (e.g., *figure*) as well as in disambiguated forms (e.g., *figure (body shape)*, *figure (graph)*, *figure (number)*, and *figure (reckon)*).

Analyses comprised evaluating the relations among the dimensions in the Glasgow Norms and comparing its results to those of other norms. Correlations between the 9 dimensions of the Glasgow Norms were generally significant (see **Table 5**) due to the large number of items. Particularly strong relationships (with rs > .5; Cohen, 1988) included the following: CNC × IMAG (concrete words are easier to imagine); VAL × DOM (positive words provoke greater feelings of dominance); FAM × AOA (familiar words are acquired earlier); and SIZE × AROU (words referring to bigger things are more arousing). The first three relationships are established in the literature (e.g., Bradley & Lang, 1999; Friendly et al., 1982; Gilhooly & Logie, 1980; Paivio et al., 1968; Stadthagen-Gonzalez & Davis, 2006; Toglia & Battig, 1978; Warriner et al., 2013). The latter is a novel finding, although it has already obtained behavioral support (Yao et al., 2013). In further analyses, we fit linear and quadratic models for all combinations of the dimensions (see **Tables S1** and **S2**, **Supplementary Materials**). We focussed on effects related to the relatively new dimensions of SIZE and GEND (see **Figure 1**). For SIZE, words referring to bigger things were more arousing, more extremely (positively or negatively) valenced, and more abstract. For GEND, feminine words were more positive.

Factor analysis of all dimensions of the Glasgow Norms yielded a four-factor solution accounting for 82% of the variance (see **Table 6**). The factors and their associated high-loading dimensions were as follows: "visualization" (CNC, IMAG), "emotion" (VAL, DOM), "salience" (SIZE, GEND, AROU), and "exposure" (FAM, AOA). Notably, both AROU and GEND also loaded moderately on "emotion" and "visualization," respectively. The lack of a one-to-one mapping between factors and dimensions highlights both the complexity of these semantic relationships as well as the need to recognize their potential influence in the design and analysis of psycholinguistic research.

The validity of the Glasgow Norms was assessed by a series of correlations with 14 different sets of English norms (see **Table 7**). All dimensions of the Glasgow Norms were tested with the exception of SIZE (to our knowledge, we are the first to obtain extensive ratings for this dimension). For any given dimension, between one and seven comparisons were made to previous sets of norms. Correlations with the prior norms were highly significant across the eight dimensions of AROU, VAL, DOM, CNC, IMAG, FAM, AOA, and GEND.

Finally, the Glasgow Norms included a set of 379 ambiguous words – presented alone or with disambiguating information. This is the first time an appreciable number of ambiguous words as well as their alternative senses have been normed across an extensive number of lexical dimensions. Informal examination of the data demonstrated that alternative senses of ambiguous words having contrasting meanings were rated appropriately (see **Figure 2A**). Ambiguous words presented in isolation were sometimes rated according to their highly dominant sense across the different dimensions (see **Figure 2B**). In general, however, the rating patterns we observed for ambiguous words and their disambiguated senses varied. We believe that these different configurations most likely depend on several factors, including the number of alternative senses, the dominance relationship among these senses, as well as the rating scales, themselves.

In conclusion, the Glasgow Norms represent a valuable resource, providing a substantial set of words normed across a large number of psycholinguistic dimensions. Key features of the norms include the evaluation of the relations between dimensions, the validation of established dimensions, the assessment of novel dimensions, and the examination of ambiguous words and their meanings. Use of the Glasgow Norms will allow both the manipulation and control of lexical variables, in particular, in studies that investigate word recognition processes, whether in the experimental context of word-based tasks or during the course of fluent reading. Establishing the semantic contingencies and interactions of these variables will inform models of language processing.

Acknowledgments

This research was supported in part by an Economic and Social Research Council (ESRC) grant RES-062-23-1900 to SCS and by a Carnegie Collaborative Research Grant (Trust Reference No. 50084) from the Carnegie Trust for the Universities of Scotland awarded to SCS, GGS, and PJO.

References

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 578–602.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, *29*(4), 639–647.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, *33*(1), 73–79.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings* (Technical Report C-1). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

British National Corpus, version 3 (BNC XML Edition). (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/.

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.

Clark, J.M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 371–383.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1084–1096.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 384–387.

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology*, *60*(8), 1072–1082.

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., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, *66*(5), 946–972.

Crawford, J. T., Leynes, P. A., Mayhorn, C. B., & Bink, M. L. (2004). Champagne, beer, or coffee? A corpus of gender-related and neutral words. *Behavior Research Method, Instruments, & Computers*, *36*(3), 444–458.

Davies, M. (2004-). *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Available online at http://corpus.byu.edu/bnc/.

Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, *32*(4), 551–559.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation*, *14*(4), 375–399.

Garnham, A., Doehren, S., & Gygax, P. (2015). True gender ratios and stereotype rating norms. *Frontiers in Psychology: Language Sciences*, *6*(1023), 1–7.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures of 1,944 words. *Behavior Research Methods & Instrumentation*, *12*(4), 395–427.

Harman, H. H. (1976). *Modern factor analysis*. Chicago: University of Chicago Press.

Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. *Behavior Research Methods*, *40*(4), 1065–1074.

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika*, *31*(3), 313–313.

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, *13*(7/8), 789–845.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131*(5), 684–712.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1312–1318.

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*(7/8), 846–863.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151.

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481.

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011) The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34.

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, *58*, 239–261.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990.

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., . . . Sczesny, S. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods*, *46*(3), 841–871.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, *50A*(3), 528–559.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

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.

Richardson, J. T. E. (1976). Imageability and concreteness. *Bulletin of the Psychonomic Society, 7*(5), 429–431.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178.

Schock, J., Cortese, M. J., & Khanna, M. M. (2012a). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods, 44*(2), 971–977.

Schock, J., Cortese, M. J., Khanna, M. M., & Toppi, S. (2012b). Age of acquisition estimates for 3,000 disyllabic words. *Behavior Research Methods, 44*(4), 971–977.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27*(5), 499–520.

Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology, 80*(1), 95–104.

Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2012). Emotion words affect eye fixations during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 783–792.

Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2014). Emotion words and categories: Evidence from lexical decision. *Cognitive Processing, 15*(2), 209-215.

Sereno, S. C., & O'Donnell, P. J. (2009). Participant and word gender in age of acquisition effects: The role of gender socialization. *Sex Roles, 61*(7-8), 510–518.

Sereno, S. C., O'Donnell, P. J., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate-bias effect. *Journal of Experimental Psychology: Human Perception and Performance, 32*(2), 335–350.

Sereno, S. C., O'Donnell, P. J., & Sereno, M. E. (2009). Size matters: Bigger is faster. *Quarterly Journal of Experimental Psychology, 62*(6), 1115–1122.

Sereno, S. C., Scott, G. G., Yao, B., Thaden, E., & O'Donnell, P. J. (2015). Emotion word processing: Does mood make a difference? *Frontiers in Psychology: Language Sciences, 6*(1191), 1–13.

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., & Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: The body-object interaction effect. *Cognition, 106*(1), 433–443.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38*(4), 598–605.

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

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.

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 143–154.

Wurm, L. H. (2007). Danger and usefulness: An alternative framework for understanding rapid evaluation effects in perception*? Psychonomic Bulletin & Review*, *14*(6), 1218–1225.

Yao, B., Keitel, A., Bruce, G., Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2017). Differential emotional processing in concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, in press*.

Yao, B., Vasiljevic, M., Weick, M., Sereno, M. E., O'Donnell, P. J., & Sereno, S. C. (2013). Semantic size of abstract concepts: It gets emotional when you can't see it. *PLoS ONE*, *8*(9), 1–10.

Zevin, J. D., & Seidenberg, M. S. (2004). Age of acquisition effects in reading aloud: Test of cumulative frequency and frequency trajectory. *Memory & Cognition*, *32*(1), 31–38.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in reading and other tasks. *Journal of Memory & Language*, *47*(1), 1–29.

Appendix A: Instructions for Rating Tasks

We have included all of the instructions for the norming task. The general instructions are presented in **Table A1**. The specific instructions for each of the 9 separate rating tasks, including how each variable was defined, are presented in **Table A2**. Finally, the specifications of each rating scale in terms of the number of scale points and scale labels are presented in **Table A3**.

In this experiment, you will be rating a set of (101/150) words on 9 different scales – familiarity, concreteness, arousal, valence, dominance, imagaeability, size, gender, and age of acquisition. These different scales assess different aspects of word meanings.

You will rate the entire word set on one scale, then rate them all again on the next scale, and so on. You will be given instructions about what each scale represents before you begin each scale.

Sometimes words have more than one meaning – for example, the word "nail" has one meaning related to fingers and one related to hammers In such cases, we will display the word in one of its meanings – "nail (finger)" or "nail" (hammer)" – or just by itself as "nail". Please rate these words according to your first impression.

It is also possible that you may be presented with a word that you don't know. If this happens, there is a button "Unfamiliar word" located below the rating scale that you can click on to proceed to the next trial.

Finally, it may sometimes be difficult to rate a word on a given scale. For example, the word "desk" might be difficult to categorise as either being a masculine or feminine thing. Likewise, the word "amusing" might be difficult to categorise as something that is big or small. When you are faced with such difficult decisions, please respond as best as you can without thinking too deeply – go with your intuitions. Note: Click on the red information button in the upper right corner at any time to refer to the rating instructions again.

Table A1. General Instructions. One version of the ratings included 101 words; the other included 150.

Arousal

Arousal is a measure of excitement versus calmness. A word is AROUSING if it makes you feel stimulated, excited, frenzied, jittery, or wide-awake. A word is UNAROUSING if it makes you feel relaxed, calm, sluggish, dull, or sleepy.

Please indicate how arousing you think each word is on a scale of VERY UNAROUSING to VERY AROUSING, with the midpoint representing moderate arousal.

Valence

Valence is a measure of value or worth. A word is POSITIVE if it represents something considered good, whereas a word is NEGATIVE if it represents something considered bad.

Please indicate the valence of each word on a scale of VERY NEGATIVE to VERY POSITIVE, with the midpoint representing NEUTRAL.

Dominance

Dominance is a measure of the degree of control you feel. A word can make you feel DOMINANT, influential, in control, important, or autonomous. Conversely, a word can make you feel CONTROLLED, influenced, cared-for, submissive, or guided.

Please indicate how each word makes you feel on a scale of YOU ARE VERY CONTROLLED to YOU ARE VERY DOMINANT, with the midpoint being neither controlled nor dominant.

Concreteness

Concreteness is a measure of how concrete or abstract something is. A word is CONCRETE if it represents something that exists in a definite physical form in the real world. In contrast, a word is ABSTRACT if it represents more of a concept or idea.

Please indicate how concrete you think each word is on a scale of VERY ABSTRACT to VERY CONCRETE, with the midpoint being neither especially abstract nor concrete.

Imageability

Imageability is a measure of how easy or difficult something is to imagine. A word is IMAGEABLE if it represents something that is very easy to imagine or picture. In contrast, a word is UNIMAGEABLE if it represents something that is very difficult to imagine or picture.

Please indicate how imageable you think each word is on a scale of VERY UNIMAGEABLE to VERY IMAGEABLE, with the midpoint being moderately imageable.

Familiarity

Familiarity is a measure of how familiar something is. A word is very FAMILIAR if you see/hear it often and it is easily recognisable. In contrast, a word is very UNFAMILIAR if you rarely see/hear it and it is relatively unrecognisable.

Please indicate how familiar you think each word is on a scale of VERY UNFAMILIAR to VERY FAMILIAR, with the midpoint representing moderate familiarity.

Age of Acquisition

A word's age of acquisition is the age at which that word was initially learned. Please estimate when in your life you think you first acquired or learned each word. That is, try to remember how old you were when you learned each word either in its spoken or written form (whichever came first).

The scale is defined as a series of consecutive 2-year periods from the ages of 0 to 12 years, and a final period encompassing 13 years and older.

Semantic Size

Size is a measure of something's dimensions, magnitude, or extent. A word represents something BIG if it refers to things or concepts that are large. A word represents something SMALL if it refers to things or concepts that are little.

Please indicate the semantic size of each word on a scale of VERY SMALL to VERY BIG, with the midpoint being neither small nor big.

Gender Association

A word's gender is how strongly its meaning is associated with male or female behaviour. A word can be considered MASCULINE if it is linked to male behaviour. Alternatively, a word can be considered FEMININE if it linked to female behaviour.

Please indicate the gender associated with each word is on a scale of VERY FEMININE to VERY MASCULINE, with the midpoint being neuter (neither feminine nor masculine).

Table A2. The Nine Rating Scales. For each scale, the definition of each variable provided in the instructions are indicated. All scale definitions ended with the closing statement, "In your ratings, please note the following: there are no correct or incorrect answers – each person will have differing opinions; trust your intuitions – do not spend a long time thinking about each word as your initial reaction is often the most accurate; and try to use the whole range of the scale.

Rating Scale	Number of Scale Points	Scale Labels (left to right)
Arousal	9	Very Unarousing, Very Arousing
Valence	9	Very Negative, Neutral, Very Positive
Dominance	9	You are very Controlled, You are very Dominant
Concreteness	7	Very Abstract, Very Concrete
Imageability	7	Very Unimageable, Very Imageable
Age of Acquisition	7	0-2, 3-4, 5-6, 7-8, 9-10, 11-12, 13+
Semantic Size	7	Very Small, Very Big
Gender Association	7	Very Feminine, Very Masculine

Table A3. Specifications of Rating Scales. For all scales, endpoints were labelled as indicated. Additional labels were as follows: Valence included a midpoint label; Age of Acquisition included labels for each scale point.