CrossMark

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

Graham G. Scott [1] · Anne Keitel [2] · Marc Becirspahic [3] · Bo Yao [4] · Sara C. Sereno [2,3]

Abstract

The Glasgow Norms are a set of normative ratings for 5,553 English words on nine 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 nine dimensions (33 participants/word, on average). Third, two novel dimensions—semantic size and gender association—are included. Finally, the corpus contains a set of 379 ambiguous words that are presented either alone (e.g., *toast*) or with information that selects an alternative sense (e.g., *toast (bread)*, *toast (speech)*). The 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 18 different sets of current psycholinguistic norms. The dimension of size was tested with megastudy data, confirming findings from past studies that have explicitly examined this variable. Alternative senses of ambiguous words (i.e., disambiguated forms), when discordant on a given dimension, seemingly led to appropriately distinct ratings. Informal comparisons between the ratings of ambiguous words and of 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

✉ Sara C. Sereno
sara.sereno@glasgow.ac.uk

Graham G. Scott
graham.scott@uws.ac.uk

[1] School of Media, Culture and Society–Psychology, University of the West of Scotland, Paisley PA1 2BE, UK

[2] Institute of Neuroscience and Psychology, University of Glasgow, 62 Hillhead Street, Glasgow G12 8QB, UK

[3] School of Psychology, University of Glasgow, 62 Hillhead Street, Glasgow G12 8QB, UK

[4] Division of Neuroscience and Experimental Psychology, University of Manchester, Manchester, UK

The Glasgow Norms provide a set of normative ratings for 5,553 English words on nine 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 (nine) was examined. Other norms typically assess only one to three dimensions. Second, the same participant provided ratings across all nine dimensions for any given word. Currently, researchers

interested in investigating more than a few lexical dimensions need to access different sets of norms that are tested on different populations of participants. Additionally, as different norms test nonoverlapping 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.

It is beyond the scope of the present investigation to catalog norms comprising fewer than 500 lexical items. Nevertheless, over the past several decades, such norms have proved valuable and have been used extensively (e.g., Morrison, Chappell, & Ellis's, 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 ensure 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, 2018). 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

Table 1 Inventory of English word norms

Dimension(s)	N Items	Participants/Item	Source
AROU, VAL, DOM	1,034	50 on average	Bradley and Lang (1999)
AROU, VAL, DOM	13,915	18–30 for most items	Warriner, Kuperman, and Brysbaert (2013)
CNC	37,058	at least 25	Brysbaert, Warriner, and Kuperman (2014)
IMAG (monosyllabic words)	3,000	31	Cortese and Fugett (2004)
IMAG (disyllabic words)	3,000	35	Schock, Cortese, and Khanna (2012)
CNC, IMAG	925	28 (CNC), 30 (IMAG)	Paivio, Yuille, and Madigan (1968)[a]
CNC, IMAG	1,080	50	Friendly, Franklin, Hoffman, and Rubin (1982)
CNC, IMAG, FAM	2,854	54–65	Toglia and Battig (1978)[b]
CNC, IMAG, FAM, AOA	1,944	35–37	Gilhooly and Logie (1980a)[c]
CNC, IMAG, FAM, AOA (homograph meanings)	905	35–37	Gilhooly and Logie (1980b)
IMAG, FAM, AOA	1,526	20	Stadthagen-Gonzalez and Davis (2006)
IMAG, FAM, AOA	629	21 (IMAG), 14 (FAM), 15 (AOA)	Juhasz, Lai, and Woodcock (2015)[d]
IMAG, FAM	2,311	16 (IMAG), 47–49 (FAM)	Clark and Paivio (2004)[e]
IMAG, AOA	2,694	78 (IMAG), 45 (AOA)	Bird, Franklin, and Howard (2001)
IMAG, AOA	2,204	277	Davies, Izura, Socas, and Dominguez (2016)
AOA (monosyllabic words)	3,000	32	Cortese and Khanna (2008)
AOA (disyllabic words)	3,000	32	Schock, Cortese, Khanna, and Toppi (2012)
AOA	30,124	18–22 for most items	Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012)
AOA (homograph meanings)	3,460	30	Khanna and Cortese (2011)
GEND	600	356	Crawford, Leynes, Mayhorn, and Bink (2004)

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. also measured meaningfulness. [b] Toglia and Battig also measured meaningfulness, pleasantness, categorizability, and number of attributes or features. [c] Gilhooly and Logie (1980a) also measured ambiguity. [d] Juhasz et al. also measured meaning dominance, semantic transparency, and sensory experience. [e] Clark and Paivio also measured an additional 13 dimensions, but only on the original set of 925 items from Paivio et al.

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 nine 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, O'Donnell, Leuthold, & Sereno, 2009; Scott et al., 2012; Scott, O'Donnell, & Sereno, 2014; Sereno et al., 2015; Yao et al., 2018).

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, 2018). 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 as compared to less imageable words (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Cortese & Schock, 2013; Yao et al., 2018). 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). More recently, however, Brysbaert (2017) demonstrated that the best predictor of objective AOA is rated AOA. 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, 2015; 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 whereas 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 present 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 a few existing norms have explicitly included ambiguous words. Clark and Paivio (2004), in their extension of the Paivio et al. (1968; *N* = 925) norms, included number of meanings as an additional measure, but did not collect 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). Gilhooly and Logie (1980a) 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. Gilhooly and Logie (1980b) collected ratings on a set of 387 ambiguous words having a total of 905 separate meanings on the scales of CNC, IMAG, FAM, and AOA. Khanna and Cortese (2011) collected AOA ratings of 1,208 ambiguous words having a total of 3,460 senses. Although 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). Ratings of homographs from previous norming studies that have *not* explicitly disambiguated their disparate senses 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 corpus of 5,553 words to participants in lists of either 101 or 150 words. For each list, participants rated words separately on all nine 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, their ages ranged from 16 to 73 years, and there were slightly more than twice the number of females than males who took part. The participants were native English speakers from the University of Glasgow community and were recruited opportunistically via an experiment

Table 2 Age and gender profile of participants

	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)	1,368 (100)	22.8

The number of participants, age, and 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 two lists, 3% did three lists, 8% did four lists, and 3% did eight lists

advertisement link on the home page of the Psychology department at the University of Glasgow. They were either paid at a rate of £6/h or given course credit for their participation. The study conformed to British Psychological Society ethical guidelines and protocols.

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 two 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, a total of 871 items in the corpus were presented with disambiguation.

Procedure

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 (eight possible lists of the 808 word set) or 150 words (32 lists of the 4,800 word set). Lists of 101 or 150 words were generated by taking every 8th or 32nd 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, the specific instructions for each of the nine different rating tasks, and the rating scale labels are presented in the supplementary materials to this article in Tables S1, S2, and S3, respectively. The same participant provided ratings across all nine 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 min 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 600 ms or if participants reported not knowing a word. For RT, examination of the trial-by-trial data revealed the presence of infrequent episodes of rapid responding by some participants, typically repeating a given rating value. In such cases, the RTs tended to be 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., 2018, used 270 words with 127 participants). A conservative lower cutoff of 600 ms was therefore implemented in the present study. No upper cutoff was imposed. While participants were encouraged to rate each word according to their initial interpretation of its meaning, there was no emphasis on speed of response. That is, participants were not instructed, for example, to respond to each item as

quickly as possible. An identical procedure of not implementing an upper cutoff has been employed by several sets of norms (e.g., Clark & Paivio, 2004; Cortese & Khanna, 2008; Khanna & Cortese, 2011; Schock, Cortese, & Khanna, 2012). Of the total number of responses recorded (N = 1,732,607), the RT distribution was as follows: 3.07% were shorter than 600 ms, 77.68% were 600–3,000 ms (with 36.03% 1,500–2,000 ms), 12.98% were 3,000–5,000 ms, and 6.27% were longer than 5,000 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 the instructions in the Table S2). This option accounted for 0.33% of all responses. On average, 33.29 responses were provided per word (SD = 3.76). A detailed profile of the numbers of responses across all nine rating scales is presented in Table 3.

The descriptive statistics for the nine 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 nine dimensions, the mean rating (M), standard deviation (SD), and number of ratings (N) for each word. Ratings for the 55 words that were included in both lists were highly correlated for all scales, ranging from r = .88 for DOM to r = .97 for VAL.

Relations between the nine dimensions of the Glasgow Norms

To provide an initial overview of the relations between all nine of the Glasgow Norms scales, we performed Spearman

Table 3 Profile of the numbers of responses across dimensions

N		AROU	VAL	DOM	CNC	IMAG	FAM	AOA	SIZE	GEND
5,553	M	33.31	33.54	33.24	33.34	33.30	32.36	33.94	33.30	33.25
	SD	3.72	3.73	3.73	3.80	3.74	3.60	3.69	3.79	3.85
	range	13–70	15–71	14–69	11–70	14–70	22–67	17–70	13–70	15–69
55	M	65.73	65.96	65.31	65.33	64.84	61.18	66.75	65.96	64.64
	SD	1.57	1.55	2.01	2.16	2.43	3.20	1.51	1.96	2.44
	range	61–70	63–71	61–69	60–70	59–70	53–67	64–70	60–70	59–69
5,498	M	32.99	33.22	32.92	33.02	32.99	32.07	33.61	32.98	32.94
	SD	1.83	1.85	1.90	2.05	2.01	2.15	1.68	1.92	2.22
	range	13–36	15–36	14–36	11–36	14–36	22–35	17–36	13–36	15–36

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

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

Table 4 Descriptive statistics of the nine dimensions of the Glasgow Norms

Dimension	Scale range	M	SD
AROU	1–9	4.63	1.10
VAL	1–9	5.10	1.55
DOM	1–9	5.07	0.91
CNC	1–7	4.64	1.42
IMAG	1–7	4.79	1.35
FAM	1–7	5.26	0.93
AOA	1–7	4.13	1.24
SIZE	1–7	4.09	1.02
GEND	1–7	4.12	0.91

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

correlations, and these are presented in Table 5. Since Spearman correlations are rank-based, this method takes into account both linear and nonlinear relations between the 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 rs > .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 the word is).

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 Tables S4 and S5, respectively.

We will highlight the effects of SIZE and GEND, as these two dimensions are relatively new and less well understood. Figure 1 shows the quadratic fits for all combinations with either SIZE or GEND that account for more than 18% of variance (see Table S5): 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, the linear fits accounted for comparable (but numerically slightly less) amounts of variance (see Table S5) 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 explained better by a quadratic ($R^2 = .19$) than by a linear ($R^2 < .01$) fit (the more extremely valenced—negative *or* positive—a word is, the semantically bigger it is).

Factor analysis of dimensions

To summarize and interpret the correlation results and explore the relative alignment of the newer scales of SIZE and GEND, we performed a factor analysis. The data for all nine 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.,

Table 5 Correlations between dimensions of the Glasgow Norms

	AROU	VAL	DOM	CNC	IMAG	FAM	AOA	SIZE	GEND
AROU	–								
VAL	.35	–							
DOM	.34	**.68**	–						
CNC	− .25	(.05)	(.05)	–					
IMAG	− .10	.10	.08	**.91**	–				
FAM	.18	.30	.23	.10	.22	–			
AOA	(.00)	− .19	− .14	− .38	− .49	− **.67**	–		
SIZE	**.51**	.12	.09	− .41	− .33	(.05)	.22	–	
GEND	− .11	− .42	− .09	.15	.07	− .21	.15	.15	–

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 rs >|±.50| (Cohen, 1988)

AROU arousal; *VAL* valence; *DOM* dominance; *CNC* concreteness; *IMAG* imageability; *FAM* familiarity; *AOA* age of acquisition; *SIZE* semantic size; *GEND* gender association
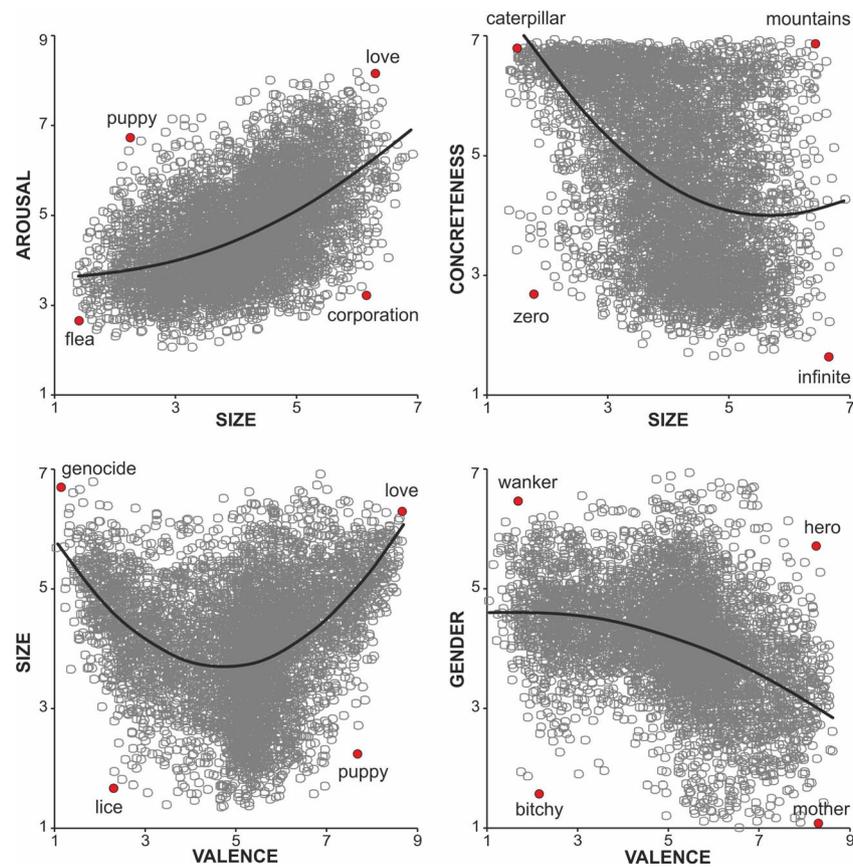
Fig. 1 Quadratic fits with semantic size or gender association that explain more than 18% of variance. R^2 values, F values, and significance for the linear and quadratic fits for all combinations of dimensions can be found in Tables S4 and S5 of the supplementary materials

varimax, quartimax, or equamax) yielded comparable results. Factors with eigenvalues greater than 1 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.

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

Table 6 Factor loadings for all dimensions of the Glasgow Norms

	Factor 1: Visualization	Factor 2: Emotion	Factor 3: Salience	Factor 4: Exposure
CNC	**.938**	.098	− .083	− .053
IMAG	**.888**	.116	− .013	− .220
VAL	.025	**.919**	− .163	− .016
DOM	.170	**.914**	.101	.103
SIZE	− .344	.086	**.753**	− .074
GEND	.446	− .283	**.683**	.238
AROU	− .242	.443	**.521**	− .211
FAM	− .026	− .036	.081	**− .935**
AOA	− .286	.068	.087	**.843**
%Var	29.77	25.87	13.18	12.97
%CumVar	29.77	55.64	68.82	81.79

Reported are loadings of an oblique rotation matrix (direct oblimin) on four factors. Loadings >|±.50| are highlighted in bold. Explained common variance (%Var) is given for individual factors, as well as the cumulative variance (%CumVar)

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

Correlations with other psycholinguistic norms

To confirm the validity of our ratings, we correlated the dimensions of the Glasgow Norms with 18 of the 20 different sets of English norms listed in Table 1. The norms that were excluded were those that were not easily accessible. Between one and ten norms were available for all dimensions except SIZE. Because 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 greater than .5, indicating a large effect (Cohen, 1988) and, therefore, sufficient validity. We do not have an explanation for why two of the 36 correlations reported in Table 7, although significant, had coefficients less than .5. The highest and most consistent correlations were achieved by VAL, CNC, AOA, and GEND (with nearly all $rs > .9$).

SIZE and GEND

One unique strength of our norms is the inclusion of the SIZE and GEND variables, which allows us to test these effects on a much larger set of words than had previously been possible. To assess the effects of SIZE, we attempted to replicate the semantic size effect reported in Sereno et al. (2009) and Yao et al. (2013). We combined our ratings and the lexical decision RT data from the English Lexicon Project (ELP; Balota et al., 2007). To avoid multiple entries of the same word, we removed items corresponding to the alternative meanings of homographs. A total of 4,568 words were entered into the analysis. We examined the effects of SIZE on RTs, with all the other variables as covariates (word frequency, word length, CNC, IMAG, AROU, VAL, FAM, AOA, DOM, and GEND). To address collinearity between the covariates (e.g., CNC × IMAG, AROU × VAL), we reduced the dimensions of covariates via a PCA using a varimax rotation. We extracted seven principal components, accounting for 93.8% of the variance, and their factor loadings are shown in Table 8. We fit a

Table 7 Correlations between the Glasgow Norms and other English word norms

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**	**.88**	**.82**	**.92**	
8	905	136				**.81**	**.84**	**.72**	**.86**	
9	1,526	1,370					**.92**	**.81**	**.94**	
10	629	61					**.94**	**.64**	**.90**	
11	2,311	1,390					.42	**.82**		
12	2,694	994					**.80**		**.86**	
13	2,204	722					**.91**		**.95**	
14	3,000	1,363							**.91**	
15	3,000	1,308							**.90**	
16	30,124	4,283							**.89**	
17	3,460	525							.20	
18	600	336								**.96**

Pearson coefficients for 18 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 ($N_{overlap}$) that were used for the correlations are indicated. Norms 8 and 17 examine different senses of ambiguous words. All correlations were highly significant ($ps < .0001$, Bonferroni corrected). Correlations with a large effect ($r > .5$, see Cohen, 1988) are printed in bold. References for the 18 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, Cortese, and Khanna (2012); 6 = Paivio et al. (1968); 7 = Gilhooly and Logie (1980a); 8 = Gilhooly and Logie (1980b); 9 = Stadthagen-Gonzalez and Davis (2006); 10 = Juhasz et al. (2015); 11 = Clark and Paivio (2004); 12 = Bird et al. (2001); 13 = Davies et al. (2016); 14 = Cortese and Khanna (2008); 15 = Schock, Cortese, et al. (2012); 16 = Kuperman et al. (2012); 17 = Khanna and Cortese (2011); and 18 = Crawford et al. (2004)

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

Table 8 SIZE Effects: Factor Loadings for the Extracted Principal Components

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Frequency			**0.98**				0.16
Length	–0.10	0.10				**0.97**	–0.13
CNC	**0.95**	–0.22					
IMAG	**0.96**						0.19
AROU		**0.81**		0.45			
|VAL|	–0.21	**0.88**		–0.22			
FAM		0.11	0.10	0.10			**0.94**
AOA	–0.35		–0.12			0.31	**–0.79**
DOM				**0.96**			0.11
GEND					**0.99**		–0.10

Note: Factor loadings ≥ |±.10| are shown and factor loadings > |±.50| are highlighted in bold. Frequency was measured in written occurrences per million as per the British National Corpus (2007; Davies, 2004) and Length in number of letters

CNC concreteness; *IMAG* imageability; *AROU* arousal; *|VAL|* absolute valence (i.e., the 1-9 scale was collapsed around its midpoint so that higher values would reflect more extreme valence, regardless of whether they were rated as positive or negative); *FAM* familiarity; *AOA* age of acquisition (negative loading); *DOM* dominance; *GEND* gender association; *SIZE* semantic size

linear model of RT with SIZE and the extracted principal components as predictors, and the results are shown in Table 9. After we had controlled for a wide range of lexical and semantic variables, SIZE negatively predicted word recognition times—that is, semantically bigger words were recognized significantly faster than semantically smaller words, replicating the findings of Sereno et al. (2009) and Yao et al. (2013).

The effects of word GEND, however, are more difficult to test. RTs in megastudies are aggregated across participant gender. Moreover, the relative proportion of male versus female participants in megastudies is typically not specified. Sereno

Table 9 SIZE Effects: Corresponding Multiple Linear Regression Results

Predictors	b	SE	t	p
SIZE	–10.51	1.21	–8.68	<.001
PC1 (CNC, IMAG)	–5.99	0.96	–6.21	<.001
PC2 (AROU, |VAL|)	4.35	1.08	4.04	<.001
PC3 (Frequency)	–7.28	0.89	–8.17	<.001
PC4 (DOM)	–3.99	0.91	–4.39	<.001
PC5 (GEND)	3.65	0.93	3.91	<.001
PC6 (Length)	40.44	0.94	43.01	<.001
PC7 (FAM, -AOA)	–45.44	0.90	–50.3	<.001

Note: Frequency was measured in written occurrences per million as per the British National Corpus (2007; Davies, 2004) and Length in number of letters

CNC concreteness; *IMAG* imageability; *AROU* arousal; *|VAL|* absolute valence (i.e., the 1-9 scale was collapsed around its midpoint so that higher values would reflect more extreme valence, regardless of whether they were rated as positive or negative); *FAM* familiarity; *AOA* age of acquisition (negative loading); *DOM* dominance; *GEND* gender association; *SIZE* semantic size

and O'Donnell (2009) examined the effects of word GEND and AOA on lexical decision times across male and female participants. All participants demonstrated AOA effects. However, females took longer to respond to male-oriented words, particularly late-AOA ones, whereas males, in contrast, demonstrated no effect of word GEND. Confirming this pattern of results with megastudy data (in combination with our norms) would entail finding a GEND × AOA interaction across participant gender. Without access to data that separately present responses from male and female participants, we were unable to directly test the effects of word GEND.

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 they were relevant to the dimension in question. Figure 2A illustrates the ratings that alternative senses of several ambiguous words received across the nine 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. Although ambiguous words typically have a dominant 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 the alternative senses of ambiguous words in our corpus, however, we are at present unable to characterize these data. Anecdotally, factors such as number of meanings, their relative strengths, and the rating scale itself seem to play distinct roles.

Two sets of prior norms contained a fairly large number of ambiguous words with clearly defined alternative meanings that were rated on a subset of dimensions (Gilhooly & Logie, 1980b; Khanna & Cortese, 2011). We correlated the shared dimensions of the disambiguated words from these samples with the disambiguated items from our norms whose meanings matched. These correlations were significant and are included in Table 7 (Norms 8 and 17).

Discussion

The Glasgow Norms examined nine semantic dimensions of words in a corpus of 5,553 words, with an average of 33 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 investigated extensively, whereas the other two dimensions (SIZE and GEND) are relatively novel and have not been examined 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 nine 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
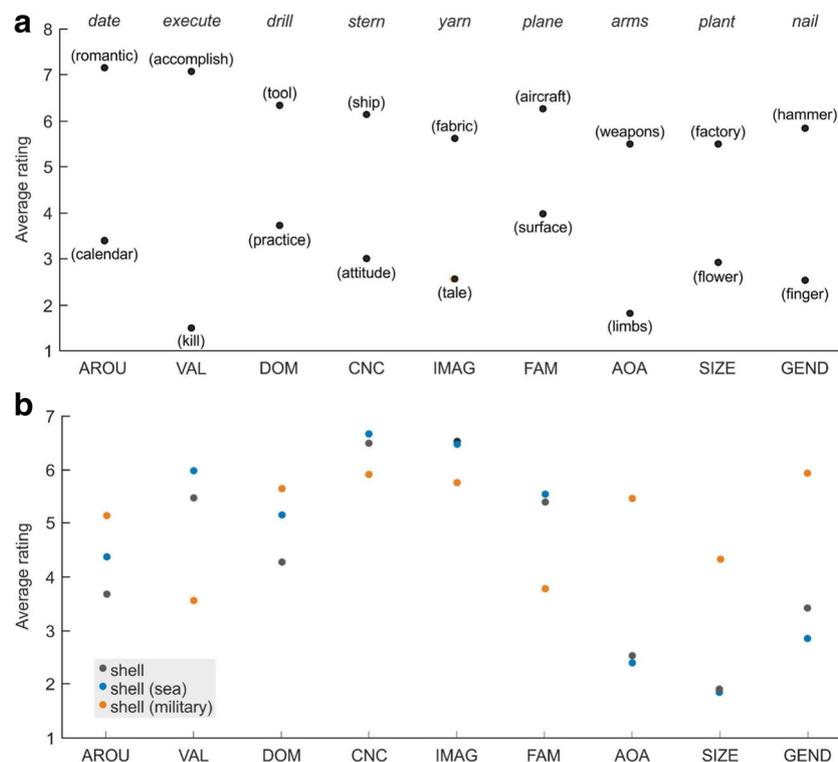
Fig. 2 (**a**) Ratings of alternative senses of ambiguous words are indicated across dimensions. For each dimension of the Glasgow Norms, an example ambiguous word is listed across the top. The ratings corresponding to the alternative senses (defined in parentheses) are indicated beneath each ambiguous word. (**b**) Ratings of the ambiguous word *shell* and its alternative senses ("sea" and "military") are indicated across all 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

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, 1980a; 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 S4 and S5 in the supplementary materials). We focused on effects related to the relatively new dimensions of SIZE and GEND (see Fig. 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. It should be noted that although all participants were native English speakers, we did not record whether they were fluent in any other languages. Knowledge of a grammatically gendered language may have potentially impacted GEND ratings of English words in some participants (Boroditsky, Schmidt, & Phillips, 2003).

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 18 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. The correlations with the prior norms were highly significant across the eight dimensions of AROU, VAL, DOM, CNC, IMAG, FAM, AOA, and GEND.

The newer dimensions of SIZE and GEND were tested against megastudy data from the ELP (Balota et al., 2007). Analyses revealed that both SIZE and GEND effects were obtained, confirming the findings from prior studies that have specifically examined these factors.

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 Fig. 2A). Ambiguous words presented in isolation were sometimes rated according to their highly dominant sense across the different dimensions (see Fig. 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.

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, 578–602. https://doi.org/10.3758/BF03200738

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, 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, *29*, 639–647. https://doi.org/10.3758/BF03200465

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459. https://doi.org/10.3758/BF03193014

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*, 73–79. https://doi.org/10.3758/BF03195349

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and cognition* (pp. 61–79). Cambridge, MA: MIT Press.

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: University of Florida, NIMH Center for Research in Psychophysiology.

British National Corpus. (2007). Version 3 (BNC XML ed.). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Retrieved from www.natcorp.ox.ac.uk

Brysbaert, M. (2017). Age of acquisition ratings score better on criterion validity than frequency trajectory or ratings "corrected" for frequency. *Quarterly Journal of Experimental Psychology*, *70*, 1129–1139. https://doi.org/10.1080/17470218.2016.1172097

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

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 371–383. https://doi.org/10.3758/BF03195584

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

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*, 1084–1096. https://doi.org/10.1037/0278-7393.16.6.1084

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36*, 384–387. https://doi.org/10.3758/BF03195585

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*, 1072–1082. https://doi.org/10.1080/17470210701315467

Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods*, *40*, 791–794. https://doi.org/10.3758/BRM.40.3.791

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, *66*, 946–972. https://doi.org/10.1080/17470218.2012.722660

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 Methods, Instruments, & Computers*, *36*, 444–458. https://doi.org/10.3758/BF03195592

Davies, M. (2004). BYU-BNC (based on the British National Corpus). Retrieved from http://corpus.byu.edu/bnc/

Davies, S. K., Izura, C., Socas, R., & Dominguez, A. (2016). Age of acquisition and imageability norms for base and morphologically complex words in English and in Spanish. *Behavior Research Methods*, *48*, 349–365. https://doi.org/10.3758/s13428-015-0579-y

Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, *32*, 551–559. https://doi.org/10.3758/BF03195846

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*, 375–399. https://doi.org/10.3758/BF03203275

Garnham, A., Doehren, S., & Gygax, P. (2015). True gender ratios and stereotype rating norms. *Frontiers in Psychology: Language Sciences*, *6*, 1023:1–7. https://doi.org/10.3389/fpsyg.2015.01023

Gilhooly, K. J., & Logie, R. H. (1980a). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures of 1,944 words. *Behavior Research Methods & Instrumentation*, *12*, 395–427. https://doi.org/10.3758/BF03201693

Gilhooly, K. J., & Logie, R. H. (1980b). Meaning-dependent ratings for imagery, age of acquisition, familiarity, and concreteness for 387 ambiguous words. *Behavior Research Methods & Instrumentation*, *12*, 428–450. https://doi.org/10.3758/BF03201694

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

Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. *Behavior Research Methods*, *40*, 1065–1074. https://doi.org/10.3758/BRM.40.4.1065

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika*, *31*, 313–313. https://doi.org/10.1007/BF02289465

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, *13*, 789–845. https://doi.org/10.1080/13506280544000066

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131*, 684–712. https://doi.org/10.1037/0033-2909.131.5.684

Juhasz, B. J., Lai, Y.-H., & Woodcock, M. L. (2015). A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods*, *47*, 1004–1019. https://doi.org/10.3758/s13428-014-0523-6

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*, 1312–1318. https://doi.org/10.1037/0278-7393.29.6.1312

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*, 846–863. https://doi.org/10.1080/13506280544000075

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151. https://doi.org/10.1177/001316446002000116

Khanna, M. M., & Cortese, M. J. (2011). Age of acquisition estimates for 1,208 ambiguous and polysemous words. *Behavior Research Methods*, *43*, 89–96. https://doi.org/10.3758/s13428-010-0027-y

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*, 14–34. https://doi.org/10.1037/a0021446

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*, 473–481. https://doi.org/10.1016/j.cognition.2009.06.007

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. https://doi.org/10.1016/j.jml.2007.09.003

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990. https://doi.org/10.3758/s13428-012-0210-4

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*, 841–871. https://doi.org/10.3758/s13428-013-0409-z

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*, 528–559. https://doi.org/10.1080/027249897392017

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: 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*, *76*(1, Pt. 2), 1–25. https://doi.org/10.1037/h0025327

Richardson, J. T. E. (1976). Imageability and concreteness. *Bulletin of the Psychonomic Society*, *7*, 429–431. https://doi.org/10.3758/BF03337237

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161-1178. https://doi.org/10.1037/h0077714

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, *44*, 374–379. https://doi.org/10.3758/s13428-011-0162-0

Schock, J., Cortese, M. J., Khanna, M. M., & Toppi, S. (2012). Age of acquisition estimates for 3,000 disyllabic words. *Behavior Research Methods*, *44*, 971–977. https://doi.org/10.3758/s13428-012-0209-x

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*, 499–520. https://doi.org/10.1016/0749-596X(88)90022-8

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*, 95–104. https://doi.org/10.1016/j.biopsycho.2008.03.010

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*, 783–792. https://doi.org/10.1037/a0027209

Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2014). Emotion words and categories: Evidence from lexical decision. *Cognitive Processing*, *15*, 209-215. https://doi.org/10.1007/s10339-013-0589-6

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*, 510–518. https://doi.org/10.1007/s11199-009-9649-x

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*, 335–350. https://doi.org/10.1037/0096-1523.32.2.335

Sereno, S. C., O'Donnell, P. J., & Sereno, M. E. (2009). Size matters: Bigger is faster. *Quarterly Journal of Experimental Psychology*, *62*, 1115–1122. https://doi.org/10.1080/17470210802618900

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. https://doi.org/10.3389/fpsyg.2015.01191

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*, 433–443. https://doi.org/10.1016/j.cognition.2006.12.011

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*, 598–605. https://doi.org/10.3758/BF03193891

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

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

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*, 143–154. https://doi.org/10.1016/S0022-5371(78)90110-X

Wurm, L. H. (2007). Danger and usefulness: An alternative framework for understanding rapid evaluation effects in perception? *Psychonomic Bulletin & Review*, *14*, 1218–1225. https://doi.org/10.3758/BF03193116

Yao, B., Keitel, A., Bruce, G., Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2018). Differential emotional processing in concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1064–1074. https://doi.org/10.1037/xlm0000464

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*, e75000:1–10. https://doi.org/10.1371/journal.pone.0075000

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in reading and other tasks. *Journal of Memory and Language*, *47*, 1–29. https://doi.org/10.1006/jmla.2001.2834

Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: Test of cumulative frequency and frequency trajectory. *Memory & Cognition*, *32*, 31–38. https://doi.org/10.3758/BF03195818