English semantic word-pair norms and a searchable Web portal for experimental stimulus creation

**Erin M. Buchanan · Jessica L. Holmes ·
Marilee L. Teasley · Keith A. Hutchison**

Abstract As researchers explore the complexity of memory and language hierarchies, the need to expand normed stimulus databases is growing. Therefore, we present 1,808 words, paired with their features and concept–concept information, that were collected using previously established norming methods (McRae, Cree, Seidenberg, & McNorgan Behavior Research Methods 37:547–559, 2005). This database supplements existing stimuli and complements the Semantic Priming Project (Hutchison, Balota, Cortese, Neely, Niemeyer, Bengson, & Cohen-Shikora 2010). The data set includes many types of words (including nouns, verbs, adjectives, etc.), expanding on previous collections of nouns and verbs (Vinson & Vigliocco Journal of Neurolinguistics 15:317–351, 2008). We describe the relation between our and other semantic norms, as well as giving a short review of word-pair norms. The stimuli are provided in conjunction with a searchable Web portal that allows researchers to create a set of experimental stimuli without prior programming knowledge. When researchers use this new database in tandem with previous norming efforts, precise stimuli sets can be created for future research endeavors.

Keywords Database · Stimuli · Semantics · Word norms

Psychologists, linguists, and researchers in modern languages require both traditional knowledge about what words mean and how those words are used when they are paired in context. For instance, we know that rocks can roll, but when rock and roll are paired together, mossy stones no longer come to mind. Up-to-date online access to word meanings will empower linguistic research, especially given language's ability to mold and change with culture. Several collections of word meanings and usages already exist online (Fellbaum, 1998; Nelson, McEvoy, & Schreiber, 2004), but several impediments occur when trying to use these stimuli. First, a researcher may want to use existing databases to obtain psycholinguistic measures, but will likely find very little overlap between the concepts present in all of these databases. Second, this information is spread across different journal and researcher websites, which makes material combination a tedious task. A solution to these limiting factors would be to expand norms and to create an online portal for the storage and creation of stimulus sets.

Concept information can be delineated into two categories when discussing word norm databases: (1) single-word variables, such as imageability, concreteness, or number of phonemes, and (2) word-pair variables, wherein two words are linked, and the variables denote when those concepts are combined. Both category types can be important when planning an experiment based on word stimuli as an area of interest, and many databases contain a mix of variables. For example, the Nelson et al. (2004) free association norms contain both single-word information (e.g., concreteness, cue-set size, and word frequency) and word-pair information (e.g., forward and backward strength). For the word-pair variables, these values are only useful when exploring the cue and target together (i.e., first word–second word, concept–feature, concept–concept), because changing word combinations result in different variable values. In this study, we have collected semantic feature production norms, which are, in essence, word-pair information. Each of the concepts was combined with its set of listed features, and the frequency of the concept–feature pair was calculated. Furthermore, we used these lists of concept and feature

E. M. Buchanan (✉) · J. L. Holmes · M. L. Teasley
Missouri State University,
Springfield, MO, USA
e-mail: erinbuchanan@missouristate.edu

K. A. Hutchison
Montana State University,
Bozeman, MT, USA

frequencies to calculate the cosine value between concepts, which created another word-pair variable. Both of these variables should be considered word-pair norms, because both words are necessary to understanding the numeric variable (i.e., frequency and cosine). Therefore, we use the term *word-pair relations* to describe any variable based on concepts that are paired with either their features or other concepts, and supplementing previous work on such norms was a major goal of this data collection.

When examining or using word pairs as experimental stimuli, one inherent problem is that some words have stronger connections in memory than others. Those connections can aid our ability to read or name words quickly via meaning (Meyer & Schvaneveldt, 1971), or even influence our visual perception for words that we did not think were shown to us (Davenport & Potter, 2005). The differences in word-pair relations can be a disadvantage to researchers trying to explore other cognitive topics, such as memory or perception, because such differences can distort experimental findings if such factors are not controlled. For example, semantic-priming research investigates the facilitation in processing speed for a target word when participants are presented with a prior related cue word, as compared to an unrelated cue. Priming differences are attributed to the meaning-based overlap between related concepts, such that activation from the cue word readies the processor for the related target word. When the target word is viewed, recognition is accelerated because the same feature nodes are already activated (Collins & Loftus, 1975; Plaut, 1995; Stolz & Besner, 1999). Meta-analytic studies of semantic priming have shown that context-based connections in memory (association) were present in stimuli for studies on meaning-based priming, thus drawing attention to the opportunity to study these factors separately (Hutchison, 2003; Lucas, 2000). Consequently, researchers such as Ferrand and New (2004) have shown separate lexical decision priming for both semantic-only (*dolphin–whale*) and associative-only (*spider–web*) connections.

The simplest solution to this dilemma is to use the available databases of word information to create stimuli for such experiments. A search of the current literature for semantic word norms illustrates the dearth of recent meaning-based information in the psycholinguistic literature (specifically, those norms accessible for download). At present, the McRae, Cree, Seidenberg, and McNorgan (2005) and Vinson and Vigliocco (2008) norms for feature production are available, along with the Maki, McKinley, and Thompson (2004) semantic dictionary distance norms (all word-pair norms). Toglia (2009) recently published an update to the original Toglia and Battig (1978) single-word norms, and he described the continued use and need for extended research in psycholinguistics. McRae et al. detailed the common practice of self-norming words for use in research labs with small groups of participants. Rosch and Mervis's (1975) and

Ashcraft's (1978) seminal explorations of category information were both founded on the individualized creation of normed information. Furthermore, Vinson and Vigliocco used their own norm collection to investigate topics related to semantic aphasias (Vinson & Vigliocco, 2002; Vinson, Vigliocco, Cappa, & Siri, 2003), to build representation models (Vigliocco, Vinson, Lewis, & Garrett, 2004), and to understand semantic–syntactic differences (Vigliocco, Vinson, Damian, & Levelt, 2002; Vigliocco, Vinson, & Siri, 2005) before finally publishing their collected set in 2008. The literature search does indicate a positive trend for non-English database collections, as norms in German (Kremer & Baroni, 2011), Portuguese (Stein & de Azevedo Gomes, 2009), and Italian (Reverberi, Capitani, & Laiacona, 2004) can be found in other publications.

The databases of semantic feature production norms are of particular interest to this research venture. They are assembled by asking participants to list many properties for a target word (McRae et al., 2005; Vinson & Vigliocco, 2008). For example, when asked what makes a *zebra*, participants usually write features such as *stripes*, *horse*, and *tail*. Participants are instructed to list all types of features, ranging from "is a"/"has a" descriptors to uses, locations, and behaviors. While many idiosyncratic features can and do appear by means of this data collection style, the combined answers of many participants can be a reliable description of high-probability features. In fact, these feature lists allow for the fuzzy logic of category representation reviewed by Medin (1989). Obviously, semantic feature overlap will not be useful in explaining every meaning-based phenomenon; however, these data do appear to be particularly useful in modeling attempts (Cree, McRae, & McNorgan, 1999; Moss, Tyler, & Devlin, 2002; Rogers & McClelland, 2004; Vigliocco et al., 2004) and in studies on the probabilistic nature of language (Cree & McRae, 2003; McRae, de Sa, & Seidenberg, 1997; Pexman, Holyk, & Monfils, 2003).

The drawback to stimulus selection becomes apparent when researchers wish to control or manipulate several variables at once. For instance, Maki and Buchanan (2008) combined word pairs from popular semantic, associative, and thematic databases. Their word-pair list across just three types of variables was only 629 concept–concept pairs. If a researcher then wished to control for pair strength (e.g., only highly related word pairs) or single-word variables (e.g., word length and concreteness), the stimuli list would be limited even further. The Maki et al. (2004) semantic distance norms might provide a solution for some research endeavors. By combining the online WordNET dictionary (Fellbaum, 1998) with a measure of semantic similarity, JCN (Jiang & Conrath, 1997), they measured semantic distance by combining information on concept specificity and hierarchical distance between concepts. Therefore, this measurement describes how much two words have in

common in their dictionary definitions. For example, *computer* and *calculator* have high relation values because they have almost identical dictionary definitions. Alternatively, several databases are based on large text collections that appear to measure thematic relations (Maki & Buchanan, 2008), which is a combination of semantic and associative measures. Latent semantic analysis (LSA; Landauer & Dumais, 1997), BEAGLE (Jones, Kintsch, & Mewhort, 2006), and HAL (Burgess & Lund, 1997) all measure a mixture of frequency and global co-occurrence in which related words frequently appear either together or in the same context.

Given the limited availability of semantic concept–feature and concept–concept information, the present collection seeks to meet two goals. The first is to alleviate the limiting factor of low correspondence between the existing databases, so that researchers will have more options for stimulus collection. The semantic feature production norms are the smallest set of norms currently available, at less than 1,000 normed individual concepts, where associative norms, dictionary norms, and text-based norms all include tens of thousands of words. Compilation of this information would allow researchers to have more flexibility in generating stimuli for experiments and allow for studies on specific lexical variables. The second goal is to promote the use of these databases to improve experimental control in fields in which words are used as experimental stimuli. These databases are available online separately, which limits public access and awareness. Consequently, a centralized location for database information would be desirable. The Web portal created in tandem with this article (www.wordnorms.com) will allow researchers to create word lists with specific criteria in mind for their studies. Our online interface is modeled after projects such as the English Lexicon Project (http://elexicon. wustl.edu/; Balota et al., 2007) and the Semantic Priming Project (http://spp. montana.edu/; Hutchison et al., 2010), which both support stimulus creation and model testing, focusing on reaction times for words presented in pronunciation and lexical decision experiments.

Method

Participants

Participant data were collected in three different university settings: the University of Mississippi, Missouri State University, and Montana State University. University students participated for partial course credit. Amazon's Mechanical Turk was used to collect final word data (Buhrmester, Kwang, & Gosling, 2011). The Mechanical Turk provides a very large, diverse participant pool that allows short surveys to be implemented for very small amounts of money. Participant answers can be screened for

errors, and any surveys that are incomplete or incorrectly answered can be rejected. These participants were paid five cents for each short survey. Table 1 includes the numbers of participants at each site, as well as the numbers of concepts and the average number of participants per concept. Common reasons for rejecting survey responses included copying definitions from online dictionary sites or answering by placing the concept in a sentence. These answers were discarded from both the university data and the paid data set.

Materials

First, other databases of lexical information were combined to examine word overlap between associative (Nelson et al., 2004), semantic (Maki et al., 2004; McRae et al., 2005; Vinson & Vigliocco, 2008), and word frequency (Kučera & Francis, 1967) norms. Concepts present in the feature production norms were excluded, and a list of unique words was created, mainly from the free association norms. Some words in the feature production norms were repeated, in order to ascertain convergent validity. This list of words not previously normed, along with some duplicates, was randomized. These norms contained several variations of concepts (i.e., *swim*, *swims*, *swimming*), and the first version that appeared after the word list was randomized was used for most words. However, as another measure of convergent validity, we included morphological variations of several concepts (i.e., *state/states*, *begin/beginning*) to examine feature overlap. After several experimental sessions, information about the Semantic Priming Project (Hutchison et al., 2010) became available, and with their provided stimuli, concepts not already normed were targeted for the completion of our investigation. For the Semantic Priming Project, cue–target pairs were selected from the Nelson et al. free association norms, wherein no concept was repeated in either the cue or target position, but all words were allowed to appear as cue and target once each. The target words were both one cue word's most common response (first associate) and a different cue word's associate (second or greater associate). The cue words from this list (1,661 concepts) were then compared to the first author's completed norming work—the previous feature production norms—and the unique words were the final stimuli selected. Therefore, our data set provides a distinctive view into concepts not previous explored, such as pronouns, adverbs, and prepositions, while also adding to the collection of nouns and verbs.

Words were labeled by part of speech using both the English Lexicon Project and the free association norms. Words not present in these databases or that had conflicting entries were labeled using Google's "Define" feature search, and two experimenters reviewed these labels. The most prominent use of the word was considered its main part of

Table 1 Data collection site statistics: words, participants, and average response N

	University of Mississippi	Missouri State University	Montana State University	Mechanical Turk
Total participants	749	1,420	127	571
Concepts	658	720	120	310
Average N per concept	67.8	71.4	63.5	60

The average participants per concept were dependent on the number of words per experimental session. "Total participants" for the Mechanical Turk is the number of unique participants across all five-word sessions.

speech for this analysis, but multiple senses were allowed when participants completed the experiment. The data set ($N = 1,808$) contains 61.3 % nouns, 19.8 % adjectives, 15.5 % verbs, 2.2 % adverbs, 0.6 % pronouns, 0.5 % prepositions, and 0.1 % interjections. Because of the small percentages of adverbs, pronouns, prepositions, and interjections, these types were combined for further analyses. Table 2 shows the average numbers of features per participant by data collection location, and Table 3 indicates the word parts of speech by the parts of speech for features produced in the experiment.

Procedure

Given the different standards for experimental credit across universities, participants responded to different numbers of words in a session. Some participants responded to 60 words during a session lasting approximately an hour (the University of Mississippi, Montana State University), while others completed 30 words within approximately a half hour (Missouri State University). Mechanical Turk survey responses are best when the surveys are short; therefore, each session included only five words, and the average survey response times were 5–7 min. The word lists implemented on the Mechanical Turk were restricted to contain 60 unique participants on each short survey, but participants could take several surveys.

In order to maintain consistency with previous work, the instructions from McRae et al.'s (2005, p. 556) Appendix B were given to participants with only slight modifications. For instance, the number of lines for participants to write in their answers were deleted. Second, since many verbs and other word forms were used, the lines containing information about noun use were eliminated (please see the Discussion below for potential limitations of this modification). Participants were told to fill in the properties of each word, such as its physical (how a word looks, sounds, and feels), functional (how it is used), and categorical (what it belongs to) properties. Examples of three concepts were given (*duck*, *cucumber*, and *stove*) for further instruction. To complete the survey, participants were given a Web link to complete the experiment online. Their responses were recorded and then collated across concepts.

Data processing

Each word's features were spell-checked and scanned for typos. Feature production lists were evaluated with a frequency-count program that created a list of the features mentioned and their overall frequencies. For example, the cue word *false* elicited some target features such as *answer* (13), *incorrect* (25), and *wrong* (30). This analysis was a slight departure from previous work, as each concept feature was considered individually. Paired combinations were still present in the feature lists, but as separate items, such as *four* and *legs* for animals. From here, the investigator and research assistants examined each file for several factors. Filler words such as prepositions (e.g., *into*, *at*, *by*) and articles (*a*, *an*, and *the*) were eliminated unless they were relevant (e.g., the concept listed was *alphabet*). Plural words and verb tenses were combined into one frequency, so that *walk–walked–walks* are all listed as the same definition for that individual word concept. Then, features were examined across the entire data set. Again, morphologically similar features were combined into one common feature across concepts, so that concepts like *kind* and *kindness* would be considered the same feature. However, some features were kept separate, such as *act* and *actor*, for the following reasons. First, features were not combined when terms

Table 2 Average numbers of features listed by participants for each data collection location

	University of Mississippi	Missouri State University	Montana State University	Mechanical Turk	Total
Noun	10.06 (6.56)	10.78 (5.70)	10.59 (4.87)	14.36 (10.26)	11.12 (7.00)
Adjective	8.01 (4.69)	8.25 (3.76)	10.90 (6.10)	10.78 (6.73)	8.67 (4.88)
Verb	6.69 (4.54)	8.18 (5.28)	8.17 (3.01)	9.86 (6.35)	7.91 (5.16)
Other	7.13 (4.94)	7.86 (3.14)	10.83 (6.40)	13.55 (14.20)	8.83 (7.48)
Total	8.93 (6.01)	10.12 (4.74)	9.78 (5.45)	13.01 (9.57)	10.03 (6.54)

Standard deviations are in parentheses.

Table 3 Concept and feature parts of speech, percentages of each feature type per cue, and average response frequencies

Cue Type	Feature Type	Percent Features	Average Frequency
Noun	Noun	65.90	17.10 (23.28)
	Verb	16.80	16.56 (20.03)
	Adjective	15.80	14.75 (16.80)
	Other	1.40	11.65 (9.79)
Verb	Noun	48.70	13.20 (19.17)
	Verb	32.60	12.80 (17.53)
	Adjective	15.10	11.67 (12.22)
	Other	3.70	12.16 (10.93)
Adjective	Noun	44.20	12.07 (14.78)
	Verb	16.90	11.24 (11.81)
	Adjective	36.10	11.41 (11.22)
	Other	2.90	12.89 (16.36)
Other	Noun	44.80	11.95 (13.85)
	Verb	17.40	10.18 (13.70)
	Adjective	19.40	12.42 (12.70)
	Other	18.30	12.88 (17.53)

Average frequency is the average number of times that a participant listed a feature for that type of cue (i.e., the feature was a noun and the cue was a noun). Standard deviations are listed in parentheses.

marked differences in the sense of a noun/verb or in gender or type of person. For instance, *actor* denotes both that the feature is a person and the gender (male) of that person (vs. *actress* or the noun form *act*). Second, similar features were combined when the cue subsets were nearly the same (80 % of the terms). Features like *will* and *willing* were not combined because their cue sets only overlapped 38 %, which implied that these terms were not meant as the same concept.

Each final feature term was given a word type, as described above. Previously, both McRae et al. (2005) and Vinson and Vigliocco (2008) analyzed features by categorizing them as animals, body parts, tools, and clothing. However, the types of words included in this database did not make that analysis feasible (e.g., pronouns would not elicit feature items that would fit into those categories). Therefore, feature productions were categorized as main parts of speech (see Table 3). Given the number and varied nature of our stimuli, idiosyncratic features were examined for each individual concept. Features listed by less than two percent of respondents were eliminated, which amounted to approximately two to five mentions per concept. An example of how features are presented in the database can be found in Appendix A below.

Cosine values were calculated for each combination of word pairings. These values were calculated by summing the multiplication of matching feature frequencies divided by the products of the vector length of each word. Equation 1 shows how to calculate cosine, which is similar to a dot-product correlation. A_i and B_i indicate the overlapping feature's frequency between the first cue (A) and the second cue (B). The subscript i denotes the current feature. When A_i and B_i match, their frequencies are multiplied together and summed across all matching features (Σ). This product-summation is then divided by the feature frequency squared for both A and B, which is summed across all features from i to n (the last feature in each set). The square root ($\sqrt{}$) of the summation is taken for both of the cue sets, and these sums are multiplied together.

$$\frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \quad (1)$$

An example of cosine values from the database can be found below in Appendix B.

The McRae et al. (2005) and Vinson and Vigliocco (2008) norms were added to the present feature norms for comparison. This procedure resulted in more than a half-million nonzero combinations of word pairs for future use. The program written to create cosine values from feature lists allowed for like features across feature production files to be considered comparable (i.e., *investigate* and *investigation* are different forms of the same word). The feature lists were analyzed for average frequencies by word type, and the cosine values were used for comparison against previous research in this field. Both are available for download or search, along with the complete database files (see Appendix C), at our website http://wordnorms.com.

Results

Data statistics

Overall, our participants listed 58.2 % nouns, 19.3 % adjectives, 19.9 % verbs, 1.7 % adverbs, 0.6 % prepositions, 0.2 % pronouns, and 0.1 % other word types. The feature file includes 26,047 features for cue words, with 4,553 unique words. The features had an overall average frequency of approximately 14 mentions ($M = 14.88$, $SD = 19.54$). Table 3 shows the different types of features, percentages by concept, and average numbers of features listed for each type of concept. Most of the features produced by participants were nouns, but more verbs and adjectives were listed as features when the cue word provided was also a verb or adjective. Corresponding shifts in features were seen when other parts of speech were presented as cue words. Interestingly, a 4×4 (word type by data collection site) between-subjects analysis of variance revealed differences in the average numbers of features listed by participants for

parts of speech, $F(3, 1792) = 15.86$, $p < .001$, $\eta_p^2 = .03$, and for the different data collection sites, $F(3, 1792) = 12.48$, $p < .001$, $\eta_p^2 = .02$, but not an interaction of these variables, $F < 1$, $p = .54$. Nouns showed a higher average number of features listed by participants, over verbs ($p < .001$), adjectives ($p < .001$), and other parts of speech ($p = .03$) using a Tukey post hoc test. Mechanical Turk participants also listed more features than did all three university collection sites (all ps $< .001$), which is not surprising, since Mechanical Turk participants were paid for their surveys. All means and standard deviations can be found in Table 2.

Cosine values show a wide range of variability, from zero to nearly complete overlap between words. The cosine Excel files include all nonzero cosine values for our stimuli. Words were examined for reliability when comparing similar concepts, as there was some overlap in word forms. For example, *begin* and *beginning* yielded a cosine overlap of .95, while a low overlap was found between *state* and *states* (.31). Examining the multiple senses of *state* (a place) versus *states* (says aloud) might explain the lower feature overlap for that pair. The average overlap for these like pairs ($N = 121$) was $M = .54$, $SD = .27$. Given that the instructions allowed participants to list features for any number of word meanings, this overlap value indicates a good degree of internal consistency.

Convergent validity

Our feature production list was compared to the data sets from McRae et al. (2005) and Vinson and Vigliocco (2008) for overlapping concepts, in order to show convergent validity. Both previous feature production norms were downloaded from the archives of *Behavior Research Methods*. Then, concepts were selected that were in at least two of the databases. Concept–feature lists were compared, and concepts with multiple word features (i.e., *four legs* in McRae et al., 2005) were separated to match our current data processing. Cosine values were then calculated between all three data sets for matching concept–feature pairs, as described above. As noted previously, the McRae et al. and Vinson and Vigliocco norms had a strong relation, even though they were collected in different countries (Maki & Buchanan, 2008; $M_{cosine} = .63$, $SD = .16$, $N = 114$). The overall relationship between the combined data sets and these norms mirrored this finding with an equally robust average ($M_{cosine} = .61$, $SD = .18$, $N = 128$). When examined individually, the McRae et al. ($M_{cosine} = .59$, $SD = .16$, $N = 60$) and Vinson and Vigliocco ($M_{cosine} = .61$, $SD = .22$, $N = 68$) norms showed nearly the same overlapping relationship.

Concept–concept combinations were combined with JCN and LSA values from the Maki et al. (2004) semantic distance database (LSA as originally developed by Landauer & Dumais, 1997). In all, 10,714 of the pairs contained information on all three variables. Given this large sample size, p values for correlations are significant at $p < .001$, but the direction and magnitude of correlations are of more interest. Since JCN is backward-coded, with a zero value showing a high semantic relation (low distance between dictionary definitions), it should be negatively correlated with both the LSA and cosine values, where scores closer to 1 would be stronger relations. The correlation between the cosine values and JCN values was small to medium in the expected negative direction, $r = -.22$. This value is higher than the correlation between the JCN and LSA values, $r = -.15$, which is to be expected, given that LSA has been shown to measure thematic relations (Maki & Buchanan, 2008). The correlation between the LSA and cosine values was a medium positive relationship, $r = .30$, indicating that feature production may have a stronger connection to themes than does dictionary distance.

Divergent validity

Finally, we examined the connection between the cue–feature list and the cue–target probabilities from the free association database. Participants were instructed to think about word meaning; however, the separation between meaning and use is not always clear, which might cause participants to list associates instead of features. Table 4 indicates the percentage of cue–feature combinations that were present as cue–target combinations in the free association norms. Nearly all of our concepts were selected on the basis of the Nelson et al. (2004) norms, but only approximately 32 % of these lists contained the same cue–target/feature combination. The forward strength values for these common pairs were averaged and can be found in Table 4. While these values showed quite a range of forward strengths (.01–.94) overall, the average forward strength was only $M = .09$ ($SD = .13$). An example of some of the very large forward strength values are combinations such as *brother–sister* and *on–off*. Additionally, these statistics were broken down by part of speech to examine how participants might list associates instead of features for more abstract terms, such as adjectives and prepositions. Surprisingly, the most common overlaps

Table 4 Percent overlap between cue–feature lists and the cue–target lists from free association norms

	Percent Overlap	M FSG	SD FSG	Minimum	Maximum
Complete database	31.68	.09	.13	.01	.94
Nouns	32.61	.09	.12	.01	.89
Verbs	32.08	.09	.13	.01	.94
Adjectives	29.44	.10	.14	.01	.94
Other	25.68	.14	.20	.01	.90

FSG = forward strength.

were found with nouns and verbs (32 % of cue–target/feature listings), with less overlap for adjectives (29 %) and other parts of speech (26 %). The ranges and average forward strengths across all word types showed approximately the same values.

The Web portal (www.wordnorms.com)

The website built for this project includes many features for experimenters who wish to generate word-pair stimuli for research into areas such as priming, associative learning, and psycholinguistics. The word information is available for download, including the individual feature lists created in this project. The search function allows researchers to pick variables of interest, define their lower and upper bounds, or enter a preset list of words to search. All of the variables are described in Table 5, and a complete list is available online with minimum, maximum, mean, and standard deviation values for each variable.

Semantic norms As described above, the original feature production norms were used to create this larger database of cosine semantic overlap values. The feature lists for the 1,808 words are available, as well as the cosine relation between the new words and the McRae et al. (2005) and Vinson and Vigliocco (2008) norms. Their feature production lists can be downloaded through the journal publication website. In cases in which word-pair combinations overlapped, the average cosine strength is given. LSA values from the Maki et al. (2004) norms are also included, as a step between the semantic, dictionary-type measures and free association measures.

Association norms Free association values contained in both the Nelson et al. (2004) and Maki et al. (2004) norms have been matched for corresponding semantic pairs. This information is especially important, given the nature of the associative boost (Moss, Ostrin, Tyler, & Marslen-Wilson, 1995), indicating that both association and semantics should be considered when creating paired stimuli.

Frequency norms Although Brysbaert and New (2009) have recently argued against the Kučera and Francis (1967) norms, those norms are still quite popular, and are therefore included as reference. Other frequency information, such as HAL and the new English SUBTLEX values from Brysbaert and New's research, are included as well.

Word information Finally, basic word information is available, such as part of speech, length, neighborhoods, syllables, and morphemes. Parts of speech (nouns, verbs, etc.) were obtained from the English Lexicon Project (Balota et

Table 5 Search variables available at Wordnorms.com

Variable Type	Variable	Definition
Semantic	Cosine	Feature overlap between word pairs
	JCN	Semantic dictionary distance taken from WordNet
	LSA	Thematic relations examined by frequency of co-occurrence in text
Associative	FSG	Forward strength: The probability of a cue eliciting a target word
	BSG	Backward strength: The probability of a target eliciting a cue word
	QSS/TSS	Cue and target set size: The number of associates for either the cue or target
	QCON/TCON	Cue and target concreteness values, ranging from 1–7
Frequency	KF	Frequency per million, Kučera & Francis, 1967
	HAL	Frequency per million, HAL norms, Burgess & Lund, 1997
	LogHAL	Log of HAL frequency
	SUBTLEX	Frequency per million, Brysbaert & New, 2009
	LogSub	Log of SUBLTEX
Word Information	Length	Number of letters
	POS	Part of speech: Noun, verb, pronoun, etc.
	OrthoN	Orthographic neighborhood size
	PhonoN	Phonological neighborhood size
	Phonemes	Number of phonemes
	Syllables	Number of syllables
	Morphemes	Number of morphemes

Descriptive statistics for nonnull values in the database are found online and in Appendix C, including minimum and maximum values, averages, and standard deviations.

al., 2007), free association norms (Nelson et al., 2004), and through Google search's Define feature, for words not listed in these databases. Multiple parts of speech are listed for each cue on the website. The order of the part-of-speech listings indicates the most common to the least common usages. For example, NN|VB for the concept *snore* indicates that *snore* is typically used as a noun, then as a verb. Word length simply denotes the numbers of letters for the cue and target words.

Phonological and orthographic neighbor set sizes are also included. Phonological neighborhoods include the set of words that can be created by changing one phoneme from the cue word (e.g., *gate→hate*; Yates, Locker & Simpson, 2004). Conversely, the orthographic neighborhood of a cue is the group of words that can be created by replacing one letter with another in the same placeholder (e.g., *set→sit*), and these neighborhood set sizes and their interaction have

been shown to affect the speed of processing (Adelman & Brown, 2007; Coltheart, Davelaar, Jonasson, & Besner, 1977). These values were obtained from WordMine2 (Durda & Buchanan, 2006), and were then crosschecked with the English Lexicon values. The numbers of phonemes, morphemes, and syllables for concepts are provided as the final set of lexical information for the cue and target words. *Snore*, for example, has four phonemes, one syllable, and one morpheme.

Discussion

The word information presented here adds to the wealth of word-norming projects available. We collected a large set of semantic feature production norms, and the semantic feature overlap between words was calculated for use in future research. A strong relationship was found between this data collection and previous work, which indicates that these norms are reliable and valid. A searchable Web database is linked for use in research design. Interested researchers are encouraged to contact the first author about addition of their information (norms, links, corrections) to the website.

Several limitations of feature production norms should be noted, especially when considering their use. First, our data-processing procedure created feature lists as single-word items. We believe that this change over some paired concepts did not change the usability of these norms, as correlations between our database and existing databases were as high as those between the existing databases themselves. However, this adjustment in feature processing may have some interesting implications for understanding semantic structure. For instance, is the concept *four legs* stored in memory as one entry, or separated into two entries with a link between them? Three-legged dogs are still considered *dogs*, which forces us to consider whether the *legs* feature of the concept is necessarily tied to *four* or is separated, with a fuzzy boundary for these instances (Rosch & Mervis, 1975).

The negative implication of this separation between features may be an inability to finely distinguish between cues. For instance, if one cue has *four legs* and another has *four eyes*, these cues will appear to overlap because the *four* features will be treated the same across cue words. However, separating linked features may provide advantages to a person when trying to categorize seemingly unrelated objects (Medin, 1989). In other words, *dog* would be more similar to other four-legged objects because the concept is linked to a *four* feature.

Second, multiple word senses can be found for many of the normed stimuli, which will invariably create smaller sublists of features, depending on participants' interpretations. While these various sense lists are likely a realistic

construal of linguistic knowledge, the mix of features can lower feature overlap for word pairs that intuitively appear to match. The feature production lists are provided to alleviate this potential problem (i.e., cosine values may be calculated for feature sublists), and future research could investigate whether sense frequency changes the production rates of certain items. Also, participant creation of norms may exclude many nonlinguistic featural representations, such as spatial or relational (i.e., *bigger than*) features. Likewise, while the features listed for a concept could match, their internal representations may vary. For example, rabbits and kangaroos both *hop*, but one would argue that the difference between their hops is not present in these types of concept features. Finally, overlap values are prone to capturing the relationship of salient features of concepts, possibly because salient features have a special status in our conceptual understanding (Cree & McRae, 2003).

Finally, our database is the first to examine feature production for abstract terms, adjectives, and other word types not typically normed. We examined the relationship of our cue–feature lists to the free association cue–target data, with approximately 32 % overlap between the lists. If participants were unable to list features for verbs and adjectives, we would expect this overlap to be higher for such cues, which it was not. Furthermore, we would expect to find many low-frequency cues, with no general agreement on the features of a concept (i.e., cues listed by all participants). Yet, most participants listed *accomplish*, *success*, and *goal* for the verb *achieve*, along with other similar, infrequent features, such as *finish*, *win*, and *work*. The adjective *exact* only showed high-frequency features, such as *precise*, *accurate*, and *strict*, indicating that most participants agreed on the featural definition of the concept. Finally, we would expect reduced correlations to other databases or lower internal overlap of pairs if participants were unable to list features for abstract terms, which did not occur.

While this project focused on word-pair relations, many other types of stimuli are also available to investigators. Although the work is out of date, Proctor and Vu (1999) created a list of many published norms, which ranged from semantic similarity to imageability to norms in other languages. When the Psychonomic Society hosted an archive of stimuli, Vaughan (2004) published an updated list of normed sets. Both of these works indicate the need for researchers to combine various sources when designing stimulus sets for their individualized purposes. Furthermore, concept values found in these norms are an opening for other intriguing research inquiries in psycholinguistics, feature distributional statistics, and neural networks.

Appendix A: Database files—Feature lists

The file Feature_lists.xlsx contains features listed for each cue word, excluding features eliminated due to low frequency (see the information in Table 6 below).

Table 6 Contents of feature_lists.xlsx

cue	cuepos	frequency	target	targetpos
abandon	2	8	away	4
abandon	2	7	behind	4
abandon	2	6	child	1
abandon	2	7	complete	3
abandon	2	10	control	2
abandon	2	11	desert	1
abandon	2	9	discard	2
abandon	2	9	discontinue	2
abandon	2	7	forsake	2
abandon	2	24	give	2
abandon	2	28	leave	2
abandon	2	8	one	3
abandon	2	9	withdraw	2
ability	1	17	able	3

The file contains the following information: (A) Cue word, (B) Cue part of speech (1 = Noun, 2 = Verb, 3 = Adjective, 4 = Other), (C) Frequency of feature (the number of times participants listed the feature word), (D) Feature word, and (E) Feature part of speech (1 = Noun, 2 = Verb, 3 = Adjective, 4 = Other). The feature lists can be viewed online at http://wordnorms.missouristate.edu/database/feature_lists.xlsx.

Appendix B: Database files—Cosine values

Cosine A-J.xlsx and Cosine K-Z.xlsx: These files contain all nonzero cosine values for every cue-to-cue combination (see the information in Table 7 below). These values were calculated as described in the Method section.

Table 7 Contents of cosine A-J.xlsx and cosine K-Z.xlsx

abandon	ability	0.04429889
abandon	able	0.0092353
abandon	absurd	0.02032996
abandon	abuse	0.03631799
abandon	achieve	0.00912561
abandon	act	0.01712004
abandon	addicting	0.07212855
abandon	adolescence	0.07282952
abandon	adolescent	0.03623845
abandon	adults	0.01119557

The columns are as follows: (A) Cue word, (B) Target word, and (C) Cosine value. These cosine values can be viewed online at http://wordnorms.missouristate.edu/database/cosine_A_J.xlsx and http://wordnorms.missouristate.edu/database/cosine_K_Z.xlsx.

Appendix C: Database files—Complete database

Dataset.zip (see Table 8): This file contains six separate space-delimited text files of all available values on the Web portal. Each file is 100,000 lines, except for the final text file. These files can be imported into Excel for sorted and searching. Please note that the files are quite large and may open very slowly. The data set can also be searched online for easier use.

Table 8 Contents of dataset.zip files

Information Type	Variable	Label	Minimum	Maximum	M	SD
	Cue word	The first word in a word pairing. For semantic variables, the order of pairings is not important. For the association variables, this value represents the first word given to participants in a free association task.				
	Target word	The second word in a word pairing. For associative variables, this word represents the first word that "came to mind when shown the cue word"				
Semantic	Cosine	The feature overlap between two words. This value ranges from 0 to 1, where 0 values indicate no overlap between words and 1 values indicate complete overlap between words.	0.00	1.00	0.13	0.12
Semantic	JCN	The dictionary distance between words. Using WordNet, Maki et al. (2004) have calculated the relationship between word pairs. This variable ranges from 0 to 32 and is reverse coded so that 0 values have a very high semantic relationship.	0.00	28.03	10.97	6.19
Semantic	LSA	Latent semantic analysis shows both the semantic and thematic relationship between word pairs. Low values are close to 0, and high values are close to 1.	0.00	1.00	0.25	0.17
Associative	FSG	Forward strength: Probability of the target word association when shown the cue word (ranges from 0 to 1).	0.00	0.94	0.07	0.11
Associative	BSG	Backward strength: Probability of the cue word association when shown the target word (ranges from 0 to 1).	0.00	20.00	0.06	0.57
Associative	QSS	Cue set size: Number of associates a cue word is connected to (neighbors).	1.00	34.00	14.66	5.04
Associative	TSS	Target set size: Number of associates a target word is related to.	1.00	34.00	14.61	5.01
Associative	QCON	Cue concreteness: Ranges from low (1) to high (7).	1.49	7.00	4.95	1.33
Associative	TCON	Target concreteness: Ranges from low (1) to high (7).	1.00	7.00	4.94	3.41
Frequency	KF1	Cue word frequency: Kučera and Francis (1967) norms.	0.00	21,341.00	105.14	505.04
Frequency	KF2	Target word frequency: Kučera and Francis norms.	0.00	1,625,073.00	114.33	2,552.49
Frequency	HAL1	Cue word frequency: Burgess and Lund (1997)	0.00	8,015,301.00	47,464.93	231,060.18
Frequency	HAL2	Target word frequency: Burgess and Lund norms	0.00	8,015,301.00	51,479.61	256,346.73
Frequency	LogHAL1	Log of cue word frequency from HAL	0.00	15.90	8.93	1.98
Frequency	LOGHAL2	Log of target word frequency from HAL	0.00	5,247.45	8.96	7.96
Frequency	Subtlex1	Cue word frequency: Brysbaert and New (2009)	0.02	18,896.31	133.79	681.69
Frequency	Subtlex2	Target word frequency: Brysbaert and New norms	0.02	41,857.12	165.83	1,263.21
Frequency	LogSub1	Log of SUBTLEX cue word frequency	0.30	5.98	2.90	0.86
Frequency	LogSub2	Log of SUBTLEX target word frequency	0.30	9.00	2.92	0.87
Lexical	Length1	Cue number of letters	2.00	16.00	5.75	1.99
Lexical	Length2	Target number of letters	2.00	16.00	5.75	2.00
Lexical	POS1	Part of speech (noun, verb, etc.) for cue word				
Lexical	POS2	Part of speech for target word				

Table 8 (continued)

Information Type	Variable	Label	Minimum	Maximum	*M*	*SD*
Lexical	Ortho1	Orthographic neighborhood size (number of neighbors that look similar) for cue word	0.00	34.00	5.57	6.54
Lexical	Ortho2	Orthographic neighborhood size for target word	0.00	50.00	5.50	6.41
Lexical	Phono1	Phonographic neighborhood size (number of words that sound the same) for cue word	0.00	59.00	12.80	14.47
Lexical	Phono2	Phonographic neighborhood size for target word	0.00	59.00	12.71	14.25
Lexical	Phonemes1	Number of phonemes for cue word	1.00	12.00	4.63	1.70
Lexical	Phonemes2	Number of phonemes for target word	1.00	12.00	4.62	1.69
Lexical	Syllables1	Number of syllables for cue word	1.00	5.00	1.70	0.80
Lexical	Syllables2	Number of syllables for target word	1.00	5.00	1.68	0.79
Lexical	Morphemes1	Number of morphemes for cue word	1.00	5.00	1.30	0.54
Lexical	Morphemes2	Number of morphemes for target word	1.00	5.00	1.30	0.53

Variables are presented in their order in the database. Some words appear twice (i.e., with a "2" or "_information") due to multiple entries in previous semantic databases. The denotations are indicators of different tenses or word meanings. These values are explained in the original databases (McRae et al., 2005; Vinson & Vigliocco, 2008). This table is also provided online. The entire data set may be downloaded at http://wordnorms.missouristate.edu/database/dataset.zip. This variable list may be downloaded at http://wordnorms.missouristate.edu/database/variables.docx.

References

Adelman, J. S., & Brown, G. D. A. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review, 14,* 455–459. doi:10.3758/BF03194088

Ashcraft, M. H. (1978). Property norms for typical and atypical items from 17 categories: A description and discussion. *Memory & Cognition, 6,* 227–232. doi:10.3758/BF03197450

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. doi:10.3758/BF03193014

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, Instruments, & Computers, 41,* 977–990. doi:10.3758/BRM.41.4.977

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. doi:10.1177/1745691610393980

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language & Cognitive Processes, 12,* 177–210. doi:10.1080/016909697386844

Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–428. doi:10.1037/0033-295X.82.6.407

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology. General, 132,* 163–201. doi:10.1037/0096-3445.132.2.163

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science, 23,* 371–414. doi:10.1207/s15516709cog2303_4

Davenport, J., & Potter, M. (2005). The locus of semantic priming in RSVP target search. *Memory & Cognition, 33,* 241–248. doi:10.3758/BF03195313

Durda, K., & Buchanan, L. (2006). WordMine2 [Online]. Available at http://web2.uwindsor.ca/wordmine

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database (Language, speech, and communication)*. Cambridge, MA: MIT Press.

Ferrand, L., & New, B. (2004). Semantic and associative priming in the mental lexicon. In P. Bonin (Ed.), *The mental lexicon* (pp. 25–43). Hauppauge, NY: Nova Science.

Hutchison, K. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review, 10,* 785–813. doi:10.3758/BF03196544

Hutchison, K. A., Balota, D. A., Cortese, M. J., Neely, J. H., Niemeyer, D. P., Bengson, J. J. & Cohen-Shikora, E. (2010). The Semantic Priming Project: A Web database of descriptive and behavioral measures for 1,661 nonwords and 1,661 English words presented in related and unrelated contexts. Available at http://spp.montana.edu, Montana State University.

Jiang, J. J., & Conrath, D. W. (1997, August). *Semantic similarity based on corpus statistics and lexical taxonomy*. Paper presented at the International Conference on Research on Computational Linguistics (ROCLING X), Taipei, Taiwan.

Jones, M., Kintsch, W., & Mewhort, D. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language, 55,* 534–552. doi:10.1016/j.jml.2006.07.003

Kremer, G., & Baroni, M. (2011). A set of semantic norms for German and Italian. *Behavior Research Methods, 43,* 97–109. doi:10.3758/s13428-010-0028-x

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240. doi:10.1037/0033-295X.104.2.211

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review, 7,* 618–630. doi:10.3758/BF03212999

Maki, W. S., & Buchanan, E. (2008). Latent structure in measures of associative, semantic, and thematic knowledge. *Psychonomic Bulletin & Review, 15,* 598–603. doi:10.3758/PBR.15.3.598

Maki, W. S., McKinley, L. N., & Thompson, A. G. (2004). Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments, & Computers, 36,* 421–431. doi:10.3758/BF03195590

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37,* 547–559. doi:10.3758/BF03192726

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology. General, 126,* 99–130. doi:10.1037/0096-3445.126.2.99

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469–1481. doi:10.1037/0003-066X.44.12.1469

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234. doi:10.1037/h0031564

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 863–883. doi:10.1037/0278-7393.21.4.863

Moss, H. E., Tyler, L. K., & Devlin, J. T. (2002). The emergence of category-specific deficits in a distributed semantic system. In E. M. E. Forde & G. W. Humphreys (Eds.), *Category-specificity in brain and mind* (pp. 115–147). Hove, East Sussex, U.K.: Psychology Press.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36,* 402–407. doi:10.3758/BF03195588

Pexman, P. M., Holyk, G. G., & Monfils, M.-H. (2003). Number-of-features effects and semantic processing. *Memory & Cognition, 31,* 842–855. doi:10.3758/BF03196439

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17,* 291–321. doi:10.1080/01688639508405124

Proctor, R. W., & Vu, K.-P. L. (1999). Index of norms and ratings published in the Psychonomic Society journals. *Behavior Research Methods, Instruments, & Computers, 31,* 659–667. doi:10.3758/BF03200742

Reverberi, C., Capitani, E., & Laiacona, E. (2004). Variabili semantico–lessicali relative a tutti gli elementi di una categoria semantica: Indagine su soggetti normali italiani per la categoria "frutta. *Giornale Italiano di Psicologia, 31,* 497–522.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605. doi:10.1016/0010-0285(75)90024-9

Stein, L., & de Azevedo Gomes, C. (2009). Normas Brasileiras para listas de palavras associadas: Associação semântica, concretude, frequência e emocionalidade. *Psicologia: Teoria E Pesquisa, 25,* 537–546. doi:10.1590/S0102-37722009000400009

Stolz, J. A., & Besner, D. (1999). On the myth of automatic semantic activation in reading. *Current Directions in Psychological Science, 8,* 61–65. doi:10.1111/1467-8721.00015

Toglia, M. P. (2009). Withstanding the test of time: The 1978 semantic word norms. *Behavior Research Methods, 41,* 531–533. doi:10.3758/BRM.41.2.531

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

Vaughan, J. (2004). Editorial: A Web-based archive of norms, stimuli, and data. *Behavior Research Methods, Instruments, & Computers, 36,* 363–370. doi:10.3758/BF03195583

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition, 85,* 61–69. doi:10.1016/S0010-0277(02)00107-5

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meaning of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology, 48,* 422–488. doi:10.1016/j.cogpsych.2003.09.001

Vigliocco, G., Vinson, D. P., & Siri, S. (2005). Semantic and grammatical class effects in naming actions. *Cognition, 94,* 91–100. doi:10.1016/j.cognition.2004.06.004

Vinson, D. P., & Vigliocco, G. (2002). A semantic analysis of noun–verb dissociations in aphasia. *Journal of Neurolinguistics, 15,* 317–351. doi:10.1016/S0911-6044(01)00037-9

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods, 40,* 183–190. doi:10.3758/BRM.40.1.183

Vinson, D. P., Vigliocco, G., Cappa, S., & Siri, S. (2003). The breakdown of semantic knowledge: Insights from a statistical model of meaning representation. *Brain and Language, 86,* 347–365. doi:10.1016/S0093-934X(03)00144-5

Yates, M., Locker, L., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review, 11,* 452–457. doi:10.3758/BF03196594