Appendix A

All of the normed words, their responses, and related information.

Table 1

Quick reference for meaning of column names

Column Label	**Definition**
Cue	Normed Word, or Q
Target	Response to Normed Word, or T
Target Normed?	Is Response Normed?
Sample Size	Number of participants
# Subjects Producing Target	Number of participants producing the target response
Forward Strength	Forward Cue-to-Target Strength, Q-T
Backward Strength	Backward Target-to-Cue Strength, T-Q
Mediated Strength	Mediated Strength, Q-M-T
Shared Associate Strength	Shared Associate Strength, Q-SA & T-SA
# Mediated Associates	Number of Mediators linking Q-T
Non-normed Potential Mediators	Number of associates of the cue that could be mediators that were not normed
# Shared Associates	Number of Shared Associates linking Q-T
Cue Set Size	Number of different cue associates, or network size
# Cue Competitors	Cue Set Size minus connections that link Q-T (minus target, # mediators, and # shared associates)
Cue Competitor Strength	The summed strengths of the cue's competitors
Cue Frequency (K&F)	Printed frequency in Kucera & Francis (1967) norms
Cue Concreteness (Scale 1-7)	Concreteness ratings for the cue word

Cue Homograph? (Letter=Yes)	Is the cue a homograph? Letter indicates norms used
Cue Part of Speech	Part of speech of the cue determined by a dictionary
Mean Cue Connectivity	Mean number of associate-to-associate links for the cue
Cue Connectivity Strength	Summed strength of cue's associate-to-associate links
Cue Resonance Probability	Probability that cue's associates produce it as an associate
Cue Resonance Strength	Summed strength of cue's associate-to-target links
Target Set Size	Number of different target associates, or network size
# Target Competitors	Target Set Size minus connections that link Q-T (minus cue and # shared associates)
Target Competitor Strength	The summed strengths of the target's competitors
Target Frequency (K&F)	Printed frequency in Kucera & Francis (1967) norms
Target Concreteness (Scale 1-7)	Concreteness ratings for the target word
Target Homograph? (Letter=Ye	Is the target a homograph? Letter indicates norms used
Target Part of Speech	Part of speech of the target determined by dictionary
Mean Target Connectivity	Mean number of associate-to-associate links for target
Target Connectivity Strength	Summed strength of target's associate-to-associate links
Target Resonance Probability	Probability that target's associates produce it as an associate
Target Resonance Strength	Summed strength of target's associate-to-target links
# of Other Associate Links	Number of links from other associates in the target's set to the target's cue
Other Associate Strength	Strength of the input from other associates in the target's set to the target's cue

Explanation of the column names

This Read Me describes the content of Appendix A. Data for 5,019 normed words, their 72,176 responses, and 33 data fields can be found in this appendix on two sheets of an Excel file.

The first column or field presents the normed words or **Cues** listed in alphabetical order, and the second field presents their responses or **Targets** alphabetized within cues**.** We refer to these items as cue-target pairs because of how such items are selected for use in research in our area of memory. Targets are selected as words to be studied in memory experiments, and cues are used to prompt the recall of these studied targets.

Given the wide variation in word properties, the norms are used for constructing lists of pairs that systematically vary in some properties while holding other properties constant. By incorporating the database into a personal database program, these norms make it feasible to control certain word pair and word attributes while varying others with greater degrees of rigor than ever before. For example, by imposing search restrictions on the targets in the pool, such as reporting only target words that occur 50 or more times per million, that have a concreteness rating of 4.8 or greater, and that have a forward strength of .15 or less can be reported. By using the flexibility that the norms provide, the "noise" resulting from uncontrolled factors can be substantially reduced whether pairs are being selected for semantic memory studies or lists are being constructed for episodic memory studies. With less noise, more subtle main effects can be detected with greater ease and reliability. In general, Appendix A can be used for selecting pairs of related words that have been produced by two or more subjects in free association. Furthermore, by re-alphabetizing the norms by target instead of cue, it is possible to determine the cue words that produce a given target word, along with all of the useful associative information.

The remaining fields present information about the pairs or the individual words comprising them. The *Target Normed?* field provides a yes-no alternative and indicates whether the target word in the pair has been normed by a separate group of participants. Of the 72,176 total targets appearing in the database, 8,557 have not been normed and therefore cells that provide normative information about these items is listed as missing. For example, backward strength values are provided for only 88% of the responses. These responses comprise the 5,019 normed words produced redundantly by different cues, e.g., 18 different words produce ABILITY as a response.

The ***Sample Size*** field provides the number of participants serving in the group norming the cue, and the ***# Subjects Producing Target*** field provides the number of participants producing a particular response. The ***Forward Strength*** field provides the free association response probabilities or what is traditionally called forward cue-to-target strength. This value is calculated by dividing # Subjects Producing Target by Sample Size, which gives the proportion of subjects in the group who produce a particular target in the presence of the cue. For example, for the word ABILITY, 17 out of the 143 participants in the group produced CAPABILITY as a response, so forward strength for this pair is .119.

Backward strength refers to the probability that the target, when independently normed produces the cue word as a associate. The term backward follows the conventional but admittedly confusing use of the term in memory research. It is measured in the same way as forward strength, except the word appearing as the "target" now serves as the "cue" to be normed instead of the reverse. If it is important for some purpose to know sample size and the # Subjects Producing Target for the backward strength index, look up the word serving as the target in a given pair as a cue. For example, for CAPABILITY in the above pairing, 35 out of a group of 124 participants produced ABILITY as a response, so backward strength in the ABILITY CAPABILITY pairing is calculated at 35/124 = .282.

The next 4 fields index indirect connections between the word pairs. Forward and backward strength represent measures of direct strength because one word directly produces the other as an associate in free association. Indirect connections index links between related words that occur through other words. Such connections are often ignored in research applications of normative data but they can be very strong and can have large effects on memory performance in certain tasks (Nelson, Bennett, & Leibert, 1997; Nelson et al., 1998). ***Mediated strength***, also called 2-step strength in the memory literature, indexes the summed strength of the mediators running from the cue word to each response. For example, ABILITY produces "competence" with a probability of .06, which in turn produces "capability" with a probability .08. The mediated strength of the ABILITY-CAPABILITY pair is calculated by cross multiplying the individual links,

and then summing the results across each link. Given that no other mediated links were detected for this pair, mediated strength was calculated as .06 * .08 = .0048. This particular pair has one 2-step mediated link, but some word pairs have no such connections whereas others have as many 17. When multiple mediators are present, the results of the multiplication process are summed across mediators. The highest calculated mediated strength in this database is .66 and it should be noted that indirect strength as indexed by this procedure can exceeds direct strength. The field ***# Mediated Associates*** represents how many mediators or 2-step connections link the cue and the target were measured. Importantly, the field called ***Non-normed Potential Mediators*** reports that number of associated responses of the cue that were never normed. Without norming these associates we cannot know whether they are mediators. This fact is important because it indicates that the data reported here for the number and strength of the mediators can underestimate these measures. When manipulating mediators, the non-normed field should indicate values of 0.0.

Shared Associate Strength arises when two words comprising a particular pair have associates in common. The cue and target may produce some of the same words as associates. For example, both ABILITY and CAPABILITY produce the same 6 words as associates, including able, strength, talent, potential, capacity, and knowledge. Shared associate strength is calculated by multiplying cue-to-shared associate strength times target-to shared strength and then summing the results when more than one shared associate is involved. Shared associate strength is calculated as a converging mediator. The field ***# of Shared Associates*** reports the number of shared associates for each word pair. It is important to note that some of the shared associates were also mediators, and these mediators were incorporated into the relevant columns. In effect, some associates were included in both mediator and shared associate counts.

The next 22 fields provide information about the cue and the targets as individual words, and these fields are subdivided into 11 descriptive statistics about the cue and 11 comparable statistics concerning the target. The ***Set Size*** indices are calculated by counting the number of different responses or targets given by two or more participants in the normative sample. Some words have set sizes of 1.00 (e.g., LEFT) whereas others have set sizes of 30 or more different

words (e.g., FARMER), and in general, set size closely approximates a normal distribution. The criterion of "two or more" participants was chosen at the beginning of the norms project on the assumption that idiosyncratic responses given by a single participant would tend to be "off the wall." The opinion was that such responses should not be counted as in the set because they would "vary with different walls" and would therefore be unreliable. Retrospectively, such responses are probably directly related weak associates, and are not as "off the wall." They are, however, unreliable because re-normings of hundreds of the same words showed that a different set of idiosyncratic responses tend to be produced each time the words was normed (Nelson & Schreiber, 1992). Words given by two or more subjects tend to be highly reliable, as is the number of different words produced by the cue, regardless of whether they are given by two or more participants or by a single participant. What is different are the specific idiosyncratic responses produced by a single participant.

We now interpret these findings to mean that most words are linked to very large numbers of other words. Discrete free association norms, provide a reliable index of the number of strongest associates, or nearest neighbors in the sense of semantic distance. Even a response that is provided by only 2 out of 150 participants is regarded as a relatively strong associate. However, because idiosyncratic responses seem to be unreliable members of the set, we concluded that words are connected strongly to some associates and that they are very weakly connected to many others. These weak associates tend to be linked to specific contexts and so are unlikely to appear in free association, which is designed to be a context free task. The lesson we take from these considerations is that discrete free association provides a very good indicator of the number of strong associates and a very poor indicator of the number of weak associates. Hence, set size provides a relative index of the set size of a particular word by providing a reliable measure of how many strong associates it has. Because it fails as an indicator of the number of weak associates, this index should never be construed as providing an index of absolute set size. Appendix B offers a total set size index that incorporates idiosyncratic responses, and Appendix D provides the actual idiosyncratic responses for each cue word.

The fields labeled ***# of Cue Competitors*** and ***# of Target Competitors*** represent a subset of the set size index because the pairing of specific words constrains both measures. Competitors are defined as those that compete with the target as a response in either free association or cued recall, and they can come from either the cue or the target. Some types of links join the cue and target and some do not, and those that do not compete. Number of competitors is computed by determining set size and then removing associates of the cue and target that link them together as pairs. Cue competitors are determined by subtracting the target as well as the numbers of mediators and shared associates from set size. The target is the associate of the cue whose recovery is in question, and both mediators and shared associates facilitate recall so such items need to be subtracted from the competitor count for the cue. It is important to note that non-normed associates of the cue are potential mediators and so they were subtracted along with known mediators. Hence, the cue competitor index underestimates the number of true competitors to the extent that non-normed associates could be competitors instead of mediators. Experiments manipulating cue competitors should rely on using cues with low numbers of non-normed cue associates, preferably 0.0.

Number of target competitors is calculated under the same principal. Number of target competitors is computed by determining set size and than subtracting the cue and the number of shared associates. Number of mediators is not subtracted because this linking connection comes from the cue, not the target. When a mediator is also a shared associate, it is subtracted as a shared associate. Note that number of target competitors does not suffer the same complications as number of cue competitors because mediators are not involved in their computation. Finally, ***Cue and Target Competitor Strength*** are computed by summing the strengths of the individual competitors. For cue competitor strength, the strengths of the cues' associates are summed, and then forward strength, cue-to-shared associate strength, and cue-to-mediator strength are each subtracted from the summed strength. For target competitor strength, the strength of the targets' associates are summed, and then backward strength to the cue, and target-to-shared associate strength are subtracted. For both cues and targets, the only associates involved in these

calculations are those produced by 2 or more participants. Idiosyncratic responses and non-responses are not involved.

Cue and Target Frequency values were borrowed from the Kucera and Francis (1967) norms for the convenience of readers. Many but not all of the ***Concreteness Ratings (Scale 1-7)*** were also borrowed. First, we looked up a given word in the Paivio, Yuille and Madigan (1968) norms. If the word was found, its concreteness was entered into our database, and if not, the word was looked up in the Toglia and Battig norms (1978), and this value was used. Finally, if the word was not in either source, we normed it using procedures described by Paivio et al. (1968). In this way, concreteness values are provided for 3,260 words. Data from homograph norms were also included, as indicated by the field ***Homograph?***. The information was also borrowed from other databases that separate the associates into two or more classes on the basis of different meanings. A blank space indicates that the word under consideration is probably not classified as a homograph, and a single letter indicates that it is a homograph or that it is likely to be one. The letters refer to the first letter of the first author associated with the homograph norms so that interested readers can pursue source if desired. This information is provided in Table 2, and it should be noted that, as with concreteness ratings, sources were used in a particular ordering. This ordering can be described by arranging the letters of the authors from first to last used: N, P, W, T, G and C. Other than selecting what was handy at the time, no particular rationale was used in determining this ordering, but it means that some words will appear in more than one set of norms and this fact is not recognized here.

Table 2

Sources of homograph norms

Letter	# Words	Source
C	6	Cramer, P., 1970
G	247	Identified as likely by Nancy Gee

N	297	Nelson et al., 1980
P	33	Perfetti, et al., 1971
T	167	Twilley et al., 1992
W	48	Wollen et al., 1980

The ***Part-of-Speech*** classification was determined by the first listing in The American Heritage Dictionary of the English Language (1980). Only a single entry is provided for each word, even when, for example a word can be classified as either a noun or a verb. Part of speech is indicated by the first letter or by two letters for each classification, and Table 3 provides the definitions.

Table 3

Definitions of parts of speech

Abbreviation	**Part of Speech**
N	Noun
V	Verb
AJ	Adjective
AD	Adverb
P	Pronoun
PP	Preposition
I	Interjection
C	Conjunction

The ***Mean Connectivity*** fields index the average associate-to-associate connectivity among the associates of the cue and of the target. This measure is obtained by norming the associates of the cue and target using independent groups of participants, counting the number of connections among the associates in the set, and then dividing by the size of the set (minus the number of non-normed missing associates, if any). This index captures the density and the level

of organization among the strongest associates of the cue. ***Connectivity Strength*** sums the strengths of the individual associate-to-associate links in a words' associative set. These indices can best be understood by examining the NXN associative matrices for over 4,000 words available in Appendix C.

Resonance Probability indexes associate-to-target links and is the probability that the target's associates produce the target as an associate. Historically, we used the resonance descriptor because such connections create a potential resonating link formed by the target-to-associate, associate-to-target loop. However, research has since shown that activation does not appear to spread (Nelson, McEvoy, & Pointer, 2003). Instead, resonant links contribute additively with associate-to-associate links to increase the activation level of the target. Resonance and connectivity have additive as opposed to interactive effects on extralist cued recall. ***Resonance Strength*** sums the strengths of the associate-to-target links in a words' associative set. The strengths are available for specific words in the NXN matrices of Appendix C.

Finally, the last two columns in Appendix A provide the number and the strengths of the associate-to-associate links that are specific to the cue word, respectively, ***# of Other Associate Links to Target*** and ***Other Associate Strength***. These statistics represent a subset of associate-to-associate connectivity. Other associate links are specific to the cue. These statistics are important in the extralist cued recall task in which participants study a list of individually presented words and are then cued by related words during testing. When the cue is an associate of the studied target, theoretically it receives activation input from the target (backward strength) as well as input from other associates in the target's set (other associate strength). According to the encoding specificity principle (Tulving & Thomson, 1973), PIER2 (Nelson, McKinney, Gee, & Janczura, 1998) as well as other models, cues with stronger backward and other associate links will be more effective, and this appears to be the case (Goodmon & Nelson, in press).