

# Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns

DAVID C. RUBIN

*Duke University, Durham, North Carolina*

and

MICHAEL FRIENDLY

*York University, Downsview, Ontario, Canada*

To investigate the properties that make a word easy to recall, we added to existing norms for 925 nouns measures of availability, goodness, emotionality, pronunciability, and probability of recall in multiple-trial free recall. Availability, imagery, and emotionality were found to be the best predictors of which words were recalled. This result, which is stable across recall data collected in three separate laboratories, argues for the importance of availability as a predictor of recall and questions the role of the correlated variables of word frequency and meaningfulness. Consistent with earlier work on a smaller sample of words, six factors describe the numerous properties of words studied by psychologists. The six factors are composed of variables based on orthography, imagery and meaning, word frequency, recall, emotionality, and goodness.

What factors determine whether a particular word will be recalled by a particular subject under a specific set of experimental conditions? Over the history of the study of memory and learning, four separate aspects of the recall task have received attention (Jenkins, 1979): (1) the nature of the material to be learned and the conditions of presentation; (2) the type of response or criterion task used; (3) the activities engaged in by the learner during study and at test, including depth of processing, rehearsal, and mnemonic strategies; and (4) the characteristics of the learner, including general knowledge, skills, and motivation.

In the study presented here, we addressed the first aspect of recall tasks, the nature of materials and the conditions of presentation, by trying to predict the probability of recall of individual words when the other three aspects of the recall task were randomized or held constant. Most of our emphasis was on the material itself because here, and in other studies (Rubin, 1980, 1985), presentation effects were minimal.

Consider a typical study investigating one or more item variables, such as frequency, concreteness, or age of acquisition. Words are selected to represent different portions of the range of the variables of interest (e.g., high

frequency vs. low frequency). The recall performance of subjects learning these words is then compared. For our purposes, this form of experiment has four disadvantages. First, in selecting fixed lists of stimulus items, the researcher must deal with the confounding relationships between the variable(s) under study and other correlated variables. Holding the correlated stimulus variables constant eliminates the confound, but may produce lists of items which are not representative of the variable being studied, or which are confounded in new ways. For example, Paivio, Yuille, and Madigan (1968) noted that varying imagery while holding concreteness constant results in a set of words that varies in emotionality. Second, so many properties of the experimental procedure (e.g., presentation time, modality, etc.) and the subject population are held constant that the generality of the results may be questioned, and varying these properties in addition to the variables of interest is tedious. Third, when items are grouped into broad ranges of a stimulus variable (e.g., high, medium, low), researchers can assess whether or not the variable is related to recall, but it is difficult to assess the form of the relation or its magnitude. Fourth, many properties of words have been shown to affect recall. Comparing more than a few of these variables and their interrelations is not practical.

To surmount these difficulties, we adopted the following strategy. First, data on recall of a set of nouns were collected from a large set of experiments which employed varying presentation and study conditions. We argue that the probability that each word was recalled, averaged over all subjects and experiments, is the most representative general measure of recallability of individual items col-

---

We wish to thank Ron Collis, Pat Franklin, David Hoffman, and Kevin Mallory for their help collecting and analyzing the data. Support was provided by National Science and Research Council of Canada Grant No. A8615 to Michael Friendly and by National Science Foundation Grant No. BNS-8410124 to David C. Rubin. Reprints are available from D. C. Rubin, Psychology Department, Duke University, Durham, NC 27706.

lected to date. Second, normative data were collected on a wide variety of word properties thought to predict recall. Finally, several correlational analyses were designed to assess the relationships among the variables and the relative contributions of each of the normative scales in predicting recall.

Three selections had to be made in pursuing this approach: first, the set of words to be normed; second, the set of independent variables on which to norm the words; and, third, the recall tasks to use. Only one large sample of words has an extensive set of previously normed variables. Paivio et al.'s (1968) listing of values of concreteness (C), imagery (I), and meaningfulness (m) for 925 nouns is one of the most widely used articles in psychology, having over 900 citations, with not less than 50 in each of the last 10 years (*Social Science Citation Index*, 1974–1985). Scales available for these 925 nouns include the originally published values of C, I, and m; easily obtained frequency measures (e.g., Kučera & Francis, 1967; Thorndike & Lorge, 1944); values for first- and second-order approximations to English; orthographic neighbor ratio (Rubin, 1981); availability (Rubin, 1983); and free recall (Christian, Bickley, Tarka, & Clayton, 1978).

The choice of the set of independent variables can be guided by earlier work. Rubin (1980) obtained 51 scales for a set of 125 words which occurred both in Paivio et al.'s (1968) and Brown and Ure's (1969) norms. A factor analysis of these 51 scales produced six factors. To obtain at least one variable to represent each of the six factors found by Rubin (1980), we added emotionality and goodness to the variables already normed for the 925 nouns. In addition, the 51 scales were divided into 34 "independent" and 17 "dependent" variables, and multiple regressions were performed. To include all the variables that were noted as being good predictors in the multiple regressions, we added pronunciability. For reasons of generality, an additional measure of availability was also included. The resulting set of existing and new independent variables contains measures of orthography, pronunciability, imagery, concreteness, meaningfulness, availability, familiarity, frequency of occurrence, goodness, and emotionality, and thus most theories that predict what kinds of words should be easy to recall are represented by a measure in the set of independent variables. In multiple regressions, these variables can be compared to each other to determine which best accounts for the relative probability of recall of a word.

The Rubin (1980) study also served to demonstrate that the exact nature of the free-recall task has little effect on the relative ease of learning words in free recall. Although some tasks produced more recall than others, the relative level of recall of the 125 words changed little. This finding is important if values for the percentage of subjects recalling each word of a set of words are to be useful. Unless such a set of values can be generalized to a fairly wide range of situations, one such set of values would be of little theoretical or empirical use. The free-

recall tasks used were: free recall with overt rehearsal, using a list length of 18 words and presentation rate of 5 sec/word; the same free-recall task, but with overt rehearsal of only the word being shown; free recall of a single list of all 125 words presented at a rate of one per second; surprise free recall following a lexical decision task; and surprise free recall following a paired-associate task. Different groups of approximately 30 subjects each took part in each task. Although there were large differences in the average number of items recalled in various conditions, the correlations among the various free-recall tasks approached their reliabilities. The average correlation among all the tasks listed was .62, with an average reliability (Cronbach's alpha; Cronbach, 1951) of .70. For these tasks, reliability can be viewed as the correlation between the data actually obtained and the hypothetical data from a replication of the task using new subjects from the same subject population. Thus, the finding that the reliabilities are nearly equivalent to the correlations among the tasks implies that the tasks correlate almost as highly with each other as they would with themselves if the same tasks were repeated with new subjects.

Similar differences in how many, but not which, units were recalled were found by Rubin (1985), using three incidental learning tasks, and by Postman and Kruesi (1977) over a range of free-recall tasks that included incidental and intentional learning following ratings of pleasantness of meaning, pleasantness of sound, frequency of occurrence of the word, and average frequency of occurrence of the syllables of the word, as well as an intentional learning condition without any ratings. Over a wide range of tasks, the relative proportion of times that each word was recalled was nearly invariant within experimental error. Thus, free-recall values obtained from one specific set of experimental conditions were at least a good approximation of the values obtained from different conditions and, therefore, are of general theoretical and empirical interest (Rubin, 1985).

In choosing a recall measure, we decided to use the data from an existing set of experiments that differed from all the previously mentioned tasks in that it used multiple-trial free recall (Friendly & Franklin, 1980).

## METHOD

### Rating Scales of Goodness, Emotionality, and Pronunciability

By factor analyzing a large number of ratings of attributes of connotative meaning using the semantic differential scale, Osgood, Suci, and Tannenbaum (1957) found good-bad evaluation to be a major factor of connotative meaning. Emotionality, as measured by the absolute distance from neutral, regardless of direction in the semantic space of connotative meaning, was also important. These dimensions of goodness (Boucher & Osgood, 1969; Zajonc, 1968) and emotionality (Jenkins, 1960; Rapaport, 1942) probably have been the most studied dimensions of connotative meaning. In the present study, goodness was rated using a 7-point scale, based on Brown and Ure's (1969) scale, of how intensely good or bad a word's meaning is. The scale ranged from *bad* (1) to *average* (4) to *good* (7). Goodness is very similar to the measure that loaded the highest on

Osgood et al.'s evaluation factor. Emotionality was rated using a 7-point scale that ranged from *not emotional* (1) to *average* (4) to *emotional* (7).

Pronunciability was introduced by Underwood and Schulz (1960) as a predictor of paired-associate learning and was later scaled by Spreen and Schulz (1966) and by Ley and Karker (1974). We adapted Underwood and Schulz's 9-point rating scale and instructions for use with words, instead of with trigrams. The scale ran from *easy* (1) to *average* (5) to *hard* (9).

A fourth rating scale was calculated from one of the three variables actually measured. Emotional goodness was formed by taking the absolute distance from neutral of a word's goodness rating. The rationale for forming this scale came from Brown and Ure's (1969) definition of emotionality, as well as from the semantic differential's measure of emotionality as the absolute distance from neutral. Empirical support for the scale came from the observation of a marked curvilinear relation between emotionality and goodness found in Brown and Ure's (1969) norms: The extreme values on goodness were rated most highly on emotionality. A scale of strength of evaluation, independent of direction, was therefore formed. In mathematical terms, the equation is: emotional goodness =  $|goodness - 4|$ , where 4 is *neutral* on a 7-point scale. Thus, the measure is identical to Jenkins's (1960)  $D_a$  measure calculated over the single evaluation dimension.

Because we felt that reliable ratings could not be obtained on all 925 words in one session, we had subjects rate either one third of the words in a half-hour session or two thirds of the words in an hour session. Three random orders of the 925 words, along with the reverse of these orders, were prepared. Three booklets were made for each complete order of 925 nouns. The subjects, who were given credit toward a course requirement, received either one or two booklets, depending on the length of their sessions. When two booklets were received, each contained different words.

For scoring purposes, the booklets were combined to form complete lists of 925 nouns. Fifty-seven complete lists were obtained for each scale. Cronbach's alpha was calculated for these lists and produced values of .983, .975, and .977 for the goodness, emotionality, and pronunciability scales, respectively. The norms obtained agree with previous values: goodness, emotionality, and pronunciability correlate .934, .921, and .940 with the values reported for 125 nouns in Rubin (1980); goodness correlates .948 with unpublished ratings of pleasantness collected by Paivio (1982) on 316 overlapping words; and pronunciability correlates .924 with unpublished ratings of pronunciability collected by Paivio (1982) for 341 overlapping words.

### Availability

Availability, or associative frequency, is the converse of Noble's (1952) meaningfulness. Instead of measuring how many associates are given to a word, availability measures the number of times a word is given as an associate to a sample of stimulus words (Ervin, 1963; Howes, 1957). Thus, availability is the ease with which a word comes to mind in free association and may be viewed as a measure of general response availability in Asch and Ebenholtz's (1962), Thorndike's (1932), and Tversky and Kahneman's (1973) terms, or accessibility in Tulving and Pearlstone's (1966) terms. Availability has been calculated (Rubin, 1980, 1983) as the  $\log_{10}$  of the number of times a word is given as an associate in the Palermo and Jenkins (1964) word-association norms. Palermo and Jenkins obtained first associations to 200 stimulus words from 250 boys and 250 girls in Grades 4, 5, 6, 7, 8, 10, and 12 and from 500 male and 500 female undergraduates. If a word was given as an associate to a given stimulus word once in at least one grade-sex group, but did not occur more than once in any other grade-sex group, its exact frequency was not recorded in the Palermo and Jenkins norms and was assumed to have a frequency of 2.

In addition to the multiple regressions based on 125 words that were reported earlier, it has been shown with the 925 words used here that high-availability words are judged to be more familiar and are easier to recall but harder to recognize than are low-

availability words, even with meaningfulness, imagery, length in letters, and frequency excluded as factors. Moreover, when used as foils in a recognition experiment, high-availability words attract more responses than do low-availability words (Rubin, 1983).

The findings to date, however, are limited in that they are all based on one measure of availability calculated from the Palermo and Jenkins (1964) word-association norms. Fortunately, another associative study can provide a second measure. Kiss (1975; Kiss, Armstrong, & Milroy, 1972) provided each of 8,400 subjects from a collection of schools in Great Britain with 100 stimulus words, resulting in 840,000 responses (i.e., tokens), of which 55,000 were unique (i.e., types). With the exception of the initial, 2,000 or so stimulus words, all stimulus words were the responses of previous subjects. The resulting semantic network lists each of the 8,400 word nodes that were used as stimuli, along with the 100 responses each node produced, and the 55,000 different responses, along with the stimuli that produced them. Availability is the  $\log_{10}$  of the number of times a word was given as a response. In Appendix A, the values before the logarithms were taken are listed to allow greater precision. There are 97 zero values for the 925 words. As with previous work (Rubin, 1983), zero values were replaced by one half before logarithms were taken. If all 925 words are included, availability based on the Kiss norms correlates .777 with that based on the Palermo and Jenkins norms. If only the 457 words with nonzero values on both measures are included, the correlation is .770.

The Kiss et al. (1972) technique of using the novel responses of one set of subjects as the stimulus for the next set of subjects could be seen as a potential problem for the Kiss availability measure introduced here. If a word was used as a stimulus early in the process of forming an expanding semantic network, that word would introduce new nodes into the network which would be likely to add to the original word's availability. Words added later in the process would not have this opportunity, because responses to these later words would not be used as stimuli. This potential problem is extremely difficult to assess empirically because the time of entry of a word into the network is confounded with availability. This same confounding, however, provides a theoretical rationale for the measure of availability obtained. The method of expansion of a semantic network is a good model of how a semantic network might be searched, or activated, to produce an availability measure: Rather than starting at a new node each time, the search, or activation, would go from node to node. Nodes never reached, or those reached only rarely, would have less effect than nodes commonly accessed. Thus, the potential problem caused by Kiss et al.'s expansion method is not seen as a limitation for the measure introduced here.

### Multiple-Trial Free Recall

Although Rubin's (1980) study provided several free-recall measures for 125 words, only one free-recall measure exists for a large sample of words. Christian et al. (1978) provided values for 899 of Paivio's nouns, each value based on the immediate recalls of 32 subjects, who were presented with lists of 25 words, one word each 6 sec. The initial and final 5 words of the list were unscored buffers. This free-recall measure has a reliability of .57.

New measures of free recall, based on Friendly and Franklin's (1980) multiple-trial free-recall experiments, are included here. The percentage of times that subjects recalled these words on the first three trials and the number of subjects involved in each estimate are reported. The data reported came from 13 different free-recall experiments run in a common environment over a 5-year period. In all, over 500 students served as subjects. The database consists of 16,932 subject-items from each of three trials. The basic parameters of these experiments are summarized in Table 1.

**Apparatus.** Although the aims and designs of these experiments varied, certain features of the method were common to all of them. All of the experiments were run under computer control using a standard CRT display terminal linked to an interactive mainframe (IBM/370 or DEC-10). The computer program used in this research was derived from a general software system developed for running memory experiments (Friendly & Franklin, 1979). The computer

**Table 1**  
**Summary of the Multitrial Free-Recall Experiments**

| Experiment | S   | W  | T | OBS   | I   | FR1  | FR2  | FR3  | FR   |
|------------|-----|----|---|-------|-----|------|------|------|------|
| 1          | 69  | 20 | 6 | 1,380 | 678 | .438 | .662 | .780 | .627 |
| 2          | 30  | 20 | 5 | 600   | 421 | .475 | .650 | .818 | .648 |
| 3          | 41  | 20 | 6 | 820   | 500 | .438 | .645 | .772 | .618 |
| 4          | 52  | 30 | 4 | 1,560 | 704 | .388 | .643 | .766 | .599 |
| 5          | 39  | 40 | 5 | 1,560 | 707 | .350 | .598 | .764 | .571 |
| 6          | 64  | 40 | 6 | 2,560 | 767 | .329 | .559 | .723 | .537 |
| 7          | 32  | 40 | 6 | 1,280 | 657 | .327 | .561 | .698 | .529 |
| 8          | 41  | 40 | 4 | 1,640 | 744 | .310 | .530 | .680 | .507 |
| 9          | 12  | 36 | 5 | 432   | 347 | .336 | .540 | .678 | .518 |
| 10         | 28  | 36 | 5 | 1,008 | 566 | .317 | .542 | .688 | .516 |
| 11         | 32  | 36 | 5 | 1,152 | 580 | .306 | .507 | .636 | .483 |
| 12         | 100 | 25 | 3 | 2,500 | 379 | .308 | .458 | .550 | .439 |
| 13         | 20  | 22 | 4 | 440   | 344 | .491 | .727 | .875 | .698 |

Note—*S* = number of subjects; *W* = number of words shown to each subject; *T* = number of trials (in this paper, only data from the first three trials are used); *OBS* = number of subject-items per trial, or *S* \* *W*; *I* = number of items used at least once; and *FR1*, *FR2*, *FR3*, and *FR* are the probabilities of recall in Trial 1, Trial 2, Trial 3, and all three trials combined.

system was designed to be completely driven by data and tables in external files. These files contained the words and experimental parameters, including selection parameters for the items, presentation rate, numbers of words, and number of trials.

For a given subject, the control program selected a list of items from the word pool as described below. For each trial of the experiment, the control program prepared an appropriate presentation order, displayed the items on the CRT screen, and accepted the subject's typed recall responses. The responses were scored interactively using a spelling correction algorithm to minimize the influence of spelling errors.

**Item Selection.** Items were sampled from the Paivio et al. (1968) norms. Ratings for imagery, concreteness, meaningfulness, frequency of occurrence, number of letters, and number of syllables were the six selection variables. The word list for each subject was a random sample of all words in the pool that fell within the range specified for each of the selection variables. The selection criteria, which were constant for each experiment, usually imposed restrictions on only two variables in the word pool: word frequency and length in letters. These served to eliminate long words (greater than 10-12 letters) and very infrequent words (less than 2-4 per million). As a result of these restrictions, in a given experiment, between 60 and 100 of the 925 words in the Paivio et al. list were eliminated from the sampling procedure.

The number of words actually used in a given experiment reflected random sampling within these limits. Variable *I* in Table 1 represents the number of items from the pool of 925 words which were selected for one or more subjects in the study. Since the sampling unit here is considered to be one subject-item, the number of observations (*OBS*) for one trial of an experiment is the product of subjects (*S*) and number of words (*W*). Thus, in Experiment 1, a total of 69 subjects each received 20 words for a total of 1,380 subject-items observations for each of three trials.

**Procedure.** The experiments in Table 1 were designed to investigate the effects of interactive presentation and recall conditions intended to enhance recall relative to the effects of standard free-recall conditions. In interactive presentation, the computer system scores each recall trial immediately, and presents the items to the subject on the next trial in a way that depends on what items the subject recalled on previous trials.

Except where noted below, items were individually presented to subjects on the computer screen for 2 sec each. In order to reduce

recall trial immediately, and presents the items to the subject on the next trial in a way that depends on what items the subject recalled on previous trials.

Except where noted below, items were individually presented to subjects on the computer screen for 2 sec each. In order to reduce recall from immediate memory, subjects added two-digit numbers for 15 sec following presentation of the last item. Subjects were allowed 8 sec/item to type all the words they could recall on each test trial. The first trial always used a random order of presentation. All subjects were given two practice trials on a short list to familiarize themselves with the procedures before starting the main experiment.

Experiments 1-5 have the same basic design as that described in Friendly and Franklin (1980), and differed only in the number of words and number of trials of learning. The design for these studies was a 2x3 factorial with the six independent groups representing a two-level "tagging" manipulation combined with three presentation orders. The tag factor represented whether or not nonrecalled words from the previous trial were printed with an asterisk (\*) to make them distinctive. In the presentation order conditions, words were presented either randomly, or all previously recalled items were presented in their previous recall order at the beginning (B), or in their previous recall order, meshed with the previously nonrecalled items. In the results presented in the Friendly and Franklin study, a few subjects were randomly eliminated to achieve equal numbers of subjects in all conditions. The present analyses made use of the original data files containing all available data.

Experiments 6, 7, and 8 are three unpublished studies designed to examine the effects of individual item presentation versus massed presentation for the same total study time, and the effects of various conditions of recall. Experiment 6 manipulated three factors in a 2 x 2 x 2 design: single item versus massed presentation, random presentation versus B plus tag, and single item recall versus 2D recall. The 2D recall was a modification of Buschke's (1973) procedure, in which subjects were asked to recall items in groups, according to which words "went together in their memory for the list." In the computer implementation, subjects typed all the words that they remembered as a group on one line, before going on to the next subjective unit.

Experiments 7 and 8 each used four groups in a 2x2 design: single versus massed presentation of items and standard recall versus 2D recall. In Experiment 7, recall was typed as usual; in Experiment 8, recall was given orally and typed by the experimenter.

Experiments 9, 10, and 11 investigated the effects of presenting items to a subject in a manner consistent with his or her own subjective memory units. For each experiment, we used the same design with three conditions. The random group received items one by one in a random order and recalled items one by one. Two other groups each recalled their memory groups with a 2D procedure. Subjects in one of those two groups were presented with their own recall units, all items in a memory group presented on the screen together. Items were assigned to memory groups using all previous recall trials. Subjects in the second group also received items presented in massed groups, but items were assigned to groups at random. The presentation time was 2.5 sec/item in these three experiments.

Experiment 13 contrasted a random condition with a B-plus-tag condition. In this experiment, however, subjects were told that they must recall all items in the list perfectly on two trials to complete the experiment. All subjects required at least four trials to reach this criterion.

As a group, the 13 experiments provide a varied collection of multiple-trial free-recall tasks. If the results were to differ from other recall studies, it would be difficult to know exactly what caused the differences. If, on the other hand, the results were similar to single-trial recall studies, some generality over the details of procedures would be gained.

## RESULTS

The Appendix presents values for Paivio et al.'s (1968) 925 nouns for goodness (G), emotionality (E), emotional goodness (EG), pronunciability (PR), availability based on Kiss et al.'s (1972) semantic network (KA), probability of recall on each of the first three trials of the multiple-trial free-recall experiments (FR1, FR2, FR3), the average probability of recall on the first three trials of the multiple-trial free-recall experiments (FR), and the combined average free recall from the first three trials of the multiple-trial free-recall and the Christian et al. (1978) recall experiments weighted by the number of attempted recalls of each word (FRA). That is,  $FRA = (3 \times NS \times FR + 32 \times FRC) / (3 \times NS + 32)$ , where NS is the number of subjects recalling each word in the three trials that led to the FR value, and 32 is the number of subjects in the FRC experiment. Table 2 lists and Table 3 presents correlations among both the newly collected and the previously published scales. Here and in all future analyses, in order to ensure stability in the values for FR, and FR1, FR2, and FR3, these values are included for a word only if at least 10 subjects attempted recalls on that word (i.e.,  $NS > 9$ ). For FR, this means that there are always 30 or more observations per word, a value similar to FRC's 32 observations per word. Similarly, FRA is included only if FR, FRC, or both are included. The correlations follow the pattern of earlier work on more limited samples (e.g., Rubin, 1980).

## A Factor Analysis

The principal-factor orthogonal, varimax procedure used by previous researchers in this area (Frincke, 1968; Paivio, 1968; Rubin, 1980; Wimer, 1963) was chosen to reduce the  $21 \times 21$  correlation matrix shown in Table 3 to an interpretable summary. The claim that is being made for this factor analysis is that variables that load on the same factor are related; in a hierarchical clustering solution they would cluster, and in a multidimensional scaling solution they would be near one another. The factor analysis solution presented in Table 4 is flawed on technical grounds, but still emerges when these flaws are corrected by removing variables so that no correlations greater than .85 remain and by removing all variables that were formed by summing other variables.

Table 4 is similar to the factor analysis that resulted when 51 properties of 125 words were factor analyzed (Rubin, 1980) and is also in good agreement with other earlier work (Paivio, 1968; Wimer, 1963). Five of the six factors found by Rubin (1980) are present. The sixth factor, goodness, is not present in Table 4 because only one variable that should load on this factor was included here. This variable of goodness does not load substantially on any of the five factors shown and has a low communality, indicating that if another related measure, such as rated pleasantness, were included, the earlier factor analysis would probably be completely matched. A factor analysis, based on the 800 words remaining when the 125 words used by Rubin (1980) were excluded, resulted

**Table 2**  
**Variables Included in the Analyses**

| Variable                                | Abbreviation | Source                  |
|---|--------------|-------------------------|
| Length in Letters*                      | L*           |                         |
| 1st-Order Approximation to English      | FOA          | Rubin, 1981             |
| 2nd-Order Approximation to English      | SOA          | Rubin, 1981             |
| Pronunciability*                        | Pr*          | Appendix A              |
| Imagery                                 | I            | Paivio et al., 1968     |
| Concreteness                            | C            | Paivio et al., 1968     |
| Meaningfulness                          | M            | Paivio et al., 1968     |
| Availability based on Kiss              | KA           | Appendix A              |
| Availability based on Palermo & Jenkins | A            | Rubin, 1983             |
| Familiarity                             | F            | Paivio, 1982            |
| K-F Frequency                           | KFF          | Kučera & Francis, 1967  |
| T-L Frequency                           | TLF          | Thorndike & Lorge, 1944 |
| Goodness                                | G            | Appendix A              |
| Emotionality                            | E            | Appendix A              |
| Emotional Goodness                      | EG           | Appendix A              |
| Free Recall                             | FR           | Appendix A              |
| Free-Recall Trial 1                     | FR1          | Appendix A              |
| Free-Recall Trial 2                     | FR2          | Appendix A              |
| Free-Recall Trial 3                     | FR3          | Appendix A              |
| Free-Recall CBTC                        | FRC          | Christian et al., 1978  |
| Free Recall Averaged                    | FRA          | Appendix A              |

*Note*—The scales followed by an \* have had their values multiplied by  $-1$  to make the positively evaluated direction of all variables the mathematically positive direction. All variables have values for 925 words except the recall variables. FRC has values for 899 words. The restriction that at least 10 subjects had to be presented with a word for that word to be included in calculations resulted in FR, FR1, FR2, and FR3's having values for 820 words and FRA's having values for 921 words.

Table 3  
Correlations Among Variables

| Variable | L*  | FOA | SOA | Pr* | I   | C   | M   | KA | A  | F   | KFF | TLF | G   | E   | EG | FR | FR1 | FR2 | FR3 | FRC |  |
|----------|-----|-----|-----|-----|-----|-----|-----|----|----|-----|-----|-----|-----|-----|----|----|-----|-----|-----|-----|--|
| FOA      | 97  |     |     |     |     |     |     |    |    |     |     |     |     |     |    |    |     |     |     |     |  |
| SOA      | 92  | 93  |     |     |     |     |     |    |    |     |     |     |     |     |    |    |     |     |     |     |  |
| Pr*      | 79  | 79  | 77  |     |     |     |     |    |    |     |     |     |     |     |    |    |     |     |     |     |  |
| I        | 35  | 32  | 28  | 43  |     |     |     |    |    |     |     |     |     |     |    |    |     |     |     |     |  |
| C        | 30  | 28  | 24  | 32  | 83  |     |     |    |    |     |     |     |     |     |    |    |     |     |     |     |  |
| M        | 33  | 32  | 29  | 46  | 72  | 56  |     |    |    |     |     |     |     |     |    |    |     |     |     |     |  |
| KA       | 55  | 55  | 57  | 66  | 44  | 24  | 50  |    |    |     |     |     |     |     |    |    |     |     |     |     |  |
| A        | 50  | 50  | 51  | 61  | 43  | 29  | 46  | 78 |    |     |     |     |     |     |    |    |     |     |     |     |  |
| F        | 27  | 28  | 31  | 55  | 34  | 10  | 48  | 74 | 63 |     |     |     |     |     |    |    |     |     |     |     |  |
| KFF      | 27  | 30  | 34  | 46  | 09  | -06 | 28  | 68 | 57 | 75  |     |     |     |     |    |    |     |     |     |     |  |
| TLF      | 41  | 43  | 46  | 59  | 33  | 16  | 43  | 78 | 66 | 76  | 77  |     |     |     |    |    |     |     |     |     |  |
| G        | 04  | 05  | 06  | 13  | 07  | -02 | 13  | 18 | 20 | 26  | 26  | 28  |     |     |    |    |     |     |     |     |  |
| E        | -05 | -06 | -05 | -05 | -19 | -51 | -08 | 06 | 00 | 04  | 08  | 05  | -02 |     |    |    |     |     |     |     |  |
| EG       | 00  | -02 | -01 | 04  | -02 | -30 | 08  | 12 | 08 | 11  | 05  | 06  | -11 | 78  |    |    |     |     |     |     |  |
| FR       | 19  | 18  | 16  | 14  | 35  | 32  | 26  | 20 | 26 | 01  | 00  | 06  | 03  | -01 | 06 |    |     |     |     |     |  |
| FR1      | 15  | 14  | 14  | 09  | 27  | 26  | 19  | 13 | 19 | -05 | -04 | 00  | 03  | 02  | 06 | 84 |     |     |     |     |  |
| FR2      | 17  | 16  | 13  | 11  | 31  | 29  | 24  | 19 | 25 | 06  | -01 | 01  | 02  | 01  | 09 | 89 | 62  |     |     |     |  |
| FR3      | 16  | 17  | 13  | 14  | 30  | 29  | 24  | 20 | 23 | 12  | 00  | 01  | 02  | -01 | 00 | 82 | 51  | 62  |     |     |  |
| FRC      | 27  | 25  | 23  | 26  | 44  | 40  | 31  | 31 | 38 | 13  | 09  | 17  | 09  | 00  | 09 | 50 | 45  | 44  | 38  |     |  |
| FRA      | 28  | 27  | 24  | 26  | 50  | 44  | 38  | 32 | 38 | 12  | 07  | 18  | 08  | 01  | 11 | 90 | 78  | 80  | 72  | 81  |  |

Note—Leading decimal points omitted.

in the same five factors shown in Table 4, indicating that the similarity in results between the two studies is not due to the overlap in some of the stimuli used.

In addition to being in agreement with earlier work and possible expectations of researchers in this field, the factor structure is extremely simple. With the exception of availability and pronounciability, variables load highly on one, and only one, factor. This implies that most of our commonly used variables can be collapsed into five (or more likely, six) factors or clusters and that any one selected variable will reasonably represent the other variables from a factor.

Factor 1 has its highest loadings from variables that could be formed without regard to whether or not the strings of letters scaled were words. Factor 2 consists of

variables related to frequency of occurrence, and Factor 3 contains the recall variables. Factor 4 contains both imagery and meaningfulness variables. Although this combination is somewhat unsatisfying, it is one with some empirical precedent (Paivio, 1968; Rubin, 1980). Factor 5 contains the emotionality variables and also has a moderate negative loading on concreteness, a finding hinted at in Rubin's (1980) study. As noted by Paivio et al. (1968), emotional words can be of any imagery value, but tend to be of low concreteness.

Multiple Regressions

For both the 925-word sample and 125-word sample used by Rubin (1980), multiple regressions were performed on four measures of free recall available for Pai-

Table 4  
A Factor Analysis

| Variable             | 1   | 2   | 3   | 4   | 5   | C  |
|----------------------|-----|-----|-----|-----|-----|----|
| Length in Letters*   | 12  | 95  | 14  | 13  | -02 | 95 |
| First Order Approx.  | 11  | 95  | 16  | 10  | -04 | 96 |
| Second Order Approx. | 09  | 93  | 20  | 06  | -02 | 92 |
| Pronunciability*     | 05  | 76  | 41  | 27  | 02  | 82 |
| Imagery              | 26  | 17  | 14  | 88  | -06 | 90 |
| Concreteness         | 26  | 17  | -05 | 81  | -39 | 90 |
| Meaningfulness       | 16  | 15  | 34  | 76  | 06  | 74 |
| Kiss Availability    | 14  | 44  | 73  | 27  | 12  | 83 |
| Availability         | 22  | 38  | 65  | 28  | 06  | 70 |
| Familiarity          | -05 | 14  | 86  | 25  | 10  | 82 |
| K-F Frequency        | -03 | 18  | 88  | -05 | 06  | 81 |
| T-L Frequency        | 00  | 30  | 83  | 18  | 06  | 82 |
| Goodness             | 08  | -09 | 51  | -13 | -22 | 34 |
| Emotionality         | 02  | -03 | 05  | -21 | 91  | 87 |
| Emotional Goodness   | 07  | -01 | 02  | 02  | 94  | 89 |
| Free Recall          | 98  | 06  | 02  | 08  | -01 | 96 |
| Free-Recall Trial 1  | 83  | 06  | -04 | 04  | 02  | 71 |
| Free-Recall Trial 2  | 86  | 05  | 02  | 08  | 03  | 76 |
| Free-Recall Trial 3  | 79  | 05  | 08  | 07  | -08 | 65 |
| Free-Recall CBTC     | 60  | 15  | 09  | 35  | 08  | 51 |
| Free Recall Averaged | 92  | 13  | 09  | 27  | 06  | 96 |

Note—Leading decimal points omitted.

vio's nouns. In both cases, all available predictor variables were included, with the exception of those that correlated greater than .85 with other predictor variables. This resulted in 13 predictor variables for the 925-word sample (the first 15 variables listed in Table 4, less first- and second-order approximations to English) and 25 predictor variables for the 125-word sample. In addition, one free-recall measure, Free Recall Grouped, was added from the Rubin (1980) study. This variable is the percentage of 214 subjects recalling each word when a collection of six incidental and intentional learning tasks are combined. Thus, for the 925-word sample, there are two completely independent recall measures (Free Recall and Free-Recall CBTC) and two recall measures partially confounded with these (Free-Recall Trial 1 and Free Recall Average). The 125-word sample has an additional independent recall measure (Free Recall Grouped). Multiple regressions for Free-Recall Trials 2 and 3 are not shown because, as can be seen from the correlations in Table 3, they differed little from the regressions for Free Recall and Free-Recall Trial 1. The simple correlations given in Table 3 indicate the ability of independent variables to predict recall in isolation. The multiple regressions indicate the ability of independent variables to predict recall when the confounding effects of other variables are removed. That is, the multiple regressions can be viewed as analyses in which many variables are simultaneously compared with each other to determine which variables can best predict the probability of recall of words. Combined, these two types of analyses contain a wealth of information when compared to analysis of variance designs containing only a few independent variables.

Table 5 presents the multiple regression. The order of entry of the variables in a stepwise regression is given by the numbers 1-9. The multiple regressions shown result from the restriction that each variable, when it is added, must have a beta weight significantly different from zero ( $p < .05$ ) under the liberal assumption that the words are independent. The multiple regression correlation coefficients resulting when this restriction is removed are shown in the last column

The results are remarkably similar for all measures in both samples, as are the results for the sample of

800 words consisting of those 925 words used here that were not used in Rubin's (1980) study. Three kinds of variable account for most of the variance explained in all regressions: imagery, as measured by I or C; availability as measured by A or KA; and emotionality, as measured by E or EG. For the more sensitive 925-word sample, goodness also has a small effect. Frequency, as measured by F, or TLF, also enters for the 925-word sample, but it appears to do so as a correction. Frequency's beta weight, from the first step in which it enters to the last step shown in Table 5, is always negative, although its simple correlation with recall is small but positive. Finding three basic determiners of which words are recalled is noteworthy, considering the differences in procedures in the various recall measures and the multitude of variables available to enter into the regression equations.

The similarity among all the multiple regressions, as well as the high correlations among the various recall measures, indicates that the words that are recalled in multiple-trial free recall tend to be the same words that are recalled in single-trial recall, and therefore, that the free-recall measures introduced here are general ones. That is, the results presented in Table 5 hold for different samples of words, for different samples of subjects, and across widely different procedures and experimental conditions.

### DISCUSSION

What have we learned about free recall? First, there appear to be three major predictors of which words tend to get recalled: imagery, availability, and emotionality. Measures such as meaningfulness, frequency, and pronounciability, which we might have expected, based on the classical literature, to be among the important predictors, were not, although they were included in the analyses.

Although imagery, and to a lesser extent emotionality, might have been expected to function as they did, except for some recent research (Rubin, 1980, 1983), availability appears as a surprise. The unexpected results argue that at our science's current level of development, exploratory work is still warranted to prevent us from holding to assumptions based on a less than adequate theory or data base. For instance, the early studies supporting

**Table 5**  
**Beta Weights from Stepwise Multiple Regressions**

| Sample | Dependent Variable   | Independent Variables |       |        |          |       |         |         |         |       | Shown | All |
|--------|----------------------|-----------------------|-------|--------|----------|-------|---------|---------|---------|-------|-------|-----|
|        |                      | 1                     | 2     | 3      | 4        | 5     | 6       | 7       | 8       | 9     |       |     |
| 925    | Free Recall          | .09I                  | +.24A | -.23F  | +.15EG   | +.23C | +.19KA  | -.15TLF | +.07G   |       | .46   | .47 |
|        | Free-Recall Trial 1  | .06I                  | -.25F | +.24A  | +.18E    | +.24C | -.17TLF | +.16KA  | +.08G   |       | .41   | .42 |
|        | Free-Recall CBTC     | .16I                  | +.29A | -.21F  | +.16EG   | +.23C | +.10G   | +.19KA  | -.15TLF |       | .55   | .56 |
|        | Free Recall Averaged | .21I                  | +.26A | -.24F  | +.11EG   | +.28C | +.18KA  | +.09G   | -.11TLF | +.11E | .60   | .60 |
| 125    | Free Recall          | .57A                  | +.46I | +.24EG | -.34AOA* |       |         |         |         |       | .69   | .78 |
|        | Free-Recall Trial 1  | .40A                  | +.46I | +.34EG | -.32AD*  |       |         |         |         |       | .57   | .72 |
|        | Free-Recall CBTC     | .38A                  | +.39I | +.24E  |          |       |         |         |         |       | .71   | .76 |
|        | Free Recall Averaged | .39A                  | +.44I | +.28EG |          |       |         |         |         |       | .78   | .83 |
|        | Free Recall Grouped  | .32A                  | +.36E | +.36I  | -.19Reh  |       |         |         |         |       | .62   | .72 |

Note—AOA\* is the rated age of acquisition of the word multiplied by -1 (Rubin, 1980). AD\* is the rated associative difficulty of the word multiplied by -1 (Rubin, 1980). Reh is the number of times the word was rehearsed in 5 sec (Rubin, 1980).

meaningfulness, frequency, and pronunciability may have been due in part to these variables confounding with availability and imagery. Similarly, availability's effect may be due to its confounding with a yet-to-be-discovered factor. Large correlational studies, such as this one, provide one way to check for such possibilities. Although the results are somewhat unexpected, they are robust. Recall from two samples of words and three laboratories support the notion that imagery, availability, and emotionality are the best predictors of which words are recalled.

Emotionality presents something of a puzzle. In all multiple regressions, a measure of emotionality enters as the third (after imagery and availability) or fourth (after imagery, availability, and frequency) variable. In multiple regressions, the emotionality variables account for considerable variance. Nonetheless, as can be seen in Table 4, the simple correlations between the six recall measures and the two emotionality measures (E and EG) are quite small, ranging from  $-.01$  to  $.11$ . The only variable with a correlation above  $.1$  with both measures of emotionality is concreteness ( $-.51$  with E and  $-.30$  with EG). This is consistent with Paivio et al.'s (1968) report as well as with the factor analysis shown in Table 5. Thus, emotionality has no simple isolated effect on recall and does not enter into multiple regressions because of a clear suppressor effect on another variable that enters ahead of it. The fact that emotionality is uncorrelated with almost all the other independent variables allows any small effects it has to be noticed. If emotionality entered into only one regression equation, this puzzle would probably be resolved by viewing emotionality's predictive powers as a chance occurrence. Its persistent effect over different recall experiments and different samples of words, however, makes it difficult to dismiss.

Availability, because of its relative lack of use as a predictor of recall deserves special attention. Availability, as measured by associative frequency, is the number of times a word is given as an associate to a sample of stimulus words. That is, it is the relative ease with which an isolated word comes to mind. In a study contrasting 34 independent variables (Rubin, 1980), it proved to be an excellent predictor of recall. In a more conventional analysis of variance design, 64 high-availability words were easier to recall than 64 low-availability words, even when both sets of words were equated for meaningfulness, imagery and length in letters, and when differences in frequency of occurrence were minimized (Rubin, 1983).

Although availability, as measured here, has a limited history as an independent variable in memory experiments, it has a long history in psychology. For instance, it is well known that the ease with which words come to mind in specific domains or categories affects recall (Asch & Ebenholtz, 1962; Dale, 1967; Deese, 1965; Leicht, 1968; Tversky & Kahneman, 1973). A general measure of availability, different from the frequency with which a word comes to mind in general usage (i.e., word fre-

quency; Howes, 1957), probably was not developed because of the general assumption that associations were essentially symmetrical and, thus, meaningfulness and availability were the same measure. Nonetheless, as a theoretical construct, availability has been used widely. The measure of availability used here was proposed, but not actually measured, by Underwood (1965) using the term "implicit associative response" and by Glanzer and Bowles (1976) to explain the word frequency effect in recognition. In more modern language, new views would hold that high-availability words are frequently activated through spreading activation. The term "availability," without the associative frequency measure used here, has been used in the same sense by Asch and Ebenholtz (1962), Thorndike (1932), and Tversky and Kahneman (1973). Tulving and Pearlstone (1966) also make use of a similar concept called accessibility which they contrast to their concept of availability.

In summary, although availability as measured by associative frequency has a limited history as a predictor of recall, it has been extremely successful when tested and has a wealth of theoretical and general empirical support. Moreover, it is hard to find a simpler, more obvious hypothesis: Some words come to mind easily; these words will be recalled (but not recognized) easily. Availability deserves serious consideration in future work.

A presupposition is contained in the question of what variables predict the memorability of words. It is necessary that certain words are recalled more often than others. The data reported here for the recall of 925 nouns from two laboratories using different procedures, as well as data from other studies (e.g., Postman & Kruesi, 1977; Rubin, 1980, 1985), indicate that the presupposition is valid for a wide range of tasks. That is, the data indicate that the results reported here should hold for a class of free-recall experiments. Careful analysis of data collected from a larger set of subjects than were used here will most likely show interesting differences in which words tend to be recalled under various conditions, but the data reported here indicate that, over a range of tasks, such differences will be small. The values reported here, therefore, can be used as approximations to the relative values that would be obtained in other experiments, and the predictors that work best here will probably generalize to new situations.

How should the finding that imagery, availability, and emotionality are the best predictors of a word's probability of recall be viewed? The simplest way is to combine the existing theories of the effects of imagery, availability, and emotionality on recall. Such theories work well together in that they tend to describe different aspects of behavior, and the combination fits the spirit of the multiple regression analyses performed here. A more ambitious solution would be to provide a single combined predictor or a theoretically motivated way to combine existing predictors. The data that need to be accounted for are in the appendix along with measures based on existing theories. The challenge is clear. Although a success-



ful post hoc solution for these 925 nouns would need to be validated on a new set of words, the data presented will allow inadequate solutions to be efficiently removed.

The norms themselves should not be overlooked. In addition to containing Free Recall Averaged (FRA), which is the most general set of recall values available for a large set of nouns, the norms, when combined with other published studies, provide values for the extensive list of variables presented in Table 2. Using correlational techniques, the norms can be used to examine other tasks besides recall, and the large number of words normed can aid in selecting levels of some variables while controlling for others in analysis of variance studies.

## REFERENCES

- ASCH, S. E., & EBENHOLTZ, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society*, **106**, 135-163.
- BOUCHER, J., & OSGOOD, C. E. (1969). The pollyanna hypothesis. *Journal of Verbal Learning & Verbal Behavior*, **8**, 1-8.
- BROWN, W. P., & URE, D. M. J. (1969). Five rated characteristics of 650 word association stimuli. *British Journal of Psychology*, **60**, 232-249.
- BUSCHKE, H. (1973). Selective reminding for analysis of memory and learning. *Journal of Verbal Learning & Verbal Behavior*, **12**, 543-550.
- CHRISTIAN, J., BICKLEY, W., TARKA, M., & CLAYTON, K. (1978). Measures of free recall of 900 English nouns: Correlations with imagery, concreteness, meaningfulness and frequency. *Memory & Cognition*, **6**, 379-390.
- CRONBACH, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**, 297-334.
- DALE, H. C. A. (1967). Response availability and short-term memory. *Journal of Verbal Learning & Verbal Behavior*, **6**, 47-48.
- DEESE, J. (1965). *The structure of associations in language and thought*. Baltimore: Johns Hopkins Press.
- ERVIN, S. M. (1963). Correlates of associative frequency. *Journal of Verbal Learning & Verbal Behavior*, **1**, 422-431.
- FRIENDLY, M., & FRANKLIN, P. (1979). Computer control of memory experiments on a large-scale timesharing system. *Behavioral Research Methods & Instrumentation*, **11**, 212-217.
- FRIENDLY, M., & FRANKLIN, P. (1980). Interactive presentation in multitrial free recall. *Memory & Cognition*, **8**, 265-270.
- FRINCKE, G. (1968). Word characteristics, associative-relatedness, and the free-recall of nouns. *Journal of Verbal Learning & Verbal Behavior*, **7**, 366-372.
- GLANZER, M., & BOWLES, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 21-31.
- HOWES, D. A. (1957). On the relation between the probability of a word as an association and in general linguistic usage. *Journal of Abnormal & Social Psychology*, **54**, 75-85.
- JENKINS, J. J. (1960). Degree of polarization and scores on the principal factors for concepts in the semantic atlas study. *American Journal of Psychology*, **73**, 274-279.
- JENKINS, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 429-446). Hillsdale, NJ: Erlbaum.
- KISS, G. R. (1975). An associative thesaurus of English: Structural analysis of a large relevance network. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory*. London: Wiley.
- KISS, G. R., ARMSTRONG, C. A., & MILROY, R. (1972). *An associative thesaurus of English (microfilm version)*. Wakefield: E. P. Microforms.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LEICHT, K. L. (1968). Recall and judged frequency of implicitly occurring words. *Journal of Verbal Learning & Verbal Behavior*, **7**, 918-923.
- LEY, R., & KARKER, J. (1974). Pronunciability ratings of 319 CVCVC words and paralogs previously assessed for meaningfulness and associative reaction time. *Bulletin of the Psychonomic Society*, **3**, 421-424.
- NOBLE, C. E. (1952). An analysis of meaning. *Psychological Review*, **59**, 421-430.
- OSGOOD, C. E., SUCL, G. J., & TANNENBAUM, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- PAIVIO, A. (1968). A factor-analytic study of word attributes and verbal learning. *Journal of Verbal Learning & Verbal Behavior*, **7**, 41-49.
- PAIVIO, A. (1982). *Imagery ratings and other norms for 2448 words*. Unpublished manuscript, University of Western Ontario.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, **76**(1, Pt. 2).
- PALERMO, D. S., & JENKINS, J. J. (1964). *Word association norms*. Minneapolis: University of Minnesota Press.
- POSTMAN, L., & KRUESI, E. (1977). The influence of orienting tasks on the encoding and recall of words. *Journal of Verbal Learning & Verbal Behavior*, **16**, 353-369.
- RAPAPORT, D. (1942). *Emotions and memory*. Baltimore: Williams & Wilkins.
- RUBIN, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning & Verbal Behavior*, **19**, 736-755.
- RUBIN, D. C. (1981). First order approximation to English, second order approximation to English, and orthographic neighbor ratio norms for 925 nouns. *Behavior Research Methods & Instrumentation*, **13**, 713-721.
- RUBIN, D. C. (1983). Associative asymmetry, availability, and retrieval. *Memory & Cognition*, **11**, 83-92.
- RUBIN, D. C. (1985). Memorability as a measure of processing: A unit analysis of prose and list learning. *Journal of Experimental Psychology: General*, **114**, 213-238.
- Social science citation index* (Pt. 2), *Citation index, 1971 to 1984 Annuals*. (1974-1985). Philadelphia: Institute for Scientific Information.
- SPREEN, O., & SCHULZ, R. W. (1966). Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning & Verbal Behavior*, **5**, 459-468.
- THORNDIKE, E. L. (1932). *The fundamentals of learning*. New York: Bureau of Publications, Teachers College.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers College Press.
- TULVING, E., & PEARLSTONE, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, **5**, 381-391.
- TVERSKY, A., & KAHNEMAN, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, **5**, 207-232.
- UNDERWOOD, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, **70**, 122-129.
- UNDERWOOD, B. J., & SCHULZ, R. W. (1960). *Meaningfulness and verbal learning*. Chicago: Lippincott.
- WIMER, C. (1963). An analysis of semantic stimuli factors in paired-associate learning. *Journal of Verbal Learning & Verbal Behavior*, **1**, 397-407.
- ZAJONC, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality & Social Psychology Monograph Supplement*, **9**(2, Pt. 2), 1-27.













| WORD        | KA   | G    | E    | EG   | PR   | NS | FRI | FR2  | FR3  | FR  | FRA |
|-------------|------|------|------|------|------|----|-----|------|------|-----|-----|
| WHEAT       | 165  | 4.45 | 1.88 | .45  | 2.41 | 27 | .33 | .74  | .70  | .59 | .56 |
| WHOLESALE   | 0    | 3.84 | 2.05 | .16  | 4.88 | 29 | .34 | .59  | .69  | .54 | .46 |
| WIFE        | 370  | 5.30 | 4.61 | 1.30 | 1.79 | 18 | .72 | .94  | .89  | .85 | .83 |
| WIGWAM      | 5    | 4.00 | 2.36 | .00  | 4.41 | 14 | .57 | .64  | .93  | .71 | .58 |
| WINDOW      | 564  | 4.38 | 2.50 | .38  | 3.18 | 24 | .54 | .63  | .83  | .67 | .61 |
| WINE        | 543  | 4.82 | 4.18 | .82  | 1.88 | 9  | .67 | .78  | 1.00 | .81 | .66 |
| WINTER      | 130  | 4.52 | 4.30 | .52  | 2.59 | 12 | .33 | .50  | .83  | .56 | .49 |
| WISTFULNESS | 0    | 3.98 | 4.21 | .02  | 6.21 | 14 | .43 | .57  | .93  | .64 | .55 |
| WOMAN       | 1042 | 5.41 | 5.11 | 1.41 | 3.07 | 9  | .67 | 1.00 | .89  | .85 | .81 |

| WORD      | KA  | G    | E    | EG   | PR   | NS | FRI | FR2 | FR3  | FR  | FRA |
|-----------|-----|------|------|------|------|----|-----|-----|------|-----|-----|
| WOODS     | 78  | 4.89 | 3.80 | .89  | 2.59 | 11 | .45 | .82 | .55  | .61 | .55 |
| WORKHOUSE | 8   | 2.80 | 2.34 | 1.20 | 3.77 | 7  | .43 | .71 | 1.00 | .71 | .53 |
| WORLD     | 368 | 4.95 | 4.25 | .95  | 3.25 | 15 | .53 | .73 | .80  | .69 | .60 |
| YACHT     | 97  | 4.80 | 3.36 | .80  | 5.23 | 16 | .25 | .69 | .81  | .58 | .54 |

(Manuscript received April 18, 1985;  
 revision accepted for publication July 23, 1985.)