

Perceived frequency of concrete and abstract words

RICHARD C. GALBRAITH and BENTON J. UNDERWOOD

Northwestern University, Evanston, Illinois 60201

Studies are reported which show that concrete and abstract words of equal objective frequency (based on available word counts) are not perceived as being equal. The abstract word has greater perceived frequency than the concrete word. The judged variety of contexts in which a word appears correlates very highly with perceived frequency. The results have relevance to the design of learning studies in which concrete and abstract words are used, and also to the interpretation of such experiments.

The studies to be reported here were prompted by the results of a minor experiment carried out by the first author as a part of the work in a graduate course in psychophysics. The S's task was to choose the word of a pair which had the higher background frequency. It was observed that when pairs contained an abstract word and a concrete word of equal background frequency as indexed by the Thorndike and Lorge (T-L) (1944) tables, the abstract words were chosen for an overwhelming majority of the pairs. It was as if the Ss had a strong bias for selecting the abstract word in a forced-choice situation in which the two words were equal on the characteristic being scaled (frequency). Such a finding, if confirmed, has several implications. These implications will be discussed later. The purpose of the first experiment was to determine if the finding could be replicated.

EXPERIMENT I

Method

Materials

A total of 595 nouns was selected initially from the list provided by Paivio, Yuille, and Madigan (1968). For the present study, concrete words were defined as those having C ratings greater than 4.0 on the 7-point scale, and abstract words were those having ratings less than 4.0. The mean value of the concrete words was 6.46, with a range of 4.26 to 7.00. The mean for the abstract words was 2.50, the range being 1.42 to 3.77. Eight levels of T-L frequency were selected for testing: 2, 10, 20, 25, 30, 40, A, and AA.

The Ss were tested by the forced-choice procedure. Two types of pairs were constructed, equivalent pairs and variable pairs. An equivalent pair consisted of a concrete word and an abstract word of equal T-L frequency, equal number of letters, and equal number of syllables. There were fewer abstract than concrete words available, particularly in the middle frequency levels. Therefore, in forming the pairs, the abstract words were selected first, being drawn randomly from the eight frequency levels. The concrete words were then grouped according to frequency level, number of letters, and number of syllables. From these groupings, a word was chosen randomly to correspond to a previously selected abstract word. Five such equivalent pairs were formed for each of the eight frequency levels. The frequency values for the two words in an equivalent pair were not always identical, particularly at the higher frequency levels,

but when a difference was present, it was small and, across pairs, unsystematic. The mean values at the eight frequency levels were very close to those specified above. The frequencies for A and AA words were determined from the Lorge count, these values being divided by 4.5 so as to represent frequency per million.

A variable pair always included one word with a frequency of 2 (actually this varied between 1 and 5), and the other word came from one of the remaining seven frequency levels. For half the variable pairs, the word with a frequency of 2 was a concrete word, the other member of the pair being an abstract word of higher frequency. For the other half, the word with a frequency of 2 was an abstract word, with the other member being a concrete word of higher frequency. Three pairs occurred for each combination of frequencies: 2-10, 2-20, 2-25, 2-30, 2-40, 2-A, 2-AA. Since these combinations occurred when the low-frequency word was both abstract and concrete, 42 pairs were required. However, for the combination concrete-2/abstract-40, no pairs could be found among the remaining words that fit when the requirements of equal number of letters and equal number of syllables were imposed. Hence, this combination was omitted and only 39 variable pairs were included on the test.

In summary, there were 79 pairs made up of 40 equivalent pairs and 39 variable pairs. The order of these 79 pairs was randomized, as was the order of the two words within each pair. They were presented to the S in a two-page booklet for the unpaced test.

Subjects and Procedure

Sixty Northwestern University students enrolled in the introductory psychology course participated in the experiment. They were told that one word in each pair occurred more frequently than the other in printed discourse (books, pamphlets, newspapers, magazines) and that their task was to go through the booklet encircling the word in each pair that they believed occurred more frequently.

Results

The percentage of times the abstract word was selected was used as the response measure. These values are shown in Table 1. Looking first at the results for the equivalent pairs, it can be seen that the abstract word was selected far more frequently than the concrete word and that there is no systematic relationship between this selection and T-L frequency. Overall, the abstract word was selected for 81.7% of the pairs. Not 1 of the 60 Ss selected the abstract word for less than 50% of the pairs.

Table 1
Percentage of Times the Abstract Word was Chosen
for Equivalent and Variable Pairs

Equivalent Pairs		Variable Pairs		
Fre- quency	Percent	Frequency Pairing	Concrete Higher	Abstract Higher
2	84.3	2-10	62.8	95.6
10	78.7	2-20	66.1	90.6
20	88.0	2-25	43.9	97.8
25	94.7	2-30	66.1	96.7
30	76.3	2-40	41.7	—
40	92.3	2-A	36.1	97.2
A	66.3	2-AA	37.2	99.4
AA	73.3			

An even more striking outcome, perhaps, is shown for the variable pairs. When the abstract word in the pair had the higher frequency, it was selected at least 90% of the time, and for all six frequency pairings the average was 96.2%. On the other hand, when the concrete word had the higher frequency, the choice of the concrete word never approached this level, even with the largest frequency difference (2-AA). Overall, the abstract word was chosen 50.6% of the time. Yet, it should be noted that the judgments do reflect the frequency differences. When the frequency was small, the abstract word was chosen approximately two-thirds of the time; when the difference was large, it was chosen a little more than one-third of the time.

Discussion

The results confirmed the earlier observation. When a S is asked to select the more frequent word in a pair consisting of an abstract and a concrete word presumed to have equal objective frequency, he is more likely to select the abstract word. There are several possible reasons for this finding, the most obvious being that the T-L count does not reflect the experienced word frequency of current college students. Many investigators, starting with the work of Howes (1954), have shown that apparent frequency of words correlates very highly with T-L frequency. So far as can be determined, however, no investigator has asked about this relationship when only concrete and abstract nouns were involved. It is possible, therefore, that the current college student may have been exposed far more frequently to abstract words (relative to concrete words) than inferred from the T-L count.

An objective word count was published by Kucera and Francis (K-F) in 1967. The relationship between the T-L count and the K-F count may be examined for the words used in Experiment I. The product-moment correlation between the T-L and K-F counts for the 79 abstract words was 0.84, and for the 79 concrete words, 0.61. Thus, for the abstract words, the relative relationship between the two sets of frequencies shows some stability, but for the concrete words, even the

relative ordering is appreciably different.

The relationship between the two counts may be viewed in another way. For all the pairs (excepting the 2-40 pairs), the total T-L frequency was essentially the same for the concrete and abstract words (3,156 and 3,255, respectively). For the K-F count, however, the values for these same words distinctly favored the abstract words, the ratio being 1.75 to 1. As a further test, the K-F values were determined for each of the 80 words used in the equivalent pairs. This distribution of 80 values was divided at the median. Of the 40 words above the median, 27 were abstract and 13 were concrete; for those below the median, of course, the values were reversed. A chi-square test showed that an equal-distribution hypothesis can be rejected ($\chi^2 = 9.80$, $df = 1$, $p < .01$).

The above data clearly lead to the conclusion that abstract and concrete words having equal frequency in the T-L count differ in the K-F count. Relative to concrete words, abstract words have higher frequencies in K-F than in T-L. These data might suggest that there has been a shift in objective frequency along the concrete-abstract dimension and that the K-F count, being more recent than the T-L count, reflects this shift. However, such an inference is questionable. Furthermore, there is no way by which one could conclude that the sources used for the K-F count are more representative of the reading sources of contemporary college students than those used for the T-L count. Nevertheless, as a further experiment, Experiment I was essentially repeated, but the K-F values were used to determine the pairings.

EXPERIMENT II

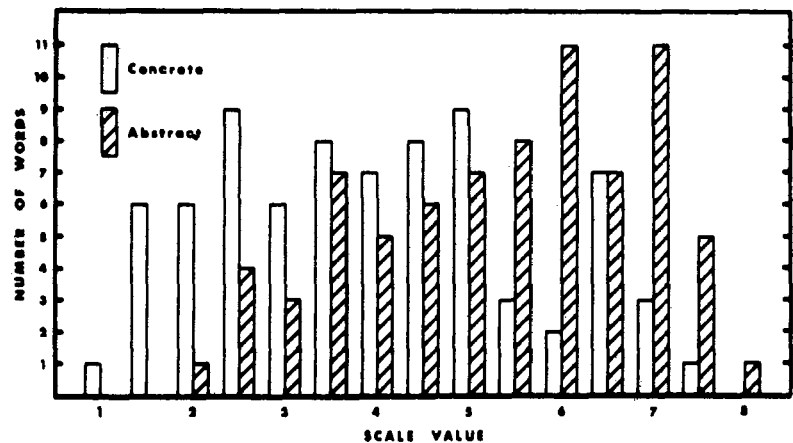
Method

A total of 578 words (380 concrete and 198 abstract) were taken from the Paivio et al (1968) norms and assigned values from the K-F count. Equivalent pairs were formed with five pairs at each of eight frequency levels: 2, 10, 20, 25, 30, 40, 50-99 (corresponding to the T-L A words), and 100 and above (corresponding to the T-L AA words). Again, length of the words in terms of number of letters and number of syllables was held constant within a pair. The variable pairs were the same as

Table 2
Percentage of Times the Abstract Word was Chosen
for Equivalent and Variable Pairs

Equivalent Pairs		Variable Pairs		
Fre- quency	Percent	Frequency Pairing	Concrete Higher	Abstract Higher
2	52.0	2-10	45.5	65.0
10	49.0	2-20	48.9	89.4
20	58.3	2-25	34.4	
25	70.0	2-30	12.2	96.1
30	69.3	2-40	25.5	90.0
40	56.3	2-A	15.0	93.9
A	52.3	2-AA	7.2	96.1
AA	60.7			

Fig. 1. Number of concrete and abstract words (76 each) assigned the various half-step scale values when judged for background frequency.



those in Experiment I, except no words were found to form the concrete 2-abstract 25 pairs.

Sixty introductory psychology students were tested by a procedure that was identical to that used in Experiment I.

Results

Table 2 gives the results in terms of the percentage of times the abstract word was chosen. For the equivalent pairs, choice of the abstract word exceeded 50% at all frequency levels except one (10). Across all levels, the abstract member of the pair was chosen 58.5% of the time. Compared with chance responding, this value is highly significant statistically [$F(1,59) = 18.87$, $p < .01$].

The variable pairs also show a more frequent choice of the abstract words than of the concrete words. As the frequency difference between the words within a pair increases, the percentages rise quickly when the abstract word has the higher frequency. When the concrete word has the higher frequency, there is a slower rise in the percentage of the choice of the concrete words or, as Table 2 shows, a slower shift away from the choice of the abstract words. Thus, the effects shown in Experiment I using the T-L count appeared in Experiment II, although it is quite apparent that the magnitude of the effect is lessened when the K-F count is used to estimate objective frequency.

Discussion

It was noted earlier that it is not possible to know the true objective frequency for the words used in the two experiments. Can both sets of results be explained solely on the basis of inadequate word counts, counts which are inappropriate for the background of current college students? This remains a possibility. However, the sheer magnitude of the differences favoring the abstract words made it seem unlikely that inappropriate background counts is the only factor involved. Since no way seemed open for determining true objective frequency, a different line of attack was instituted. The first step was

to obtain measures of subjective or phenomenal frequency.

EXPERIMENT III

Method

The 158 words used in Experiment I were randomly ordered and presented in a two-page booklet, three columns per page. The words were judged individually by 100 introductory psychology students for background frequency, being sorted along a 9-point scale. To the left of each word was a short line for the S to record a number indicating his judgment of frequency. The instructions, printed on a cover sheet, informed the Ss that words differed widely in their frequency of use in printed discourse and that the task was to rate each word on the 9-point scale according to how frequent they thought it appeared in printed form. The cover sheet contained a 12-cm line, marked by nine equally spaced slashes, with the numbers 1 through 9 printed in correspondence with the slashes. This scale line also appeared at the top of both pages containing the words. Four anchor words, two abstract and two concrete, two of high frequency, two of low frequency, were given with scale values: *water*-9, *figment*-1, *belfry*-1, *thought*-9. The instructions, of course, indicated that words judged to have intermediate frequencies should receive intermediate values. The Ss were further instructed to use all 9 points on the scale but that there was no requirement to use each point equally often. The test was unpaired.

Results and Discussion

The mean scale value assigned each of the 152 words (omitting the 6 words used in the abstract 2-concrete 40 pairs of Experiment I) by the 100 Ss was determined. These values were grouped according to half unit scale steps and the number of words in each group counted. The results are plotted in Fig. 1. It is quite apparent that these two frequency distributions differ. The overall mean value for the abstract words was 5.58, and for the concrete words, 4.19 ($F = 223.62$). Of the 100 Ss, only 4 gave a higher mean to the concrete words than to the abstract words.

Correlations were calculated between the scale values and the T-L and K-F counts. For the abstract words, the correlation with T-L was 0.49, and with K-F, 0.66. For

the concrete words, the corresponding values were 0.55 and 0.60, and for both types combined, 0.48 and 0.63.

Experiments I and II have shown that by a forced-choice procedure, abstract and concrete words of equal frequency in the T-L and K-F tables were not judged to have equal frequency. One possibility for this outcome is that objective counts of the frequency of occurrence of concrete and abstract words are not appropriate for current college students. The data shown in Fig. 1, and the relatively low correlations with available counts, might be used to support this conclusion. Other interpretative possibilities remain, however. In a forced-choice situation, when the S is faced with a pair with near-equal apparent frequency, irrelevant characteristics may determine his decision. Some unknown characteristic of abstractness might be responsible for a response bias, so that with equal apparent frequency the S does not respond randomly. When judging pairs with different apparent frequency, intermingled with pairs in which the words within a pair have equal frequency, the same bias might be operative. In category scaling, the effect of such a bias might be minimized, although it probably would not be ruled out completely. In any event, as a further study, 19 pairs, each consisting of a concrete and an abstract word, were formed such that the scale values (as determined in Experiment III) for the words within a pair were equivalent. These pairs varied widely in scale value. Twenty other pairs, differing in frequency within the pair, were also included. These frequency differences were not systematic across abstract and concrete words and were included merely to make the nature of the task comparable to the ones used in the first two experiments. Sixty Ss were given this forced-choice test. The abstract word was selected over 50% of the time for 13 of the 19 pairs. Overall, however, the difference was not significant [$t(59) = 1.70$, $p > .05$]. Thus, if there is a response bias based on some irrelevant characteristic of the abstract words, it appears to operate both in category scaling and in forced-choice decisions. That is, it must operate when the S judges the frequency of a single word (category scaling) and when he chooses the more frequent word from a pair of words (forced choice).

An experiment by Freund (1970) is relevant to the issue of response bias. Concrete and abstract words were presented with varying frequencies in the laboratory. On the test, some of the pairs included a concrete and an abstract word that had been presented with equal frequency, and the S was forced to decide which member of the pair had been presented more frequently. With situational frequency equivalent for the two words, the judgments did not deviate significantly from 50%. If a response bias was associated with abstract words, it was not found in this situation. We are inclined to discount any appreciable influence of a response bias in the present experiments.

Let us assume that, in fact, the S overestimates the

frequency of abstract words because of a correlated frequency property of abstract words. More particularly, a working assumption would be that abstract words occur on the average in more varied linguistic contexts than do concrete words. A word such as *skillet* appears in a very limited number of contexts, whereas a word like *mood* appears in many different contexts. If contextual frequency is at least a partial determinant of word-frequency judgments per se, and if abstract words occur in more varied contexts than do concrete words, the overestimation of the frequency of the abstract words might be accounted for. Experiment IV was a test of this hypothesis. In this study, the Ss were required to make judgments of contextual variety.

EXPERIMENT IV

Method

The two-page booklet containing the 158 words used in Experiment III was employed, with a new set of printed instructions. These instructions informed the Ss that words differed widely in the variety of contexts with which they are used in printed discourse. Two specific examples were given, both involving concrete words. The word *skillet* (the Ss were told) had a very limited number of contexts, since the word would probably nearly always have reference to a kitchen. The word *water*, on the other hand, may be used in several different contexts, such as references to water in the well, water bed, mineral water, and so on. These instructions had been pilot-tested quite carefully to make sure that the Ss would not be confused and judge word frequency as such. A 9-point scale was used and four anchor words were provided, two concrete and two abstract: *water-9*, *sobriety-1*, *barber-1*, *thought-9*. The instructions on the use of the scale were the same as for Experiment III. The Ss were 100 college students taking the introductory psychology course.

Results

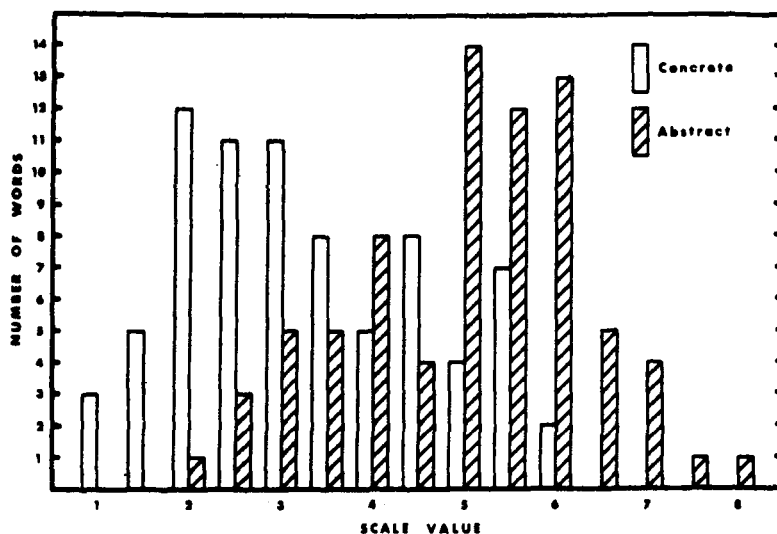
The data were analyzed exactly as were those in Experiment III. The resulting plot is seen in Fig. 2. Figure 2 bears a close resemblance to Fig. 1, although the bimodality is somewhat sharper. For all abstract words, the mean was 5.26, and for the concrete words, 3.56 ($F = 227.32$). Out of the 100 Ss, only 5 rated the concrete words higher than the abstract words.

Since the words used in Experiments III and IV were the same, the relationship between the two sets of scale values resulting from scaling for apparent frequency and scaling for contextual variety could be determined. For the concrete words, the product-moment correlation was 0.89, and for the abstract words, 0.95. For all words combined, the value was 0.92.

GENERAL DISCUSSION

The fourth experiment would seem to implicate contextual variety as a factor influencing the judgments of word frequency as such. However, the data are correlational in nature, and any firm conclusion is not

Fig. 2. Number of concrete and abstract words (76 each) assigned the various half-step scale values when judged for contextual variety.



warranted. The magnitudes of the correlations are such that it seems highly probable that a concrete and an abstract word equated on contextual variety would be judged to have equivalent frequency. Furthermore, if contextual variety is a critical factor, it is probably not a unique property of the concrete-abstract dimension.

The data have indicated that when abstract and concrete words are involved, there is a considerable discrepancy between objective and perceived frequencies. This finding is at variance with a number of previous findings (e.g., Carroll, 1971), but these previous findings did not inquire about the relationship along the abstract-concrete dimension. As mentioned earlier, there seems to be no way to resolve these apparent contradictions, since objective counts of concrete and abstract words may be inappropriate for college students.

The present results have implications for verbal learning studies in which the variable is concrete vs abstract words. First, it is quite apparent that to match concrete and abstract words on T-L (or K-F) frequency is not necessarily to match them on phenomenal frequency. If phenomenal frequency influences learning in the particular task being used, the concrete-abstract variable is confounded with frequency and the outcome of the experiment is difficult to interpret.

A second implication stems from the finding that contextual variety is correlated with the abstract-concrete dimension. Superiority in learning concrete (compared to abstract) words had previously

been attributed to imagery differences, and an impressive amount of evidence has been brought together in support of this position (Paivio, 1971). The fact that contextual variety is greater (on the average) for abstract than for concrete words might suggest alternative interpretations.

REFERENCES

- Carroll, J. B. Measurement properties of subjective magnitude estimates of word frequency. *Journal of Verbal Learning & Verbal Behavior*, 1971, 10, 722-729.
- Freund, J. S. Variability of subject's judgments of frequency and their effect on verbal-discrimination learning. Doctoral dissertation, Northwestern University, 1970.
- Howes, D. On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology*, 1954, 48, 106-112.
- Kucera, H., & Francis, W. N. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press, 1967.
- Paivio, A. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1970.
- Paivio, A., & Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, 76(No. 1, Part 2).
- Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

(Received for publication August 23, 1972;
accepted August 30, 1972.)