JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 10, 382–385 (1971)

Conjoint Frequency, Category Size, and Categorization Time

ARNOLD J. WILKINS[1]

Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QY, England

Two experiments which studied the time to decide whether or not a word was a member of a verbal category demonstrated that the results of several recent studies of categorization may be interpreted in terms of *conjoint frequency* (the frequency of cooccurrence of category and instance in English). The experiments showed that (*a*) instances with high conjoint frequency were categorized faster than instances of similar Thorndike–Lorge frequency but low conjoint frequency, and (*b*) when conjoint frequency was controlled, Thorndike–Lorge frequency did not reliably affect categorization time. It was also found that instances of large categories took longer to categorize than instances of small categories and that negative instances took longer to categorize when they were related to the category in terms of set.

Several recent studies (Landauer & Freedman, 1968; Collins & Quillian, 1969, 1970; Meyer, 1970) have investigated the latency with which *S*s classify object names into verbal categories. In all these experiments it has been the practice to control for the Thorndike–Lorge frequency of the object names used. In this paper it will be argued that such a procedure is inadequate to control for all frequency effects and that the frequency of cooccurrence of category and instance in English usage is a critical variable. This *conjoint frequency* can, of course, be independent of the individual frequencies of the category and instance, and its measurement consequently presents some problems.

One method of measuring conjoint frequency is to use the Connecticut norms (Cohen, Bousfield, & Whitmarsh, 1957). The norms were obtained by presenting category words to a large group of *S*s and asking them to generate instances of the categories. It is assumed that the frequency of *S*s' generation of instances reflects the conjoint frequency of these instances and their categories. The principal aim of the first experiment was to examine the effect of conjoint frequency on categorization by using the Connecticut norms.

There were two further questions with which the present study was concerned. Since these problems were independently investigated by Collins and Quillian (1970) whose paper appeared after the present experiments were conducted, they will be only briefly discussed here.

The first issue, which was investigated in Experiment I, was whether category size affects categorization time when the categories concerned are not nested. This was prompted by Landauer and Freedman's (1968) finding that category size affects categorization time when pairs of nested categories (e.g. Animal–Dog) are used to manipulate category size.

The second question was suggested by the fact that Landauer and Freedman used only instances which were either positive for both nested categories or negative for both nested categories. One possible result of this selection procedure is that the omission of positive instances of the superset category from the selection of negative instances of the subset category decreased the mean latency of negative instances of the subset category.

[1] This work was conducted during the tenure of a Medical Research Council Scholarship. The author thanks Dr. A. D Baddeley, Professor N. S. Sutherland, and Dr. A. R. Jonckeere for their advice.

Consider the nested pair Animal–Dog; all negative instances of Dog were nonanimals as well as nondogs. It is conceivable that the decision that a nonanimal word like Burial is not a dog is easier than the decision that an animal word like Horse is not a dog. The second question was therefore whether the categorization time of negative instances is affected by the proximity of category and instance in terms of set. This was investigated in Experiment II.

EXPERIMENT I

Method

Subjects. Twenty-five male and six female students at the University of Sussex served as unpaid *S*s.

Stimulus material. Eight large and eight small categories were selected from the Connecticut norms (Cohen et al., 1957). The large categories were: a part of the human body, a bird, a substance used for flavoring food, a part of a building, a weapon, a crime, a disease, an occupation or profession. The small categories were: a unit of time, a fruit, a metal, an article of clothing, a vegetable, a musical instrument, a vehicle, a sport. Two positive instances, one with a high frequency of generation (median frequency 202, minimum frequency 91) and the other with a low frequency of generation (median frequency 15, maximum frequency 50) were chosen from each category. The median Thorndike–Lorge frequency of the instances with high frequency of generation was 20 per million and the median Thorndike–Lorge frequency of the instances with low frequency of generation was 28 per million. The corresponding median Howes (1966) frequencies were 2 and 3 per 250,000 respectively.

For each category two nouns which were not obvious associates of the category name were selected as negative instances. The median Thorndike–Lorge and Howes frequencies of these nouns were 24 per million and 2 per 250,000 respectively.

The number of different meanings of each word in the final sample was obtained using both the Oxford and Penguin English dictionaries and in neither count was there a systematic relation between the number of meanings and any of the experimental conditions.

Procedure. The stimulus words were typewritten on cards which were mounted in a stack behind a screen in which was a window. Through the left-hand side of the window *S*s could read the category name and to the right of this a shutter concealed the instance. The *S*s were allowed 1.5 sec to study the category name after which *E* alerted *S* and pressed a key which started an electric clock and raised the shutter. The *S* was instructed to use one index finger to press a response key (marked "Yes") if the instance was positive and the other index finger to press a second response key (marked "No") if the instance was negative. The instructions stressed accuracy and speed, and the use of the left and right hands for positive and negative instances was balanced across *S*s. A session commenced with four practice items which were not used in the subsequent analysis. The remaining 64 stimuli were arranged in four blocks of 16 stimuli in which each category appeared once, paired at random with one of its four possible instances. Each type of instance appeared equally often within each block, and the blocks were presented in random order.

Results

The results are summarized in Table 1. A natural logarithmic transformation of the latencies preceded all analyses and reduced the positive skew of the latency distribution. Errors and missing observations amounted to 8.0% of the data and were replaced by individual *S*'s means.

TABLE 1

EXPERIMENT I: CATEGORIZATION TIMES[a]

	Category size	
	Small	Large
Positive instances		
High conjoint frequency	582 (567, 597)	597 (578, 616)
Low conjoint frequency	622 (604, 641)	643 (621, 665)
Negative instances	626 (612, 640)	657 (640, 673)

[a] The geometric mean categorization times are tabulated in mseconds. The 95% confidence limits of each mean are shown in parentheses.

A two-way analysis of variance of the latencies of positive instances in terms of generation frequency and category size revealed that words with high generation frequency were categorized more quickly than those with low generation frequency, $F(1, 958)$

$= 32.4$, $p < .001$. An analysis of variance of the entire data in terms of category size and sign of instance indicated that instances of small categories were categorized more quickly than instances of large categories, $F(1, 1950) = 17.8$, $p < .001$, and that negative instances took longer to categorize than positive instances, $F(1, 1950) = 28.9$, $p < .001$.

The interaction terms were not significant in either analysis but the error term in the second analysis was inflated since the further factor of generation frequency was involved. There was no significant difference between the latencies of negative instances and the latencies of positive instances with low conjoint frequency, $t < 1.0$.

Experiment II

Method

Subjects. Eight male and eight female students at the University of Sussex and nine male and seven female pupils (aged 16–18) from two local schools served as unpaid *S*s.

Stimulus material. Sixteen categories were selected from the Connecticut norms (a color, a unit of time, a precious stone, a metal, an article of footwear, a military title, a vegetable, an alcoholic beverage, a member of the clergy, an insect, a building for religious services, a tree, a flower, a bird, a fish, a male first name) and from each category two positive instances of similar generation frequency were chosen. The instance with the higher Thorndike–Lorge frequency (or, in marginal cases, the higher Howes frequency) was allocated to one group of positive instances and the instance with the lower Thorndike–Lorge frequency was allocated to a second group. As a result, the median Thorndike–Lorge frequency of the first group was within the A classification and the median frequency of the second group was 11 per million. The median generation frequencies of the two groups were very similar, 96 and 106, respectively.

For each category two negative instances were selected. One negative instance was a positive instance of an arbitrary category which formed a superset of the category with which *S*s were presented. The other instance was a negative instance of this superset category, and in the majority of cases was not an object name. The Thorndike–Lorge frequencies of the two negative instances of each category did not differ by more than 4 per million and the median Thorndike–Lorge frequency for all negative instances was 15 per million.

Procedure. The procedure differed from that of Experiment I only in so far as the number of practice trials was increased to six and the trials were subject-paced, *S*s pressing a footswitch and initiating a warning tone of 1-sec duration.

Results

As in Experiment I the data were transformed prior to analysis, and errors and missing observations (8.2% of the data) were replaced by *S*'s means. The geometric mean latency of the positive instances with high Thorndike–Lorge frequency was 691 mseconds (95% confidence limits = 671, 711) and the geometric mean latency of the positive instances with low Thorndike–Lorge frequency was 702 mseconds (95% confidence limits = 682, 721). An analysis of variance of the positive instances showed that the difference in the mean latencies of the two groups did not approach significance, $F(1, 991) = 1.34$, $p > 0.2$.

The geometric mean of the negative instances which were positive instances of a superset category was 749 mseconds (95% confidence limits = 726, 771) and the geometric mean of the negative instances which were negative instances of the superset was 707 mseconds (95% confidence limits = 688, 726). An analysis of variance of the negative instances revealed a significant difference between the two groups, $F(1, 991) = 17.05$, $p < .001$, indicating that negative instances take longer to process when they are related to the category.

Discussion

The large effect of generation frequency found not only in Experiment I but also in the recent study by Freedman and Loftus (1971) supports the contention that conjoint frequency is an important variable in its own right and has effects which are independent of Thorndike–Lorge frequency. This has implications for other work on categorization as will now be shown. The experimental designs used by Collins and Quillian (1969, 1970, Experiment II) and Meyer (1970) have in common with Landauer and Freedman's

first experiment the fact that nested categories were used in conjunction with object names which were positive instances of more than one category per nest. Collins and Quillian (1969), for example, compared sentences such as "A cedar is a tree" with sentences such as "An elm is a plant." Meyer (1970) compared "universal affirmatives" such as "All thrones are chairs" with others such as "All thrones are furniture". Both studies found that sentences of the latter type took longer to process and their findings were attributed to memory structure. The effect of conjoint frequency provides a compatible nonstructural interpretation of the above results since instances like Elm and Thrones doubtless have a relatively low conjoint frequency with such supersets as Plant and Furniture. This is a reflection of *specificity* in language usage; words are normally classified in terms of their immediate supersets except when some particular contrast is desired.

The absence of a significant effect of Thorndike–Lorge frequency in Experiment II and the large effect of conjoint frequency in Experiment I suggest that conjoint frequency may be more important than Thorndike–Lorge frequency in studies of this kind. To demonstrate this with certainty would, however, require a design in which conjoint frequency and Thorndike–Lorge frequency were varied simultaneously.

The significant effect of category size in Experiment I is inconsistent with the results of the first experiment in Collins and Quillian's (1970) study which compared three categories and found no effect of category size. One possible reason for this discrepancy, as Meyer (personal communication) has argued, is that in Collins and Quillian's experiment, *Ss*' idiosyncratic interpretation of the category names may have invalidated the assumed differences in the sizes of the categories. For example, Collins and Quillian refer to an *S* who apparently interpreted the category Animal as Mammal. The present study circumvents this particular difficulty by using a subject-based measure of category size. Freedman and Loftus (1971), who also used *Ss*' generation of instances to measure category size failed to find an effect of size in a paradigm which involved the latency of *S*'s production of instances. This procedure may be sufficiently different from that of the present experiments to explain the absence of a size effect.

The finding that negative instances take longer to process if they are related to the category in terms of set supports the argument that Landauer and Freedman's (1968) selection of negative instances may have decreased the mean latency of negative instances of the smaller of the nested categories, and so contributed to the effect which they attributed to category size.

The difference in the processing times of negative instances is also of interest as regards the "list" models discussed by Landauer and Freedman. Models such as these which explain categorization in terms of the processing of lists of positive instances are insufficient to account for the difference in the two groups of negative instances.

References

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, University of Connecticut, Contract Nonr 631(00), Office of Naval Research.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240–247.

Collins, A. M., & Quillian, M. R. Does category size affect categorization time? *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 432–438.

Freedman, J. L., & Loftus, E. F. The retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 107–115

Howes, D. A word count of spoken English. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 572–604.

Landauer, T. K., & Freedman, J. L. Information retrieval from long-term memory: Category size and recognition time. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 291–295.

Meyer, D. E. On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1970, **1**, 242–300.

(Received November 17, 1970)