

Event-frequency judgments as a function of the linguistic frequency and single or paired presentation of target words: II. Task requiring judgment of linguistic frequency

MELVIN H. MARX

Georgia State University, Atlanta, Georgia

A total of 80 college students in four classes rated a series of English words as high or low in linguistic frequency. Words high or low in linguistic frequency were presented 0, 1, 2, 3, or 4 times, singly or in mixed pairs. The students were then given a previously unannounced event-frequency test. The results implicated two major determinants of event-frequency judgments: (1) an item-strength factor and (2) multiple memory traces. Words high in linguistic frequency were given consistently higher event-frequency estimates, especially when they had been presented singly. Words low in linguistic frequency were more accurately judged for event frequency when they had been paired rather than singly presented; this result suggests the role of multiple memory traces because of the greater number of such traces with unique contextual cues available for retrieval at test in the paired condition. The results are compared with those earlier obtained with the same experimental materials and general procedure but a different cover task. Interpretation in terms of a dual-factor hypothesis for event-frequency judgment is discussed.

In a previous experiment (Marx, 1986a), college students estimated the frequency with which words of high (hi) or low (lo) linguistic frequency (LF) had occurred. The words were presented singly or in mixed pairs (one hiLF and one loLF word), and the task was to compose short sentences that included the one or two words presented. With single presentations, the event frequencies of loLF words were more accurately judged; with paired presentations, loLF words were given higher, but not more accurate, frequency estimates.

The present experiment used the same word lists but a different task. College students were asked to judge the linguistic frequency of each word by marking it high or low on that dimension. This task offers the advantage of manipulating both preexperimental strength of words (the LF variable) and the availability of contextual cues (present in the paired but not in the single presentation). Subjects were given the same unannounced event-frequency test as had previously been used. All other experimental conditions were essentially the same as those in the previous experiment (Marx, 1986a).

METHOD

Subjects

A total of 80 students from four introductory psychology classes at a small midwestern liberal arts college participated. Because the school

The data were collected while the author was supported by a Research Career Award from the National Institute of Mental Health. Thanks are due to Miriam Mueller and Gale Fuller of Westminster College at Fulton, MO for making available their introductory psychology classes. Address reprint requests to Melvin H. Marx, Department of Psychology, Georgia State University, Atlanta, GA 30303.

had only recently become coeducational, it had a predominantly male student body and there was a disproportionately high number of male subjects participating in this study; the gender variable, therefore, was disregarded. The subjects were tested in their normal class settings, with class sizes of 19, 20, 20, and 21.

Experimental Design

Event frequency (0, 1, 2, 3, and 4 occurrences) was manipulated within list, as were linguistic frequency (hi, lo) and word format (single, paired). Word specificity was controlled for the word-format variable by reversing the single or paired condition of each word in the second of the two lists. Each list form was given to two classes. There was no control for word specificity in the event-frequency variable.

Word Lists

Each list contained 30 critical slides, 20 with single words and 10 with paired words. The paired words were mixed, one from the hiLF set and one from the loLF set. In addition to the 30 critical slides, there were two primacy and two recency buffers, each with one single word (*paint, greed*) and one pair of mixed words (*power-cycle, badge-value*). The second list contained the same buffers and the same target words, but as described above the target words were reversed with respect to their single or paired format. Further details on the construction of the lists are available in the earlier report (Marx, 1986a).

Procedure

The subjects were told that they would be shown a series of English words, one or two on each slide, and that their task was to "decide whether the words are of high or low frequency as used in ordinary English." Examples of high (*boy, force*) and low (*cuisine, holster*) frequency were given. Subjects were instructed to put an H or L in the appropriate space on the answer sheets provided for each word shown. They were told that some of the words would be shown more than once, either singly or with another word. All of the other instructions and procedures were the same as previously described (Marx, 1986a), except that an 8-sec projector rate was used.

After the subjects completed the study task on the 34 slides and the answer sheets were collected, a second set of answer sheets was passed out. The subjects were told that they would now be shown some slides

with single words and that their task was to "estimate as closely as possible how many times each word occurred during the first part of the experiment." They were told that some of the words had not been shown, and should therefore be given a "0" score in their marking of the frequencies, but no upper limit was indicated. The projector was run at an 8-sec rate.

RESULTS

As in the previous study (Marx, 1986a), chi-square tests were used to evaluate differences between proportions of subjects showing various tendencies. Size and accuracy of event-frequency (EF) judgments were also subjected to analysis of variance (ANOVA). Because the word-list variable was ineffective in the ANOVAs ($F_s < 1.00$), it is disregarded in the description of the results.

Size of Frequency Judgments

An ANOVA of the frequency estimates resulted in reliable effects and interactions for all three within-subjects variables (LF, EF, and word format), including a reliable triple interaction [$F(1,316) = 20.71, p < .001$]. Separate ANOVAs were then computed for the various word frequencies in order to determine the basis of the reliable effects and interactions. The mean estimates for the four levels of EF occurrence (1 through 4) are shown in Table 1, along with their respective standard deviations and the results of ANOVAs comparing hiLF and loLF scores, separately for single and paired words. Because the latter variable (word format) is meaningless for words that were not actually presented (F0s), those data are treated separately below.

Note that the consistently higher estimates made for single hiLF words from F2 through F4 were all reliable, as were the higher mean estimates for paired hiLF words at the F1 and F2 levels. Although the reversal of this trend for F3 paired words, with loLF words now surpassing hiLF words, was not reliable, the larger difference in the same direction at F4 was reliable.

F0 errors. Very few errors were made on the test words that had not actually been shown on the study slides. Only 26 of the 80 subjects made an error on these F0 words, and only 6 of the total of 41 such errors were frequency

estimates greater than 1. Only hiLF F0 words were misjudged by 17 subjects, only loLF/F0 words by 4 subjects, and 5 subjects missed at least one of each type. If the latter subjects are distributed evenly between the two LF categories (2.5 to each category) and an a priori 50:50 distribution of erring subjects is assumed, the differential proportion of hiLF misses to loLF misses in the sample of 26 erring subjects is reliably different from chance [$\chi^2(1) = 6.50, p < .05$]. A similar even distribution of ties and assumption of 50:50 chance baseline is made on subsequent chi-square analyses.

A reliably higher mean estimate for hiLF compared to loLF words (.24 to .11) was also obtained by ANOVA [$F(1,79) = 4.56, p < .05$].

Number of deviations. In order to eliminate the influence of very large deviations of estimates from actual frequencies, those resulting from unreasonably high estimates, the numbers of positive and negative deviations from actual frequencies were counted and compared for each subject. The numbers of subjects who showed more positive than negative deviations, compared with the numbers of subjects who showed more negative than positive deviations, were 65:10, 46:18, 33:27, and 37:22 for hiLF single, hiLF paired, loLF single, and loLF paired words, respectively. Although neither of the loLF ratios was reliably different from a 50% chance baseline, both of the hiLF ratios were [$\chi^2(1) = 38.81, p < .01$ for single hiLF words, and $\chi^2(1) = 9.80, p < .01$ for paired hiLF words].

When the numbers of subjects who showed relatively more positive deviations were then counted, directly comparing the single and paired conditions for hiLF words, 47 subjects showed more single-word positive deviations and 13 showed more paired-word positive deviations [$\chi^2(1) = 14.45, p < .01$]. A similar direct comparison of hiLF and loLF scores for the single words revealed that 50 subjects had relatively more positive hiLF deviations compared with 9 subjects having relatively more positive loLF deviations [$\chi^2(1) = 21.01, p < .01$].

There seems to be little question but that there was an extremely strong general tendency for the singly presented hiLF words to elicit higher estimations in EF judgments.

Correct Frequency Judgments

The probabilities of correct frequency judgments as a function of single- and paired-word presentation are shown in Table 2 for hiLF and loLF words and for the various EF levels, again excluding F0 scores. The apparently consistent superiority of loLF words in the paired-word over the single-word format was reflected in the fact that, overall, a higher number of subjects (44) made more correct judgments in the paired format for the loLF words than those subjects (15) who showed the opposite tendency [$\chi^2(1) = 10.51, p < .01$]. The corresponding figures in the hiLF data were 31 and 21 subjects, for paired and single format respectively [$\chi^2(1) = 1.25, p > .05$].

Only small differences between hiLF and loLF words in correct judgments occurred. Overall, for single words, there were 24 subjects with higher loLF and 32 with

Table 1
Mean Size of Event-Frequency Judgments for Words Presented One to Four Times, with Statistical Tests of hiLF Versus loLF Words in Single and Paired Format

	Word Frequency							
	1		2		3		4	
	hiLF	loLF	hiLF	loLF	hiLF	loLF	hiLF	loLF
Single-Word Format								
Mean	1.30	1.49	3.60	2.41	4.59	2.78	5.00	4.45
SD	.68	.83	1.46	1.09	1.72	1.37	1.85	1.66
$F(1,79)$	3.08		40.48		86.44		6.60	
p	>.05		<.001		<.001		<.01	
Paired Format								
Mean	1.30	1.13	2.66	2.08	3.75	3.91	3.80	4.36
SD	.85	.64	1.19	1.08	1.20	1.54	1.42	1.44
$F(1,79)$	4.25		14.60		.87		13.01	
p	<.05		<.001		>.05		<.001	

Table 2
Probabilities of Correct Event-Frequency Judgments
for Single and Paired hiLF and loLF Words

Word Format	hiLF Words					loLF Words				
	Word Frequency					Word Frequency				
	1	2	3	4	Total	1	2	3	4	Total
Single	.75	.23	.26	.16	.35	.58	.33	.19	.16	.31
Paired	.58	.41	.28	.44	.43	.73	.43	.38	.38	.48

higher hiLF scores; for paired words, the corresponding figures were 30 and 25.

ANOVA of these data, with F0 scores excluded and the EF variable collapsed, indicated that the mean number of correct judgments for the paired format (1.69) was reliably greater than the corresponding mean for the single format (1.29) [$F(1,79) = 13.78, p < .001$]. Only a negligible difference occurred for the LF variable (1.49 mean correct judgments for hiLF words, 1.44 for loLF words; $F < 1.00$). There was no interaction of word format and LF variables [$F(1,79) = 1.31, p > .05$].

DISCUSSION

There are two major conclusions to be drawn from these data. First, there is a clear indication of the pervasive bias in the making of frequency estimates resulting from the preexperimental strength, or familiarity, of the target words. This bias is particularly evident in the higher estimates made for the hiLF words when they were singly presented. Second, there is also a very strong suggestion of the helpful role played by multiple traces in improving the accuracy of frequency estimates; this role is indicated in the general advantage of paired over single words.

Word Strength

The consistency of the heightened hiLF estimates compared with the loLF estimates for singly presented words suggests that, in this respect at least, there is no need to worry about the lack of control of word specificity over event frequencies. The effect was apparently a very general one, as shown in the three statistically reliable differences for F2, F3, and F4 categories (Table 1). Moreover, the same relationship occurred for paired words at the lower event-frequency levels, although a reversal was found for the higher frequency levels. Apparently, the biasing effect of the word-strength factor was countered only by the most frequent pairing condition, F4, for which equal experimental (study) presentations resulted in reliably higher estimates for the loLF words, as had also occurred in the prior experiment (Marx, 1986a).

Confidence in the robustness of the word-strength effect is also strengthened by the elimination of any artifactual effect from a small number of extreme scores, by the analysis that simply counted positive or negative deviations, with clearly greater numbers of subjects showing relatively more positive scores for hiLF than for loLF words.

Another indication of the power of the word-strength factor is provided by the results of the F0 analyses. Because these words were never actually presented, their frequency estimates make possible an uncontaminated test. The statistically greater overestimation of hiLF F0 words supports the conclusion drawn from the data of Table 1, and in a way that is not affected by any confounding with event frequency. Although only about one third of the subjects made such errors, above-zero estimates were both more likely to be made and were larger on hiLF words than on loLF words. Poorer recognition (more false alarms) of words high in linguistic frequency was also reported by Rao (1983).

The consistently higher estimation of hiLF words observed in the present experiment did not occur in the previous one utilizing the same word lists. Rather, reliably more subjects in the previous experiment gave higher estimates to paired loLF words than to single loLF words. These differential results can be explained by assuming that the task in the present experiment alerted subjects to the high/low linguistic dimen-

sion and thus maximized its influence. Rao's (1983) use of the typical vague "memory test" instructions apparently had a somewhat similar effect, even though attention was not drawn specifically to the linguistic dimension itself.

Why did such an effect not occur in the previous experiment (Marx, 1986a)? Perhaps because associating loLF words with hiLF words in the pairing operation led to an enhancement in their frequency estimates, as it seemed to do in the present experiment for the highest EF level. Such an effect is entirely consistent with the earlier reports of superior recall of words low in linguistic frequency when they were mixed with high-frequency words (Duncan, 1974; May & Tryk, 1970) but not when they were presented in the usual homogeneous list (May, Cuddy, & Norton, 1979).

Multiple Traces

Several lines of analysis in the present data support the commonly accepted role of multiple traces in the making of frequency judgments (Hintzman, 1976). In this experiment, the multiple traces were enhanced only for the paired words—by the different word that accompanied each repeated presentation of any given target word. There was, thus, a much greater distinction between paired- and single-word presentations (the single repetitions having no contextual accompaniment other than that provided by the repeated event itself) than in the previous experiment (Marx, 1986a). The task in that experiment, composing short sentences from presented words, ensured that single as well as paired words were richly endowed with contextual multiple traces.

The contrasting results of these two experiments may be interpreted in accordance with their contrasting methodologies. Singly presented words were more accurately estimated for frequency in the prior experiment (Marx, 1986a), as they had been in experimentation by Rao (1983), whereas there was a general advantage of all *paired* words in the present experiment, and by the ANOVA test an interaction with the LF variable. The suggestion in the chi-square proportional test of a reliable superiority of paired loLF words but not hiLF words does not contradict this argument. Such an effect in the present experiment can be accounted for by assuming that numerical frequency estimates of the preexperimentally "weaker" loLF words were more improved by the availability of contextual cues, providing better memory traces, in the paired condition.

Accuracy in event-frequency estimation of verbal items may thus be viewed, in some part at least, as a result of the sometimes conflicting and sometimes compatible factors associated with (1) the *strength* of the words, resulting both from their preexperimental status and the frequency of their experimental occurrence, and (2) the availability of multiple study cues or *traces* for retrieval, as by counting at test. In this respect, it is perhaps suggestive that the greatest advantage of paired over singly presented words (.44 - .16 = .28) occurred for the F4 hiLF words (Table 2), where the potentially greatest combined effect of strength and multiple-traces factors would be expected to occur.

Dual-Factor Hypothesis

The major general theoretical implication of the data from this experiment and the previous one (Marx, 1986a) is their support of a dual-factor interpretation of the making of frequency estimates. This hypothesis holds that when subjects are able to make more or less precise event-frequency judgments on the basis either of a direct running count or a retrospective count based on the retrieval of multiple memory traces, they will do so. However, when such a count is not feasible, or when for some reason a gross approximation is attempted, various biasing factors become more effective, and the strength of representation of the target item can have a relatively great influence on the event-frequency judgment.

Implication of both strength and multiple-trace factors in the making of event-frequency judgments is supported not only by the trends in the present data, described earlier, but also by the various retrospective reports obtained from subjects in these and other frequency-judgment experiments (Marx, 1985, 1986b). There is direct evidence in these reports for the counting of traces, especially for the task with more plentiful multiple traces and for the younger subjects. The reports are equally clear, however, in implicating also some kind of a more global, or impressionistic, factor, especially for the older, and presumably more sophisticated, subjects.

Teasing out the complex interplay of these, and probably other as yet uncovered, determining factors remains a challenging task for future research on event-frequency judgments.

REFERENCES

- DUNCAN, C. P. (1974). Retrieval of low-frequency words from mixed lists. *Bulletin of the Psychonomic Society*, **4**, 137-138.
- HINTZMAN, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10, pp. 47-93). New York: Academic Press.
- MARX, M. H. (1985). Retrospective reports on frequency judgments. *Bulletin of the Psychonomic Society*, **23**, 309-310.
- MARX, M. H. (1986a). Event-frequency judgments as a function of the linguistic frequency and single or paired presentation of target words: I. Task with unique multiple traces. *Bulletin of the Psychonomic Society*, **24**, 245-247.
- MARX, M. H. (1986b). More retrospective reports on event-frequency judgments: Shift from multiple traces to strength factor with age. *Bulletin of the Psychonomic Society*, **24**, 183-185.
- MAY, R. B., CUDDY, L. J., & NORTON, J. N. (1979). Temporal contrast and the word frequency effect. *Canadian Journal of Psychology*, **33**, 141-147.
- MAY, R. B., & TRYK, H. E. (1970). Word sequence, word frequency, and free recall. *Canadian Journal of Psychology*, **24**, 299-304.
- RAO, K. V. (1983). Word frequency effect in situational frequency estimation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 73-81.

(Manuscript received for publication April 14, 1986.)

Announcement

**The Fourth European Conference on Eye Movements
Göttingen, FR Germany
September 21-24, 1987**

Call for Papers

The Fourth European Conference on Eye Movements (E.C.E.M. 4) will take place in Göttingen, FR Germany, September 21-24, 1987. In the tradition of its predecessors, the conference will be open and interdisciplinary.

The following topics will be included in E.C.E.M. 4: the neurophysiology of eye movements, the ocular-motor system, eye movements in perceptual and cognitive tasks (visual search, word perception, reading, etc.), and applied research. The selection of papers to be presented will be made by an international committee. Papers and posters at E.C.E.M. 4 should be in English, and the presentation of the paper should not exceed 20 minutes.

The conference proceedings will be published in two volumes. The first volume, which will be available at the beginning of the conference, will consist of the abstracts of all the papers and posters to be presented. The second volume will consist of selected papers—possibly centered around a particular topic.

For consideration of a paper to be presented at E.C.E.M. 4, a one-page abstract must be received at the address given below by January 31, 1987, at the very latest. For further information, please contact:

G. Lüer and Uta Lass
European Conference on Eye Movements
Gosslerstrasse 14
D-3400 Göttingen, FR Germany