

Retrospective reports on frequency judgments

MELVIN H. MARX

Georgia State University, Atlanta, Georgia

Two groups of college students were given an unexpected event-frequency test following the completion of a simple cognitive task (composing sentences from slide-presented words or estimating the linguistic frequency of the same words). All of the subjects were then asked to describe how they had gone about making the event-frequency judgments. The predominant trend in the retrospective reports of the subjects who had composed unique sentences supported a multiple-traces factor (counting of the individual sentences), whereas the predominant trend in the reports of the subjects who had estimated linguistic frequencies supported a strength factor (general impression, familiarity). A dual-factor hypothesis of frequency processing is suggested, with counting of memory traces operative mainly for the lower frequencies and the more highly differentiated traces, and some kind of strength factor operative mainly for the higher frequencies, for which counting becomes difficult, and the less clearly differentiated traces.

Following two recent experiments involving event-frequency judgments, the data of which have not yet been fully analyzed, the subjects were asked to provide brief retrospective reports on the manner in which they had made their judgments. Here the categorizing of these reports is summarized, and some suggestions are drawn from them with respect to the theoretical interpretation of event-frequency processing.

METHOD

Different "cover" tasks with the same stimulus materials were used in the two experiments. In neither case were the subjects informed beforehand that there would be an event-frequency test.

Experiment 1

A total of 26 subjects from two introductory psychology laboratory sections at a state university volunteered to serve as subjects. They were tested in groups of 10 and 16. Their task was to compose a short sentence incorporating the 1 or 2 words shown by slide projector. Lists of five-letter English words were used. Ten words were high in linguistic frequency (AA, Thorndike & Lorge, 1944), and 10 were low in linguistic frequency (not more than 9 occurrences in one million words). Five of the high- and 5 of the low-frequency words were shown singly, with frequencies of 0, 1, 2, 3 and 4. The other 5 words in each linguistic-frequency class were shown in pairs, which always consisted of a high- and a low-frequency word, randomly ordered, with no duplication of the same pairings. Two filler words were presented singly at the beginning of the list, and two were presented at the end of the list. Twenty single-word slides and 10 paired-word slides were presented, making a total of 34 slides. The list of words given to the first group was altered for the second group, so that words presented singly were now paired, and vice versa, with no change in list frequencies. The sub-

The data were collected while the author was supported by a Research Career Award from the National Institute of Mental Health to the University of Missouri-Columbia. Thanks are due James Broyles for his assistance in the recruitment of student subjects, Michael Johnson for making available the student volunteers in his two laboratory sections, and Miriam Mueller and Gale Fuller of Westminster College at Fulton, MO, for the availability of their students. Address reprint requests to: Melvin H. Marx, Department of Psychology, Georgia State University, Atlanta, GA 30303.

jects were instructed to compose different sentences for repeated single words. The projector was set on a 15-sec rate.

The event-frequency test was administered with no advance notice immediately after the presentation of the study slides. The 20 critical words were shown singly, in a random order except that the first half of the list had 2 words at each event-frequency level and 5 words from the single and 5 from the paired format. Also, no more than 2 successive high- or low-frequency words or single or paired words were permitted. An 8-sec slide-presentation rate was used.

Following the event-frequency test, the subjects were given an additional answer sheet with these instructions: "How did you decide on the frequency estimates? For example, did you use any special procedures? If so, please write a brief description of how you did proceed."

Experiment 2

A total of 80 introductory psychology students from four classes in a small liberal arts college served as subjects. They were tested in their normal classroom setting, 19 to 21 per class. The same verbal materials were used that had been used in Experiment 1, with the first and third classes being tested on one list and the second and fourth classes on the other.

The subjects were told that they would be shown a series of common English words, one or two at a time. Their task was "to decide whether the words are of low or high frequency as used in ordinary English." Examples of high- ("bag," "force") and low-frequency ("cuisine," "holster") words were shown, and the subjects were instructed to write "H" or "L" for each word in the appropriate place on the answer sheet. Following presentation of the study slides, they were given the same event-frequency test and retrospective-report instructions that had been given in Experiment 1 and were cautioned on the distinction between linguistic and event frequencies.

RESULTS

The major objective of the request for retrospective reports was to compare the incidence of indications that the subjects had made their event-frequency judgments by means of multiple memory traces, clearly the most favored theoretical account (Hintzman, 1976), or by means of some kind of strength factor, over the two differing cover tasks.

Retrospective responses were categorized as *multiple traces* if there was any indication that the subjects had

made an effort to recall and count separate occurrences of the items. They were categorized as *strength* if the subjects indicated that some kind of unitary impression was used in making the judgments.

The results of the categorizations of the protocols from the two experiments are summarized in Table 1.

It is quite clear that the predominant trend in Experiment 1 was in support of the multiple-traces factor. Representative excerpts from the reports illustrate this conclusion. One of the participants stated that she "tried to remember the sentences written and count them up." Another stated that she "tried to remember how many and what sentences I wrote." A third "tried to remember how many times or different ways I had used the word." Another wrote "I tried to remember how many sentences I made with each word, recalling the sentences." A slightly different response was, "First the sentences that I wrote would flash through my mind so that I could count how many times these were."

In evaluating the near unanimity of these responses, it needs to be noted that the event-frequency range was restricted (0-4) and that the cover task used, composition of unique sentences, lent itself very nicely to this kind of retrieval by virtue of the richly processed and highly differentiated materials generated. Even so, there were some signs of a strength factor, especially within a comparative framework. For example, "the actual number I gave was just based on the frequency of appearance of other words, low if I didn't think as much, higher if I remembered it more in relation to others."

The results from Experiment 2, also shown in Table 1, reveal an entirely different picture. The predominant trend in these reports was toward the strength and the related comparison-framework categories. The difference in ratios of multiple-trace to strength categorizations, 20:3 in Experiment 1 and 11:36 in Experiment 2, was reliable [$\chi^2(1) = 25.28, p < .01$]. Certainly the subjects' responses, if not the actual cognitive mechanisms used, varied markedly with the nature of the task.

Illustrative excerpts from reports categorized under the strength rubric are: "the first number that came to mind"; "how familiar the words looked"; "got a feeling for the number"; "estimated according to strength of impression"; "estimated on the basis of impressions"; and "if a word seemed very familiar."

Illustrative excerpts for the category of comparison framework are: "made a rough estimate of the maximum times . . . then tried to relate the words to this figure";

"used the number of words possible for an educated guess"; and "most times probably about five . . . based judgment relative to five."

There was a marked difference between the two experiments in the number of nonusable retrospective reports. There were only 3 such (of 26) in Experiment 1 (2 essentially rewording the question, 1 indicating pure guesswork), but there were 33 subjects (of 80) in Experiment 2 who either "just guessed" (7), simply reworded the question (17), apparently misunderstood the question (4), or did not attempt to answer (5). This difference in ratios (3:23 vs. 33:47) was reliable [$\chi^2(1) = 7.46, p < .05$]. The difference seems to reflect the relatively greater definition and consequent ease of process identification of the cover task in which unique sentences were composed.

DISCUSSION

These reports are of value mainly for their suggestiveness. One very important suggestion is apparent in them. Two major processes seem to be involved in the frequency-judgment task: (1) an attempt to make some kind of count, when conditions permit, especially for the lower frequency items; and (2) a much more tentative estimate, particularly for the higher frequencies, in which impressions, rather than precise counts, are used and in which some kind of comparison framework often seems to be involved. The former process is, of course, that identified by the multiple-traces factor. A dual-process hypothesis is thus suggested, with counting of memory traces operative generally for the lower frequencies and strength operative for the higher frequencies.

The important point about this rather obvious interpretation of the reports obtained is not that it is novel—clearly, it is not. Earlier suggestions that different processes may be operative in frequency judgments as frequencies change have been made by Hintzman (1976) and by Voss, Vereb, and Bisanz (1975). But these suggestions have been largely ignored and their implications not tested. The present results indicate that researchers should pay greater attention to implications of a strength factor, perhaps prematurely dismissed by Hintzman (1976). In any event, whatever particular underlying factors are ultimately uncovered, the present data do clearly indicate the dependence of retrospective reports, and to some extent also the basic memory processes themselves, on the type of task offered the subjects.

REFERENCES

- 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.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Teacher's College Press, Columbia University.
- VOSS, J. F., VEREB, C., & BISANZ, G. (1975). Stimulus frequency judgments and latency of stimulus frequency judgments as a function of constant and variable response conditions. *Journal of Experimental Psychology: Human Learning and Memory*, 1, 337-350.

Table 1
Numbers (No.) and Percentages (%) of Usable Reports in Categories

	Categories							
	Traces		Strength		Comparisons		Total	
	No.	%	No.	%	No.	%	No.	%
Experiment 1 (Unique sentence)	20	87	1	4	2	9	23	100
Experiment 2 (Linguistic judgment)	11	23	28	60	8	17	47	100