

Retrieval of item frequency information in a continuous memory task

WILLIAM E. HOCKLEY

University of Toronto, Toronto, Ontario, Canada

The accuracy and response latency of absolute frequency judgments were measured as a function of test lag (the number of intervening items between presentations of a test item) in a continuous memory task. Frequency was varied from one to three presentations in Experiments 1 and 2 and from one to five presentations in Experiment 3. The proportion of correct responses decreased as frequency increased, and correct mean response time tended to increase with frequency. Both accuracy and correct mean response time were found to be largely a function of the most recent test lag. The lag-latency functions were best described by piecewise linear functions, with the breakpoint occurring between lags 1 and 2. Continuous frequency estimation was also shown to improve with extended practice. The implications of the results are discussed with reference to trace strength, numerical-inference, and multiple-trace theories of frequency discrimination.

Since Sternberg's (1966) now classic study, response time has become an important dependent variable in the study of retrieval processes in recognition. Response time has been measured in numerous recognition paradigms, and a variety of models have been developed to account for both the accuracy and the latency of recognition performance (e.g., Atkinson & Juola, 1973; Murdock & Anderson, 1975; Pike, Dalglish, & Wright, 1977; Ratcliff, 1978).

In contrast, studies that have investigated the storage and retrieval of item frequency information have not, in general, employed response time as a dependent variable. There are, however, two exceptions. Voss, Vereb, and Bisanz (1975) found that the latency of absolute frequency judgments tended to be an inverted U-shaped function of frequency for frequencies varying from 2 to 32 presentations. Hintzman, Grandy, and Gold (1981) measured response time in a relative judgment of frequency task. They found that response time depended on the frequencies of both the chosen and the unchosen alternatives and that response time was longer the more similar were the frequencies of the two alternatives. Also, in contrast to the results of Voss et al.,

Hintzman et al. obtained a decrease in correct mean response time with frequency.

Howell (1973a) suggested that the measurement of response time for frequency judgments might serve to distinguish between theories of frequency discrimination. However, these theories are not, as yet, sufficiently precise to enable predictions concerning response latency. Hintzman (1976, pp. 59-62), for example, argued that it is not clear what kinds of latency functions the multiple-trace and the proposition-encoding theories predict. Depending on the assumptions made, the multiple-trace theory would appear to be able to accommodate either an increase or a decrease in response time with frequency (Hintzman et al., 1981)

The inability of the extant frequency theories to make specific predictions of latency functions does not diminish the potential importance of response time in the study of frequency discrimination. The measurement of response time may serve to provide limits on the nature of the assumptions that can be made for particular frequency theories and thus more clearly delineate the frequency theories.

The present study measured the accuracy and response latency of absolute frequency judgments in a variant of the continuous memory task (cf. Shepard & Teghtsoonian, 1961). In this paradigm, items are repeated within a long, continuously presented list of items. For each presented item, subjects were required to indicate the number of times the item had occurred in the list. The principal independent variables were test lag (the number of intervening presentations between each test of a given item) and item frequency. Because the processes that underlie frequency estimation may differ for frequencies of approximately 2 to 8 and for frequencies greater than 8 (Voss et al., 1975), only a limited range of frequencies was tested. In Experi-

This research was conducted at the University of Toronto and was supported by Research Grant APA 146 from the National Research Council of Canada to B. B. Murdock, Jr. Experiments 1 and 2 were presented at the Eighth Annual Australian Experimental Psychology Conference, Adelaide, May 1981, at which time the author was supported by a post-doctoral scholarship from the University of Auckland, New Zealand. The author is indebted to M. C. Corballis and B. B. Murdock, Jr. and to the reviewers of earlier versions of this manuscript for their comments and criticisms. Requests for reprints should be sent to W. E. Hockley, Department of Psychology, University of Toronto, Toronto, Ontario M5S 1A1, Canada.

ments 1 and 2, frequency was varied from one to three presentations, and in Experiment 3, frequency was varied from one to five presentations.

EXPERIMENT 1

Method

Subjects. In all the experiments, the subjects were right-handed University of Toronto undergraduate students who were paid for their participation. All subjects were tested individually. In Experiment 1, four subjects each completed 1 practice session, 12 experimental sessions, and 1 choice reaction time session.

Apparatus and Stimuli. In each experiment, list generation, display, and response recording were controlled by a PDP-12A laboratory computer. The stimuli were presented on a cathode-ray screen that was approximately 75 cm from the subject. The stimulus words were presented in uppercase letters, and each letter was approximately 8 mm high. The subjects responded on a three-key response panel (Experiments 1 and 2) or a five-key response panel (Experiment 3) connected to the computer via the sense lines. The keys were equally spaced from a rest position. For each response, the key pressed and the latency of the response (measured from the onset of the probe to the keypress) were recorded on magnetic tape. Response latency was measured in units of 5 msec.

All words in each session of Experiment 1 were selected randomly, without replacement, from the Toronto Wordpool, a collection of 1,080 common two-syllable words not more than eight letters long and with homophones, contractions, archaic words, and proper nouns excluded. (For a complete description of this wordpool, see Friendly, Franklin, & Hoffman, 1980.)

Procedure. At the beginning of each session, a seed value was initialized for the random number generator; the seed was based on the subject's name and session number. The random number generator was used to scramble the wordpool and to determine the order in which the lags between repeated items were represented in each list. With a different start value for each subject for each session, a unique set of lists could be generated for each subject-session.

Each judgment-of-frequency session consisted of three lists of words. Within each session, no word was presented in more than one list. The lists were constructed with an algorithm that attempts to produce a list consisting of 110 words that are presented three times, for a total list length of 330 items. The number of items intervening between the first and the second presentations of a repeated word (P1-P2 lag) and between the second and the third presentations (P2-P3 lag) were manipulated independently. The possible P1-P2 and P2-P3 lags were: 0, 1, 2, 4, 6, 8, 16, 24, 32, and 40. Thus, each of the 10 possible P1-P2 and P2-P3 lags could be represented 11 times within each list.

In practice, not all repeated items could be fit into every list. Averaging across all lists of Experiment 1, 3% of the second presentations and 7% of the third presentations of repeated words could not be fit into the lists. In these instances, once-

presented filler words were shown. In order to determine whether item repetitions were distributed reasonably uniformly throughout the lists, all of the lists of the experimental sessions were arbitrarily subdivided into six consecutive blocks of 55 items each. The proportions of first presentations of words within each list block were .441, .345, .353, .351, .356, and .360, with an overall list mean of .368. The proportions of second presentations for blocks 1 to 6 were .326, .323, .321, .330, .319, and .319, with a list mean of .323. The proportions of third presentations within each block were .233, .332, .325, .319, .325, and .321, with a list mean of .309.

The subjects were instructed to press Key 1 to the first presentation of a word (1P tests), press Key 2 to the second presentation (2P tests), and press Key 3 to the third presentation (3P tests). The subjects used the index finger of the right hand in responding. The subjects initiated the presentation of each list. Each word in a list was presented for 2,500 msec, with a 500-msec blank interval between words. If a subject failed to respond in the time available (3 sec), a Key 4 "response" was recorded.

Each subject was encouraged to respond both as accurately and as quickly as possible, but the emphasis was placed on accuracy. The subjects received feedback on the total number of correct responses in the form of a summary table printed at the end of each session. Each session took less than 1 h to complete, and no subject did more than one session per day.

The final session was a choice reaction time task. This session also consisted of three lists. Each list was composed of the digits 1, 2, and 3, each presented 110 times in random order. Each digit was presented until a response was made, and there was a 500-msec blank interval between each response and the presentation of a digit. The subjects were instructed to press the appropriate key as quickly as possible for each digit presented. This session was included to determine if choice reaction time for each key on the response panel was comparable.

Results

The choice reaction time task showed only small differences between responses. The mean reaction times (in milliseconds) and their standard errors were 462 (3.7), 442 (3.4), and 454 (3.4) for Keys 1, 2, and 3, respectively.

Of the 47,520 total possible responses for the experimental sessions, there were 354 Key 4 "responses." There were also 46 responses that were recorded with a response latency of zero. A zero response latency could occur when a subject had a finger resting on a response key when a word was presented. All of these responses were excluded from the following data analyses.

The proportions of each response and the mean response for each test type are presented in Table 1. These results show that the subjects were quite accurate in discriminating among 1P, 2P, and 3P tests. The

Table 1
The Proportion of Each Response, Mean Response Time (RT, in Milliseconds), and the Mean Response for Each Test Type in Experiment 1

Test Type	Response										
	1			2			3			Mean Response	Total N
	Pro-portion	Mean RT	RT SE	Pro-portion	Mean RT	RT SE	Pro-portion	Mean RT	RT SE		
1P	.971	868	2.0	.025	1277	18.1	.003	1227	58.5	1.03	17248
2P	.073	1075	11.7	.854	970	2.5	.073	1149	11.0	2.00	15250
3P	.024	1135	22.3	.234	1142	6.1	.742	968	2.7	2.72	14622

Table 2
Mean Response Time (RT, in Milliseconds) and the Proportion of Each Response for 2P Tests Conditionalized on the Previous 1P Response and 3P Tests Conditionalized on the Previous 2P Response

Previous Response	Response										Total N
	1			2			3			Mean Response	
	Pro-portion	Mean RT	RT SE	Pro-portion	Mean RT	RT SE	Pro-portion	Mean RT	RT SE		
2P											
1	.073	1069	11.9	.866	968	2.5	.061	1164	12.1	1.99	14614
2	.102	1200	68.0	.422	1148	27.7	.476	1082	28.8	2.37	372
3	.086	1396	167.7	.371	1257	92.1	.543	1084	60.1	2.46	70
4	.046			.902			.052			2.01	194
3P											
1	.076	1104	40.6	.697	1042	12.7	.227	1123	21.8	2.15	1068
2	.019	1130	27.6	.205	1170	7.1	.776	966	2.9	2.76	12426
3	.026	1307	99.6	.102	1197	35.0	.872	951	9.2	2.85	1073
4	.145			.400			.455			2.31	55

Note—A "response" of 4 refers to a test in which the subject failed to respond within 3 sec.

proportion of correct responses decreased as frequency increased. For 2P tests, incorrect responses were evenly divided between 1P and 3P judgments. The majority of incorrect responses to 3P tests were 2P judgments. The mean latency for each response for each test type is also included in Table 1. The mean latencies for correct responses were faster than the latencies of incorrect responses. The mean latency of correct 1P responses was 100 msec faster than the mean latencies of correct 2P and 3P responses. The mean latency of correct 2P and 3P responses did not differ.

The results of the responses to 2P and 3P tests presented in Table 1 were derived without consideration of the previous responses to the items. To determine the effect of the prior response on the subsequent response, the responses to 2P tests were conditionalized on the previous 1P response, and the responses to 3P tests were conditionalized on the previous 2P response. The results of this analysis are presented in Table 2. Accuracy was highest and response latency was fastest for correct responses when the previous response was correct. However, there was also a strong consistency between frequency judgments. In general, the subjects' frequency judgment for a current test item was more likely to be consistent with the previous judgment for that test item. Thus, for a 2P test item, when the previous presentation was incorrectly identified as a second presentation, the second presentation was more likely to be judged incorrectly as a third presentation. Similarly, for 3P test items, when the previous 2P presentation was incorrectly identified as a first presentation, the 3P test item was more likely to be incorrectly judged as a second presentation. The consistency of the frequency judgments can also be seen in the pattern of mean responses—the mean response increased as a function of the previous response. The mean response times presented in Table 2 are consistent with this response pattern. The response times for consistent frequency judgments were faster than the response times for

inconsistent frequency judgments. The exceptions to this pattern were the 3P responses conditionalized on a previous (incorrect) 3P response. However, the subjects were not able to make a 4P response.

List block analysis. Each list was arbitrarily subdivided into six consecutive blocks of 55 tests each so that performance across list position could be evaluated. The proportion of correct responses for 1P, 2P, and 3P tests and the mean latencies of correct responses are presented in Figure 1. Accuracy and correct response latency did not vary systematically as a function of list block, which indicates that performance was reasonably stable across list position. Overall accuracy declined as item frequency increased. Mean response latency did not differ appreciably between correct 2P and 3P responses, whereas mean response time for correct 1P responses was consistently faster.

Lag analyses. The proportion of each response for 2P tests as a function of P1-P2 lag and for 3P tests as a function of P2-P3 lag is presented in Figure 2. The proportion of correct 2P and 3P responses decreased as a function of the most recent test lag. For 2P tests, the proportion of incorrect Key 1 responses increased as a function of P1-P2 lag, whereas the proportion of Key 3 responses stayed relatively constant. For 3P tests, the proportion of both incorrect Key 1 and Key 2 responses increased as a function of P2-P3 lag.

The mean latencies of correct 2P and 3P responses as a function of their most recent test lag are presented in Figure 3. Both 2P and 3P response latencies were negatively accelerated monotonic functions of the most recent test lag. These functions were also derived for each individual subject, and all subjects showed similar patterns.

The data presented in Figure 3 were fit by exponential, hyperbolic, logarithmic, and piecewise linear¹ functions by the method of least squares. (For the piecewise linear function, the two limbs of the function were fit separately for all possible breakpoints, and the best

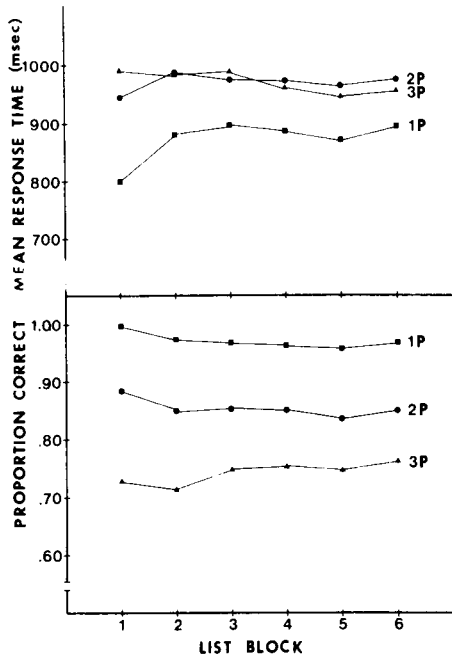


Figure 1. Correct mean response time and proportion of correct responses for each test type for each consecutive list block of 55 tests for Experiment 1.

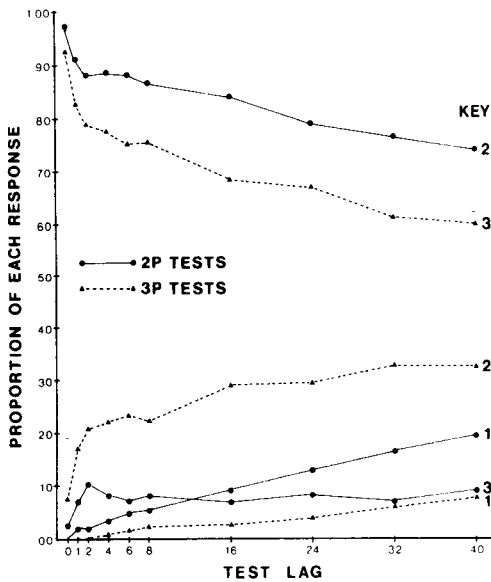


Figure 2. Proportion of each response for 2P tests as a function of P1-P2 lag and for 3P tests as a function of P2-P3 lag for Experiment 1.

overall fit was determined.) The respective root-mean-square deviations for these functions were 73.45, 23.30, 47.57, and 7.12 for the 2P functions, and 78.03, 29.70, 54.85, and 3.93 for the 3P functions. The piecewise linear function was clearly the best fit. For the upper limb of the functions, the slopes of the 2P and 3P functions were significantly greater than zero [$t(6) =$

9.594, $p < .001$, and $t(6) = 13.506$, $p < .001$, respectively], and the 2P and 3P slopes were not significantly different, the ratio of the difference to the standard error of the difference being 1.98 ($p > .05$). The best-fitting piecewise linear function averaged over both 2P and 3P response latencies is also presented in Figure 3.

It is possible that sequential response effects may be at least partially responsible for the piecewise linear function presented in Figure 3. For 2P tests with a P1-P2 lag of zero, the majority of responses to the immediately preceding test item would be Key 1 responses. Similarly, for 3P tests with a P2-P3 lag of zero, the majority of previous responses would be Key 2 responses. For lags greater than zero, the prior response would be approximately evenly distributed among the response alternatives. Thus, if there is an advantage to responding n when the previous response was $n-1$, then such an effect would decrease response latencies for lags of zero. To test for such a sequential response effect, all correct 2P responses were conditionalized on a prior 1P response, and all correct 3P responses were conditionalized on a prior 2P response. Linear functions were then fit to the conditionalized mean response times averaged over 2P and 3P responses. The linear function for lags 0 and 1 was $RT = 711.5 + 176.9L$, and the linear function for lags 2 to 40 was $RT = 985.6 + 2.30L$, where L refers to test lag. The linear functions are similar to the functions given in Figure 3. Thus, it does not seem likely that sequential response effects greatly influenced the pattern of results presented in Figure 3.

To determine whether P1-P2 lag affected 3P frequency judgments, the accuracy and latency of 3P responses were calculated as a function of P1-P2 lag collapsed across P2-P3 lag. These results are given in Table 3. Overall, P1-P2 lag had only a small effect on 3P responses. It has been previously observed (e.g., Proctor, 1980; Wells, 1974) that the spacing of repeated items

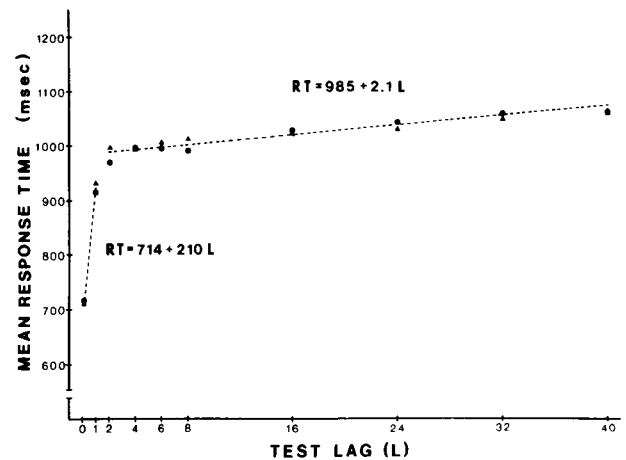


Figure 3. Mean response times and best-fitting piecewise linear functions for correct 2P responses as a function of P1-P2 lag and correct 3P responses as a function of P2-P3 lag for Experiment 1.

influences performance such that massed presentations (P1-P2 lag = 0) result in poorer performance than do spaced presentations (P1-P2 lag > 0). This effect was observed in the pattern of results for proportion correct and for mean response, but not in the pattern of results for mean response time.

Discussion

Both the accuracy and latency of 2P and 3P continuous judgments of frequency were a function of test lag. For 3P tests, the most recent test lag was the principal determinant of performance as the earlier P1-P2 test lag had only a small effect on 3P accuracy. The latency of correct 2P and 3P responses did not differ appreciably over the lag range tested, although the level of accuracy was higher for 2P judgments.

The striking discontinuity of the 2P and 3P lag-latency functions suggests a discontinuity in the retrieval of frequency information between lags 1 and 2. This discontinuity may be the result of the availability of different sources of information across lag on which to base frequency judgments. Ratcliff and Hockley (1980) argued for the availability of two different types of information in item recognition: item information concerning previous occurrences of an item, and response information. Response information allows the subject to remember what response had been made to an earlier test of the item. Response information is lost much more quickly than item information. Thus, it is possible that, in Experiment 1, the subjects could utilize response information at short lags to make a frequency judgment. At longer lags, when response information is no longer available, decisions of item frequency may have to be based solely on item information.

The discontinuity may also result from the differential availability of information from short-term and long-term memory. At short lags, frequency information in short-term memory may be the most accessible. At longer lags, frequency information from long-term memory may have to be retrieved. It may also be the case that response information is available only in short-term memory.

Experiment 2 was designed to test further the discontinuous nature of the 2P and 3P lag-latency functions. It was thought that, if a variable could be found that had no effect on one limb of the piecewise linear function but did affect the other limb of the function, further support for the discontinuity of the lag-latency functions would be provided. The additional variable included in Experiment 2 was a noun/nonnoun stimulus manipulation. Hockley (1982) and Hockley and Corballis (1982) found that nonnouns were responded to more slowly than were nouns in item recognition. It is reasonable to assume that this manipulation affected the availability of item information, because it does not seem likely that response information would differ for nouns and nonnouns. Thus, it was predicted that, if judgments of frequency are based on response information at short lags, then there would be no difference between nouns and nonnouns at short lags. However, if frequency judgments are based on item information at longer lags, then frequency judgments for nonnouns would be expected to be slower than frequency judgments for nouns.

In Experiment 1, item presentation time was fixed. Consequently subjects only had a limited time in which to respond. This was done in order to equate presentation time per item. Even though there were relatively few tests in which subjects failed to respond in the time available, the response deadline may have influenced the results. Thus, in Experiment 2, subjects were required to respond to every test item and were given unlimited time in which to respond.

EXPERIMENT 2

Method

Subjects. Four subjects each completed 1 practice and 12 experimental sessions.

Stimuli. The 1,024-word version of the Toronto Wordpool, which is evenly divided between nouns and nonnouns, was used as the source of stimulus items.

Procedure. The procedure of Experiment 2 was identical to the procedure of Experiment 1, with two exceptions. First, each list was composed of words from the noun and nonnoun portions of the wordpool. Second, the presentation rate was partly subject paced in order to ensure that the subjects would re-

Table 3
Correct Mean Response Time, Proportion of Correct Responses, and Mean Response for 3P Tests
as a Function of P1-P2 Lag for Experiment 1

P1-P2 Lag	Response Time		Proportion Correct	Mean Response	Total N
	Mean	SE			
0	944	7.8	.625	2.58	1491
1	981	7.7	.743	2.72	1502
2	957	7.3	.777	2.76	1494
4	968	7.4	.774	2.76	1477
6	956	6.8	.759	2.74	1475
8	982	7.3	.780	2.76	1470
16	984	7.6	.752	2.73	1452
24	980	7.6	.751	2.73	1440
32	964	7.6	.732	2.71	1423
40	964	7.1	.723	2.70	1398

spond to each list item. Each word appeared for a minimum duration of 2,500 msec. However, the display persisted until the subject had responded. There was a 500-msec blank interval between each presentation of a word.

Results

There were 32 responses recorded with a latency of zero that were excluded from the data analyses. In Experiment 2, the subjects had unlimited time in which to respond. To facilitate comparisons between Experiments 1 and 2, all responses with a latency greater than 4 sec were also excluded from the data analyses. This resulted in excluding 446 responses to nouns and 471 responses to nonnouns. (subject 3 contributed the majority of these responses, with 437 responses to nouns and 458 responses to nonnouns greater than 4 sec.)

The proportions of each response, the mean response, and the mean response time for each test type are presented in Table 4. With regard to accuracy, the pattern of results is similar to that obtained in Experiment 1. Nouns and nonnouns did not differ with respect to the distribution of responses for 1P tests. However, for 2P and 3P tests, the subjects underestimated the frequencies of nonnouns slightly more than those of nouns.

Correct responses were again faster than incorrect responses for each test type. Mean correct response time increased as item frequency increased. However, the large difference between correct 1P responses and correct 2P and 3P responses observed in Experiment 1 was attenuated in Experiment 2. There was a 20- to 30-msec difference between correct responses to nouns and nonnouns. However, there were no systematic differences between incorrect responses to nouns and nonnouns. When the responses to each test type were conditionalized on the previous response to the item, the pattern of results was similar to that found in Experiment 1 (Table 2).

Lag analyses. The proportion of each response for 2P tests as a function of P1-P2 lag and for 3P tests as a function of P2-P3 lag for both nouns and nonnouns is presented in Figure 4. The pattern of results is very

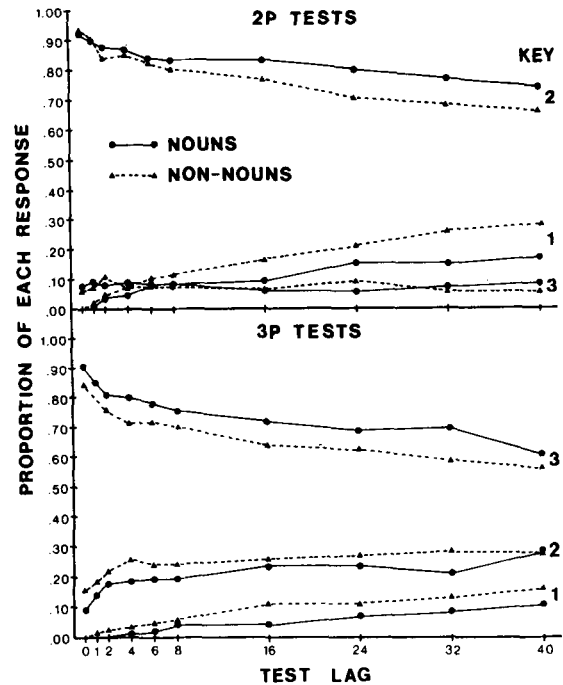


Figure 4. Proportion of each response for 2P tests as a function of P1-P2 lag and 3P tests as a function of P2-P3 lag for Experiment 2.

similar to the pattern of results obtained in Experiment 1.

Correct mean response time for 2P and 3P tests as a function of the most recent test lag for both nouns and nonnouns was calculated. An analysis of variance based on subject means showed that the main effects of test lag and stimulus type (nouns vs. nonnouns) were significant [$F(9,117) = 34.79, p < .001$, and $F(1,117) = 4.10, p < .05$, respectively]. The difference between 2P and 3P response latency was not significant [$F(1,117) < 1$]. None of the interactions approached significance.

Mean response times for nouns and nonnouns averaged over 2P and 3P responses as a function of the

Table 4
The Proportion of Each Response, Mean Response Time (RT, in Milliseconds), and the Mean Response for Each Test Type in Experiment 2

Test Type	Response										Total N
	1			2			3			Mean Response	
	Pro-portion	Mean RT	RT SE	Pro-portion	Mean RT	RT SE	Pro-portion	Mean RT	RT SE		
Nouns											
1P	.942	931	4.4	.044	1437	39.2	.014	1446	73.2	1.07	8554
2P	.081	1297	28.2	.839	950	4.7	.080	1153	21.8	2.00	7565
3P	.039	1269	38.1	.194	1233	15.1	.768	964	4.9	2.73	7288
Nonnouns											
1P	.940	962	14.1	.046	1431	38.9	.014	1249	56.9	1.07	8459
2P	.128	1273	20.9	.799	976	4.8	.073	1172	23.9	1.95	7500
3P	.066	1360	35.3	.236	1188	12.7	.698	986	5.3	2.63	7205

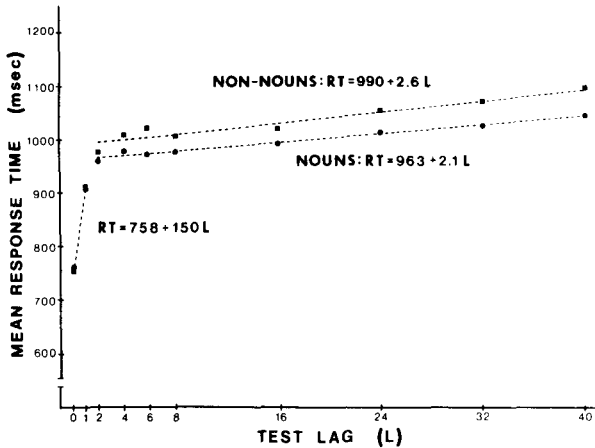


Figure 5. Mean response times and best-fitting piecewise linear functions for correct 2P responses as a function of P1-P2 lag and correct 3P responses as a function of P2-P3 lag for Experiment 2.

most recent test lag are presented in Figure 5. The lag-acy latency functions were again discontinuous. The best-fitting piecewise linear functions are included in Figure 5. The root-mean-square deviations of these functions were 3.41 for nouns and 9.51 for nonnouns. Mean response time did not differ greatly between nouns and nonnouns for lags 0 and 1. However, for lags 2 to 40, mean reaction time was greater for nonnouns, and this difference was reflected largely as an intercept difference between the two limbs of the linear functions. For the upper limbs of these functions, the slopes of the

noun and nonnoun linear functions were significantly greater than zero [$t(6) = 17.72, p < .001$, and $t(6) = 7.975, p < .001$, respectively]. The slopes were not significantly different [$t(12) = 1.56$], whereas the difference between the intercepts was significant [$t(12) = 2.60, p < .05$].

In Experiment 1, P1-P2 lag had only a small effect on the responses to 3P tests. A similar result was obtained in Experiment 2. Correct mean response time and the proportion of correct responses and the mean response for 3P noun and nonnoun tests for each P1-P2 lag averaged over P2-P3 lag are presented in Table 5. As in Experiment 1, the accuracy results of Experiment 2 show the typical advantage of spaced versus massed presentation. However, mean response time did not reflect this spacing effect.

Discussion

Experiment 2 replicated the principal findings of Experiment 1. In addition, Experiment 2 offered further support for the marked discontinuity of the 2P and 3P lag-latency functions. The noun/nonnoun stimulus comparison employed in Experiment 2 showed that response time did not differ between nouns and nonnouns for lags 0 and 1 (the lower limb of the function). However, nonnouns were responded to more slowly than were nouns across lags 2 to 40. This difference was reflected principally as a difference in intercept between the upper limbs of the linear functions.

It was suggested previously that the discontinuity of the lag-latency function reflects a difference in the

Table 5
Correct Mean Response Time, Proportion of Correct Responses, and Mean Response for 3P Noun and Nonnoun Tests as a Function of P1-P2 Lag for Experiment 2

P1-P2 Lag	Response Time		Proportion Correct	Mean Response	Total N
	Mean	SE			
Nouns					
0	1006	19.7	.690	2.63	765
1	986	16.0	.768	2.73	729
2	976	16.0	.784	2.75	751
4	977	17.1	.793	2.76	726
6	969	13.3	.782	2.75	737
8	960	14.6	.778	2.75	743
16	964	15.1	.777	2.73	689
24	938	13.7	.774	2.74	729
32	953	15.8	.772	2.75	703
40	910	13.0	.761	2.72	716
Nonnouns					
0	993	19.7	.625	2.53	718
1	996	17.4	.703	2.62	743
2	1019	17.6	.723	2.65	718
4	975	14.1	.718	2.66	731
6	1012	17.0	.714	2.66	723
8	978	15.2	.723	2.66	711
16	979	16.3	.721	2.65	770
24	965	16.3	.689	2.63	700
32	974	15.3	.697	2.65	715
40	971	17.0	.667	2.61	676

availability of response versus item information over test lag. The pattern of the differences between nouns and nonnouns indirectly supports this suggestion, because there is no reason to assume that response information differs between nouns and nonnouns, and thus there is no reason to expect a difference between nouns and nonnouns when the response is based on response information.

The results of Experiments 1 and 2 offer a reasonably clear and consistent description of performance in the continuous task. However, because of the limited range of frequencies tested, it is not clear how generalizable the results of Experiments 1 and 2 are. To make a correct 2P judgment, subjects must retrieve the information that the test item had been presented once previously. To make a correct 3P frequency judgment, subjects need only to retrieve enough information to ensure that the item had been repeated previously. However, to distinguish among 3P, 4P, and 5P test items, subjects would have to retrieve specific information on the number of previous presentations.

The purpose of Experiment 3 was to measure frequency estimation over a greater range of item frequency. Items were presented five times within each list, and subjects were required to distinguish between item presentations over this range.

EXPERIMENT 3

Method

Subjects. Four subjects each completed 1 practice and 12 experimental sessions, and 2 choice reaction time sessions.

Stimuli. For each frequency-judgment session, all words were selected randomly without replacement from the 512-word noun pool used in Experiment 2.

Procedure. Each judgment-of-frequency session consisted of three lists of words. Each list was 320 items long and consisted of approximately 64 words, each presented five times. The four lags between repetitions were manipulated independently. The possible lag values were 0, 1, 2, 4, 8, 12, 16, and 20.

In practice, not every repeated item could be fit into each list at the specified lag intervals. In these instances, filler items were presented. Filler items were also presented a maximum of five times. However, the lags between repeated filler items could not be controlled.

In order to determine whether item repetitions were distributed uniformly throughout the lists, the total number of each test type was calculated for each consecutive list block of 64 items within each list. The proportion of each test type for each list block is given in Table 6. With the exception of Block 1, the proportion of each test type within each list block was comparable. The proportions necessarily had to differ in Block 1, because repetitions had to occur in order at the beginning of a list.

The presentation rate was partly subject paced. Each word in a list appeared for a minimum duration of 2,500 msec. However, the display persisted until the subject had responded. There was a blank interval of 500 msec between presentations of a word.

The subjects indicated their frequency decision for each test item by pressing one of five keys on a response panel. The keys were arranged in a semicircle equidistant from a central rest position. The keys were arranged in order from Key 1 on the left to Key 5 on the right. The subjects used the index finger of the right (dominant) hand in responding. Each subject was en-

Table 6
The Proportion of Each Test Type for Each Consecutive List Block of 64 Items for all Experimental Lists of Experiment 3

Test Type	List Block					Total
	1	2	3	4	5	
1P	.253	.199	.199	.201	.201	.210
2P	.225	.202	.199	.200	.200	.205
3P	.199	.202	.199	.200	.197	.200
4P	.174	.200	.199	.201	.199	.195
5P	.149	.196	.204	.198	.203	.190

couraged to respond both as accurately and as quickly as possible, but the emphasis was placed on accuracy. The subjects received feedback on the total number of correct and incorrect responses for each test type in the form of a summary table printed at the end of each list.

The first and final sessions consisted of a choice reaction time task. These sessions consisted of four lists of the digits 1 to 5, each presented 64 times in random order. Each digit was presented until a response had been made, and there was a 500-msec blank interval between each response and the presentation of a digit. The subjects were instructed to press the appropriate key as quickly as possible for each digit presented. These sessions were included to determine the average choice reaction time for each key on the response panel. It was also hoped that the first choice reaction time session would help familiarize the subjects with the response panel prior to the beginning of the frequency-judgment sessions.

Results

Choice reaction time. Mean reaction time for each response for each of the two choice reaction time sessions is presented in Table 7. Inspection of the means in Table 7 shows that mean reaction times differ between the response keys. Because of these differences, comparisons between mean response times for different frequency judgments must be made with caution.

Judgment of frequency. There were 46,080 responses in the experimental sessions. The results of one list (320 tests) were lost due to a programming error. Of the remaining 45,760 responses, there were 9 (.02%) responses recorded with a latency of zero, 1,243 (2.72%) responses with a latency greater than 4 sec, and 6,489 (14.18%) responses to repeated filler items. (The lags of repeated filler items were not manipulated experimentally.) All of these responses (16.92% of the total possible responses) were excluded from the following data analyses.

Table 7
Mean Choice Reaction Time for Each Response Key for Experiment 3

Response Key	Session 1		Session 15		Mean	SE
	Mean	SE	Mean	SE		
1	550	17	487	6	519	9
2	556	6	511	8	534	5
3	536	6	501	10	519	6
4	569	7	499	6	534	5
5	586	11	552	11	569	8

Table 8
The Proportion of Each Response, Mean Response Time, and the Mean Response for Each Test Type of Experiment 3

Test Type	Response															Mean Response	Total N
	1			2			3			4			5				
	Pro-portion	Mean RT	RT SE	Pro-portion	Mean RT	RT SE	Pro-portion	Mean RT	RT SE	Pro-portion	Mean RT	RT SE	Pro-portion	Mean RT	RT SE		
1P	.959	1031	6	.028	1776	51	.008	1785	98	.005	1936	139	.001	1494	239	1.06	9362
2P	.028	1295	55	.774	1114	8	.149	1490	20	.041	1459	42	.007	1629	111	2.22	7483
3P	.006	1372	110	.092	1346	26	.659	1211	9	.206	1387	17	.037	1461	44	3.17	7253
4P	.003	1730	189	.020	1489	63	.214	1424	18	.617	1275	10	.146	1374	21	3.88	7043
5P	.001	1628	227	.008	1535	104	.075	1424	32	.412	1333	12	.504	1219	10	4.41	6878

The proportion of correct and incorrect responses, the mean frequency response, and the mean response time for each test type are presented in Table 8. The subjects were able to discriminate between item frequencies with reasonable accuracy. The proportion of correct responses decreased as item frequency increased. The subjects tended to slightly overestimate item frequency for 1P, 2P, and 3P tests and to underestimate frequency for 4P and 5P tests. Mean response time was faster for correct responses than for incorrect responses for each test type.

In Experiments 1 and 2, it was found that frequency judgments tended to be consistent with the previous response to a given item. In order to examine response consistency in Experiment 3, the responses to repeated items were conditionalized on the response to the previous test of each item. The results of this analysis are presented in Table 9. (Due to the small number of observations in many cells of Table 9, mean response times are not included.) The results of Experiment 3 also provide evidence for response consistency. In

general, responses tended to be consistent with the previous response to a given item regardless of the accuracy of the previous response. (When the previous response was a 5P response, the subsequent response also tended to be a 5P response. However, the subjects could not make a 6P response.) The tendency toward response consistency can also be seen in the mean frequency responses included in Table 9. The mean frequency response increased in direct relation to the previous response.

List block analysis. Each list was subdivided into five consecutive blocks of 64 tests each in order to evaluate performance across list position. The proportion of correct responses and the mean latency of correct responses for each test type are presented in Figure 6. Performance was reasonably stable across list position.

Lag analyses. The proportion of correct responses and the proportions of the two most frequent incorrect responses for each test type are presented in Figure 7 as a function of the most recent test lag. The proportion

Table 9
Proportion of Each Response and Mean Response for Each Test Type Conditionalized on the Previous Response for Experiment 3

Test Type	Previous Response	Response					Mean Response	Total N
		1	2	3	4	5		
2P	1	.028	.801	.136	.031	.004	2.18	7155
	2	.036	.182	.573	.186	.023	2.98	220
	3	.051	.153	.203	.492	.102	3.44	59
	4	.054	.189	.162	.297	.297	3.59	37
	5	.083	.500	.167	.083	.167	2.75	12
3P	1	.014	.599	.309	.063	.014	2.46	207
	2	.005	.084	.781	.122	.009	3.05	5592
	3	.005	.050	.257	.619	.069	3.70	1101
	4	.013	.064	.211	.366	.346	3.97	298
	5	.018	.036	.127	.145	.673	4.42	55
4P	1	.000	.405	.333	.214	.048	2.90	42
	2	.002	.053	.656	.269	.021	3.26	665
	3	.003	.015	.192	.740	.050	3.82	4670
	4	.001	.009	.107	.454	.429	4.30	1417
	5	.008	.012	.040	.241	.699	4.61	249
5P	1	.000	.333	.286	.286	.095	3.14	21
	2	.007	.043	.489	.418	.043	3.45	141
	3	.001	.009	.136	.749	.105	3.95	1506
	4	.001	.005	.048	.345	.600	4.54	4212
	5	.002	.005	.031	.189	.773	4.73	998

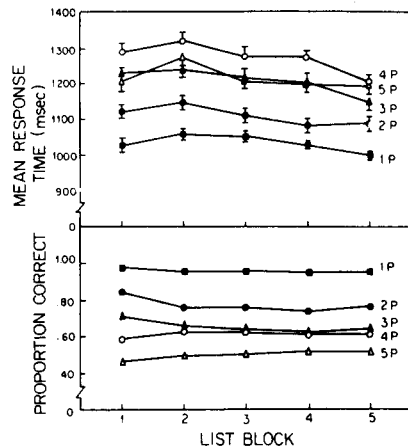


Figure 6. Correct mean response time and proportion of correct responses for each test type for each consecutive list block of 64 tests for Experiment 3. The vertical bars represent the standard errors of the response time means.

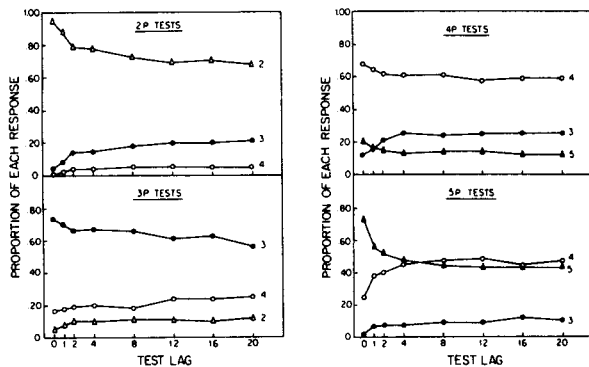


Figure 7. Proportion of correct responses and the two most frequent incorrect responses for each test type as a function of the most recent test lag for Experiment 3.

of correct responses decreased as a function of test lag.

Mean response times for correct responses as a function of the most recent test lag are presented in Figure 8. The lag-latency functions for each test type were again discontinuous. The best-fitting piecewise linear functions are included in Figure 8. The root-mean-square deviations of the 2P, 3P, 4P, and 5P functions were 11.52, 26.99, 23.61, and 23.36, respectively. The slopes of the upper limbs of the linear functions for 2P and 3P responses were significantly greater than zero [$t(4) = 7.902, p < .01$, and $t(4) = 3.40, p < .05$, respectively]. The slopes of the 4P and 5P functions failed to reach significance [$t(4) = 1.43$ and $t(4) = 2.63$, respectively].

Practice effects. Hasher and Chromiak (1977) failed to observe an effect of practice over two study-test trials in an absolute judgment of frequency task, and Zacks, Hasher, and Sanft (1982) did not obtain a practice effect over four study-test trials in a forced-choice frequency discrimination task. Because of the multi-session procedure used in the present study, the subjects had ample opportunity to benefit from practice. Thus,

the present study provided an ideal situation in which to evaluate practice effects in frequency estimation.

To evaluate practice effects, consecutive blocks of sessions of Experiment 3 were analyzed separately: Sessions 1 to 4 (Session 1 being the practice session), Sessions 5 to 9, and Sessions 10 to 13. The proportion of correct responses and the mean correct response times for each test type for each block of sessions are presented in Table 10. The proportion of correct responses increased and the mean response time decreased over session blocks. These results demonstrate that continuous frequency estimation does improve with extended practice.

Discussion

Experiment 3 was designed to test the generality of the results of Experiments 1 and 2 over a greater range of item frequency. The overall pattern of results obtained in Experiment 3 was consistent with the general findings of the previous experiments. There are, however, two differences between the results of Experiment 3 and the previous experiments. First, in Experiments 1 and 2, mean response time for correct 1P responses was faster than correct 2P and 3P response times, whereas there was little difference between 2P and 3P response times. In Experiment 3, mean correct response time increased as frequency increased for frequencies 1 to 4. However, 5P response time was faster than 4P response time. [Because of the differences between response times for each key in the choice reaction time task, it is difficult to compare judgment of frequency response times in Experiment 3. However, if the mean choice reaction time for each key is subtracted from the mean response time for each frequency judgment, the obtained "corrected" response times (in milliseconds) are 512, 580, 692, 741, and 650 for 1P, 2P, 3P, 4P, and 5P responses, respectively.] Taken together, the results of Experiments 1, 2, and 3 suggest that response times for correct frequency judgments

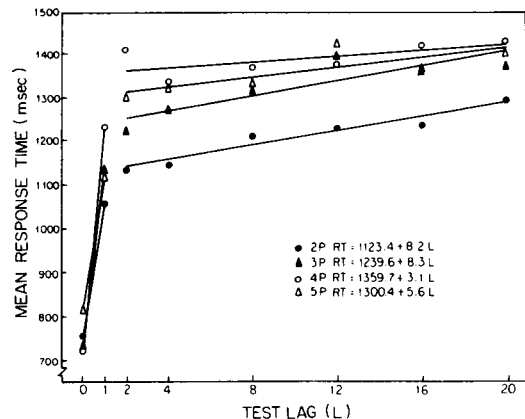


Figure 8. Correct mean response times and best-fitting piecewise linear functions for each test type as a function of the most recent test lag for Experiment 3.

Table 10
Proportion of Correct Responses and Mean Correct Response Times for Consecutive Blocks of Sessions for Experiment 3

Test Type	Proportion Correct			Mean Response Time					
	Sessions			Sessions					
	1-4	5-9	10-13	1-4		5-9		10-13	
			Mean	SE	Mean	SE	Mean	SE	
1P	.931	.966	.970	1093	10	1032	10	968	11
2P	.673	.788	.816	1200	15	1123	12	1045	12
3P	.565	.673	.701	1320	17	1206	14	1139	13
4P	.554	.624	.650	1387	19	1274	15	1188	15
5P	.465	.523	.511	1338	20	1216	16	1117	16

increase as frequency increases. The exception to this pattern is the response time for the maximum frequency tested. For whatever reason, the maximum frequency tested within an experiment appears to benefit from an advantage that is reflected in reduced response times.

The second difference concerns the lag-latency functions for correct responses. In all experiments, the lag-latency functions were discontinuous. However, in Experiment 3, although the slopes of the upper limbs of the 4P and 5P functions were positive, they did not differ significantly from zero. This suggests that frequency estimation becomes less of a function of the most recent test lag as frequency increases. However, accuracy was a function of the most recent test lag for all frequencies. The nonsignificant 4P and 5P slopes could be the result of the increased variability of response times as frequency increased (and, consequently, as the difficulty of frequency discrimination increased).

GENERAL DISCUSSION

The principal findings of the present study can be summarized as follows:

(1) Frequency judgments tended to be consistent with the previous response to a given item, regardless of the accuracy of the previous response.

(2) The proportion of correct responses decreased as frequency increased. Mean response time for correct frequency judgments increased as frequency increased, with the exception of the maximum frequency tested. Voss et al. (1975) also found that mean response time for absolute frequency judgments increased for frequencies varying from 2 to 8.

(3) The accuracy of frequency judgments and the latency of correct frequency judgments were found to be largely a function of the most recent test lag. As test lag increased, the proportion of correct responses decreased, whereas mean correct response time increased. Wells (1974) also observed a decrease in the probability of a correct frequency judgment as a function of test lag in a continuous task. A novel finding of the present study is the striking discontinuity of the lag-latency functions. As discussed previously, this discontinuity could result from a difference in the avail-

ability of different sources of information (i.e., response vs. item information) or a difference in the accessibility of information (short-term vs. long-term memory).

(4) The analysis of performance \times blocks of sessions in Experiment 3 demonstrated that subjects were able to benefit from extended practice in continuous frequency estimation. Practice effects have not been observed in frequency discrimination studies using a study-test procedure for absolute frequency judgments (Hasher & Chromiak, 1977) and for forced-choice frequency judgments (Zacks et al., 1982). However, these studies examined practice effects over only a small number of trials.

The failure to obtain a benefit from practice in frequency discrimination has been used as one of the criteria to suggest that the encoding of frequency information is an automatic process (Zacks et al., 1982). Thus, the finding of a practice effect in the present study suggests that the encoding of frequency information in the continuous task is not an automatic process.

A variety of hypotheses regarding how event frequency information is stored and retrieved have been distinguished (Begg, 1974; Flexser & Bower, 1975; Howell, 1973a). The three principal classes of theory are trace strength theory (Hintzman, 1969), multiple-trace theory (e.g., Hintzman, 1976), and the theories that propose that subjects attempt to encode frequency information directly by counting. The counting theories include the propositional-encoding hypothesis (Hintzman, 1976; Hintzman & Stern, 1978), which assumes that frequency information is encoded in a propositional form during study of the list, and the numerical-inference hypothesis (Howell, 1973a, 1973b), which assumes that subjects may attempt to encode frequency information in actual numerical terms. The results of the present study do not distinguish between these theories of frequency discrimination. However, the present results do pose constraints on the application of these theories to continuous judgments of frequency. The implications of the present results for these theories of frequency discrimination are discussed briefly below.

Trace strength theory (Hintzman, 1969) assumes that each presentation of an item increments the strength of a single memory trace that represents the item. Judg-

ments of frequency are based on the relative strength of the memory trace. There is now abundant evidence that subjects are capable of making finer discriminations than are possible on the basis of cumulative trace strength (e.g., Hintzman & Block, 1971). However, under certain circumstances, trace strength may provide a more convenient and readily available basis on which to estimate item frequency. This may be especially true when response time is important.

Wells (1974) has effectively argued against a simple strength theory account of frequency estimation, which assumes that repetition increments the mean of the strength distribution without changing the variance of the distribution. However, as Wells noted, a strength model that assumes that both the mean and the variance increase proportionally with the number of repetitions could be developed. Such a model is presented in Figure 9. (A similar model was proposed by Flexser, 1976.) Each presentation of an item is assumed to increment the strength of the distribution, although the size of the increment decreases with repetition. The variance of the distribution increases with each repetition. The strength axis is divided by a series of equally spaced cutoffs corresponding to the possible frequency judgments. Thus, a frequency judgment is determined by the bounded region in which the strength of the item falls. It can be seen from Figure 9 that the proportion of correct frequency judgments would decrease as frequency increases.

In deriving response latency, it is assumed that latency is a function of the distance from a frequency boundary—the closer an item's strength value is to a boundary, the slower the response. The filled circles in Figure 9 represent estimated mean response time for each correct frequency judgment. The means are based on the proportion of items for each point on the re-

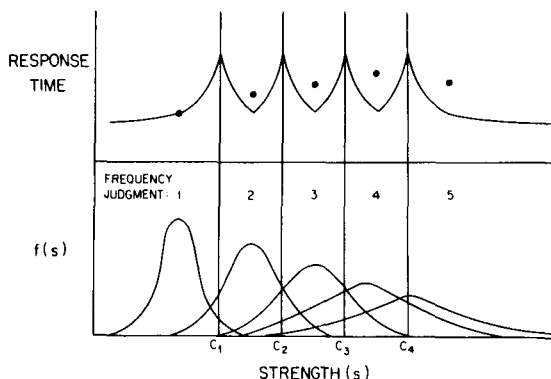


Figure 9. The lower panel presents hypothetical strength distributions corresponding to frequencies 1 to 5, and equally spaced cutoffs dividing the strength axis into frequency categories. The upper panel shows a hypothetical mapping between response time and distance from a frequency category boundary. The filled circles represent estimated mean correct response time for each frequency judgment based on the proportion of items for each point on the response latency function.

sponse latency function. Correct mean response time increases for each frequency judgment because, as frequency increases (and the variance of the strength distribution increases), the proportion of item strength values lying closer to the frequency boundary increases. The exception is the maximum frequency tested. Because this frequency category is not bounded on one side, there would be a decrease in correct mean latency. The mean latency of incorrect responses is longer than that of correct responses because the strength of items incorrectly classified would be, on average, closer to the frequency boundaries.

To account for the decrease in accuracy and the increase in response time as a function of test lag, the mean and variance of the strength distributions would also have to vary with lag. It is reasonable to assume that mean strength would decrease and the variance of the distributions would increase as test lag increased. The marked discontinuity of the lag-latency functions could also be accounted for if the additional assumption is made that, at very short lags, the availability of response information mediates frequency judgments.

According to the numerical-inference hypothesis proposed by Howell (1973a), subjects may attempt to keep a running count of item frequency under certain conditions. Zacks et al. (1982) obtained a small benefit in performance in forced-choice frequency discrimination by instructing subjects to adopt a counting strategy. The continuous frequency estimation procedure may encourage subjects to adopt such a counting strategy, because it may maximize the benefits of such a strategy (Flexser & Bower, 1975). Begg and Rowe (1972) suggested that, in the continuous frequency task, subjects may form paired associates in which the stimulus term is the item and the response term is the list frequency. On each presentation, subjects retrieve the relevant paired associate, increment the numerical associate by one, and give the new response. Such a counting strategy would effectively transform the continuous frequency judgment task into a continuous paired-associate task.

A paired-associate interpretation of continuous frequency estimation can account for many of the present results. First, the finding of response consistency between successive frequency judgments would be expected if subjects attempted to maintain a running count of the number of presentations of each item. Second, the finding that the accuracy and response latency of frequency judgments are largely functions of the most recent test lag also would be expected if it is assumed that the difficulty of retrieval of the last frequency count is a function of the number of intervening items. Finally, the discontinuity of the lag-latency functions is also easily accommodated if it is assumed that retrieval of the frequency count is easier from short-term memory than from long-term memory.

It is not obvious, however, how a paired-associate view can account for the increase in mean response time as a function of frequency. Why should the time to

retrieve and increment the last frequency count vary with the numerical value of the count? It could be assumed that prior frequency counts associated with an item interfere with the new updated frequency count associated with the item and that this interference would increase as frequency increased. In other words, continuous frequency estimation is analogous to an A-B, A-C, A-D paired-associate task. However, such an assumption could not explain the finding that the response time for the maximum frequency tested does not increase.

The multiple-trace theory (Hintzman, 1976; Hintzman & Stern, 1978) maintains that each occurrence of an item is represented in memory by its own memory trace. The multiple traces of an item coexist in memory and may be distinguishable on the basis of contextual or temporal information. Frequency judgments are derived from an estimate of the number of available traces. Thus, like strength theory, frequency information is not encoded per se but may be derived from the available information in memory.

The tendency for mean response time to increase with frequency is compatible with the multiple-trace hypothesis if it is assumed that retrieval time increases with the number of traces that are available. However, such an assumption implies that the memory search is terminated as soon as all relevant traces have been retrieved, and such a self-terminating process would require either foreknowledge of the number of traces available or information that would act as a reliable stop cue (Hintzman et al., 1981).

If response time is dependent largely on the time required to retrieve the available memory traces of an item, then mean response time for correct frequency judgments would not be expected to be largely a function of the most recent test lag. The present finding that response time is largely a function of the most recent test lag would thus strongly argue against a serial counting of memory traces. The present results also pose constraints on a search process that assumes memory traces are retrieved in parallel. If it is assumed that memory traces are retrieved in parallel and that retrieval time is independent for each memory trace, then again mean response time would not be expected to be principally a function of the test lag of the most recent memory trace. The present results thus suggest that the retrieval of the individual traces must, in some way, be interdependent.

The multiple-trace hypothesis cannot account for the discontinuous nature of the lag-latency functions on the basis of the distinction between short-term and long-term memory, because the search process must necessarily search both memory stores regardless of test lag. However, at short lags, subjects would not have to count memory traces if information regarding the previous response to an item was readily available.

It is apparent from the above discussion that the present results do not distinguish between the proposed theories of frequency discrimination. However, the present results do serve to delineate more narrowly

versions of these theories in their application to continuous frequency estimation.

Frequency information is assumed to underlie at least partially the ability to distinguish old items from new ones (e.g., Underwood, Zimmerman, & Freund, 1971). Thus, it is of interest to compare frequency judgments and recognition. Investigators that have compared performance on these two tasks have reached differing conclusions. Hintzman and Stern (1978) and Proctor (1977) have argued that, to a certain extent, recognition decisions and judgments of frequency must be based on different types of stored information or on different retrieval processes. However, Harris, Begg, and Mitterer (1980) concluded that both tasks are highly dependent measures of memory retention and use the same information in a similar way. Harris et al. suggested that frequency estimation includes a recognition process in which test items are first assessed as new or old prior to numerical estimation. This view assumes that judgments of frequency are controlled by at least two separable processes: memory for the prior occurrences of an item, and decisional processes that translate this memory into a frequency judgment.

Performance on recognition and judgment of frequency tasks have only been compared on the basis of measures of accuracy. The use of response time as a dependent variable permits another basis of comparison. The procedure used in the present study is very similar to the procedure used by Hockley (1982) in a study of continuous recognition performance. In both studies, accuracy and correct response time were found to be largely a function of the most recent test lag. However, the forms of the lag-latency functions were different in the two tasks. For correct frequency judgments, the lag-latency functions were discontinuous. Over the same range of test lags, the functions for correct 2P and 3P recognition decisions were approximated best by logarithmic functions. It is tempting to infer that, because both the latency and accuracy of recognition decisions and frequency judgments are largely functions of the most recent test lag, the same information is being retrieved in a similar fashion in the two tasks. The fact that the form of the lag-latency functions differs between the two tasks thus reflects differences in the decision processes underlying each task. However, before such a suggestion can be considered seriously, it is necessary to compare the two tasks directly within the same experiment.

The present study examined continuous frequency estimation using both accuracy and response time as dependent variables. Although the obtained results do not distinguish between current theories of frequency discrimination, they do provide constraints for any theoretical interpretation of frequency estimation. Further use of response time as a dependent variable in the study of frequency discrimination and in comparisons between judgment of frequency and recognition performance may provide greater insight into how repetitions of events are stored and retrieved.

REFERENCES

- ATKINSON, R. C., & JUOLA, J. F. (1973). Search and decision processes in recognition memory. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press.
- BEGG, I. (1974). Estimation of word frequency in continuous and discrete tasks. *Journal of Experimental Psychology*, **102**, 1046-1052.
- BEGG, I., & ROWE, E. J. (1972). Continuous judgments of word frequency and familiarity. *Journal of Experimental Psychology*, **95**, 48-54.
- FLEXSER, A. J. (1976). *Reaction times for comparing a frequency judgment with a preassigned criterion number*. Unpublished manuscript, University of Toronto.
- FLEXSER, A. J., & BOWER, G. H. (1975). Further evidence regarding instructional effects on frequency judgments. *Bulletin of the Psychonomic Society*, **6**, 321-324.
- FRIENDLY, M. L., FRANKLIN, P., & HOFFMAN, D. (1980). *Imagery and concreteness ratings for 1080 verbal items* (Tech Rep. No. 94). Toronto: York University, Department of Psychology.
- HARRIS, G., BEGG, I., & MITTERER, J. (1980). On the relation between frequency estimates and recognition memory. *Memory & Cognition*, **8**, 99-104.
- HASHER, L., & CHROMIAK, W. (1977). The processing of frequency information: An automatic mechanism? *Journal of Verbal Learning and Verbal Behavior*, **16**, 173-184.
- HINTZMAN, D. L. (1969). Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology*, **80**, 139-145.
- HINTZMAN, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10). New York: Academic Press.
- HINTZMAN, D. L., & BLOCK, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, **88**, 297-306.
- HINTZMAN, D. L., GRANDY, C. A., & GOLD, E. (1981). Memory for frequency: A comparison of two multiple-trace theories. *Journal of Experimental Psychology: Human Learning and Memory*, **7**, 231-240.
- HINTZMAN, D. L., & STERN, L. D. (1978). Contextual variability and memory for frequency. *Journal of Experimental Psychology: Human Learning and Memory*, **4**, 539-549.
- HOCKLEY, W. E. (1982). Retrieval processes in continuous recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **8**, 497-512.
- HOCKLEY, W. E., & CORBALLIS, M. C. (1982). Tests of serial scanning in item recognition. *Canadian Journal of Psychology*, **36**, 189-212.
- HOWELL, W. C. (1973a). Representation of frequency in memory. *Psychological Bulletin*, **80**, 44-53.
- HOWELL, W. C. (1973b). Storage of events and event frequencies: A comparison of two paradigms in memory. *Journal of Experimental Psychology*, **98**, 260-263.
- MURDOCK, B. B., JR., & ANDERSON, R. E. (1975). Encoding, storage, and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium*. Hillsdale, NJ: Erlbaum.
- PIKE, R., DALGLEISH, L., & WRIGHT, J. A. (1977). A multiple-observations model for response latency and the latencies of correct and incorrect responses in recognition memory. *Memory & Cognition*, **5**, 580-589.
- PROCTOR, R. W. (1977). The relationship of frequency judgments to recognition: Facilitation of recognition and comparison to recognition confidence judgments. *Journal of Experimental Psychology: Human Learning and Memory*, **3**, 679-689.
- PROCTOR, R. W. (1980). The influence of intervening tasks on the spacing effect for frequency judgments. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 254-266.
- RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.
- RATCLIFF, R., & HOCKLEY, W. E. (1980). Repeated negatives in item recognition: Nonmonotonic lag functions. In R. S. Nickerson (Ed.), *Attention and performance VIII*. Potomac, MD: Erlbaum.
- SHEPARD, R. N., & TEGHTSOONIAN, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, **62**, 302-309.
- STERNBERG, S. (1966). High speed scanning in human memory. *Science*, **153**, 652-654.
- UNDERWOOD, B. J., ZIMMERMAN, J., & FREUND, J. S. (1971). Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology*, **87**, 149-162.
- 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*, **104**, 337-350.
- WELLS, J. E. (1974). Strength theory and judgments of recency and frequency. *Journal of Verbal Learning and Verbal Behavior*, **13**, 378-392.
- ZACKS, R. T., HASHER, L., & SANFT, H. (1982). Automatic encoding of event frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **8**, 106-116.

NOTE

1. This function typically has been referred to as a bilinear function in the literature (e.g., Hockley & Corballis, 1982). However, as one reviewer pointed out, this function should be referred to as a piecewise linear function as it describes two different linear functions of only one variable.

(Manuscript received April 27, 1983;
revision accepted for publication September 27, 1983.)