Journal of Clinical and Experimental Neuropsychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ncen19

Evaluating evidence for automaticity in frequency of occurrence judgments: A bias for bias?

James R. Tweedy [a] & Eliyahu Vakil [b]
[a] Queens College, City University of New York ,
[b] University Medical Center and Queens College, City University of New York , New York
Published online: 04 Jan 2008.

Journal of Clinical and Experimental Neuropsychology
1988, Vol. 10, No. 5, pp. 664-674.

0168-8634/88/1005-0664 $3.00

COMMENT

Evaluating Evidence for Automaticity in Frequency of Occurrence Judgments: A Bias for Bias ?*

James R. Tweedy
Queens College, City University of New York

Eliyahu Vakil
New York University Medical Center
and
Queens College, City University of New York

ABSTRACT

Frequency of occurrence judgments were evaluated in young adults recovering from closed-head injuries, normal elderly, and young adult controls. Impaired performance was observed in both head-injured and elderly subjects, a result contrary to Hasher and Zacks' 1979 hypothesis that this information accumulates in memory via automatic processes which are unaffected by age, but supporting their conjecture that damage to the central nervous system would be sufficient to interfere with this function. The head-injured subject's performance on the frequency judgment task was correlated with effortful memory capacity as measured by several widely used memory tests. Whether the obtained group differences reflect differences in memory capacity or response criteria effects is discussed, and several methods of analyzing the data are compared.

The distinction between automatic and effortful memory processes has given rise to much recent research involving judgment tasks tapping memory for frequency of occurrence information. In their initial formulation Hasher and Zacks (1979) postulated that processes governing the encoding of frequency of occurrence information are automatic, and unaffected either by subject variables (e.g., age, ability) or task variables (e.g., instructions, practice) in contrast to the effortful processes required for performance in tasks requiring recognition or recall of item information.

* The authors thank Drs. Yehuda Ben-Yishay and Eugene Piasetsky of the New York University Head Trauma Program for their helpful advice and their cooperation in arranging for the participation of the head-injured subjects. The second author is now at the National Institute for Rehabilitation of Head Injured Persons, Tel-Aviv, Israel. Requests for reprints should be addressed to J. R. Tweedy, Department of Psychology, Queens College, Flushing, New York, 11367, USA.

Accepted for publication August 11, 1987

Proponants of the automatic-effortful distinction have amassed considerable evidence supporting this distinction in experiments with normal subjects, mostly in the form of null results such as failure to demonstrate age-related performance declines, lack of benefit from training, feedback, or practice, and the absence of a relationship between frequency of occurrence judgment accuracy and academic achievement. This evidence has recently been summarized by Hasher and Zacks (1984).

Studies involving depressed (Hasher & Zacks, 1979, exp. 3) and diabetic (Lichty, 1982) subjects also support the view that frequency of occurrence judgments are much less easily disrupted than performance on effortful recall tasks. Even in their early report, however, Hasher and Zacks speculated that damage to the central nervous system might be sufficient to interfere with automatic memory functions. They remark (1979, p. 372) that "automatic processes should be functioning under all conditions of consciousness, except perhaps where brain damage has occurred."

To investigate this possibility, we tested a group of young adults recovering from traumatic closed-head injuries (CHI) and an age-matched control group. In an attempt to replicate the age invariance result we included a control group of normal elderly. The effect of instructions was evaluated by showing each subject two lists, the first under very general memory instructions, the second immediately afterward, under instructions that the judgments required would be identical to those which followed the first list.

RESPONSE BIAS EXPLANATIONS

The literature on automatic memory processes contains several studies demonstrating group differences which dismiss the result as a response bias effect rather than a difference in accuracy, attributing both main effects for subject group and interactions of subject group and presentation condition to differing response biases between groups of subjects with essentially equivalent memory capacity. For example, Hasher and Zacks (1979) and Attig (1981) presented data indicating that elderly subjects made frequency estimates that were generally lower than young controls, but ascribed it to the tendency for elderly subjects to respond conservatively on such tasks. Similarly, McCormack (1981) presented functions relating mean judged temporal position to actual position which were further from ideal performance in the elderly, and an accompanying age by list position interaction which was significant, but he interpreted his finding as resulting from the elderly subjects' tendency to make greater use of responses near the center of the permissible set, and supported the interpretation with the results of a second experiment in which adjusting scores for the presence of such a response bias resulted in no age by list position interaction.

Interpreting group differences as resulting from differing response criteria preserves the unique nature of the automatic processes, but requires the assumption, currently without an empirical basis, that differences in response criteria are the cause for observed differences in performance rather than a result of diminished capacity. We feel it premature to take a position on the causal relationship of response criteria and capacity limitations, or on the contribution each makes to between-group differences on these tasks. Instead we present, via three parallel analytic procedures, each sensitive to a different aspect of the data, a more complete description of the empirical effects of age and traumatic head injury on frequency of occurrence judgments.

Hasher and Zacks' (1979, 1984) method of analyzing frequency judgments involves averaging a subject's frequency of occurrence estimates, and is therefore insensitive to the variability in the subject's responses to items in this set. Analyses based on the likelihood of a correct estimate or the actual size of judgment errors have been directly compared with analyses based on mean judgment values in several frequence estimation tasks. In studies by Alba, Chromiak, Hasher, and Attig (1980) and Brooks (1985), the results of the different analyses were essentially equivalent, but in a developmental study by Ghatala and Levin (1973), the analysis based on the variability of an individual subject's response revealed an age interaction that was not evident in the analysis based on average judgment values.

METHOD

Subjects

Group CHI consisted of 45 adults, 33 males and 12 females, who had suffered closed-head injury with loss of consciousness at least 1 year prior to testing. All attended an outpatient rehabilitation program at the Institute of Rehabilitation Medicine of the New York University Medical Center. The age of the group ranged from 18 to 49 years (M = 28.67 years, SD = 8.70 years), and they had an average of 14.0 years of education. All had either a VIQ or a PIQ of greater than 80 on the Wechsler Adult Intelligence Scale -Revised (Wechsler, 1981). All were ambulatory, and none were employed at time of testing. All the patients had cognitive impairments sufficient to require participation in a 4 day per week program, though none had a language-based deficit sufficient to interfere with understanding the requirements of the task.

Each CHI subject, on admission to the rehabilitation program, had been given a battery of neuropsychological tests which included the WAIS-R and several widely used memory tests: the Logical Memory and Associate Learning subtests of the Wechsler Memory Scale (Wechsler, 1945), the Auditory Verbal Learning Test (Rey, 1964), the Visual Retention Test (Benton, 1974) and the Lateralized Asymmetries in Visual Attention Test (Piasetsky, 1981).

The young control group consisted of 45 volunteers, 22 males and 23 females, All were undergraduates at Queens College, several from adult education classes. They ranged in age from 17 to 39 years (M = 26.8 years, SD = 5.49 years). The normal elderly group consisted of 25 people, 16 females and 9 males. All were volunteer workers in the rehabilitation program attended by the CHI subjects. The age range was 55 to 85 years (M = 67.0 years, SD = 8.84 years). None of the young controls or elderly subjects had a history of alcohol abuse, or of significant neurological or psychiatric illness. None had serious medical conditions or took medications with known effects on mentation at the time of testing.

Stimuli

Two sets of 52 high-frequency words from the Thorndike and Lorge (1944) AA list norms were selected. Each was used to construct two different versions of a 72-item presentation list. In each of the four different lists, the initial and final four items were fillers used to counteract primacy and recency artifacts. The body of each list consisted of 64 items, divided into four 16-item quarters, each containing 10 trials devoted to the presentation of four words later used in the frequency of occurrence test. One word was presented once, one twice, one three times, and one four times. The remaining trials in each quarter were devoted to six singly-presented words which were later employed in tests for memory of temporal order and spatial position.

Testing Procedure

Subjects were tested individually. Each was seated before a TRS-80 Color microcomputer. During the presentation of an acquisition list, the display screen was divided into quadrants by bars 6 mm wide. Each of the words in an acquisition list appeared in the center of one of the quadrants. They were presented at a 5-s rate, printed in uppercase characters 15 mm high. Each quadrant was used equally often in each of the four quarters of a list. Multiply presented words appeared only once in any quadrant, and all presentations of these words were confined to a single temporal quarter of the list.

After completion of the tests on items from the first list, the effect of instructions was evaluated by repeating the entire procedure. Initially, subjects had been *uninformed* about the specific types of memory tests they would encounter, being told only to "pay close attention to what is presented on the screen because later your memory will be tested." Prior to the presentation of a second acquisition list of new words in an identical structure and format, subjects were specifically *informed* that the tests which would follow it "were exactly the same as those used previously: tests of when, where, and how often" the words had been presented. This variation in instructions is of course confounded with whatever practice and/or fatigue effects operate within the experiment.

Following the presentation of each list, half of the subjects in each group made 20 frequency of occurrence judgments, followed by 24 temporal and spatial judgments. For the other half, the frequency judgments followed the temporal and spatial tasks. To test retention of frequency of occurrence information, the 16 words assigned to the frequency test (4 which had been presented four times, 4 three times, 4 twice, and 4 once) plus 4 novel foils were combined into a list of 20 test items which were presented in a random order one at a time, at the center of the display screen. Each remained in view until the subject responded verbally with a number between *0* and *4* to indicate the estimated number of times the word had been presented. After the experimenter entered this information on the computer's keyboard, the next test word appeared.

The temporal and spatial tests were both based on the same set of 24 singly presented words (six from each quarter) in the acquisition list which were not used in the frequency of occurrence tests. These temporal and spatial judgment data will not be presented here.

For the first 20 CHI and the first 20 young controls tested, retention of temporal-order information was evaluated by combining test words from different quarters of the acquisition list into 12 pairs, presenting them in a random order and requiring the subject to make a relative recency judgment on each pair by pointing to the word which had appeared last in the acquisition list. Memory for spatial information was then assessed by presenting each of the test words singly at the center of the display screen, and requiring the subject to identify the quadrant of original presentation on a small facsimile of the display screen as it appeared during the presentation of the acquisition list. Each quadrant of the facsimile was labled with the letters A through D, and it appeared randomly either left or right of the test word.

The temporal and spatial tests were both modified for the last 25 CHI subjects, the last 25 controls, and for the 25 elderly subjects tested. For these subjects retention of temporal information was evaluated by presenting each test word singly in the center of the screen, and requiring the subject to estimate the quarter of the acquisition list in which it had originally appeared. Memory for spatial information was tested by presenting each test word in its original quadrant and in the quadrant diagonal to it, and requiring the subject to indicate which of the two was the quadrant in which it had originally been presented.

Analysis Procedures

Average scores analysis. Frequency of occurrence estimates were first analyzed by the method popularized by Hasher and Zacks (1979, 1984), in which each subject contributes one average judgment value for every presentation frequency. This analysis will indicate the typical response to a set of items in a particular presentation condition, and the extent to which judged frequency increases with actual presentation frequency, but not the number of correct estimates or the amount of variability in an individual's judgments about the items in a particular condition.

Error analysis. An alternative analytic procedure can be based on the proportion of trials in which the subject gives an exactly correct estimate. In this analysis every subject contributes a set of error rates, one for each presentation frequency used in the experiment. This procedure, like the average score analysis, is insensitive to the size of the errors of estimation which are made. An additional difficulty is that, as the number of available response alternatives increases, the likelihood of a correct response declines. In tasks with many response alternatives, a success-error classification may result in success rates that are so low, and so variable over conditions, that effects contained in the original estimates are obscured by "floor" effects and error variance.

Deviation analysis. This method is based on the size of the unsigned difference between each item's actual judged frequency. In this procedure every subject contributes a difference score for every item on which a judgment is made. The difficulty with this analysis is that deviation scores can be larger for items presented at the extremes of the frequency variable than for items presented at a central value. (For example, when the items presented for testing have previously appeared 0, 1, 2, 3, or 4 times, the deviation scores for items presented either 0 or 4 times could be as high as 4, while for items presented twice the maximum deviation is 2.) This imbalance can be corrected by expressing performance at each level of the frequency variable as a ratio of the obtained deviation score and the one that would result from random responding. Under this transformation, scores range from 0 (perfect responding) to 1.0 (performance at the chance level).

RESULTS

The results of three different analytic procedures assessing performance accuracy will be compared: the average scores procedure, with an accompanying analysis of response category utilization like the one done by McCormack (1981), an error analysis based on the proportion of correct judgments observed, and a deviation analysis based on the actual size of judgment errors. Then correlations within each group between performance in the informed and uninformed conditions will be presented. Finally, the correlation between-frequency of occurrence judgments and several traditional measures of memory function in the head-injured group will be examined.

Average scores analysis

Figure 1a plots the functions relating judged to actual frequency of occurrence for each subject group, in a format used by Hasher and Zacks (1979, 1984). The data points are group means of each subject's averaged scores data, collapsed over task order and instructions, since in this analysis neither variable had an effect on performance.

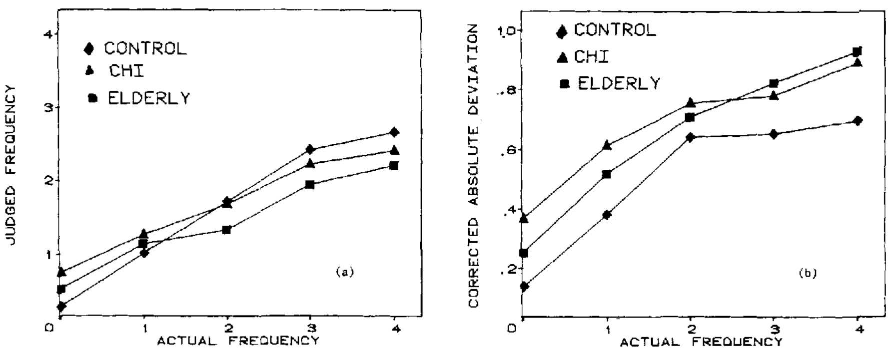

Fig. 1. Performance of three subject groups on the frequency of occurrence judgment task: averaged frequency judgments (a), mean sums of corrected deviation scores (b), plotted against actual presentation frequency.

The functions for each subject group are quite similar in shape, and no main effect for group was obtained. A highly significant ($F(4, 448) = 381.7, p < .001$) frequency effect indicates that for all groups mean frequency judgments increased with actual presentation frequency. All subjects overestimated the presentation frequencies of the novel foils, and underestimated the frequencies of items presented 2, 3, or 4 times.

Closer inspection of Figure 1a reveals that the young control subject's function starts lower and rises more steeply than those obtained from the elderly and CHI groups. This difference is reflected in a significant group-by-frequency interaction, $F(8, 448) = 8.61, p < .001$. Analyses using contrasts (Kirk, 1982) confirmed that the interaction resulted from control group's scores being generally closer to ideal performance than scores from the elderly and CHI groups, which could not be distinguished from each other.

The category utilization analysis examined the number of times each permissible response was used, irrespective of whether it was correct. It revealed no effect for group or for instructions, but a strong effect for response category ($F(4, 448) = 53.27, p < .001$), reflecting the tendency for all subjects to overuse responses *0, 1,* and *2,* and underuse *3* and *4*. The effect was somewhat heightened for the elderly subjects, relative to the young control and CHI subjects, which resulted in a subject group-by-response category interaction which just reached statistical significance $F(8, 448) = 1.96, p = .05$. All subjects were less inclined to use the *0* response in the informed condition (second list) than in the uninformed (first list) condition, which resulted in a significant interaction of instructions and response category, $F(4, 448) = 4.51, p < .001$.

Error analysis

An analysis ofvariance on the data coded as successes or errors was performed, and in contrast to the average scores analysis, a significant group difference was obtained, $F(2, 112) = 18.46, p < .001$. The young controls had a lower

Table 1

Correlations Between Several Effortful Tasks and Frequency of Occurrence Judgments in the Informed and Uninformed Condition and for the CHI Subjects

	Instructional Condition	
Test	Uninformed	Informed
Verbal Memory Measures		
Rey Auditory Verbal Learning Test		
Total recalled	.60**	.30
Improvement score	.56**	.21
Recognition score	.43	.84**
Wechsler Memory Scale		
Logical Memory subtest	.35**	.24
Associate Learning subtest	.22	.09
Nonverbal Memory Measures		
BentonVisual Retention Test	.41**	.12
Lateral Asymmetries in Visual Attention Test		
Match-to-sample admin.	.78**	-.10
Memory admin.	.43*	.09
WAIS-R Intelligence Measures		
Verbal IQ	.13	.12
Performance IQ	.25*	.04

$* = p < .05$ $** = p < .01$

proportion of errors (.55) than either CHI (.67) or elderly (.65) subjects. No instructions effect or subject group-by-instructions interaction was observed. A highly significant frequency effect ($F(4, 448) = 75.18, p < .001$) indicated that errors were more common on multiply presented items than on the novel foils and once-presented items. Differences between CHI subjects and controls and between elderly subjects and the young controls were larger for the novel foils than for items which were actually presented in the acquisition list, resulting in a significant interaction of frequency with group, $F(8, 448) = 2.83, p < .005$. A significant interaction of frequency and instructions ($F(4, 448) = 4.32, p < .002$) reflects the fact that error rates increased selectively for the novel foils between the initial uninformed and the subsequent informed portions of the experiment.

Deviation analysis

Figure 1b presents the functions relating group means of the sum of the unsigned deviation scores for every subject at every value of the presentation frequency variable, corrected for the greater error possible on items at the extremes of the frequency of presentation continuum. When the size of the judgment error is considered, the young controls clearly performed better than the CHI and elderly subjects. This is evident in the generally lower values of the function presented in Figure 1b, and was confirmed in a significant main effect for group, $F(2, 112) = 16.60, p < .001$.

A significant effect for presentation frequency ($F(4, 448) = 66.52, p < .001$) indicates that the dispersion of frequency estimates produced by an individual for items given a particular number of presentations increased with the number of actual presentations those items received. The deviation analysis also suggests an effect for instructions ($F(1, 112) = 3.58, p < .06$) not found in the other analyses. The instructions effect is paradoxical. Deviations increased for all groups between the first and second list, indicating a decline in accuracy.

Correlations

The association between subject's error scores in the initial (uninformed condition) and the final (informed condition) list, while not large, was positive for every group tested. The Pearson product-moment correlations obtained were +.28 for the young control subjects, +.25 for group CHI, and +50 ($p < .01$) for the elderly subjects.

In the group CHI data, significant correlations were also obtained between performance on the frequency of occurrence judgment task and all but one of the traditional memory tests in the evaluation battery administered on admission to the rehabilitation program. Seven of the 16 memory test values listed in Table 1 are significant at the .05 level, six of these at the .01 level. In contrast, WAIS-R VIQ was uncorrelated with frequency of occurrence judgments. A total recall measure for the Rey list, the sum of correct responses on trials 1 through 5, and a measure of improvement in recall between the 1st and 5th trial were both significantly correlated with the initial (uninformed) frequency judgments, while performance on the recognition phase of the Rey list procedure was significantly correlated only with informed (second list) performance on the frequency judgment task.

DISCUSSION

The average scores analysis indicated no main effect for subject group, but did reveal a significant interaction of group and frequency, reflected in the flatter judgment functions (Figure 1a) of the CHI and elderly subjects, and suggesting an attenuated relationship between judged and actual frequency relative to the young controls.

Results like these were obtained in a temporal order judgment task by McCormack (1981), who attributed the effect to a bias in his elderly subjects for responses at the center of the permissible set. As in the McCormack study, our analysis of the use of the available response categories indicated a group-by-response category interaction, reflecting a tendency of the elderly subjects to overuse alternatives near the low end and underuse of the high end of the set of permissable responses.

By McCormack's logic, this difference should be attributed to a difference in bias rather than in memory capacity. It is not clear, however, how response bias effects can be distinguished from real differences in memory for frequency information. Differences in response criteria which exist before the fact may indeed cause differences in performance which obscure the essential equivalence in capacity of various subject groups, but differences in the utilization of the available responses may also be the specific result of impaired memory capacity.

In the frequency judgment task, a rational but memory-impaired subject required to respond to items that seemed unfamiliar would be expected to overuse low values of the permissible response set, since items presented only once (or novel foils) would be more likely to be absent from memory than multiply presented items. This "response bias" would be most pronounced in subjects with the poorest memory for the presented items, so the presence of response category by subject group interactions can be expected whenever between-group differences in memory capacity are present.

The interpretation of the group-by-frequency interaction in the average scores analysis is problematic because of a significant group-by-response category interaction in the category utilization analysis, but both the deviation analysis and the error analysis indicate clear group differences, and they do so with a main effect for group rather than by tests of the group-by-frequency interaction. The error analysis indicates that elderly and head-injured subjects made more errors, and deviation analysis indicates that they made larger errors. These results suggest that a task tapping putatively automatic memory processes may reflect the effects of normal aging and traumatic injury when analyzed by methods that are sensitive to response consistancy or error magnitude, and therefore that automatic processes are not as fundamentally different from effortful proceses in terms of their response to varying age and integrity of the central nervous system as prior studies (see Hasher & Zacks, 1984) suggest.

The large number of significant correlations between frequency of occurrence judgments and effortful memory tests obtained among the CHI subjects provides further support for this view. The relationship between frequency-of-occurrence judgment accuracy and effortful memory measures was stronger in the initial uninformed condition, when subjects would be most likely to be attending to the items themselves, rather that to their frequency of occurrence. Under the more specific instructions which preceded the second list, the subjects might have been expected to devote more effort to attending to the number of presentations each item received, yet the obtained correlations declined in value for all but one of the effortful memory measures, and only one significant correlation was obtained. These declines may reflect range restriction artifacts stemming from the decline in accuracy between the first and second lists.

In sum, alternative methods of analyzing frequency judgment data indicate decreased accuracy associated with age and traumatic injury in a task tapping putatively automatic function. The correlations between head-injured subjects' performance on this task and several traditional measures of their effortful memory function suggests that traumatic injuries to the brain have effects on frequency-of-occurrence judgment accuracy that are to some extent predictable from their effects on tests requiring the recall or recognition of item information.

Each of the analytic techniques employed is ultimately neutral regarding the relative importance of (and the causal connections between) differences in response criteria and differences in memory capacity in the overall differences in performance which were observed. We see no reason to assume a logical or empirical primacy for bias effects relative to capacity differences in the explanation of these results. Performance decrements should not be dismissed as reflections of differing response criteria which are superimposed on equal underlying information-processing capacities unless evidence is presented which establishes that these bias effects are pre-existing, and not simply accomodations to decreased memory for the material.

REFERENCES

Alba, J. W., Chromiak, W., Hasher, L., & Attig, M. S. (1980). Automatic encoding of category size information. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 370-378.

Attig, M. S. (1981). *An investigation of the effects of experimental and subject variables on automatic and effortful memory procession in young and elderly adults.* Unpublished doctoral thesis, Temple University, Philadelphia.

Benton, A. L. (1974). *Revised Visual Retention Test: Clinical and experimental applications* (4th ed.). New York: The Psychological Corporation.

Brooks, J. E. (1985). Judgments of category frequency. *American Journal of Psychology, 98,* 363-372.

Ghatala, E. S., & Levin, J. R. (1973). Developmental differences in frequency judgments of words and pictures. *Journal of Experimental Child Psychology, 16,* 495-507.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356-388.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of frequency information: The case of frequency of occurrence. *American Psychologist, 39,* 1372-1388.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences.* Belmont, CA: Brooks/Cole Division of Wadsworth Publishing Company.

Lichty, W. (1982). *Automatic and effortful memory processes in type I diabetics versus nondiabetics.* Unpublished master's thesis, University of Missouri, Columbia, MO.

McCormack, P. D. (1981). Temporal coding by young and elderly adults: A test of the Hasher-Zacks model. *Developmental Psychology, 17,* 509-515.

Piasetsky, E. B. (1981). *A study of pathological asymmetries in visual-spatial attention in unilaterally brain-damaged stroke patients.* Unpublished doctoral dissertation, City University of New York.

Reu, A. (1964). *L'examen clinigue en psychologie.* Paris: Presses Universitaires de France.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's book of 30,000 words.* New York: Columbia University Press.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19,* 87-95.

Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale Revised.* New York: The Psychological Corporation.