Judgments of category size: Now you have them, now you don't

JOEL S. FREUND
University of Arkansas

LYNN HASHER
Duke University

To provide at least a partial explanation for the level-of-processing effect in frequency judgments, in three experiments we explored the role of covert elicitations of events on subsequent frequency judgments. Subjects were asked to judge the number of category instances that had been presented, when cued at retrieval by the category name. In these experiments we hoped to carry the argument regarding the importance of covert occurrences to its extreme. That is, we sought to show that in the absence of covert occurrences of the category label during encoding, judgments of category size would be impossible. Alternatively, when covert occurrences of those category labels were encouraged during encoding, frequency judgments would be possible. We were successful in two experiments, but only moderately successful in another.

Hasher and Zacks (1979) proposed that frequency of occurrence is one of a limited number of attributes of experience that have high priority access to memory, or are "automatically" encoded. Given attention to an event (Zacks, Hasher, & Hock, 1986), such attributes are thought to be encoded inevitably, without regard to such variables as the intention, training, abilities, or age of the person, and without regard to such variables as the difficulty of the task and the existence of simultaneous demands. Although there is research (Hasher & Zacks, 1984) that supports this view, there are exceptions to the predicted patterns, one of which is the focus of the experiments reported here. In particular, manipulation of the level of processing during encoding influences the magnitude of the frequency judgments. Deep encoding tasks result in higher frequency judgments than do shallow encoding tasks (e.g., Greene, 1984; Maki & Ostby, 1987; Naveh-Benjamin & Jonides, 1986; Rose & Rowe, 1976; Rowe, 1974). On the surface, such findings clearly contradict the "invariance" in encoding that the automaticity view suggests.

Hasher and Zacks (1984) have dealt with these findings only in-

AMERICAN JOURNAL OF PSYCHOLOGY
Fall 1989, Vol. 102, No. 3, pp. 333–352

directly (see their Footnote 5). Essentially, they acknowledge that invariance in judgments will not be found under some circumstances. One such circumstance occurs whenever experimental conditions give rise to different degrees of covert elicitations of events (e.g., by rehearsing, visually imagining, reconsidering, and so on). Subjects who engage in greater amounts of covert activities will produce higher frequency estimates than subjects who do not. In parallel fashion, items that receive these covert activities will be judged as more frequent than items not receiving them. In fact, there ought to be a direct relationship between the number of covert activities and the inflation of the estimate. These inflated estimates are a direct consequence of two facts: (a) People are unable to distinguish perfectly between memory traces that represent actual environmental events and memory traces that represent self-initiated covert events, resulting in inflated estimates of the frequency judgments of actual events; and (b) the greater the number of covert events, the greater the degree of inflation. This argument is based on a considerable empirical literature, showing just such confusions (see e.g., Johnson, 1977; Johnson & Raye, 1981; Johnson, Raye, Wang, & Taylor, 1979; Johnson, Taylor, & Raye, 1977; Raye, Johnson, & Taylor, 1980). Individuals or groups of individuals who engage in high levels of covert elicitation of presented events are more likely to give higher estimates of frequency of occurrence than are those who engage in very little covert activity. Hence, equal exposure to overt occurrences coupled with differential rates of covert occurrences can lead to differential frequency judgments.

Hasher and Zacks (1984) suggested on the basis of work reported by Postman and Kruesi (1977) that some versions of the classic levels-of-processing task encourage differential covert occurrences. Often, subjects given a semantic or "deep" task are asked to scale items on some dimension (e.g., pleasantness) while subjects given the "shallow" task are asked to judge individual items for the presence or absence of some characteristic (e.g., a particular letter or letters). Postman and Kruesi argued that typically there are more displaced rehearsals in scaling tasks than in evaluation tasks, because in rating tasks subjects try to keep a uniform scale across all items. This process results in a sizable number of "displaced rehearsals" of items, each of which will presumably leave a memory trace. In evaluation tasks, each item can be judged on its own merit, giving rise to relatively few if any displaced rehearsals.

The thrust of Postman and Kruesi's (1977) argument was twofold. First, some tasks require (or at least elicit) comparisons among list items, and thus elicit differential covert occurrences. Some tasks do not encourage differential covert occurrences. Second, deep or semantic tasks occur primarily in the first category, whereas shallow tasks occur primarily in the second. This distinction applies to all of the studies cited above (i.e., Greene, 1984; Maki & Ostby, 1987; Naveh-Benjamin & Jonides, 1986; Rose & Rowe, 1976; Rowe, 1974).

By this reasoning, it is not all surprising that subjects give higher frequency-of-occurrence estimates for words in semantic types of cover tasks than in other cover tasks. Indeed, subjects should give higher frequency ratings whenever tasks induce them to engage in higher levels of rehearsals, assuming of course that subjects are unable to distinguish perfectly between traces representing directly experienced items and traces representing self-generated items.

Here we try to demonstrate that different degrees of covert occurrences will give rise to different levels of frequency judgments. In the present studies, subjects were asked to judge the number of category instances that had been presented, when cued at retrieval by the category name. Alba, Chromiak, Hasher, and Attig (1980) demonstrated that subjects are able to make such a discrimination, at least for familiar instances of familiar taxonomic categories. They reasoned that for such materials, the presentation of each instance covertly elicited its superordinate category label, thereby enabling subjects to judge category size on a rapidly paced, surprise test (see also Barsalou & Ross, 1986; Brooks, 1985).

In these experiments we hoped to carry the argument regarding the importance of covert occurrences to its extreme. That is, we sought to show that in the absence of covert occurrences of the category label during encoding, judgments of category size would be impossible. Alternatively, when covert occurrences of those category labels were encouraged during encoding, frequency judgments would be possible. To anticipate our findings, we were successful in two experiments, but only moderately successful in another.

EXPERIMENT 1

To manipulate the covert occurrence of the category labels while keeping the particular instances constant, we used words which are not normally conceptually related. However, by appropriate cues or instructions, their relatedness becomes manifest. The sensory-impression norms (Underwood & Richardson, 1956) consist of just such words. Not only do these words elicit sensory-impression responses (e.g., "blood" elicits redness) only under particular circumstances, but Underwood (1965) found no evidence that these words elicited sensory impressions in the absence of any cue or special instruction.

Because of the nature of these words, we expected that only subjects informed about the categorical nature of the list prior to seeing the items would covertly produce the requisite category names. If judgments of the number of instances per category are based on covert occurrences of the category name (i.e., called an *implicit associative response* or IAR by Underwood, 1965), informed subjects would be able to make reliable judgments; not-informed subjects (who do not produce the category name as an IAR) would not.

METHOD

Design and lists

The basic design was a 2 × 2 independent groups factorial, with two test tasks (frequency judgments, free recall), and two types of instructions given prior to the list (informed, not informed) regarding the categorical nature of items in the list. In many respects the design and procedures were similar to those used by Kausler, Hakami, and Wright (1982) and by Tzeng (1976).

The critical items were 8 words chosen from each of 10 categories in the Underwood and Richardson (1956) sensory-impression norms (called "property" categories by Barsalou and Ross, 1986). The categories chosen were black, green, red, round, sharp, small, smelly, soft, white, and yellow. An additional 12 words were chosen from three categories (2 from slimy, 4 from shiny, and 6 from dark) to serve as filler items at the beginning and end of the list. Words were chosen so as to minimize the associative overlap among categories.

Presentation lists were constructed so that within a list two categories were represented by either 0, 1, 3, 5, or 8 instances. For example, one list included all eight instances of the categories smelly and white, whereas no instances of either sharp or yellow appeared (frequency = 0). When fewer than eight instances of a category appeared in a list, the instances with the highest association value were used. Five presentation lists were constructed so that across all five, each category appeared equally often at each frequency. Items from each category were distributed throughout the list. Six filler items were assigned to each end of the list so that there were an equal number of items from each category at each end. The fillers were the same across all lists.

Subjects and procedure

Subjects were 80 students in General Psychology serving in experiments as part of a course requirement. Subjects were assigned to conditions as they arrived at the laboratory, according to a predetermined block randomized schedule that assigned 4 subjects to each list under each condition.

Subjects were tested individually. Upon arrival subjects read and signed an informed consent form, after which they were read standard free-recall instructions modified for the particular instructional condition to which they were assigned. All subjects were instructed that after presentation of the list they would be asked to recall as many of the words as possible. Subjects assigned to the informed conditions were given, in addition, a card containing all 13 category names which they could study for as long as they liked. They were further instructed that trying to rehearse the items by category had been found to be helpful. Subjects in the not-informed conditions were given no such instructions.

Words were presented at a 3-s rate on a Stowe memory drum. After all items had been presented, the recall subjects were given a sheet containing all 13 category names with eight blank lines beneath each name. They were instructed to recall as many items as they could, placing them in the appropriate category. If they recalled an item but could not place it in a category, they were to write it at the bottom of the sheet. Subjects in the noncued conditions were given additional instructions concerning the categorical nature of the list.

Subjects in the frequency-judgment conditions were given an unexpected frequency-judgment test in the form of a typed sheet containing the 13 category names. They were asked to write a number next to each name, indicating how many instances of each category had been presented. Subjects were told they would have only a brief time in which to finish, but were actually allowed to take as much time as necessary. After completing the frequency-judgment task, they were given the recall sheet and were asked to recall as many of the items as they could.

RESULTS

Recall

For purposes of these analyses, a word was counted as a correct recall whether or not it was listed under the correct category label. Mean recall is presented in Table 1. Analysis of these data included instructions, number of instances per category, and task as factors. Because all subjects were given a recall test, task refers to the presence or absence of the frequency-judgment task prior to the recall. The only significant effects were instructions, $F(1, 76) = 4.62$, $MS_e = 1.83$, and instances per category, $F(3, 228) = 119.56$, $MS_e = 1.83$. Informed subjects recalled more than not-informed subjects, and recall increased with number of instances per category.

The finding that the Instruction × Frequency interaction, $F(3, 228) = 1.18$, $MS_e = 1.83$, was not significant replicates the results of other investigators (Maki & Ostby, 1987; Wood & Underwood, 1969).

Frequency judgments

The dependent variable was mean judged frequency for each category size. The results of the frequency judgments are presented in

Table 1. Mean total recall as a function of instructions, task, and number of items per category

Category instructions	Task: Recall only, Items per category 1	3	5	8	Task: Frequency judgment and recall, Items per category 1	3	5	8
Informed								
Mean	.75	2.30	3.05	4.55	.70	2.30	3.60	5.15
SD	.72	1.45	1.79	2.89	.73	1.53	1.85	1.93
Not informed								
Mean	.60	1.10	2.80	3.90	.55	1.65	2.85	4.40
SD	.60	1.25	1.82	2.50	.69	1.14	1.39	2.48

Note. Means represent recall summed over the two categories represented at each level of items per category.

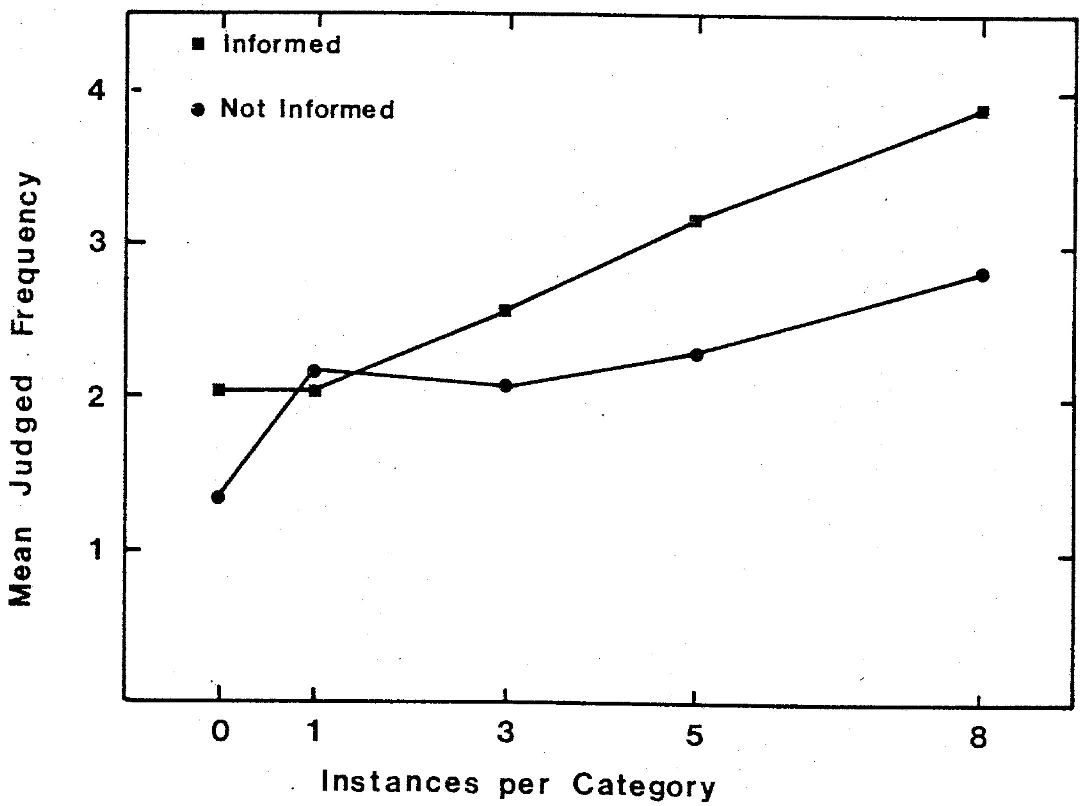

Figure 1. Mean judged frequency as a function of presented number of instances per category for Experiment 1

Figure 1. Analysis of these data, including lists as a factor, but excluding the zero point, indicated that the subjects not given category

information during input were poorer at judging the number of instances in each category than were the informed subjects, $F(1, 30) = 4.88$, $MS_e = 11.07$. In addition, the effect of frequency and the Frequency × Instruction interaction were significant, with Fs = 12.69 and 2.92, respectively. (Both Fs had $df = 3/90$ and $MS_e = 3.82$.) Analysis of the linear trend across frequency using unequal intervals yielded significant effects of frequency, $F(1, 30) = 36.96$, and Frequency × Instruction, $F(1, 30) = 7.72$, $MS_e = 3.87$.

Follow-up analyses within each instructional condition using the Newman-Keuls test indicated that subjects in the not-informed condition could not, as expected, discriminate frequency differences among the input categories. That is, the frequency judgments for categories with eight items did not differ from categories having one or three (ps > .05). Subjects in the informed input condition showed reliable discriminations among categories that differed in number of exemplars. Although the difference in judgments given to categories with one or three items per category was not significant, all other differences were significant. (The difference between the judgments in Conditions 3 and 5 was marginally significant. The calculated difference was 1.20 with 1.205 required for significance.) Trend analyses within each instructional condition yielded significant linear trends with Fs(1, 15) = 29.27 and 8.26, for informed and not-informed conditions, respectively.

A Pearson product-moment correlation between the judged frequency of items per category and the actual number of instances was computed for each subject. These correlations excluded nonpresented items. The means of these correlations,[1] calculated separately for informed and not-informed groups, were .81 and .51, respectively. Both correlations differed from zero, ts(19) = 5.11 and 2.78. Of the correlations for the 20 subjects in the informed condition, only 7 were significantly greater than zero ($p < .05$, one-tailed), but 12 were above .70. Only 3 of the correlations from the not-informed group were significant, and only 4 were above .70.

Analysis of the not-presented (zero frequency) items alone indicated no effect of instructions, $F(1, 30) = 2.54$.

DISCUSSION

The manipulation used here to control the implicit occurrences of the category label (the presence vs. absence of the label at encoding) was successful. When the labels were provided, and therefore presumably covertly elicited, frequency judgments that were sensitive to differences in category size were possible; without the implicit occurrence of the superordinates, frequency discrimination was extremely difficult, if not almost impossible. (The finding of a significant linear trend in the not-informed condition probably reflects the fact that complete elimination of IARs is not possible.) Similar results have been reported by Barsalou and Ross (1986), who called the attention of subjects to properties of items (e.g., "hot" for sauna and fire) by blocking related items or by having properties rated for centrality for a particular exemplar. Frequency discrimination of exemplars sharing a property was reliable. When attention was not called to properties, subjects were unable to discriminate exemplar number. These results support the hypothesis that frequency estimates of category occurrence are based upon information concerning the label that accrues on the basis of covert occurrence of a label during encoding. At retrieval, information is then available to make frequency decisions. In our view (see also Alba et al., 1980; Barsalou & Ross, 1986; Brooks, 1985), frequency decisions regarding exemplars are possible because of encoding processes rather than retrieval processes. To take this position, several competing alternatives must be considered and eliminated.

Alternative explanations involve retrieval processes such as activation of traces or covert retrieval of items (availability) followed by a counting process. That is, at the time of test, subjects either retrieve as many items from each category as they can and then count them (see Williams & Durso, 1986), or the superordinate traces are activated and provide the basis on which a count, or at least a judgment, is made (Brooks, 1985).

Two sources of evidence argue against versions of the retrieve-and-count or availability hypothesis. First, if a version of the retrieve-and-count or availability hypothesis is correct, then manipulation of a particular variable must produce the same pattern of results on both recall and frequency judgments. Similar patterns of results have not always been found. Pitz (1976, Exp. II) found that relative frequency judgments of category membership were not affected by a manipulation that did affect the probability of recall. Similarly, Alba et al. (1980) found that frequency discrimination did not vary even when they manipulated the category structure of the list (blocked vs. random), a variable that ordinarily does influence recall (Bourne & Parker, 1964; Puff, 1966). Recall is better with blocked than with randomly ordered category instances. In the present experiment, judgments of category frequency increased substantially more in the informed than in the not-informed condition, whereas recall increased equally with frequency in both conditions. Hence, in contrast to data reported by

Williams and Durso (1986), frequency judgments do not inevitably reflect the ease with which items can be recalled. Additionally, Underwood (1969a, 1969b) reported that recall and frequency judgments are at least somewhat independent, as did Hock, Malcus, and Hasher (1986).

Second, Alba et al. (1980) showed that subjects were able to make accurate frequency discriminations of category labels when there was not sufficient time for a retrieve-and-count strategy to be effective. Although subjects in the present experiment were allowed more time for making judgments than were those of Alba et al., most of the current subjects completed the task in under one minute.

The results of the present experiment support the notion that covert elicitations of category labels enable exemplar frequency judgments. When a nonobvious category exemplar is used (e.g., coal), it is not likely to elicit its superordinate (e.g., black) without some situationally provided cue. Without this covert elicitation, little or no information permitting a subsequent exemplar frequency judgment is encoded.

EXPERIMENT 2

Although significant, the Frequency × Instructions interaction for judged frequency in Experiment 1 was not large. In addition, the follow-up analyses indicated only small differences in discriminability in the informed condition. The relatively small effect we obtained may have been due to the circumstances under which subjects were given the information about category labels, coupled with the relatively low salience of the sensory impression labels themselves. As a result, some labels might not have been elicited during actual item presentation, and some items may have been misclassified.

To test this explanation, we conducted a second experiment designed to maximize the elicitation of the superordinates during instance presentation and thus to maximize the discriminability of their frequencies. This experiment used the same materials and procedure as Experiment 1, with one major exception. Instead of presenting the category names for study prior to presentation of the instances, the appropriate category label was actually presented with each instance. Thus, each time an instance of a concept (e.g., black) occurred, the concept label (e.g., BLACK) was presented next to it. Such a procedure ensured that the concept name was overtly elicited, and if the subjects desired, they could actually *count* the number of times each concept was presented.

METHOD

Design and lists

Because the results of Experiment 1 indicated that inserting a frequency-judgment task before the recall test did not significantly affect recall, we gave each subject a surprise frequency-judgment task followed by a free-recall test. Thus we employed a 2 × 2 mixed design with instructions (informed, not informed) manipulated between subjects, and task (frequency judgment, recall) manipulated within subjects.

The lists and the method of balancing categories across lists used were the same as those used in the first experiment. However, an additional set of lists was created for the informed condition. These lists differed from those used in the not-informed condition in that the category name was typed in parentheses next to each category instance.

Procedure

Subjects were 40 students in General Psychology serving in experiments as part of a course requirement. Subjects were assigned to conditions in the same manner as in Experiment 1.

The procedure was essentially the same as in the first experiment with four exceptions. In this experiment all subjects were given the unexpected frequency-judgment test, the category exemplars were presented for 2.5 s instead of 3.0 s, both the category instances and superordinates were presented during the learning phase to subjects in the informed condition, and recall was uncued.

All subjects were told that they would be asked to learn a list a words for later recall. Subjects in the informed condition were additionally told that "In parentheses next to each of the target words there will be a second word that describes the first word. For example, if the word you are to study is CRYSTAL, the word in parentheses might be CLEAR. These descriptors are present to help you organize and learn the words you will be tested on. You will *not* be asked to recall the words in parentheses."

The final changes occurred at the time of the frequency test. Now subjects were told that they would have only 30 s in which to complete the frequency-judgment task, although they were allowed additional time if necessary. Also, the recall sheet contained only blank lines; the category names were not included. Subjects were asked to recall as many words as possible in any order.

RESULTS AND DISCUSSION

Recall

The mean recall is plotted in the top half of Figure 2. Analysis of these data indicated that the effect of instructions was marginally significant, $F(1, 90) = 3.26$, $MS_e = 3.04$, $p = .07$. Mean recall was 9.66 and 8.40 for the informed and not-informed conditions, re-

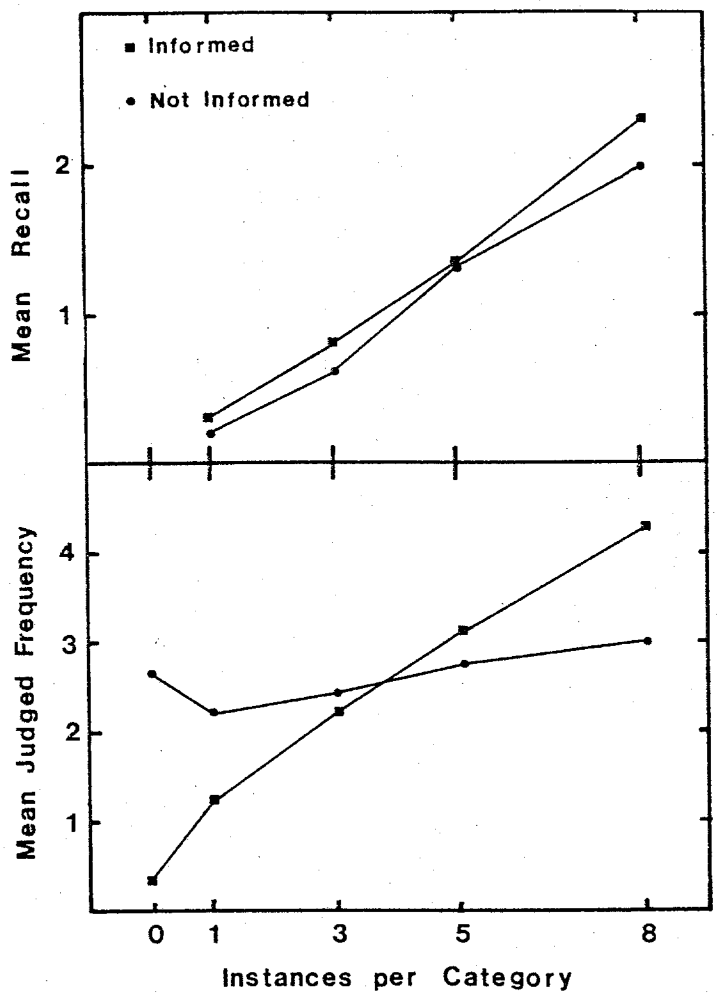

Figure 2. Mean recall and mean judged frequency as a function of presented number of instances per category for Experiment 2

spectively. The effect of frequency, $F(3, 270) = 195.30$, $MS_e = 1.38$, was significant. As in Experiment 1, the Frequency × Instructions interaction did not reach significance, $F(3, 270) = 1.07$.

Frequency judgments

The basic data are presented in the bottom half of Figure 2. Analysis of these data, including lists as a factor, but excluding the zero point, indicated that the judgments of the informed subjects did not differ from those of the not-informed subjects, $F(1, 90) < 1$, $MS_e = 27.14$. The effect of frequency, and the Frequency × Instructions interaction were significant, with *F*s = 44.86 and 29.29, respectively. (Both *F*s had $df = 3/270$ and $MS_e = 6.09$.) Analysis of the linear trend across frequency using unequal intervals yielded significant effects of frequency, $F(1, 90) = 89.02$, and Frequency × Instructions, $F(1, 90) = 28.67$, $MS_e = 9.17$.

Follow-up analyses within each instructional condition using the Newman-Keuls test indicated that subjects in the not-informed condition could not, as expected, discriminate frequency differences among the input categories ($ps > .05$). The only significant comparison for subjects in the not-informed condition was that between frequency of one versus eight. By contrast, subjects in the informed input condition showed reliable discriminations (all $ps < .01$) among categories that differed in number of exemplars. Trend analyses within each instructional condition yielded significant trends in both conditions, $Fs(1, 45) = 149.62$ and 6.56 in the informed and not-informed conditions, respectively.

As in Experiment 1, we computed individual correlations between frequency judgments and presented frequency. The average correlations were .89 and .32 for the informed and not-informed groups, respectively. Both correlations differed from zero, $ts(49) = 12.69$ and 2.43. Twenty-four of the 50 correlations in the informed group were significant ($p < .05$, one-tailed), whereas only 7 of those in the not-informed group were.

Analysis of the nonpresented items indicated that the judgments of the subjects in the informed condition were considerably more accurate than those in the not-informed condition, $F(1, 90) = 60.19$, $MS_e = 8.79$.

Examination of the data indicated that part of the large difference between the informed and not-informed conditions at zero frequency can be accounted for by extreme judgments by a few subjects. (One subject judged the two zero-frequency categories to have a total of 25 instances presented.) However, much of the difference is more pervasive. The median judgments of frequency for the nonpresented categories were 0 and 2.2 for the informed and not-informed conditions, respectively. Thirty-five of the subjects in the informed conditions correctly assigned judgments of zero to both nonpresented categories, whereas only 2 of the subjects in the not-informed conditions did so. Median judgments of frequency in the one-item per category conditions were .59 and 1.52 for the informed and not-informed conditions, respectively.

Because Experiments 1 and 2 differed primarily in the manner in which the superordinate names were presented, we compared the frequency judgments from the two experiments using an unweighted means analysis. The factors included were experiment, instructions, list, and frequency. The effect of frequency, $F(3, 360) = 38.89$, and the Frequency × Instructions interaction, $F(3, 360) = 10.92$, were significant, $MS_e = 5.53$. The Experiment × Instructions × Frequency interaction was not significant, $F(3, 360) = 1.07$.

Although the triple interaction involving experiment, instructions, and frequency was not significant, comparison of the curves representing the frequency judgments for the informed conditions in Figures 1 and 2 indicates that overtly presenting the category labels resulted in a steeper judgment function and more accurate judgments. To test the significance of this difference, we calculated the slope of the line relating frequency judgments to actual frequency for the presented items as suggested by Naveh-Benjamin and Jonides, (1986).[2] Unweighted means analysis of these data yielded a significant effect of experiment, $F(1, 60) = 6.601$, $MS_e = .240$. Mean slopes were .533 and .866 for Experiments 1 and 2, respectively. Explicitly presenting the category names resulted in more accurate judgments of category size (i.e., steeper slopes).

In this experiment we ensured the elicitation of superordinate category labels by explicitly presenting them during encoding. The manipulation had the intended effect of sharpening the discrimination among categories. The second experiment differed from the first in three ways. First, the Frequency × Instructions interaction in this experiment was more pronounced. Second, the average slope of the judgment functions in the informed condition of the second experiment was higher than in the first experiment. Finally, the average correlation between judged and actual frequencies was higher in the second experiment.

EXPERIMENT 3

A third experiment was undertaken to extend the generalizability of the findings of the first two experiments to taxonomically organized categories. Here we tried to influence the likelihood of a covert elicitation of a category name by using only items that are uncommon exemplars, and then by again varying whether or not subjects were explicitly informed about the presence of such items in the list. We anticipated a repetition of the pattern of results seen for Experiment 1.

METHOD

Design and lists

We used a 2 × 2 mixed design with instructions (informed, not informed) manipulated between subjects, and task (frequency judgment, recall) manipulated within subjects.

The words used in this experiment were relatively infrequent instances from 13 categories in the Battig and Montague (1969) norms. The following categories were used: body parts, professions (occupations), furniture, animals, fruits, sports, vehicles, musical instruments, clothing, and birds. Instances from the categories of fish, insects, and flowers were used as fillers.

The number of lists (five), list construction, ordering of the items within each list, and the method of balancing categories across lists were the same as in Experiment 1.

Procedure

Subjects were 40 students in General Psychology serving in experiments as part of a course requirement. Subjects were assigned to conditions in the same manner as in Experiment 1.

The procedure was essentially the same as in the first experiment with two exceptions. In this experiment all subjects were given the unexpected frequency-judgment test, and the category exemplars were presented for 2.5 s instead of 3.0 s.

RESULTS

Recall

As in the first experiment, a word was counted as a correct recall whether or not it was listed under the correct category. The mean recall is plotted in the top half of Figure 3. Analysis of these data indicated that the effect of instructions was not significant, $F(1, 38) = 1.62$, $MS_e = 3.24$. Mean recall was 10.15 and 8.55 for the informed and not-informed conditions, respectively. The effect of frequency, $F(3, 114) = 35.78$, $MS_e = 1.98$, was significant, although the Frequency × Instructions interaction was not, $F(3, 114) = 1.61$.

Frequency judgments

The basic data are presented in the bottom half of Figure 3, and in most respects the analysis of the frequency data, using only Frequencies 1, 3, 5, and 8, supports what can be observed in the figure. Judgments of the informed subjects were higher and more accurate than those of the not-informed subjects, $F(1, 34) = 19.10$, $MS_e = 3.54$. (Inspection of the data indicated that this difference was primarily caused by one list in which there were a few very high frequency estimates.) The only other significant effect was that for frequency, $F(3, 114) = 37.15$, $MS_e = 3.43$.

As in Experiment 1, we computed individual correlations between frequency judgments and presented frequency. The average correlations were .85 and .76 for the informed and not-informed groups, respectively. Both of these correlations differed from zero, $ts(19) =$

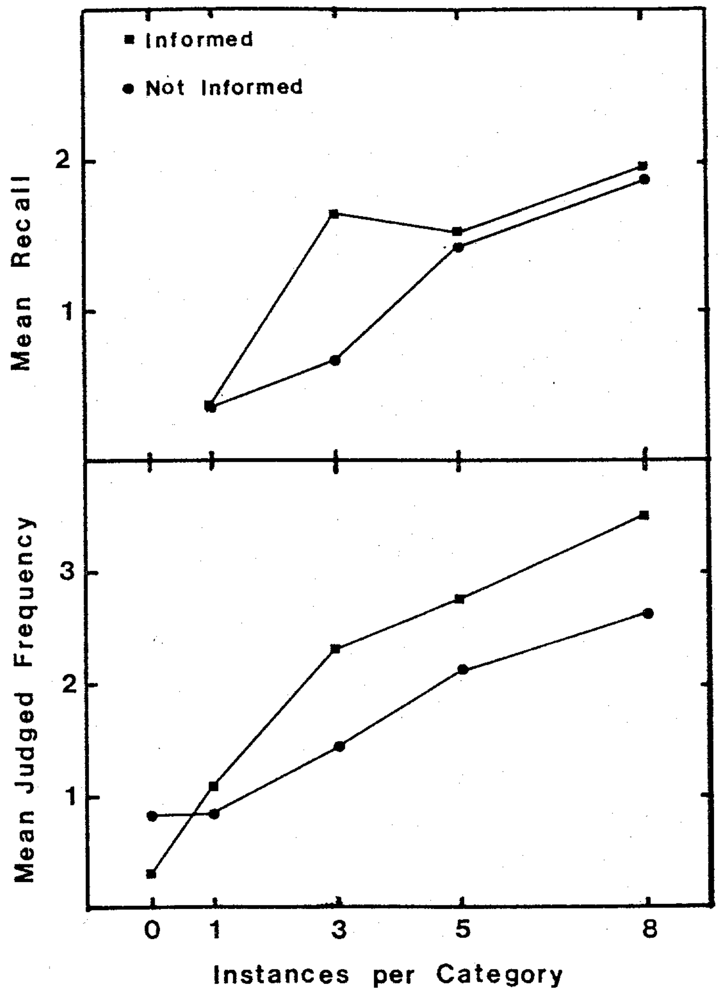

Figure 3. Mean recall and mean judged frequency as a function of presented number of instances per category for Experiment 3

7.01 and 7.11. Eight of the 20 correlations in the informed group and 6 of those in the not-informed group were significant ($p < .05$, one-tailed).

DISCUSSION

There are two points to be made regarding these results. First is the failure to replicate, for judgments of frequency, the Frequency × Instructions interaction found in the first two experiments. Such a finding can be interpreted as evidence against the hypothesis that frequency judgments of category labels are mediated by covert elicitations of the label that occur during encoding. The overall pattern of results, however, does not support this interpretation. Informing subjects about the categorical nature of the list affected frequency judgments, but not recall. Although both groups were able to discriminate among the various levels of frequency, judgments were higher (and more accurate) under informed conditions. The means of the individual correlations were similar for the two groups, as were the numbers of significant correlations. In Experiment 1, informing subjects about the nature of the list affected frequency judgments, recall, and the correlations between presented and judged frequency. We attribute these different patterns of results to the differential availability of the two types of categories (taxonomic in Exp. 3 and sense impression in Exps. 1 and 2) in the not-informed condition. Although low frequency exemplars of taxonomic categories were used here in an attempt to reduce the spontaneous elicitation of category labels in the not-informed condition, it is likely that some of the labels were spontaneously elicited often enough to produce reliable frequency estimates. A mixed list using these low frequency taxonomic category exemplars interspersed among noncategory items might have prevented subjects from detecting the categorical nature of some items and also reduced the category label elicitations. Alternatively, the levels of familiarity that college students have with taxonomic categories may extensively limit, especially when presented in a categorized list, the situations in which instances of a category fail to elicit their category label.

A second point concerns the hypothesis that frequency judgments are based on retrieval of exemplars and a subsequent judgment of their number based on their availability. The finding that instructions about category membership affected frequency judgments but not recall suggests that retrieval of exemplars and judgment of frequency based on an availability heuristic cannot be solely responsible for differences in frequency estimates of category size (cf. Williams & Durso, 1986).

GENERAL DISCUSSION

The present experiments addressed the issue of the role that covert rehearsals play in inflating frequency judgments. To assess this, we manipulated the likelihood of elicitation of covert occurrences of category labels during encoding of exemplars of those categories. At retrieval, we gave an unanticipated test of category size judgments, using category labels as cues. Reliable discrimination among categories of different sizes was expected when category labels were elicited during the presentation of exemplars. No discrimination was expected when category labels were not elicited. The data from Experiments 1 and 2, in which items were instances of sense impression norms, were in agreement with this prediction. That is, when subjects were informed of these categories (Exp. 1) or were given the category labels during item presentation (Exp. 2), they were sensitive to experimental differences in category size. When given no such hints (and so when subjects did not spontaneously think of the category label), subjects were relatively insensitive to differences in category size. This same lack of knowledge does not reliably influence overall recall or the relation between category size and recall.

Equivalent recall might be the result of the use of a generate-and-recognize strategy for retrieval (Bahrick, 1970) by subjects not informed about the list structure prior to recall. The use of such a strategy could account for how a cue (the category label) that did not spontaneously occur to subjects during encoding could actually serve as an effective retrieval cue, in contrast to expectation from the large encoding-specificity literature (e.g., Tulving & Osler, 1968; Watkins & Tulving, 1975), suggesting that cues provided only at retrieval tend to be relatively impotent compared with cues provided at both encoding and retrieval (Tulving & Thompson, 1973).

Note, however, that the provision of the same cue at retrieval for not-informed subjects asked to make category size judgments did not have the same salutary effect: Subjects who did not have this cue available at encoding were not able to benefit from having it at retrieval to make frequency judgments. This logic rather explicitly suggests that frequency judgments (at least in this case) are not necessarily made with the use of a generate-and-count, or availability, strategy. Taken together with data reported elsewhere (e.g., Alba et al., 1980; Hock et al., 1986; Maki & Ostby, 1987), the present data at least constrain the limits of the item-availability heuristic as the sole procedure for judging frequency of events.

Furthermore, the present findings suggest, as argued in the introduction, that a plausible account of the levels-of-processing effect on frequency judgments (e.g., Greene, 1984; Maki & Ostby, 1987; Naveh-Benjamin & Jonides, 1986; Rose & Rowe, 1976; Rowe, 1974) can be tied to increased covert elicitations of items encoded "deeply" during list presentation. Clearly, covert elicitations influence frequency judgments, as the present data and those of others (e.g., Barsalou & Ross, 1986; Hanson & Hirst, 1988) suggest. The levels effect may indeed be the product of heightened elicitations (rehearsals) during encoding, coupled with imperfect discriminations at test between traces representing actual occurrences and traces representing covert occurrences.

Notes

The research reported in this article was supported by a grant from the Marie Wilson Howells Fund. Portions of the research reported here were reported at the 27th annual meeting of the Psychonomic Society, November 1986, New Orleans.

We gratefully acknowledge the assistance of Rose Marie McCahon for collecting the data in Experiments 1 and 3, and Beth Shillingford and Scarlett Newby for collecting the data in Experiment 2.

Correspondence concerning this article should be addressed to Joel S. Freund, Department of Psychology, University of Arkansas–Fayetteville, Fayetteville, AR 72701. Received for publication May 17, 1988; revision received November 1, 1988. Accepted by the previous editor.

1. Means of all correlations reported were calculated by converting the correlations to *z*-scores, averaging the *z*-scores, and reconverting the mean *z*-score to an *r*.

2. Slopes were also calculated for all subjects in both experiments. For Experiment 1, the average slopes were .533 and .199 for the informed and not-informed conditions, respectively. For Experiment 2, the means were .866 and .239. Unweighted means analysis yielded significant main effects for experiment, $F(1, 120) = 3.02$, and instructions $F(1, 120) = 20.03$. The Experiment × Instructions interaction was not significant, $F(1, 120) = 1.86$.

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.

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review, 77,* 215–222.

Barsalou, L. W., & Ross, B. H. (1986). The roles of automatic and strategic processing in sensitivity to superordinate and property frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 116–134.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs, 80*(3, Pt. 2).

Bourne, L. E., Jr., & Parker, B. K. (1964). Inter-item relationships, list structure, and verbal learning. *Canadian Journal of Psychology, 18,* 52–61.

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

Greene, R. L. (1984). Incidental learning of event frequency. *Memory & Cognition, 12,* 90–95.

Hanson C., & Hirst, W. (1988). Frequency encoding of token and type information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 289–297.

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 fundamental information: The case of frequency of occurrence. *American Psychologist, 39,* 1372–1388.

Hock, H. S., Malcus, L., & Hasher, L. (1986). Frequency discrimination:

Assessing global-level and element-level units in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 232–240.

Johnson, M. K. (1977). What is being counted nonetheless? In I. M. Birnbaum & E. S. Parker (Eds.), *Alcohol and human memory* (pp. 43–57). Hillsdale, NJ: Erlbaum.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.

Johnson, M. K., Raye, C. L., Wang, A. Y., & Taylor, T. H. (1979). Fact and fantasy: The roles of accuracy and variability in confusing imagination with perceptual experiences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5,* 229–240.

Johnson, M. K., Taylor, T. H., & Raye, C. L. (1977). Fact and fantasy: The effects of internally generated events on the apparent frequency of externally generated events. *Memory & Cognition, 5,* 116–122.

Kausler, D. H., Hakami, M. K., & Wright, R. E. (1982). Adult age differences in frequency judgments of categorical representations. *Journal of Gerontology, 37,* 365–371.

Maki, R. H., & Ostby, R. S. (1987). Effects of level of processing and rehearsal on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 151–163.

Naveh-Benjamin, M., & Jonides, J. (1986). On the automaticity of frequency coding: Effects of competing task load, encoding strategy, and intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 378–386.

Pitz, G. F. (1976). Memory and frequency estimation processes in a decision task. *Memory & Cognition, 4,* 132–138.

Postman, L., & Kruesi, E. (1977). The influence of orienting tasks on the encoding and recall of words. *Journal of Verbal Learning and Verbal Behavior, 16,* 353–369.

Puff, C. R. (1966). Clustering as a function of the sequential organization of stimulus word lists. *Journal of Verbal Learning and Verbal Behavior, 5,* 503–506.

Raye, C. L., Johnson, M. K., & Taylor, T. H. (1980). Is there something special about memory for internally generated information? *Memory & Cognition, 8,* 141–148.

Rose, R. J., & Rowe, E. J. (1976). Effects of orienting task and spacing of repetitions of frequency judgments. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 142–152.

Rowe, E. J. (1974). Depth of processing in a frequency judgment task. *Journal of Verbal Learning and Verbal Behavior, 13,* 638–643.

Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology, 77,* 593–601.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Tzeng, O. J. L. (1976). A precedence effect in the processing of verbal information. *American Journal of Psychology, 89,* 577–599.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70,* 122–129.

Underwood, B. J. (1969a). Attributes of memory. *Psychological Review, 76,* 559–573.

Underwood, B. J. (1969b). Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 8,* 83–94.

Underwood, B. J., & Richardson, J. (1956). Some verbal materials for the study of concept formation. *Psychological Bulletin, 53,* 84–95.

Watkins, M. J., & Tulving, E. (1975). Episodic memory: When recognition fails. *Journal of Experimental Psychology: General, 104,* 5–29.

Williams, K. W., & Durso, F. T. (1986). Judging category frequency: Automaticity or availability? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 387–396.

Wood, G., & Underwood, B. J. (1967). Implicit responses and conceptual similarity. *Journal of Verbal Learning and Verbal Behavior, 6,* 1–10.

Zacks, R. T., Hasher, L., & Hock, S. (1986). Inevitability and automaticity: A response to Fisk. *American Psychologist, 1986, 41,* 216–218.