Memory & Cognition
1995 23 (2), 213–226

When encoding fails: Instructions, feedback, and registration without learning

DOUGLAS L. HINTZMAN and TIM CURRAN
University of Oregon, Eugene, Oregon

Four experiments replicated and extended the registration-without-learning effect, in which there is little improvement in the ability to discriminate an old target (X) from a highly similar test item (Y) after the first few presentations of X, even though judgments of frequency continue to rise in an open-ended fashion. Forced-choice testing revealed the anomalous form of the learning curve for X–Y discrimination (faster and then slower than the exponential). Effects of several different learning instructions were compared, but these appeared to affect only the level of initial learning, and to do little to promote X–Y discrimination learning on later presentations. The opportunity for self-testing with feedback during study provided no benefits when responding was covert, but did when overt anticipation was required. The findings are discussed in relation to the roles of bottom-up and top-down processing in memory encoding, and to the importance of error-correcting feedback in further structural learning of materials, once the materials have become familiar.

It is a truism of everyday life that practice makes perfect, and a truism of psychology that, with repetition, the learning curve approaches 100%. However, recent evidence suggests that these beliefs are not always correct. Hintzman, Curran, and Oppy (1992) discovered that repetitions of a stimulus can "register"—as revealed in increasing judgments of frequency—without the subject becoming better able to discriminate it from another, highly similar but distinct, stimulus. In this article we present evidence that the latter failure reflects a powerful cognitive bias against learning more about an already familiar item's structure, and that it may take overt responding and error-correcting feedback to overcome this bias.

In the experiments of Hintzman et al. (1992), people studied a long list in which individual stimuli were shown various numbers of times, and then they judged the frequencies with which test stimuli had appeared in the list. The subjects were instructed to give frequency judgments of zero to stimuli that were highly similar to, but not identical to, those actually seen. Let us denote the studied targets by X and the similar stimuli by Y. In some of the experiments, the X items were singular and plural nouns (e.g., BELLS, DEAL) and the Y items were their alternate forms (e.g., BELL, DEALS); in other experiments, the X items were asymmetrical bit-mapped drawings of objects, and the Y items were their right–left (mirror-image) reversals. Using both types of materials, Hintzman et al. (1992) found that the ability to discriminate between X and Y "stalled" after 1 or 2 presentations, remaining more or less steady at an intermediate level, even when X had been seen as many as 25 times. Despite this apparent learning failure, judgments of frequency (JOFs) to targets increased with repetition in an open-ended fashion. The result seems to reveal a dissociation between registration of an item's occurrence, on the one hand, and learning more about the item's structure, on the other. Subjects learned that X was repeated without learning more of the detail needed to discriminate X from Y.

Two additional kinds of evidence support the conclusion that registering an item's occurrence and learning the item's structure are served by different processes. First, Hintzman et al. (1992) found that JOF distributions for the similar Y items were bimodally distributed, with one mode at zero and another mode that tended to track the frequency of the corresponding target. The mode at zero presumably reflects subjects' awareness that the structure of the test item Y was different from that of the target item, X, so that the frequency of Y had to be zero. The mode that shifts with target frequency presumably reflects cases when the subject does not notice that Y is different from X and bases Y's judged frequency on its apparent familiarity. Second, when Hintzman and Curran (1994) used the response-signal method to examine the time course of retrieval, they found biphasic retrieval functions for Y items (nouns with changed plurality). Early in the retrieval episode, subjects tended to call such test items old, but about 90 msec later, this trend was abruptly reversed. Both the bimodal distributions of JOFs to Y and the biphasic retrieval curves for Y support the hypothesis of two distinct underlying processes. These are hypothesized to be a fast, undifferentiated strength or familiarity signal and a slower recall process that supports retrieving the content of an experience. The former reflects recent registration, and the latter reflects learning of structure.

The second author is now at Harvard University. This is based on work supported by the National Science Foundation under Grants BNS-90-08909 and SBR-93-19265. Correspondence should be addressed to D. L. Hintzman, Department of Psychology, University of Oregon, 97403 (e-mail: hintzman@oregon.uoregon.edu).

This evidence for dual processes poses a problem for several memory models, all of which make assumptions that imply, in one way or another, that repetition and similarity should have multiplicative effects (see discussion by Hintzman et al., 1992, pp. 667–669).[1] That is, the models imply that *familiarity* (Y) = $\beta \times$ familiarity (X), where $\beta = 0$ indicates an absence of similarity between X and Y and $\beta = 1$ indicates that X and Y are identical. This equation implies that, for high β, increasing the frequency of X should cause the familiarity of Y to increase, lagging the familiarity of X by a fixed proportion. If one assumes that judgments of frequency reflect the test item's familiarity (Hintzman, 1988), this proportionality prediction is disconfirmed by the finding that frequency judgments to X are unimodal and those to Y are bimodal. Nevertheless, in spite of this disconfirmation, the upper mode of the Y distribution appears to lag the mode of X proportionally, just as these models predict (Hintzman et al., 1992 , Figure 6). Thus, the multiplicative property that is basic to these models may also hold for human memory, but only for the mechanism underlying undifferentiated familiarity or strength.[2]

The primary purpose of the present experiments was to examine effects of instructions and of feedback on the type of structural learning needed to differentiate between targets and highly similar distractors (X vs. Y). We wanted to determine what manipulations, if any, would keep this type of learning from slowing or stalling far short of the accuracy ceiling of 100%. Hintzman et al. (1992) found essentially flat X–Y discrimination functions in two experiments (their Experiments 2 and 4), and gradually upward-trending discrimination functions in two others (their Experiments 3 and 5). They speculated (p. 677) that the crucial difference may have been in the amount of foreknowledge subjects had about the nature of the test. Therefore, in the present Experiment 1, we manipulated the instructions given prior to the study list. We manipulated instructions also in Experiment 2, but replaced the JOF test with a forced-choice test of recognition memory. Both of these experiments were done in two versions: one testing the ability to remember the right–left orientations of drawings, and the other testing the ability to remember the pluralities of nouns. The remaining two experiments were designed to study the effects of feedback during learning on memory for the pluralities of nouns. Responses during learning were covert in Experiment 3 and overt in Experiment 4—a difference that turned out to be important. A secondary purpose of these experiments was to learn more about the registration, or familiarity process. Experiments 1 and 3 provide information on the effects of study instructions on frequency judgments, and on the proportionality of nonzero judgments given to X and to Y.

EXPERIMENT 1

This experiment compared the effects of three types of study instruction on memory for both the pluralities of nouns (Experiment 1A) and the orientations of drawings (Experiment 1B). Target items were presented from 0 to 20 times and were tested in either the target form (X) or similar form (Y). There were three instruction conditions: *neutral* instructions, which did not specify the type of memory test that would follow; *frequency* instructions, which told subjects they would be asked to remember presentation frequencies of the stimuli; and *structure* instructions, which told subjects they would have to remember either the plurality of each noun (Experiment 1A) or the left–right orientation of each drawing (Experiment 1B). The rationale for the frequency instructions was that they should make subjects attend to repetitions; and for the structure instructions, that they alert subjects to the particular features we wanted them to learn. The subjects served in Experiments 1A and 1B during the same session, and a given subject was assigned to the same instruction condition in both experiments. (Study lists for both experiments were given first, and then both tests were administered.)

Method

Subjects. A total of 211 University of Oregon undergraduates participated for course credit. The data of 6 additional subjects were dropped, owing to failure to follow instructions. The subjects were tested in groups of up to 12 persons. Approximately equal numbers of subjects served in the three instruction conditions.

Design and Materials. The same experimental design was used for both words (1A) and pictures (1B). The words were common, four-letter English nouns and their plurals. Only nouns pluralized by adding an *s* were used. The pictures were bit-mapped line drawings suitable for presentation on a Macintosh computer, taken from various sources, including Snodgrass and Vanderwart (1980). Nearly all depicted distinctly different single objects, and all were characterized by left–right asymmetry and an absence of written symbols.

For both 1A and 1B, the study list consisted of 24 distinct stimuli, with 6 assigned to each of the presentation frequencies: 1, 3, 8, and 20. Four different study lists were created by rotating each stimulus through the 4 presentation frequencies, with approximately equal numbers of subjects assigned to each list. An additional 78 items served as once-presented fillers, to ensure correct spacing of repetitions (see the following). Presentation order was randomly determined for each study list, with the following restrictions: (1) At least 5 different stimuli intervened between repetitions of the same item; (2) stimuli of each presentation frequency were represented equally in each third of the list; and (3) at least 3 out of every 12 consecutive stimuli were fillers. Six more items served as buffers at the beginning and 6 at the end of each list. Including buffers, fillers, and repetitions, each presentation list had a length of 282 items.

All repetitions of a stimulus in the study list were identical; that is, a given noun was always of the same plurality, and a given drawing was always shown in the same right–left orientation. Half of the nouns within each frequency condition were presented in the plural form and half were presented in the singular form.

One test list was constructed for Experiment 1A and one for 1B. In each list, 24 new stimuli were presented along with the 24 experimental items from the study list. Test order was determined randomly, with the constraint that no more than 3 items from the same condition occur consecutively. Two words within each presentation frequency were tested with the study list plurality (X), and 4 were tested with the plurality changed (Y). Likewise, 2 drawings from each presentation frequency were presented in the study (X) orientation, and 4 were tested in the mirror-reversed (Y) orientation. This imbalance between X and Y was intended to increase the number of observations in the more interesting (Y) cells of the experimental design.

The stimuli were presented by a Macintosh Plus computer running PsychLab software (Gum & Bub, 1988). They were projected via a computer-controlled overhead projection panel onto the front wall of the testing room. The words were presented in 28-point bold Geneva font, and the pictures covered an area ranging from 6 to 40 cm^2 when displayed on the Macintosh screen.

Procedure. Each subject was presented with both study lists, one of 282 words and the other of 282 pictures. Approximately half the subjects saw the words (1A) first, and half saw the pictures (1B) first. Each group of subjects was randomly assigned to one of three different instruction conditions: (1) In the neutral condition, subjects were told, "Your task is simply to try your best to remember each word [picture]. Your memory for these words [pictures] will be tested later in the session." (2) In the frequency condition, subjects were told, "Your task is simply to try your best to remember each word [picture]. Some of the words [pictures] will be presented more than once. Later in the session you will be asked to *estimate the number of times* each of these words [pictures] appeared." (3) In the structure condition of 1A, subjects were told, "Your task is simply to try your best to remember each word. Some of the words will be presented in the singular form and others will be plural. Be sure to note whether each word is singular or plural. The *plurality of the words will be important* when your memory for these words is tested later in the session." A similar instruction was given for 1B, recouched in terms of the orientations of the pictures.

The subjects were assigned to the same instruction condition for both types of stimuli. The stimuli in each study list were exposed for 2,860 msec, with a 17-msec interstimulus interval (ISI).

Following presentation of both study lists, two test lists of 48 stimuli each (24 new and 24 old) were presented. For each subject, the order of the word and picture test lists was the same as the order of the study lists. During testing, each stimulus was shown for about 4 sec, with a 1.5-sec ISI. The subjects were asked to write down a numerical frequency judgment for each stimulus. They were explicitly instructed to pay particular attention to the plurality of the word or orientation of the picture and to only write down the number of times they had seen the test stimulus exactly as shown. For the word test, the following example was given:

> Only count the times that the word was presented exactly as shown in the test list. In particular, pay attention to whether the word is singular or plural. For example, in the presentation list you could have seen the word "cat" (the singular form) twice, but never "cats" (the plural form). The correct answer depends on whether "cat" or "cats" is shown to you in the test. If "cat" (the original singular form) is in the test, you should answer "two". If "cats" (the plural form) is in the test, you should answer "zero."

Similar instructions were given prior to the picture test. Subjects were also told that no stimulus was ever presented more than 25 times, so that all frequency judgments should be between 0 and 25.

Results

Mean JOFs. Mean frequency judgments made to targets (X) and to similar items (Y) are shown as a function of frequency and instruction condition in Figure 1. As was found by Hintzman et al. (1992), the distributions of judgments made to X and Y items had different forms. Example histograms, for pictures with frequency = 8, collapsed over instruction conditions, are shown in Figure 2.[3] In general, X distributions appeared to reflect strong response biases (e.g., favoring multiples of 5) superimposed on an underlying unimodal form. Y distributions, in contrast, tended to be bimodal, with the lower mode at zero.

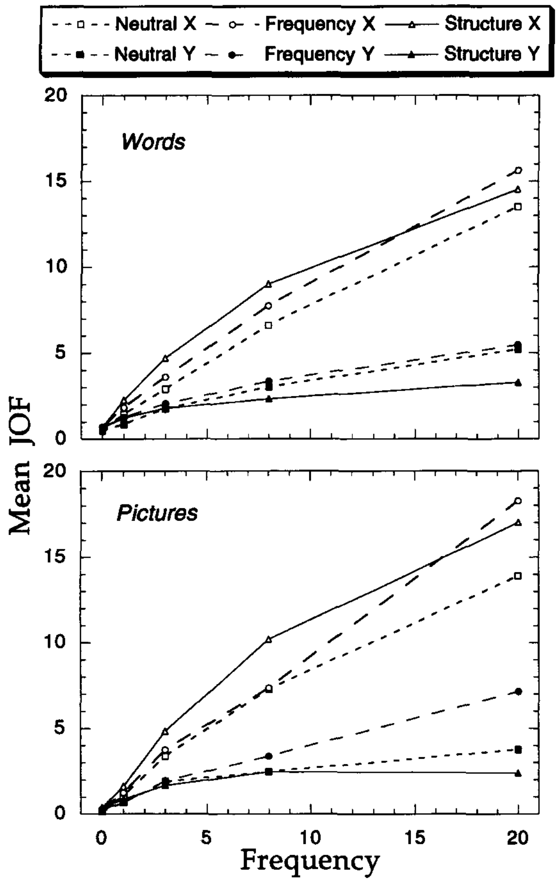

Figure 1. Mean judgments of frequency for the conditions of Experiment 1. X, old targets; Y, similar items.

Because means of bimodal distributions are uninformative, further analyses of mean JOFs were restricted to those made to X items. Separate 3 (instruction conditions) × 5 (frequencies) ANOVAs were done on the mean JOFs for words and for pictures, using proportionally spaced (linear) contrast coefficients to specify frequency. Aside from the main effect of frequency, which was of course highly reliable, mean JOFs showed reliable main effects of instructions [$F(2,208) = 6.09$, $MS_e = 24.7$, for words; $F(2,208) = 11.12$, $MS_e = 22.4$, for pictures; both $ps < .01$]. The instruction × frequency interaction was reliable only for pictures [$F(2,208) = 9.50$, $p < .001$]. In general, however, the patterns for words and pictures were quite similar: Neutral instructions led to lower judgments than did either frequency or structure instructions; however, mean judgments tended to track frequency in a more linear fashion under frequency instructions than under structure instructions.

JOF accuracy. We measured the accuracy of judgments to X items in two ways: (1) by computing each subject's Pearson r between mean judgment and actual frequency, and (2) by computing each subject's absolute

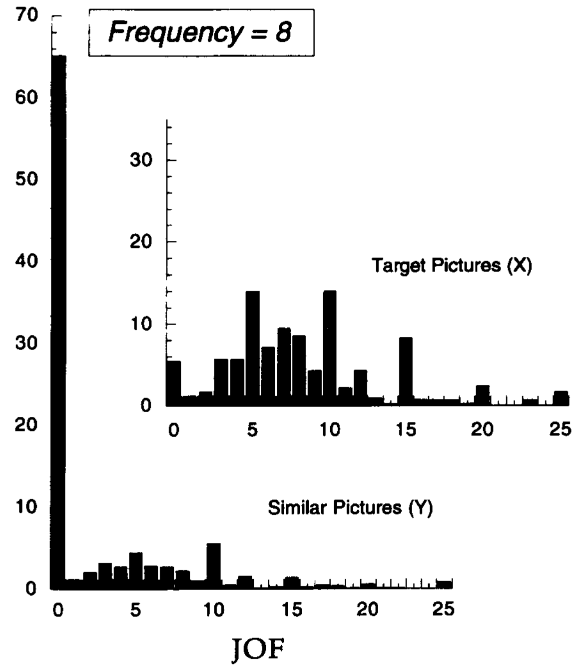

Figure 2. An example comparing histograms of frequency judgment distributions for old (X) items and similar (Y) items. Data are for pictures with frequency = 8, Experiment 1. (Response percentage on ordinate.)

error for each combination of stimulus type and frequency. Both measures led to essentially the same conclusion, so only the analysis of Fisher-transformed *r*s is presented in Table 1. For both words and pictures, frequency instructions yielded the most accurate JOFs overall, and structure instructions yielded the least accurate judgments, with neutral instructions falling in between. The instruction effect was reliable for both stimulus types (see Table 1 for *F* ratios and MS_es). Scheffé tests showed that frequency and neutral instructions were superior to structure instructions for words, and that frequency instructions were superior to both neutral and structure instructions for pictures (all $ps < .02$). Mean absolute errors showed that the superiority of frequency instructions lay primarily at frequency = 20.

Nonzero JOFs. A comparison of the X and Y distributions for pictures with frequency = 8, shown in Figure 2, suggests that the high incidence of zeros is not the only difference between JOFs given to highly similar items and those given to targets: in addition, the nonzero judgments given to similar items appear to be shifted downward in comparison with those made to targets. A number of memory models predict such a shift (ignoring JOF = 0), because the models imply multiplicative effects of frequency and similarity. The effect of the multiplicative relationship is that JOFs made to Y items should be proportionally lower than those made to X.

Table 1
Mean Fisher-Transformed Pearson *r*s Between JOF and Frequency, Experiment 1

	Materials			
	Words		Pictures	
Instructions	Fisher *r*	Mean *r*	Fisher *r*	Mean *r*
Neutral	1.99	.964	2.16	.974
Frequency	2.06	.968	2.55	.988
Structure	1.64	.928	2.03	.966
F(2,208)	6.22*		8.44†	
MS_e	0.57		0.61	

$^*p < .002$. $†p < .001$.

To avoid subject selection artifacts in statistics done on nonzero judgments to X and Y items, we deleted all subjects who failed to contribute at least one nonzero judgment in each target and nontarget cell at each frequency greater than 0. This left 92 subjects (44%) contributing to the word data and 78 subjects (38%) to the picture data, collapsed over instruction conditions. Separate ANOVAs on words and pictures showed that nonzero JOFs were higher for X items than for Y items [$F(1,86) = 13.19$, $MS_e = 9.00$, for words; and $F(1,72) = 17.72$, $MS_e = 8.66$, for pictures; both $ps < .001$] and that this difference interacted with frequency [$F(3,258) = 10.53$, $MS_e = 9.17$, for words; and $F(3,216) = 5.87$, $MS_e = 6.87$, for pictures; both $ps < .001$]. For both words and pictures, the nature of the interaction was that the X–Y difference increased with presentation frequency.

Hintzman et al. (1992) examined evidence for proportionality by plotting scatter diagrams relating nonzero JOFs for Y against those for X. If proportionality holds, the points should fall on a straight line with slope less than 1, passing through the 0,0 intercept. Figure 3 shows proportionality plots of nonzero JOFs for each combination of materials and instruction, with straight lines fitted by eye (data from all subjects are included, not just those used in the ANOVA). These graphs resemble those for four different experiments, displayed in Figure 6 of Hintzman et al. (1992). The data for words under structure instructions may deviate from the general pattern, in that the intercept of a best-fitting line is reliably greater than 0, and those for pictures under frequency instructions have a slope only slightly less than unity. In general, though, these data sets are consistent with those of Hintzman et al. (1992) and with the proportionality prediction.

JOF = 0. The height of the zero bar in the histogram of Y judgments shown in Figure 2 is a measure of subjects' ability to reject highly similar test items as new. The major puzzle turned up by Hintzman et al. (1992) was that, even though it was not near the 100% ceiling, the height of that bar rose slowly or not at all beyond presentation frequencies of about 2. Figure 4 shows the percentage of JOF = 0 for each of the conditions of the present experiment. The upper three curves of each graph show correct rejections of Y items, and the lower three curves show incorrect rejections of X items (the leftmost point on each curve represents correct rejections of items that are entirely new).

Data from the neutral instruction condition are very similar to those obtained by Hintzman et al. (1992), show-

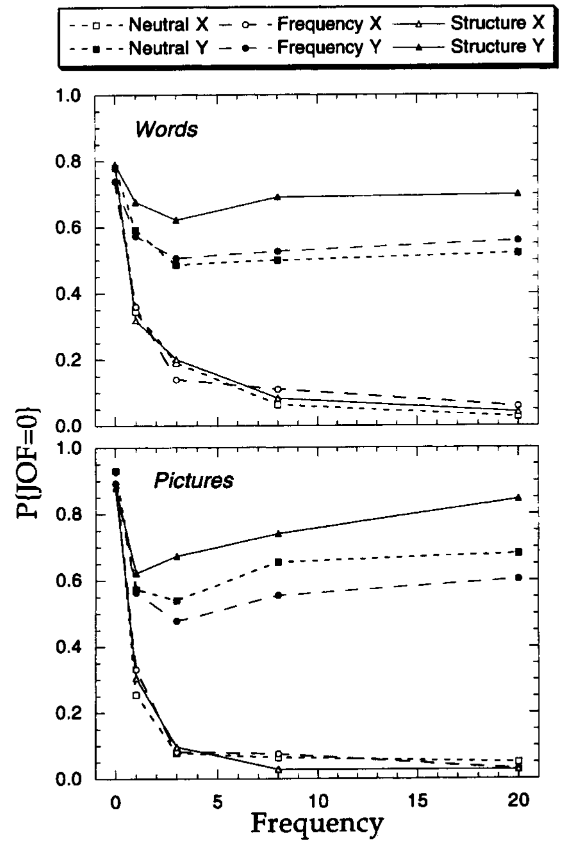

Figure 3. Proportions of judgments of frequency of 0, for Experiment 1. X, old targets; Y, similar items.

ing considerable learning of the X–Y discrimination in the first 3 presentations, but little additional learning beyond frequency = 3. The new question addressed here concerns the effect of instructions on this later learning. To approach this question statistically, we performed separate ANOVAs on the word and picture data, including only frequencies 3, 8, and 20. Both analyses revealed main effects of instructions [$F(2,208) = 8.77$, $MS_e = 0.20$, for words; and $F(2,208) = 11.18$, $MS_e = 0.21$, for pictures; both ps $< .001$]. Post hoc tests showed that structure instructions led to better performance than did neutral or frequency instructions. For neither set of materials did the frequency and neutral conditions differ reliably. A test for linear trend across frequencies 3, 8, and 20 was marginally reliable for words [$F(1,208) = 5.61, p < .02$] and highly reliable for pictures [$F(1,208) = 49.01, p < .001$]. In neither case, however, did the trend interact reliably with instructions (both Fs < 1).

Could the relative flatness of the X–Y discrimination curves of Figure 4 be due to subject or item differences? In principle, one could get a suboptimal ceiling on performance if some subjects learned the crucial features while other subjects never did, or if some items could support such learning while other items could not. We scrutinized subject and item differences in several ways and found no evidence for the extreme differences that would produce this result as an artifact. Distributions of the JOF = 0 proportion across subjects appeared normal, with standard deviations around .23. Moreover, these proportions correlated only about $r = .14$ between words (Experiment 1a) and pictures (Experiment 1b), showing little consistency among subjects across stimulus materials. Distributions of the JOF = 0 proportion across items were also normal, with even smaller standard deviations (.09 for words and .14 for pictures). It seems very unlikely, therefore, that subject or item differences could have created a low performance ceiling.

Discussion

Several features of these data are of interest. First, we have replicated the findings previously reported by Hintzman et al. (1992). We again obtained bimodal frequency-judgment distributions for stimuli highly similar to old targets; we again found that nonzero judgments to similar items were roughly proportional to those to targets; and we again found little or no learning of the feature necessary for X–Y discrimination past the first few study trials, even as judgments of frequency continued to increase. There was some late learning of picture orientation, but it was very slow. Hintzman et al. (1992), too, found very slow differentiation of X and Y in some conditions. Such learning may be more evident for the orientation of pictures than for the plurality of

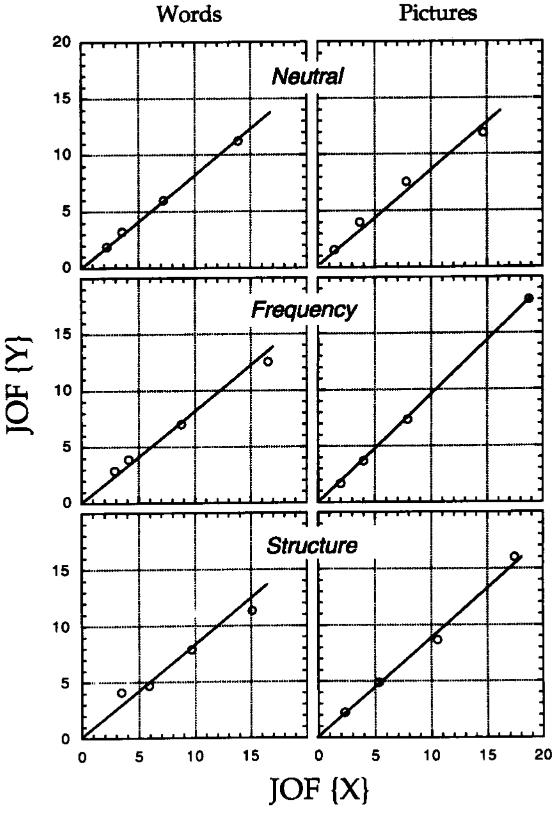

Figure 4. Scatter diagrams relating mean nonzero judgments of frequency for old targets (X) and similar items (Y), Experiment 1.

nouns in the present data, but this is not consistent across studies. For example, Hintzman et al. (1992) found more late discrimination learning with nouns than with pictures in their Experiments 2 and 3.

Second, while instructions did affect the learning of X–Y discriminations, the effects were largely confined to Presentations 1–3. If subjects were alerted to the aspects of the stimuli that would be important, as they were by the structure instructions, they did a better job of encoding the crucial features. This learning advantage effectively disappeared, however, after the first few exposures of an item. One might suspect that this result was due to subjects' changing their encoding strategy during list presentation—as might happen if they gradually forgot the instruction—but such a trend would also have affected items with frequency = 1 and 3, because low-frequency items were dispersed throughout the study lists. It appears, therefore, that telling subjects what feature to learn affects initial encoding but has little effect on how later repetitions are processed. This result is contrary to our original expectations.

Third, the instructional manipulation had opposite effects on X–Y discrimination and on judgments of frequency. That is, the instruction to prepare for a frequency-judgment test yielded the most accurate JOFs, whereas the instruction to learn pluralities of words and orientations of pictures yielded the best discrimination between old and similar test items. On an intuitive level this is not surprising, and it is consistent with the effects of compatibility between encoding process and retrieval task that are routinely reported in the memory literature. Nevertheless, it poses problems for models that assume only one kind of learning, a topic to which we return later.

Fourth, the effect of instructions on memory for frequency is of interest in its own right, because the literature suggests that there is no such effect. In a number of studies, frequency judgments after explicit warnings of the nature of the upcoming test have been compared with such judgments after less explicit or misleading instructions. These studies have uniformly reported an absence of instruction effects (Attig & Hasher, 1980; Flexser & Bower, 1975; Greene, 1984; Hasher & Chromiak, 1977; Howell, 1973; Kausler & Puckett, 1980; Rose & Rowe, 1976). However, in only one of these studies were the frequencies greater than 6. A look at the JOF means reported by Howell (1973) suggests that frequency instructions led to more accurate judgments than did free recall instructions at frequency = 10. This would be consistent with the present results, in which the superiority of frequency over both neutral and structure instructions emerged only at frequency = 8, and in which it was pronounced at frequency = 20. It is not clear why warnings of a frequency judgment task should be effective only at relatively high frequencies. One possibility is that appropriately instructed subjects try to code frequency in an associative or propositional form (explicit counting would be one example of such a code). This could provide an advantage primarily at high frequencies, where there is the most ambiguity concerning how to map familiarity onto JOFs. The literature is mixed on whether subjects spontaneously engage in direct coding of frequency when they are uninformed about nature of the upcoming test (Hintzman, 1982; Hintzman, Nozawa, & Irmscher, 1982; Jonides & Jones, 1992), but it is plausible that they would employ such a strategy when told that memory for frequency will be tested.

EXPERIMENT 2

We do not claim that subjects fail to learn stimulus structure beyond the first few presentations. In some conditions (e.g., words in the present Experiment 1; and pictures in Experiment 1 and words in Experiment 4 of Hintzman et al., 1992), improvement is not detectable, but in others (e.g., pictures in the present Experiment 1, words in Experiment 3 and pictures in Experiment 4 of Hintzman et al., 1992), it is. Even in the latter cases, however, later learning seems to be anomalously slow. Traditionally, learning theorists have debated whether the typical learning curve is exponential (Bush & Mosteller, 1951; Estes, 1950), or sigmoid (Culler & Girden, 1951). The present data suggest learning curves that deviate from exponential in the opposite fashion: The initial learning rate is high, but it slows drastically, far short of the 100% ceiling.

The measure used in Figure 4 somewhat obscures the shape of the learning curve, however, because the proportion of JOF = 0 given to Y items should both start high, at frequency = 0, and end high, at frequency = 20. One goal of Experiment 2 was to demonstrate the anomalous shape of the learning curve by using a forced-choice test, where chance performance is 50%. This test also enabled us to determine whether the registration without learning phenomenon was somehow an artifact of the subject's seeing only one version of the stimulus (either X or Y, but not both) on the test.

A second goal was to compare the learning of the X–Y discrimination with acquisition of an arbitrary category assignment. In principle, learning that BELL was shown in plural form and DEAL in singular form should be no more difficult than learning that BELL belongs to category A and DEAL to category B. We therefore compared forced-choice performance after learning in three conditions: neutral instructions and structure instructions, both as in Experiment 1, and category instructions, in which each stimulus had to be associated with one of two category names. The instructional manipulation was used with words, in Experiment 2A, and with pictures, in Experiment 2B.

Method

Subjects. A total of 240 University of Oregon undergraduates participated for course credit. These subjects were tested in groups of 2–11 persons. Each group was randomly assigned to one of the three instruction conditions: neutral, structure, or category. The numbers in each condition ranged from 74 to 85.

Procedure. The subjects were presented with the same word and picture study lists as in Experiment 1, at the same presentation rate. Again, approximately half the subjects saw the pictures first and half saw the words first, and the subjects were assigned to the same instruction condition for both types of stimuli. Subjects in the category condition were presented with either an A or a B directly above each to-be-remembered stimulus. These subjects were told that each stimulus would be assigned to the A category or the B category, and that they should try to remember the category assignment for a later memory test. In the category condition, all words were presented in singular form. Up until the test, the neutral and structure subjects were treated identically to those in the corresponding conditions of Experiment 1.

Following the study lists, two test lists of 24 stimuli each were presented. As in Experiment 1, words and pictures were tested separately, following the same order as the word and picture study lists. Each stimulus was presented for approximately 4 sec, with a 1.5-sec ISI, and responses were recorded by subjects on computer-scored bubble sheets. Subjects in the category group were asked to mark the category (A or B) to which each test item belonged. Subjects in the neutral and structure conditions were given a forced-choice recognition test. Each test displayed two stimuli side by side. In the word test list, the 2 stimuli were the singular and plural forms of one of the nouns, and in the picture test list, the 2 stimuli were a target picture and its mirror-reversed counterpart. Right and left assignments of the correct alternative (and of singular and plural, in the case of the words) were counterbalanced. The subjects were instructed to mark their form "A" if they had originally studied the stimulus to the left and "B" if they had studied the one on the right.

Results

Figure 5 shows the mean performance on the forced-choice test separately for words and for pictures. All curves have been anchored at .5 for frequency = 0 (which was not tested), and exponential curves with asymptotes of 1 are shown for comparison in both panels. Every empirical curve overshoots the exponential at low frequencies and undershoots it at high frequencies. Interestingly, this was true for the category-learning task, as well as for the other two instruction conditions.

The difference between the neutral and structure conditions is apparent in both panels of Figure 5, and it is in the same direction as in Experiment 1. However, neither the main effect of neutral versus structure nor its interaction with frequency was statistically reliable for either set of materials. Performance in the structure condition was virtually identical to that in the category condition for words, but reliably better than the category condition for pictures [$F(1,157) = 12.17$, $MS_e = .065$, $p < .001$]. The latter difference also interacted reliably with the linear trend on frequency [$F(1,157) = 5.65$, $p < .02$].

Discussion

This experiment showed that the result of fast initial learning and slow later learning of structural features of stimuli is also observed in a forced-choice recognition test. Direct comparisons with the exponential function reveal the anomalous shape of the learning curve.

However, we obtained a curve of the same shape for the acquisition of arbitrary category memberships. This suggests that the instruction to learn word plurality may induce subjects to treat that as a category-learning task (singular vs. plural). A somewhat different outcome was found with picture stimuli. The picture category data deviated less strikingly from the exponential, and there was considerably less category learning than structure learning on Presentations 1–3, where even the neutral instruction subjects performed better than those learning whether stimuli belonged to category A or category B. This result may reflect the very different natures of the picture orientation and category tasks. Learning the orientation of a picture involves familiarization with the stimulus itself, whereas learning the category assignment of the picture requires associating it with an arbitrary category name. By contrast, familiarization was presumably not a problem in the case of the nouns. In addition, the nouns naturally fall into two classes, singular and plural, whereas the drawings contain no surface cue, like plurality, around which a categorization strategy might be built.

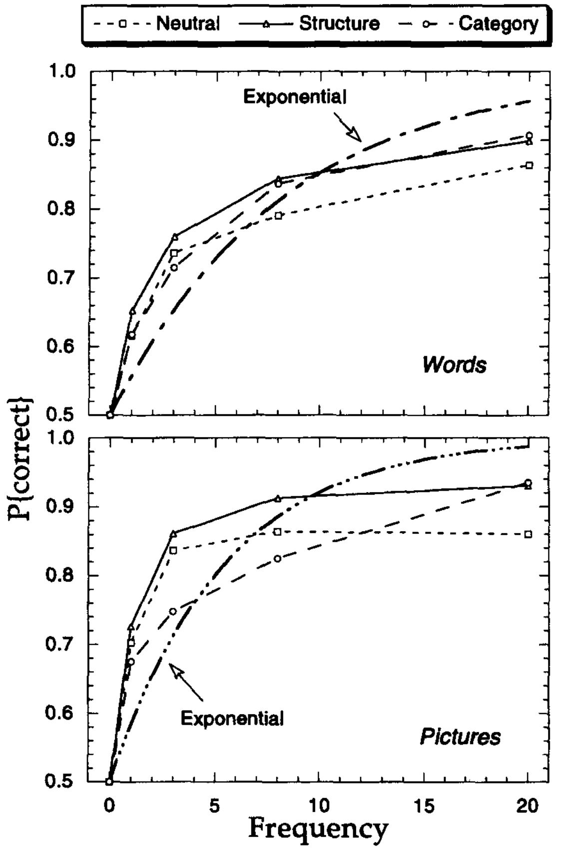

Figure 5. Forced-choice recognition learning curves from Experiment 2. Exponential curves are shown in each panel for comparison with the data.

It is puzzling, nevertheless, that the category-learning data deviated from the exponential. The category task was essentially that of paired-associate learning with two response alternatives. The literature shows that this task routinely yields learning curves that are closely approximated by the exponential (e.g., Bower, 1961). However, it has been customary in experiments on paired-associate learning to test subjects as part of the acquisition process. The typical experiment uses the anticipation method, where each trial is composed of a test phase immediately followed by a study phase. In our experiment, in contrast, presentation trials were dispersed throughout a long study list and subjects were not tested until the study list was over. Under these conditions, subjects may not be motivated to learn more—about either category membership or structure—because they may not realize that their knowledge is incomplete. This seems especially likely when the study trial presents information that already seems highly familiar, so that the subject believes that the stimulus as a whole is well known. Put differently, global familiarity would be a poor basis for assessing one's knowledge of an individual feature of a stimulus. Thus, the subject who repeatedly sees BELLS during the study list may have no way of realizing that the word's plurality has not been encoded, because the word as a whole seems highly familiar, and its plurality is evident in the stimulus itself.

EXPERIMENT 3

Experiment 3 was done to determine whether we could make subjects realize during study that their knowledge of stimulus structure was incomplete. Singular and plural nouns were studied by subjects in two conditions. In one, subjects were told to remember pluralities, as in the structure condition of the previous experiments. In the other condition, subjects were first shown the word stem and then the completion (e.g., BEL__ , BELLS), with the instruction to try to anticipate whether the completion was singular or plural before the complete word was shown. We reasoned that subjects who tried to anticipate a word's plurality and failed would become aware of their lack of knowledge and therefore would learn.

Method

Subjects. A total of 165 University of Oregon undergraduates participated for course credit. They were tested in groups of up to 12 persons, and each group was assigned to either the structure condition or the completion condition. Altogether, 77 subjects served in the structure condition and 79 served in the completion condition (the data of an additional 4 structure and 5 completion subjects were deleted, owing to failure to follow instructions).

Procedure. The presentation list was the same as that used in the word conditions of the previous experiments. Subjects in both instruction groups were informed prior to the study list that their memory for the words and the pluralities of the words would be tested. Subjects in the structure condition saw each word for 4 sec. Subjects in the completion condition were presented for 2 sec with a word stem that was truncated before the final letter of the singular form (e.g., BEL__). Then, for the remaining 2 sec, the complete word was revealed (either BELL or BELLS). The interword interval was 500 msec for both conditions. Thus presentation rate was equated at 4.5 sec/word for the two groups. Subjects in the completion group were further instructed to try to complete each word upon seeing its incomplete form. They were told that trying to guess a word's completion would be an especially helpful method for learning the word's plurality. Following the study list, there was a frequency-judgment test identical to the one used in Experiment 1.

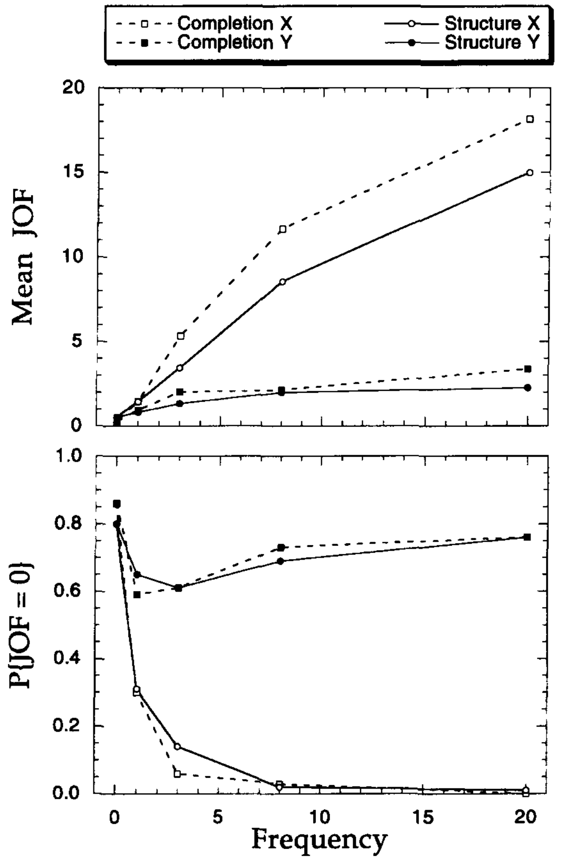

Figure 6. Data from Experiment 3. Top panel: mean judgments of frequency. Bottom panel: proportions of judgments of frequency of 0. X, old targets; Y, similar items.

Results

Mean JOFs and proportions of JOF = 0 are shown in Figure 6. The structure condition data were virtually identical to the word data obtained with the use of the same instructions in Experiment 1. Completion instructions led to higher frequency judgments on X items than did structure instructions [$F(1,153) = 24.30$, $MS_e = 24.6$, $p < .001$], and the effect of instructions interacted with frequency [$F(1,153) = 13.86$, $p < .001$]. However, correlations between JOF and frequency did not differ reliably between conditions, and—as can be seen in the bottom panel of Figure 6—the JOF = 0 proportions in the two instruction conditions were virtually the same. The proportions in both conditions did, however, increase reliably over frequencies 3, 8, and 20, as shown by a test for linear trend [$F(1,155) = 45.4$, $p < .001$].

Proportionality plots for mean nonzero JOFs to items with frequencies of 1, 3, 8, and 20 are shown in the first two panels of Figure 7. In both instruction conditions, the scatter diagrams are well described by a straight line passing through the origin and having a slope less than 1.

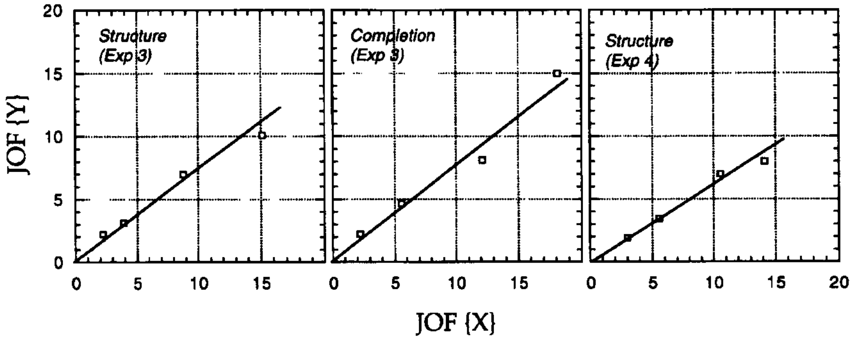

Figure 7. Scatter diagrams relating mean nonzero judgments of frequency for old targets (X) and similar items (Y), Experiments 3 and 4.

The nonzero judgments therefore appear to be in accordance with the proportionality prediction. To do statistical tests unbiased by subject selection, we discarded data on all subjects who failed to contribute at least one nonzero judgment to each X and Y cell with frequency > 0. This left 18 subjects in the structure condition (21%) and 19 in the completion condition (23%). An ANOVA of the data from both groups combined showed that judgments were higher to X than to Y items [$F(1,35) = 19.69, p < .001$] and that this difference interacted with frequency [$F(1,35) = 16.26, p < .001$]. Neither effect interacted reliably with instruction, although the tendency for completion subjects to make higher judgments than those made by structure subjects was marginally reliable [$F(1,35) = 5.46, MS_e = 75, p < .05$].

Discussion

Asking subjects to mentally complete word stems with either the singular or the plural form before seeing feedback did not have the expected effect. The failure to continue learning plurality beyond frequency = 3 was just as pronounced for the completion subjects as for the structure subjects. The only discernible effect of stem completion on performance in this experiment was to raise the mean of the nonzero judgments of frequency. This may be a manifestation of the effect of generation on frequency judgments observed by Greene (1988). If generation of the complete word enhances its episodic familiarity, one might expect it to produce higher judgments of frequency.

What accounts for the failure of subjects to learn plurality even when given an opportunity for self testing? One possibility is that group testing situations, as used in Experiments 1–3, lead to lax performance. Another possibility is that the feedback that subjects obtained on their knowledge states was ambiguous, because their knowledge states were ambiguous. That is, not having to face the reality of overt errors, subjects might have deluded themselves into overestimating their knowledge of the target nouns' pluralities.

EXPERIMENT 4

This experiment was essentially a replication of Experiment 3, with two changes: All subjects were tested individually, and subjects in the completion condition were required to respond overtly, by indicating with a keypress whether the word completing the stem was singular or plural.

Method

There were 48 subjects, recruited as before, and tested individually. Half received the structure instructions and half received the completion instructions. In the previous experiments, stimuli had been projected on the wall of the experimental room; in Experiment 4, subjects saw the stimuli on the screen of a Macintosh Plus computer. All timing and instructions were the same as in Experiment 3, except that subjects in the completion condition were required to respond on the computer keyboard during the anticipation phase of each study trial, by pressing one key if the completion of the word was singular and another key if it was plural.

Results

One subject, in the completion condition, failed to follow instructions on the frequency judgment task. That subject's data were eliminated, leaving 23 subjects in one condition and 24 in the other.

Figure 8 shows the performance of completion subjects during anticipation trials, on those words that were presented 20 times. An exponential curve with an asymptote of 1 has been fitted to the data. The data conform well to the exponential, except for fluctuations in performance just short of ceiling on the later trials. Such fluctuation is not unusual in paired-associate learning, although it is sometimes absent from published learning curves—particularly if training has been terminated once a subject reaches a criterion such as one or two errorless cycles through the list. The graph shows that subjects are capable of learning the pluralities of nouns when they are tested individually and required to make overt responses. Group testing per se is not the cause of registration without learning, because the structure subjects showed the same pattern as was found in previous experiments.

Mean JOFs, shown in the top panel of Figure 9, are similar to those of Experiment 3. An ANOVA done only on the judgments for X items showed that completion subjects gave marginally higher JOFs than did structure subjects [$F(1,45) = 3.38, MS_e = 24.9, p < .08$], and that this difference interacted with the linear trend on frequency [$F(1,45) = 4.18, p < .05$].

The JOF = 0 proportions are shown in the bottom panel of Figure 9. The failure to demonstrate learning of the X–Y discrimination has been essentially eliminated by requiring overt completion of word stems during learning, although—even after 20 presentations—performance still fell short of 100%. An ANOVA on $P(\text{JOF} = 0)$ for Y items showed no reliable main effect of instruction condition ($F < 1$), but instruction condition did interact with the linear trend on frequency [$F(1,45) = 5.81, p = .02$]. The performance of completion subjects was reliably worse than that of structure subjects at frequency = 1 [$t(45) = 2.06, p < .05$], but better at frequency = 8 [$t(45) = 2.21, p < .05$]. By two-tailed test, the instructional difference at frequency = 20 was short of significance.

Only the structure subjects contributed enough nonzero judgments on Y items to allow examination of the proportionality of nonzero X and Y judgments. A scatter diagram plotting mean nonzero judgments for Y against those for X is shown in the right-hand panel of Figure 7. Again, a straight line passing through the origin and having a slope less than 1 does a good job of fitting the data. Subjects giving nonzero judgments to at least one item in each cell with frequency > 0 were too few to allow a meaningful ANOVA on these data.

Discussion

The outcome of this experiment was similar to that of Experiment 3, except that the subjects given the stem completion task showed virtually complete learning of plurality. After a single exposure, however, the ability of completion subjects to correctly reject Y items was below that of structure subjects. This was probably because in the stem-completion task the entire word was exposed for 2 of the 4 sec allotted to stimulus presentation, so that exposure duration for completion subjects was half that for structure subjects. (There was a nonsignificant difference in the same direction at frequency = 1 in Experiment 3.)

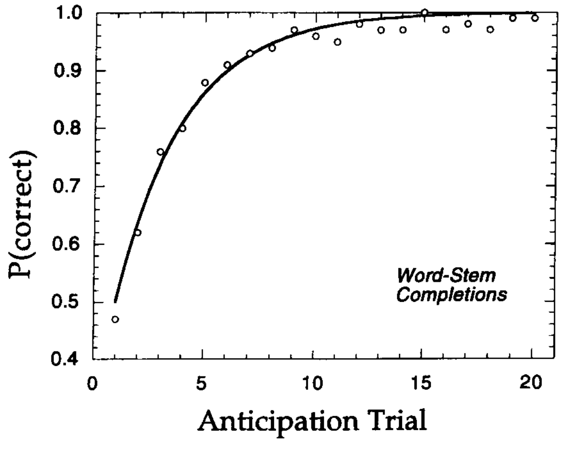

Figure 8. Proportion of correct responses to the frequency = 20 items during learning, for the completion subjects of Experiment 4.

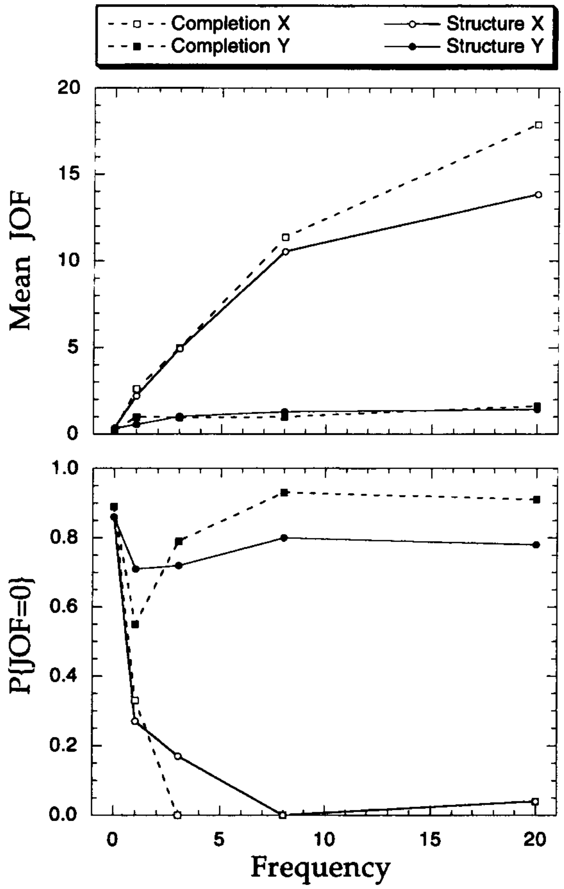

Figure 9. Data from Experiment 4. Top panel: mean judgments of frequency. Bottom panel: proportions of judgments of frequency of 0. X, old targets; Y, similar items.

Completion-subject performance fell just short of the 100% ceiling even after 20 anticipation trials. This failure to perform perfectly could reflect careless responding, or a combination of forgetting and incomplete learning. The hypothesis that learning was incomplete is supported by the fluctuation seen at ceiling in Figure 8. Incomplete learning could be a consequence of reliance on corrective feedback, because with only two response alternatives (singular and plural) the probability of guessing correctly is 0.5. Thus a subject who correctly guessed on an anticipation trial that the completion of BEL__ was BELLS might not realize that the response was a guess, and thereby pass up an opportunity to learn the completion. The learning of paired associates has been found to be slower when there are only two response alternatives than when there are several (Hintzman, 1967; Smith, Jones, & Thomas, 1963), and this too may reflect subjects' reliance on corrective feedback for learning.

GENERAL DISCUSSION

These experiments replicated the outcome that Hintzman et al. (1992) called registration without learning:

the tendency for frequency judgments to continue to increase even after learning of the discrimination between old (X) and very similar (Y) items has effectively stalled. Frequency judgments to Y were again bimodally distributed, with one mode at zero and the other mode tracking—but lagging proportionally behind—frequency judgments to X.

Forced-choice testing (Experiment 2) revealed the anomalous shape of the learning curve for the X–Y discrimination, overshooting the exponential on initial exposures and undershooting it on later repetitions. Surprisingly, however, the learning curve for an arbitrary categorization of the same stimuli took a similar form.

There appears to be a cognitive bias against learning structure on later repetitions, as contrasted with early ones, which is hard to overcome. Instructions to learn the discriminating feature raised the initial learning rate but failed to overcome this late-repetition bias (Experiment 1). Instructing subjects to covertly test their own knowledge had no apparent effect beyond telling them what feature to learn (Experiment 3). Only when subjects were required to respond overtly prior to feedback did they show a fairly constant rate of learning over Presentation Frequencies 1–20 (Experiment 4). This shows that the bias against learning can be overcome and helps explain why studies using the traditional anticipation procedure fail to reveal the bias. A general conclusion is that repeated presentation without overt testing is a relatively ineffective method of learning.

This conclusion is consistent with other findings. The importance of overt responding and explicit feedback are generally acknowledged in educational psychology and have been documented in experimental research going back to the early part of the century. For example, in his classic study, Gates (1917) compared learning through recitation with learning through repeated reading, by presenting both nonsense and meaningful materials to both school children and adults. On the basis of his results and a review of the existing literature, Gates concluded, "In general, recitation, after a few initial readings, is of much more value than more reading" (p. 61). Operant-conditioning accounts of the effectiveness of overt responding, which have been popular in educational psychology, emphasize the importance of confirmation (reinforcement) of correct responses, as opposed to the identification and elimination of errors (Skinner, 1968). We suspect, however, that overt responding is effective primarily because it confronts subjects with the gaps in their knowledge.

Despite the early empirical evidence on recitation versus rereading, as well as the growing evidence of the effectiveness of self-testing, in educational research (e.g., Hamaker, 1986), most memory models still ignore overt responding and feedback as factors in encoding. This may be a legacy of the models of the 1960s and 70s, in which cognition was the computer-like processing of perceptual inputs, and learning was simply a matter of storing the products of that processing. There seemed to be no need in such conceptions for overt responding and feedback to play a role in learning, except for the apparently unusual (and therefore largely ignored) case of motor skills. Modern connectionism has reintroduced error-correcting feedback, in the form of supervised learning models (see, e.g., Grossberg, 1987; Rumelhart & McClelland, 1986), and there is a closed-loop version of Murdock's TODAM (Lewandowsky & Murdock, 1989; McDowd & Murdock, 1986); but these models have not been applied widely to data from standard memory tasks.

It is tempting to relate the present findings to two other well-known phenomena, although the similarities may be superficial. One phenomenon is the generation effect (see, e.g., McDaniel, Waddill, & Einstein, 1988), in which subjects are shown to remember words better when they generate them in response to highly constraining cues (e.g., completing the pair BOY–GI__) than when they merely read them aloud (BOY–GIRL). Generation effects, however, show up in a single presentation. They also show up in judgments of frequency (Greene, 1988; see also the present Experiments 3 and 4). In contrast to this pattern, our results suggest a bias against learning that is specific to late repetitions and that has little effect on judgments of frequency (repetitions continue to register after learning has ceased). The other phenomenon that might be related to our findings is the spacing effect, in which massed repetitions of an item lead to poorer memory than do spaced repetitions, presumably because encoding is deficient on presentations that closely follow the first (see, e.g., Hintzman, 1974). Spacing effects, however, show up in many memory tasks, including judgment of frequency; and the present results were obtained even though repetitions were spaced. We conclude that, if there is a connection between registration without learning and either the generation effect or the spacing effect, the connection is not obvious.

Why should the learning of structure slow drastically, short of the 100% asymptote, in the absence of overt responding and feedback? Hintzman et al. (1992) suggested that failing to extract all available information from an item, despite numerous encounters with it, might be characteristic of the way humans and other animals interact with the environment. Once the cognitive system knows enough about an object to deal with it effectively, the system may resist wasting resources by analyzing the object further. This idea is generally consistent with mismatch theory, as presented recently by Johnston and Hawley (1994). Having considered a variety of cognitive phenomena, these authors propose that the mind is simultaneously biased toward top-down processing of expected objects and bottom-up processing of unexpected objects. An effect of the former bias ("equilibration") is that small discrepancies between the expected input and the actual input are not noticed. An effect of the latter bias ("transformation") is that large discrepancies are singled out for attention. This is consistent with our learning curves, which show rapid learning of structure on the first one or two presentations, but little learning on subsequent repetitions (where the stimulus is more "expected" than it was initially).

A possible basis for the distinction between expected and unexpected stimuli is the familiarity or strength signal that is assumed in one form or another in most current models of recognition memory. This signal, variously called echo intensity (Hintzman, 1988), similarity (Eich, 1982; Murdock, 1982), matching (Humphreys, Bain, & Pike, 1989), and familiarity (Gillund & Shiffrin, 1984), is seen in these models as the primary basis of recognition judgments. These models have been called global recognition models, and the proposed familiarity signal is "global" in two senses: (1) it is based on the match of the memory probe with *all* items in memory (or at least with all items learned in a particular context), and (2) it is based on the degree of match between stored information and the probe *as a whole*. It is because of the second of these properties that global matching models predict a proportional relationship between frequency judgments to Y and frequency judgments to X, as discussed earlier. (A more complete discussion of this point can be found in Hintzman et al., 1992.)

Our argument is that a weak familiarity signal could be the cue to do bottom-up processing of stimulus structure, and a strong familiarity signal could be the cue to let top-down processing suffice. Top-down processing could result in storing information that the item has been repeated (registration), without adding information about details such as those needed to distinguish X from Y. Because the familiarity signal is global in the second sense—memory is matched to the stimulus as a whole—the system is fooled into thinking it has stored adequate information about the repeated stimulus, even when the distinguishing features are not known.

Such an account helps explain why structure instructions—telling subjects to learn the orientations of drawings or the pluralities of nouns—only help in the beginning. Unless subjects are required to overtly predict the distinguishing feature and are thereby confronted with their errors, they may be oblivious to their failure to master the task. This account applies equally to learning the structure of an individual stimulus and to associating items with arbitrary category labels like A and B (Experiment 2). If the subject sees TRUCK–A, and if both TRUCK and A are highly familiar in the experimental context, there may be no basis for realizing that the association has not been learned. In the anticipation-learning procedure, by contrast, because overt testing is obligatory, one's lack of knowledge cannot easily be ignored.

Recent research suggests that people are poor at judging the degree to which a perceptual experience is driven by bottom-up as opposed to top-down processing. One consequence of this inability is that perception can be biased by expectations, or by past experience, in ways of which the perceiver is unaware (see, e.g., Jacoby, Allan, Collins, & Larwill, 1988; Ste-Marie & Lee, 1991). A less appreciated consequence is that the sufficiency of information available to the senses can mask the insufficiency of information available from memory. Our hypothesis is that subjects who study material simply by repeating the perceptual experience (e.g., rereading) are unable to identify gaps in their top-down knowledge, because the bottom-up information fills in, so that they do not experience the gaps. Thus in the absence of explicit errors and corrective feedback, structural learning slows or comes to a halt.

The type of learning mechanism we have in mind is similar to that hypothesized by Johnston and Hawley (1994), outlined earlier, and to Grossberg's ART model (e.g., Grossberg, 1987). There is an important difference between the present idea and the ART model, however. That model assumes that perceptual experience drives two kinds of learning: (1) if the mismatch between bottom-up and top-down sources of information is above a certain threshold, this triggers creation of a new template; and (2) if the mismatch is below the threshold, the old template is tuned. Our proposal, in keeping with Johnston and Hawley (1994), is that when the mismatch is relatively minor—as when a small percentage of bottom-up features are missing from the template—even tuning of the old template may not occur. Indeed, to the extent that the system is unable to differentiate between bottom-up and top-down sources of information, a reliable signal for tuning presumably would not exist.

In contrast to an adaptive-learning system such as ART, most models that have evolved directly to account for data from the memory laboratory assume that repetition improves the reliability of stored information in an open-ended fashion—either by increasing redundancy (Hintzman, 1986, 1988) or by increasing strength (Gillund & Shiffrin, 1984; Humphreys et al., 1989; Murdock, 1982) . Such models can explain why judged frequency continues to increase with repetition, but they predict a corresponding improvement in X–Y discrimination. On the other hand, an adaptive learning model might be made to predict the slow-down in learning of the X–Y discrimination—for example, through the assumption in ART that tuning is very slow, or through the assumption in the closed-loop version of TODAM that learning stops before the error feedback signal reaches zero. Such models, however, would not explain why judged frequency continues to grow after discrimination learning has stopped.

The key assertion here (and in Hintzman et al., 1992) is that there is a decoupling between judgments of frequency on the one hand, and learning the X–Y discrimination on the other. The present data reveal that decoupling in at least two ways. First, Experiments 1 and 3 revealed continued increases in mean JOF with repetition, in the absence of substantive improvement in the ability to discriminate X from Y. This replicates the registration-without-learning effect reported by Hintzman et al. (1992). Second, study instructions sometimes affected mean JOF and X Y discrimination differently.

In Experiment 1, structure instructions led to better initial X–Y discrimination learning than did neutral instructions. Structure instructions also led to higher JOFs. This difference emerged on early presentations in a way consistent with the view that JOF and X–Y discrimination are manifestations of the same learning. However, contrary to that view, frequency instructions led to the highest JOFs at frequency = 20, and JOF accuracy was greatest under frequency instructions overall. In Experiment 3, completion instructions led to higher frequency judgments than did structure instructions, but there was no corresponding difference in X–Y discrimination.[4] In Experiment 4, the same instructional effect emerged later for JOF than for X–Y discrimination, and overt responding and feedback had a large effect on X–Y discrimination and only a small one on JOF.

On intuitive grounds, such dissociations do not appear hard to explain: Learning that an item was repeated seems different from learning more about the nature of that item. To explain our results, memory models may have to incorporate such a difference. It is unclear how easy such models will be to construct.

REFERENCES

Attig, M., & Hasher, L. (1980). The processing of frequency of occurrence by adults. *Journal of Gerontology*, **35**, 66-69.

Bower, G. H. (1961). Application of a model to paired-associate learning. *Psychometrika*, **26**, 255-280.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, **58**, 313-323.

Culler, E., & Girden, E. (1951). The learning curve in relation to other psychometric functions. *American Journal of Psychology*, **64**, 327-349.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review*, **89**, 627-661.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, **57**, 94-107.

Flexser, A. J., & Bower, G. H. (1975). Further evidence regarding instructional effects on frequency judgments. *Bulletin of the Psychonomic Society*, **6**, 321-324.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, **6** (No. 40).

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

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

Greene, R. L. (1988). Generation effects in frequency judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 298-304.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, **11**, 23-63.

Gum, T., & Bub, D. (1988). *PsychLab Manual*. Montreal: PsychLab.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, **56**, 212-242.

Hasher, L., & Chromiak, W. (1977). The processing of frequency information: An automatic mechanism? *Journal of Verbal Learning & Verbal Behavior*, **17**, 173-184.

Hintzman, D. L. (1967). Some tests of a discrimination net theory: Paired-associate learning as a function of stimulus similarity and number of responses. *Journal of Verbal Learning & Verbal Behavior*, **6**, 809-816.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77-99). Potomac, MD: Erlbaum.

Hintzman, D. L. (1982). Are presentation frequency and spatial numerosity distinct attributes of memory? *Bulletin of the Psychonomic Society*, **20**, 196-198.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory & Language*, **33**, 1-18.

Hintzman, D. L., Curran, T., & Oppy, B. (1992). Effects of similarity and repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 667-680.

Hintzman, D. L., Nozawa, G., & Irmscher, M. (1982). Frequency as a nonpropositional attribute of memory. *Journal of Verbal Learning & Verbal Behavior*, **21**, 127-141.

Hintzman, D. L., Summers, J. J., Eki, N. T., & Moore, M. D. (1975). Voluntary attention and the spacing effect. *Memory & Cognition*, **3**, 576-580.

Howell, W. C. (1973). Storage of events and event frequencies: A comparison of two paradigms in memory. *Journal of Experimental Psychology*, **98**, 260-263.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, **96**, 208-233.

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 240-247.

Johnston, W. A., & Hawley, K. J. (1994). Perceptual inhibition of expected inputs: The key that opens closed minds. *Psychonomic Bulletin & Review*, **1**, 56-72.

Jones, C. M., & Heit, E. (1993). An evaluation of the total similarity principle: Effects of similarity on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 799-812.

Jonides, J., & Jones, C. M. (1992). Direct coding for frequency of occurrence. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 368-378.

Kausler, D. H., & Puckett, J. M. (1980). Frequency judgments and correlated cognitive abilities in young and elderly adults. *Journal of Gerontology*, **35**, 376-382.

Kortge, C. A. (1990). Episodic memory in connectionist networks. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 764-771). Hillsdale, NJ: Erlbaum.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, **96**, 25-57.

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory & Language*, **27**, 521-536.

McDowd, J. M., & Murdock, B. B., Jr. (1986). Mathematical models and the problem of stimulus variation: A comparison of MINERVA 2 and TODAM. *Acta Psychologica*, **62**, 177-188.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89**, 609-626.

Rose, R. J., & Rowe, E. J. (1976). Effects of orienting task and spacing of repetitions on frequency judgments. *Journal of Experimental Psychology*, **2**, 142-152.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing*. Cambridge, MA: MIT Press.

Skinner, B. F. (1968). *The technology of teaching*. Englewood Cliffs, NJ: Prentice-Hall.

Smith, T. A., Jones, L. V., & Thomas, S. (1963). Effects upon verbal learning of stimulus similarity, number of stimuli per response, and concept formation. *Journal of Verbal Learning & Verbal Behavior*, **1**, 470-476.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, famillarity, and visual complexity. *Journal of Experimental Psychology Human Learning & Memory*, **6**, 174-215.

Ste-Marie, D. M., & Lee, T. D. (1991). Prior processing effects on gymnastic judging. *Journal of Experimental Psychology Learning, Memory, & Cognition*, **17**, 126-136.

NOTES

1. Such models include those described by Gillund and Shiffrin (1984), Hintzman (1988), Kortge (1990), and Murdock (1982).

2. Jones and Heit (1993) found no evidence for multiplicative effects of repetition and similarity on either JOFs or recognition judgments to Y items. However, their X and Y items were members of the same taxonomic category and were therefore of much lower similarity—both semantically and physically—than the materials used in the studies under discussion here.

3. JOF distributions for Figure 2 and the other picture and word conditions are available from the first author.

4. This JOF difference might be a kind of rehearsal effect, owing to attempted anticipations. Effects of rehearsal on mean JOF were reported by Hintzman, Summers, Eki, and Moore (1975).

(Manuscript received February 25, 1994,
accepted for publication March 9, 1994.)