# A bias in favor of the positive response to high-frequency words in recognition memory

## YUJI HOSHINO

Tokyo Metropolitan University, Tokyo, Japan

A relatively liberal response bias for high-frequency words and a violation of the mirror effect for hit and false-alarm rates were found in a yes-no recognition-memory test. Subjects more frequently responded "old" to high-frequency words than to low-frequency words. Four experiments were conducted to determine the causes of the different response biases and of the violation of the mirror effect. The word-frequency effect on hit rates did not appear, whereas the false-alarm rate for low-frequency words was lower than that for high-frequency words. When low- or high-frequency words were presented separately in distinct halves of a recognition-memory test, the relatively liberal response bias for high-frequency words was diminished. A model for recognition judgment is proposed that assumes the use of a common criterion for low- and high-frequency words.

Low-frequency words are more accurately recognized as having been seen before than are high-frequency words (Balota & Neely, 1980; Earhard, 1982; Eysenck, 1979; Glanzer & Adams, 1990; Glanzer & Bowles, 1976; Gregg, 1976; Mandler, Goodman, & Wilkes-Gibbs, 1982; McCormack & Swenson, 1972; Rao & Proctor, 1984; Schulman & Lovelace, 1970; Underwood & Freund, 1970). In a yes-no recognition-memory test, the hit rate is higher and the false-alarm rate is lower for low-frequency words than for high-frequency words. Glanzer and Adams (1985) named such symmetrical patterns of hit and false-alarm rates the *mirror effect* and revealed that 23 of 24 studies of the word-frequency effect in recognition memory showed the mirror effect.

In contrast to previous studies, Experiment 1 of the present investigation did not show the mirror effect. The false-alarm rate was lower for low-frequency words than it was for high-frequency words, but the hit rates were not different with respect to word frequency. Instead of indicating the mirror effect, this pattern indicates a more liberal response bias for high-frequency words with respect to low-frequency words. The purpose of this study was to investigate the causes of the difference in response bias between low- and high-frequency words and of the violation of the mirror effect.

Part of this research was presented at the 47th, 50th, and 53rd Annual Conventions of the Japanese Psychological Association held in Tokyo in September 1983, in Nagoya in October 1986, and in Tsukuba in December 1989, respectively. This research was partly supported by a Grant-in-Aid for Scientific Research, Ministry of Education, Science and Culture, No. 63790026. I thank David Balota, Arthur Glenberg, Robert W. Proctor, Joan Snodgrass, and an anonymous reviewer for their useful comments. Requests for reprints should be sent to Yuji Hoshino, Department of Psychology, Faculty of Social Sciences and Humanities, Tokyo Metropolitan University, Minamiosawa 1-1, Hachioji, Tokyo 192-03, Japan.

# Measurement of Response Bias

Signal-detection theory has been used as a model for a process of recognition judgment (Banks, 1970; Lockhart & Murdock, 1970; Murdock, 1965; Parks, 1966). In signal-detection theory,  $\beta$  is used as a measure of response bias (Green & Swets, 1966; McNicol, 1972).  $\beta$  is a likelihood ratio of a signal distribution to a noise distribution. An increment in hit rates and/or false-alarm rates results in a decrease in  $\beta$ , indicating relatively liberal response bias.

Values of  $\beta$  depend partly on values of d' (Banks, 1970; Snodgrass & Corwin, 1988; von Winterfeldt & Edwards, 1982), which indicate the degree of discrimination between the noise and the signal distributions. The range of  $\beta$  values decreases as d' decreases. Consequently, Snodgrass and Corwin proposed an alternative measure of response bias, C, as follows:

$$C = z_{\rm FA} - d'/2,$$

where  $z_{\text{FA}}$  is the z score corresponding to the false-alarm rate. Unlike  $\beta$ , the value of C is not limited as discrimination becomes worse. Because C is independent of d', it is used as the reported measure of bias in the present study. However, similar patterns of response bias were obtained when  $\beta$  was used as the index of response bias.

# Word-Frequency Effect on Response Bias

In an experiment by McCormack and Swenson (1972), subjects given high-frequency words exhibited a reliable  $\beta$  decrement from the first to the sixth section of the test list, whereas subjects given low-frequency words did not. In the first section of the test list, there was no difference in false-alarm rates between low- and high-frequency words, although the hit rate was higher for low-frequency words than for high-frequency words. However, as subjects proceeded through the test lists, the false-alarm rate increased more for high-frequency words than for low-

frequency words, with no reliable changes in the hit rates. Thus, the criterion for high-frequency words became relatively more liberal.

Direct evidence for a more liberal response bias for high-frequency words with respect to low-frequency words is found in Experiment 2 of McNicol and Ryder (1971). They examined recognition memory for pairs of words that were rated as familiar associations and as unfamiliar ones. The liberal response bias for familiar pairs was apparent, whereas the discriminability of familiar pairs was not different from that of unfamiliar pairs. The different response biases seem to be obtained when discriminability of unfamiliar materials is not much greater than discriminability of familiar materials.

To minimize the difference in discriminability between low- and high-frequency words, words having a frequency of 16 occurrences per million were used as low-frequency words in the present study. The frequency of these words is higher than that of the words used as low-frequency words in previous experiments that obtained the mirror effect. For example, Mandler et al. (1982) used words that had a frequency of 1 occurrence per million as lowfrequency words; in Earhard (1982), frequencies of lowfrequency words ranged from 1 to 5 occurrences per million words. In other studies (Balota & Neely, 1980; Duchek & Neely, 1989; Glanzer & Bowles, 1976; Rao & Proctor, 1984; Underwood & Freund, 1970), frequencies of low-frequency words were lower than 5 occurrences per million words. The higher frequency of the low-frequency words in the present study should reduce their typical discriminability advantage.

The relatively liberal response bias for high-frequency words and the violation of the mirror effect were examined. In Experiment 1, the relatively liberal response bias for high-frequency words and the violation of the mirror effect were shown after incidental learning. In Experiment 2, the word-frequency effect on response bias and the violation of the mirror effect appeared under fast-and slow-test conditions. In Experiment 3, low- and high-frequency words were presented separately during a test phase. The separate presentations of the two levels of word frequency diminished the effect of word frequency on response bias. In Experiment 4, a lexical-decision task was performed during a study phase. The mirror effect emerged, and the different response biases disappeared.

# **EXPERIMENT 1**

For incidental learning, the subjects performed semantic and acoustic orienting tasks in which low- and high-frequency words were presented. Whether or not the effect of the orienting tasks on recognition memory interacted with the levels of word frequency was examined. In Experiment 2 of Mandler et al. (1982), the word-frequency effect on accuracy in recognition memory was reduced after a semantic orienting task, whereas in experiments by Duchek and Neely (1989) and Rao and Proctor (1984), a deeper level of processing facilitated recog-

nition memory for low-frequency words with respect to high-frequency words. A relatively rapid rate of presentation (about 1 sec/item) was used for both study and test lists.

## Method

**Design**. There were two within-subjects variables: orienting task (acoustic or semantic) and word frequency (low or high).

Subjects, Materials, and Apparatus. Twelve graduate and undergraduate students at Tokyo Metropolitan University participated in this experiment. Each subject performed the two orienting tasks and the recognition-memory test individually.

Two separate lists, each consisting of 20 low- and 20 highfrequency words, as well as 20 medium-frequency words for use as fillers, were constructed. All the words were nouns composed of two Kanji characters, which are Chinese characters that are used in Japan. Each list was used both in the acoustic and in the semantic orienting tasks. Approximately half of the low-, medium-, and high-frequency words in each list were likely to elicit responses from one of the two response categories used in each orienting task. All of the low-frequency words had a frequency of 16 occurrences per million words (The National Language Research Institute, 1962). The high-frequency words had a mean frequency of 343 occurrences per million words, and the medium-frequency words had a mean frequency of 97 occurrences per million words. The differences between the means of the four attributes other than word frequency for the low- and high-frequency words were minimized. According to the measurements of Ogawa and Inamura (1974), the mean values of imagery, concreteness, meaningfulness, and ease of learning were 4.40, 4.15, 3.37, and 4.39, respectively, for the lowfrequency words and 4.14, 3.80, 3.38, and 4.73, respectively, for the high-frequency words. Eighty additional words were used as distractors in the recognition-memory test. Half of the words had a frequency of 16 occurrences per million words (low-frequency distractors); the other words had a mean frequency of 343 occurrences per million words (high-frequency distractors).

A Fujitsu FM-8 microcomputer controlled all aspects of this experiment.

Procedure. The general procedure for the stimulus presentation was as follows. The subject was seated in front of the microcomputer and, at a comfortable distance, viewed a display controlled by the microcomputer. Before the item to be responded to was presented, "Ready" was shown for about 600 msec. The item then was displayed until the subject made a response by pressing one of the two keys on the keyboard. The interval between each response and the appearance of the next item was about 1 sec. The two keys' assignments to response categories were shown in the lower region of the display throughout the task. Each subject received a unique and random sequence of stimuli.

Each subject received both study lists, one with the acoustic orienting task and one with the semantic orienting task. In the acoustic orienting task, the subjects judged whether or not any of the consonants "s," "sh," "z," or "zh" were included in the presented word. In the semantic orienting task, the subjects judged whether the word was an abstract or a concrete noun. No subjects were told of the later recognition-memory test. The order of the two orienting tasks and the assignments of the two lists to the orienting tasks were counterbalanced across subjects.

Following the two orienting tasks, a lexical-decision task was performed to lower overall performance on the recognition-memory test. The subjects judged whether or not the presented string of two Kanji characters was a word when it was read in an ordinary way, that is, from the left character to the right one. The 40 medium-frequency words presented previously in the orienting tasks were used as word stimuli. Forty other medium-frequency words consisting of two Kanji characters with the order of the two characters

reversed were used as nonword stimuli. It took about 5 min to complete the lexical-decision task. Because the lexical-decision task served only as a filler, performance on it is not examined in the following data analyses.

Following the lexical-decision task, the recognition-memory test was administered. The 40 low- and 40 high-frequency words shown in the orienting tasks and the 40 low- and 40 high-frequency distractors were presented. The subjects were told that half of the items that would appear in this task were the words they had seen in either of the previous orienting tasks and that the other items were words that had not been seen in the experimental context. The subjects used their dominant hand for responding "old" and their other hand for responding "new." They were asked to respond as fast and as accurately as possible.

#### Results and Discussion

Hit rates, false-alarm rates, response times for hits and correct rejections, and values of d' and C in the recognition-memory test were analyzed. The significant level for subsequent statistical tests was set at .05. Response times were measured in about 9-msec units. Hit and false-alarm rates were corrected by adding 0.5 to the number of hits or false alarms and dividing by N+1, where N was the number of new or old items (Snodgrass & Corwin, 1988). This transformation is convenient and necessary for calculating values of d' and C, because hit rates of 1.0 and/or false-alarm rates of 0 cannot define the values of these measures. The transformation of hits and false alarms was applied to all of the data reported in the present study.

Table 1 shows the mean values of these measures for the two levels of word frequencies and the two types of orienting tasks. Because the words presented in the two orienting tasks were shown together in the recognition-memory test, distractors could not be classified under the types of orienting tasks. Thus, in Table 1, the false-alarm rates and the response times for correct rejections are the same under both the acoustic and the semantic orienting tasks.

For hit rates, the main effect of word frequency and its interaction with orienting task was not significant, but orienting task showed a significant effect  $[F(1,11) = 107.8, MS_e = .008]$ . The mean hit rate for the words presented in the semantic orienting task was .81, whereas for the words presented in the acoustic orienting task, it was .56.

Table 1
Recognition Memory for Low- and High-Frequency Words After
Acoustic and Semantic Orienting Tasks in Experiment 1

		•			
Word Frequency	Acoustic Task Sem		Seman	antic Task	
	Low	High	Low	High	
Hit rate	.54	.59	.78	.85	
False-alarm rate	.12	.36	.12	.36	
d'	1.31	0.62	2.11	1.57	
C	0.55	0.07	0.15	-0.40	
RT for hit	1.02	0.89	0.96	0.95	
RT for CR	0.92	1.06	0.92	1.06	

Note—RT = response time; CR = correct rejection. Response times are given in seconds. False-alarm rates and response times for false alarms are common to acoustic and semantic orienting tasks.

There was a significant effect of word frequency on response times for hits  $[F(1,11) = 8.8, MS_e = 77.0]$ . Hits for the high-frequency words were faster than hits for the low-frequency words (0.92 sec vs. 0.99 sec). The main effect of orienting task and its interaction with word frequency were not significant.

False-alarm rates and response times for correct rejections were subjected to a one-way analysis of variance (ANOVA) with word frequency as a within-subjects factor. The mean false-alarm rate for the low-frequency words was significantly lower than that for the high-frequency words  $[F(1,11) = 148.5, MS_e = .002]$ . Correct rejections for the low-frequency words were significantly faster than those for the high-frequency words  $[F(1,11) = 16.5, MS_e = 28.5]$ .

Hit and false-alarm rates for each subject were transformed to values of d' and C. Note that because the same false-alarm rates were used for calculating the values for both the acoustic and the semantic orienting tasks, statistical tests for the main effect of orienting task depended mainly on hit rates. For d' values, the main effects of word frequency and orienting task were significant [F(1,11) =48.0,  $MS_e = 0.10$ , and F(1,11) = 63.1,  $MS_e = 0.18$ , respectively]. The mean values of d' were 1.71 for the low-frequency words and 1.09 for the high-frequency words. The mean values of d' for the acoustic and the semantic orienting tasks were 0.97 and 1.84, respectively. For C values, there were significant main effects of word frequency and orienting task  $[F(1,11) = 47.7, MS_e = .07,$ and F(1,11) = 65.4,  $MS_c = .04$ , respectively]. The mean values of C for the low- and the high-frequency words were 0.35 and -0.17, respectively. The mean values of C for the acoustic and the semantic orienting tasks were 0.31 and -0.13, respectively. For d' and C, there were no significant interactions between word frequency and orienting task. The mean values of C indicated the inclination of the subjects to make the positive response more frequently to the high-frequency words than to the lowfrequency words.

There were no significant interactions between orienting task and word frequency for all dependent measures. The abstractness-concreteness judgment in the semantic orienting task elicited equally extensive processing for the low- and high-frequency words.

In recognition-memory tests, responses to low-frequency words have been faster than responses to high-frequency words (Duchek & Neely, 1989; Mandler et al., 1982). However, faster hits for the high-frequency words and faster correct rejections for the low-frequency words were seen in Experiment 1. Neely, Schmidt, and Roediger (1983) found that response time in a speeded yes-no recognition-memory test is faster when the preceding item is from the same category as the test item. This finding would explain the faster hits for the high-frequency words if the short interstimulus interval at the test phase emphasizes semantic relations between high-frequency words. However, the faster correct rejections for the low-frequency words could not be explained in this manner.

The faster hits for the high-frequency words and the faster correct rejections for the low-frequency words implied that recognition judgment partly relied on the levels of word frequency. The dependency of recognition judgment on the levels of word frequency might be the cause of the liberal response bias for the high-frequency words and the violation of the mirror effect.

## **EXPERIMENT 2**

In Experiment 1, the study and the test lists were presented at fast rates of 1 item/sec. It is possible that word frequency affected recognition judgment only because a speeded response had to be made. In most previous studies, test items either were presented with relatively long intervals (Duchek & Neely, 1989) or were unpaced (Earhard, 1982). To examine whether or not relatively short interstimulus intervals caused the violation of the mirror effect and the different response biases, fast-and slow-test conditions were compared in Experiment 2.

In the fast-test condition, time for recognition judgment was limited, and each item was presented at a relatively fast rate. In the slow-test condition, time for decision was unlimited, and the subjects could respond to each item at their own pace. In both conditions, low- and high-frequency words were presented. These words had the same levels of word frequency as they had in Experiment 1. The question of interest was whether the effect of word frequency on response bias would be diminished in the slow-test condition.

#### Method

Design. Test type (fast or slow) and word frequency (low or high) were within-subjects variables. In both test conditions, low- and high-frequency words appeared. After one type of recognition-memory test was performed, the subjects were asked to learn other low- and high-frequency words and were given the other type of recognition-memory test. The order of the two types of recognition-memory tests was a between-subjects variable.

Subjects, Materials, and Apparatus. Twenty-four undergraduate and graduate students at Tokyo Metropolitan University participated in the experiment. They were tested individually.

One hundred sixty nouns composed of two Kanji characters were collected. Half of them were low-frequency words with a frequency of 16 occurrences per million words; the other half were high-frequency words with a mean frequency of 323 occurrences per million words (The National Language Research Institute, 1962). Four lists of 20 words each were constructed for each frequency level, with each list for use as to-be-remembered words and distractors. Two lists of 10 medium-frequency words that were nouns composed of two Kanji characters were constructed for use as fillers in the beginning of the learning phases.

A Fujitsu FM-7 microcomputer controlled all aspects of the experiment. A green CRT display was used. The 1-msec counter and the controller of the CRT display and keyboard communication were programmed in the 6809 machine language.

**Procedure.** On arriving at the laboratory, the subjects were asked to sit down in front of the display of the microcomputer. They were told that the words were to be presented on the display one at a time and that they would be asked later whether or not they remembered seeing the words. Each word was shown for 800 msec, following a fixation dot that was shown in the proximate center of

the display. The interval between the disappearance of the word and the appearance of the next word was 3.2 sec. Twenty low- and 20 high-frequency words were shown after 10 medium-frequency words. The medium-frequency words were not shown in the following recognition-memory test.

A letter-matching task was interpolated between the learning and the test phases. The subjects were told that a pair consisting of a lowercase letter and an uppercase letter of the English alphabet would appear on the display. They judged whether or not the name of one letter of the pair and the name of the other letter were the same. Responses were made by pressing one of two keys on the keyboard with either the index or the middle finger of the dominant hand. Each pair was shown for 80 msec, followed by a masking stimulus. Forty-eight pairs were presented to each subject. It took about 5 min to complete this task.

In the recognition-memory test, the subjects judged whether each word had been presented in the more recent study phase. Responses were made by pressing one of the two keys on the keyboard with either the index or the middle finger of the dominant hand. The index finger was used to denote that the presented item was old; the middle finger was used to denote that the presented item was new. In the fast test, a fixation dot was shown for 500 msec prior to the presentation of each word. The word disappeared either when the subjects made a response or when 1.2 sec had passed after the presentation of it. When 1.8 sec had elapsed following the presentation of the word, the subjects heard a brief buzzer tone. No response was accepted after the buzzer tone came on. After the passing of 1.8 sec following a response, the next word appeared. The limitation of time for the fast judgment was not set so severely that the overall level of performance was lowered drastically. In the slow test, the time for making a recognition judgment was not limited. The fixation dot was shown for 700 msec prior to the presentation of each word. Each word was displayed until the subjects made a response. A confidence rating about each judgment was then made by pressing one of the two keys that were designated as high and low confidence. The subjects were able to correct their judgment and confidence rating before the carriage-return key was pressed. The next stimulus did not appear until 5 sec had passed following the recognition judgment.

Half of the subjects first performed the fast test; the other half first performed the slow test. Before the subjects saw the words for the second recognition-memory test, they performed a detection task. The subjects judged whether or not a luminous point was displayed. The luminous point was shown for 14 msec in one of four positions on the display, followed by a mask stimuli. There were 25 trials, and it took about 4 min for a subject to complete this task.

Four combinations of the eight lists were made so that words used as to-be-remembered words in either the fast or the slow test were used as distractors in the other test. One of the four list combinations was assigned to each subject. Each subject received a unique random order of items in the study and the test phases.

### **Results and Discussion**

Hit rates, false-alarm rates, response times for hits and correct rejections, and values of d' and C in the recognition-memory test were analyzed. Confidence ratings in the slow-test condition were not analyzed. Table 2 shows the mean values of these measures for the low- and the high-frequency words in the slow and the fast tests.

The mean response times for hits and correct rejections in the fast test were about 3 times as fast as in the slow test  $[F(1,22) = 240.5, MS_e = 0.20, \text{ and } F(1,22) = 280.2, MS_e = 0.22, \text{ respectively}]$ . Word frequency had a significant effect on response times for correct rejections and

Table 2	
Recognition Memory for Low- and High-Frequency Wor	ds
Under Fast- and Slow-Test Conditions in Experiment 2	

Word Frequency	Fast Test		Slow Test	
	Low	High	Low	High
Hit rate	.69	.74	.74	.76
False-alarm rate	.15	.26	.15	.24
d'	1.67	1.43	1.93	1.53
$\boldsymbol{c}$	0.31	0.03	0.25	0.01
RT for hit	0.788	0.783	2.127	2.284
RT for CR	0.812	0.845	2.284	2.565

Note—RT = response time; CR = correct rejection. Response times are given in seconds.

interacted with test type  $[F(1,22) = 8.7, MS_e = 0.07,$  and  $F(1,22) = 5.9, MS_e = 0.06,$  respectively]. The correct rejections for the low-frequency words were faster than those for the high-frequency words, and this difference in response time was reduced in the fast condition.

For hit rates, the main effect of test order was significant  $[F(1,22) = 5.9, MS_e = .031]$ . The mean hit rate for the subjects who first performed the slow test was .79, whereas for the subjects who first performed the fast test, it was .70. The main effect of test type also was significant  $[F(1,22) = 4.9, MS_e = .008]$ . The mean hit rates for the slow and the fast tests were .75 and .71, respectively. The main effect of word frequency was not significant, and there were no significant interactions.

For false-alarm rates, word frequency showed a significant effect  $[F(1,22) = 29.6, MS_e = .009]$ . The mean false-alarm rates were .25 for the high-frequency words and .15 for the low-frequency words. The main effects of test type and test order did not approach significance. There were no significant interactions.

For d' and C, the main effect of word frequency was significant  $[F(1,22) = 15.6, MS_e = 0.16, \text{ and } F(1,22) = 19.7, <math>MS_e = .08$ , respectively]. The mean values of d' for the low- and the high-frequency words were 1.80 and 1.48, respectively. The mean C values for the low- and the high-frequency words were 0.28 and 0.02, respectively. For both measures, the effect of test type did not approach significance, and there were no significant interactions.

The liberal response bias for the high-frequency words was not diminished in the slow-test condition. The violation of the mirror effect and the different response biases cannot be ascribed to the fast judgments that rely on the levels of word frequency. One possible reason for the relatively liberal response bias for the high-frequency words is the different subjective probabilities of appearance of low- and high-frequency old words. The subjective probability of appearance of old items affects the location of a cutoff (Parks, 1966). However, why low- and high-frequency words would differ in subjective probability is unclear.

Alternatively, the use of a criterion common to the lowand the high-frequency words would produce the relatively liberal response bias for the high-frequency words. Familiarity could be considered as a decision axis for recognition judgment (Lockhart & Murdock, 1970). If the familiarity of low-frequency old words is not different from that of high-frequency old words, the use of a common criterion would produce the different response biases and the violation of the mirror effect.

The patterns of distributions of low- and high-frequency words that produce the different response biases and the violation of the mirror effect are depicted in Panels 2 and 3 of Figure 1, which correspond to Panel 2 of Figure 1 in Glanzer and Adams (1990). Distributions depicted in Panels 2 and 3 indicate equal familiarity of low- and high-frequency old words and higher familiarity of high-frequency new words with respect to low-frequency new words. The use of a common criterion placed at any point on the familiarity axis would give rise to the relatively liberal response biases for high-frequency words and the violation of the mirror effect.

On the other hand, Panels 1 and 3 of Figure 1 show the mirror effect. The distribution of high-frequency old words is placed lower than that of low-frequency old words, and the distribution of high-frequency new words is placed higher than that of low-frequency new words. If the subjects used a common criterion for low- and high-frequency words, the distribution patterns would produce the mirror effect.

Thus, the violation of the mirror effect could not be ascribed solely to the relatively liberal placement of a criterion for high-frequency words. Rather, it is likely the

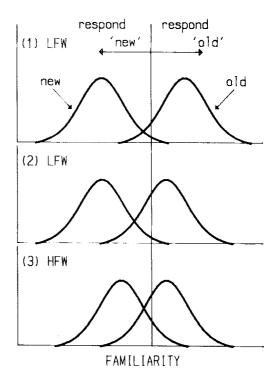


Figure 1. The order of new and old distributions: LFW = low-frequency words; HFW = high-frequency words. Panels 1 and 3 show the mirror effect. Panels 2 and 3 show the violation of the mirror effect.

use of a common criterion and the particular order of familiarity distributions that cause the different response biases for low- and high-frequency words and the violation of the mirror effect. Gillund and Shiffrin (1984) assumed that a criterion appropriate to each level of word frequency is selected by subjects. However, the violation of the mirror effect and the different response biases suggest that when low- and high-frequency words are intermixed, subjects do not discriminate the two levels of word frequency to choose a criterion appropriate to each level of word frequency.

The effect of word frequency on response bias would disappear if the placement of a criterion was set according to the levels of linguistic frequency. An experiment by Krueger (1985) indicated that in a letter-matching task, appropriate criteria can be adopted when stimuli of different types are presented in distinct blocks of trials. Thus, if the word-frequency effect on response bias is due to the use of a single criterion, the effect should be diminished when low- and high-frequency words are tested in distinct blocks.

## **EXPERIMENT 3**

Experiment 3 was conducted to examine the hypothesis that the use of a common criterion accounts, at least in part, for the difference in bias between low- and high-frequency words. Separate presentations of low- and high-frequency words should reduce the effect of word frequency on response bias by enabling adoption of criteria that are specific to the particular word set. Words having one level of word frequency were presented during each half of a recognition-memory test.

Previous studies have found that presenting pure lists of low- or high-frequency words did not diminish the better performance of low-frequency words (Gregg, 1976; Underwood & Freund, 1970). In their experiments, however, the pure list was shown during the study phase, as well as during the test phase. In Experiment 3, the pure-frequency list was used only at the test phase, to examine criterion effects at retrieval.

# Method

**Design**. The design included test condition (pure or mixed) as a between-subjects variable and word frequency (low or high) as a within-subjects variable. In the pure-test condition, words with the same level of frequency were presented continuously during a test phase. In the mixed-test condition, low- and high-frequency words were intermixed. In both conditions, low- and high-frequency words were presented randomly during the study phase.

Subjects, Materials, and Apparatus. Eighty graduate and undergraduate students at Tokyo Metropolitan University participated in the experiment. They were tested individually. Half of the subjects were allocated to the mixed-test condition; the other subjects were allocated to the pure-test condition. The data from 1 subject who was allocated to the mixed-test condition were excluded from the following analyses, because the d' value for high-frequency words in the second half of the recognition-memory test was lower than 0.

The 80 low- and 80 high-frequency words used in this experiment were the same ones that were used in Experiment 2. There

were eight lists consisting of 20 words each. Four lists comprised low-frequency words; the other four lists comprised high-frequency words. Six medium-frequency words were used as fillers in the study phase.

An NEC PC-9801 microcomputer and a CONTEC TIR-6(98) counter module controlled all aspects of the experiment.

Procedure. Each subject sat in front of a display controlled by the microcomputer. The subject was told that words would be presented one at a time on the display and that he or she would be asked later if he or she remembered seeing those words. Six medium-frequency words, which did not appear in the following recognition-memory test, were shown prior to the presentation of 20 low- and 20 high-frequency words. Each word appeared for 1.2 sec, and the interstimulus interval was 1.3 sec.

A letter-matching task was performed between the study phase and the recognition-memory test. The subjects judged whether or not a presented pair of the English alphabet letters had the same name. Fifty-two pairs were presented. Each pair consisted of one lowercase letter and one capital letter. It appeared for 80 msec and was followed by a masking stimulus. Responses were made by pressing one of two arbitrary keys on the keyboard. It took about 5 min to complete this task.

In the recognition-memory test, the subjects judged whether or not they had seen the word in the study phase. Before each item was presented on the display, an underline was shown for 1 sec. The subjects made a response by pressing one of the two keys on the keyboard; then they pressed the carriage-return key to proceed to the next word. The subjects were told that they could correct the response before they pressed the return key and that they need not make a response immediately. Feedback with regard to correctness of a response was shown on the display. On the completion of 40 responses, the subjects took a rest for a minimum of 15 sec. By pressing the return key, they proceeded to the remaining 40 items.

Half of the subjects allocated to the pure-test condition saw the low-frequency words in the first half of the test and the high-frequency words in the second half of the test. The other subjects of the pure-test condition were given the high-frequency words in the first half of the test and the low-frequency words in the second half of the test. The subjects in the pure-test condition were informed about the level of word frequency that appeared in each half of the test. In the mixed-test condition, all of the subjects encountered 20 low-frequency words and 20 high-frequency words during each half. In both conditions, 20 to-be-remembered words and 20 distractors were shown during each half.

Four combinations of two lists of the low-frequency words and two lists of the high-frequency words were constructed out of the eight lists. Each of the eight lists was used as a list of to-be-remembered words in one of the four combinations and as a list of distractors in another combination. One of the four combinations was assigned to each subject.

# **Results and Discussion**

Hit rates, false-alarm rates, and values of d' and C in the recognition-memory test were analyzed. The data from the subjects in the pure-test condition were subjected to an ANOVA to determine the effects of the order of the pure tests. For all dependent measures, the order of the pure tests did not show significant main effects, and its interaction with word frequency did not approach significance. In the following analyses, the order of the pure tests was ignored. Table 3 shows the means of hit rates, false-alarm rates, and values of d' and C for the low- and the high-frequency words in the first and the second halves of the mixed- and the pure-test conditions.

Table 3

Recognition Memory for Low- and High-Frequency Words of the First and Second Halves in the Mixed- and Pure-Test Conditions of Experiment 3

	First Half		Second Half	
Word Frequency	Low	High	Low	High
	Mixed-Te	st Condition	1	
Hit rate	.76	.79	.74	.73
False-alarm rate	.09	.28	.12	.26
d' .	2.48	1.65	2.12	1.53
C	0.38	-0.12	0.31	0.03
	Pure-Tes	t Condition		
Hit rate	.74	.66	.68	.71
False-alarm rate	.14	.22	.13	.27
d'	1.92	1.34	1.79	1.30
$\boldsymbol{C}$	0.24	0.21	0.37	0.07

For hit rates, the main effect of word frequency did not approach significance, but for false-alarm rates, it was significant  $[F(1,77) = 85.7, MS_e = .008]$ . The mean false-alarm rate was .12 for the low-frequency words and .25 for the high-frequency words. For hit and false-alarm rates, the main effect of test condition and its interaction with word frequency did not approach significance.

For d' and C, test condition did not show significant effects, but the main effect of word frequency was significant  $[F(1,77) = 73.5, MS_e = 0.20; F(1,77) = 39.0, MS_e = .07]$ . The mean values of d' for the low- and the high-frequency words were 2.00 and 1.40, respectively. The mean value of C was 0.32 for the low-frequency words and 0.05 for the high-frequency words. For d' values, the interaction between word frequency and test condition was not significant, but for C values, it was significant  $[F(1,77) = 6.1, MS_e = .07]$ . This interaction indicates that the difference in response bias was reduced in the pure-test condition. Its marginal strength is due to the reduction being limited primarily to the conditions tested first, as shown by the following analyses.

Data from the pure-test condition were analyzed to compare the performance in the first half of the test with the performance in the second half of the test. In the first half, the differences between the means of hit rates and C values for the low- and the high-frequency words did not approach significance, whereas word frequency showed significant effects on false-alarm rates and d'values [t(38) = 2.1, t(38) = 2.6]. In the second half, word frequency had significant effects on false-alarm rates, d'values, and C values [t(38) = 3.1, t(38) = 2.3, and t(38)]= 2.6, respectively], but the word-frequency effect on hit rates did not approach significance. Thus, no difference in response bias between the low- and the highfrequency words was found for the first test, although the more liberal bias for the high-frequency words was present for the second tests.

Data in the first and second halves of the mixed-test condition were subjected to a similar ANOVA. Word frequency had significant effects on false-alarm rates, d' values, and C values  $[F(1,38) = 73.1, MS_c = 0.01;$ 

F(1,38) = 52.9,  $MS_e = 0.38$ ; and F(1,38) = 34.1,  $MS_e = 0.17$ , respectively]. For hit rates, the main effect of word frequency was not significant. There was a significant difference between the means of d' values for the first and the second halves of the test  $[F(1,39) = 4.5, MS_e = 0.50]$ . The mean d' value for the first half was 2.06, whereas for the second half, it was 1.83. For C values, the interaction between word frequency and portion of test (first half or second half) was significant  $[F(1,39) = 4.7, MS_e = 0.10]$ . This interaction indicated that the difference between the mean C values for the low- and the high-frequency words was slightly reduced in the second half of the test.

The use of a common criterion is suggested by the results of the pure-test condition. The criterion set in the first half of the pure-test condition is not adjusted to another level of word frequency in the second half. In the usual blocked experimental design, a set of experimental parameters is kept constant for 100 or more trials (Sperling & Dosher, 1986). Because the number of items presented in the second half of the pure-test condition was 40, the subjects apparently did not have enough trials to shift the placement of a criterion according to the level of word frequency.

## **EXPERIMENT 4**

According to the model shown in Figure 1, it should be possible to obtain the mirror effect by increasing discriminability between low-frequency old and new words. Rao and Proctor (1984) revealed that recognition memory for low-frequency words that is more accurate than recognition memory for high-frequency words is enhanced after a lexical-decision task, compared with performance after general memory instructions and after a syllablecounting task. Consequently, the subjects were required to perform a lexical-decision task in the study phase of Experiment 4. In addition, blocked- and mixed-test conditions were examined, as in Experiment 3. If the lexicaldecision task induces a situation such as that in Panels 1 and 3 of Figure 1, the blocking manipulation should make little difference. This is because the single criterion that is used when word frequencies are mixed would be equally appropriate for both low- and high-frequency words.

#### Method

Design. In a study phase, a lexical-decision task was performed. There were two within-subjects variables: word frequency (low or high) and test condition (pure or mixed). Under the pure-test condition, low-frequency words were presented in the first half of the recognition-memory test, and high-frequency words were presented in the second half. The order of the pure- and the mixed-test conditions was a between-subjects variable.

Subjects, Materials, and Apparatus. Twenty graduate and undergraduate students at Tokyo Metropolitan University participated individually in this experiment.

The 80 low- and the 80 high-frequency words used in this experiment were the same ones that were used in Experiments 2 and 3. There were four lists consisting of 20 low-frequency words each

and four lists consisting of 20 high-frequency words each. Eighty medium-frequency words composed of two Chinese characters were collected to construct four nonword lists for use in the lexical-decision task. Each Chinese character in these words was paired with a character of another word to make nonsense strings. By using 10 words from the 80 medium-frequency words, another 10 nonsense strings were made. These nonsense strings and another 10 medium-frequency words were used as fillers in the lexical-decision task.

An NEC PC-9801 microcomputer and a CONTEC TIR-6(98) counter module controlled all aspects of the experiment.

Procedure. The subject was asked to sit in front of a display controlled by the microcomputer and to judge whether or not the presented string was a word. The subject was told about the later recognition-memory test and was required to respond as fast as possible. Before each string was shown, an underline was displayed for 1 sec. The string was displayed until a response was made. Responses were made by pressing one of the two keys on the keyboard. When an incorrect response was made, the subject heard a short beep. The 5 medium-frequency words and the 5 nonwords were always presented prior to the 20 low- and 20 high-frequency words and the 40 nonwords. The medium-frequency words were not presented in the subsequent recognition-memory test. After 50 responses were made, the subject took a rest for 1 min.

On completion of the lexical-decision task, the subjects were asked to solve mathematical and logical problems for about 15 min to avoid ceiling effects in recognition memory. In the subsequent recognition-memory test, the subjects judged whether or not they had seen the item in the lexical-decision task that was performed more recently. Each item appeared on the display, preceded by an underline displayed for 800 msec. Feedback about correctness of a response was not shown. The subjects were not told about the test conditions. The other procedures for the recognition-memory test were the same ones that were used in Experiment 3.

Half of the subjects first performed the recognition-memory test under the pure-test condition; the other subjects first performed the test under the mixed-test condition. They took a rest for about 3 min before they proceeded to the second study-test trial. Under the pure-test condition, 40 low-frequency words were presented in the first half of the recognition-memory test, and 40 high-frequency words were presented in the second half. In the mixed-test condition, 20 low-frequency words and 20 high-frequency words were shown during each half of the test. In both conditions, 20 to-be-remembered words and 20 distractors were shown during each half.

Each subject received a unique assignment of the four lists of low-frequency words and the four lists of high-frequency words. The words the subjects saw in the first study-test trial were not seen in the second study-test trial. The order of the items presented in the lexical-decision tasks and in the recognition-memory tests was randomized for each subject.

## Results and Discussion

In the lexical-decision task, 95% of the low-frequency words and 99% of the high-frequency words were identified as words. Response times in the lexical-decision task were subjected to an ANOVA. Incorrect responses were eliminated from the analysis. The mean response time for the low-frequency words was 0.717 sec, whereas for the high-frequency words, it was 0.604 sec. This difference was significant  $[F(1,18) = 104.6, MS_e = 0.002]$ . The interaction of word frequency, test condition, and order of test condition also was significant  $[F(1,9) = 11.1, MS_e = 0.001]$ . The difference in response time between the low- and high-frequency words was enlarged when the

Table 4
Recognition Memory for Low- and High-Frequency Words Under
Conditions of Pure- and Mixed-Test Conditions in Experiment 4

Word Frequency	Mixed Test		Pure Test	
	Low	High	Low	High
Hit rate	.76	.70	.80	.67
False-alarm rate	.08	.13	.09	.15
ď'	2.38	1.85	2.50	1.76
$\boldsymbol{C}$	0.39	0.34	0.34	0.38

subjects who first encountered the mixed-test condition performed the second lexical-decision task.

Hit and false-alarm rates and values of d' and C for each subject were subjected to an ANOVA. Table 4 shows the means of these measures for the low- and high-frequency words under the mixed- and the pure-test conditions.

For hit and false-alarm rates, the main effect of word frequency was significant  $[F(1,18) = 16.5, MS_e = .01,$  and  $F(1,18) = 11.6, MS_e = .006$ , respectively]. The mean hit rate was .78 for the low-frequency words and .68 for the high-frequency words. The mean false-alarm rate for the low-frequency words was .08, whereas for the high-frequency words, it was .14. Word frequency had a significant effect on d' values  $[F(1,18) = 29.6, MS_e = 0.27]$ . The mean d' value for the low-frequency words was 2.44, whereas for the high-frequency words, it was 1.80. For C values, the main effect of word frequency did not approach significance. For all dependent measures, the effect of test condition and its interactions with other independent variables was not significant.

After a lexical-decision task, the different response biases disappeared, and the mirror effect was shown. The higher hit rate for the low-frequency words with respect to the high-frequency words is due to an increment of the discriminability between the low-frequency old and new words. The discriminability could be enhanced by processing items extensively at a study phase. In a lexical-decision task, low-frequency words, which are relatively difficult to identify as words, are processed more extensively than are high-frequency words (Rao & Proctor, 1984).

## **GENERAL DISCUSSION**

Both a more liberal response bias for high-frequency words with respect to low-frequency words and a violation of the mirror effect were found in Experiments 1, 2, and 3. Although the false-alarm rate for low-frequency words was lower than that for high-frequency words, the hit rate for low-frequency words was not higher than that for high-frequency words. The causes of the violation and the different response biases were investigated.

The absence of the word-frequency effect on hit rates is not ascribed to the relatively liberal placement of a criterion for high-frequency words, because there was no incentive to respond "old" more frequently to high-frequency words. If subjects need to select a criterion or likelihood ratio appropriate to each level of word frequency prior to recognition judgment (Gillund & Shiffrin, 1984; Glanzer & Adams, 1990), the failure in the discrimination between the two levels of word frequency would cause the different response biases and the violation of the mirror effect. In the present study, the difference between the two levels of word frequency was relatively small. However, the results from Experiment 4 suggest that the mirror effect can occur when the difference between the two levels of word frequency is relatively small.

The appearance of the mirror effect in Experiment 4 is due to a lexical-decision orienting task, which increases the magnitude of the word-frequency effect on accuracy in recognition memory (Rao & Proctor, 1984). Because the low- and the high-frequency words that were used in Experiment 4 were the same as those used in Experiments 2 and 3, it is unlikely that the subjects discriminated between the two levels of word frequency to select a criterion appropriate to each level. If subjects use a common criterion for low- and high-frequency words, the results of Experiment 4 will occur when the distributions of familiarity for low-frequency words are relatively more discriminable than those for high-frequency words (see Panels 1 and 3 of Figure 1).

According to the model shown in Figure 1, the blocking manipulation at retrieval has effects on response bias when the familiarity of low-frequency old words is similar to that of high-frequency old words. In Experiment 3, the manipulation diminished the relatively liberal response bias for high-frequency words. The absence of the blocking effect in Experiment 4 suggests an increment of the discriminability between low-frequency old and new words. The single criterion that was used in the mixed-test condition is as appropriate for low- and high-frequency words as the criteria that were used in the pure-test condition, in which the discriminability of low-frequency words is enhanced.

The results from Experiment 3 also suggest the use of a common criterion. When words with one level of word frequency were presented during each half of the recognition-memory test, the effect of word frequency on response bias was not apparent in the first half of the test but was apparent in the second half. Although the subjects were informed of the word-frequency level presented in each half, the criterion set in the first half was not adjusted to another level of word frequency in the second half. Gillund and Shiffrin (1984) assumed that subjects choose a criterion relevant to each level of word frequency. However, the different response biases for low-and high-frequency words indicate that once a criterion is set, its placement is less likely to be shifted in accordance with the change of the level of word frequency.

The findings of the present study indicate that subjects use a common criterion. The violation of the mirror effect and the different response biases are due to both the

similar familiarity of low- and high-frequency old words and the use of a common criterion. If the degree of discrimination of low-frequency old and new words is sufficiently high, the use of a common criterion will produce the mirror effect.

Models for recognition memory or recognition judgment should be able to predict the violation of the mirror effect: Although the word-frequency effect on false-alarm rates is shown, the effect on hit rates disappears. The models of Gillund and Shiffrin (1984) and Glanzer and Adams (1990) assume that subjects distinguish low- and high-frequency words prior to recognition judgment. In those models, the failure of the discrimination between low- and high-frequency words causes the violation. However, the following model for recognition judgment does not presume that the item is classified into two levels of word frequency, because the use of a common criterion for low- and high-frequency words is assumed.

Recognition judgment is based on the familiarity of the item. A false alarm is due to the confusing familiarity caused by the traces that do not include the context in which the item was seen or heard. Previous narrow or routine encodings of the items do not yield the traces that include the context. The higher false-alarm rate for high-frequency words implies that the narrow encodings of high-frequency words are more frequent than those of low-frequency words.

The higher hit rate for low-frequency words with respect to high-frequency words indicates that the familiarity of low-frequency old words is higher than the familiarity of high-frequency old words. Processing lowfrequency words more extensively at the study phase increases the familiarity of those words, because extensive encodings produce traces that include the context of the study phase. Another possible reason for the lower familiarity of high-frequency old words is that the confusing familiarity inhibits an increment in familiarity. The familiarity that is caused by the traces that include the context of the study phase is reduced by the confusing familiarity. The familiarity of low-frequency old words that is caused by the traces that include the relevant context is little inhibited by the confusing familiarity when the level of word frequency is sufficiently low.

The accuracy of recognition memory is determined by the degree to which the item is processed during a study phase and the level of the confusing familiarity. Extended processing of low-frequency words at a study phase and the lower level of the confusing familiarity produce the higher hit and lower false-alarm rates. Processing low-frequency words less extensively at a study phase gives rise to the violation of the mirror effect.

The application of the above model for recognition judgment is not restricted to the word-frequency variable. Glanzer and Adams (1985) revealed that the mirror effect holds for other variables such as concreteness, meaningfulness, pictures versus words, and so forth. However, the violation of the mirror effect appeared in Experiment 1 of Peterson and McGee (1974). They ex-

amined the effects of imagery ratings of words on recognition memory and used mixed lists of low- and high-imagery words. The mean number of false alarms for low-imagery words is about two times larger than that for high-imagery words, whereas the mean number of hits for low-imagery words is very close to that for high-imagery words. These findings suggest the relatively liberal response bias for low-imagery words. Further investigations into the mirror effect and its violation would cast new light on the processes underlying recognition judgment and recognition memory.

### REFERENCES

- BALOTA, D. A., & NEELY, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning & Memory*, 6, 576-587.
- BANKS, W. P. (1970). Signal detection theory and human memory. Psychological Bulletin, 74, 81-99.
- DUCHEK, J. M., & NEELY, J. H. (1989). A dissociative word-frequency × levels-of-processing interaction in episodic recognition and lexical decision tasks. *Memory & Cognition*, 17, 148-162.
- EARHARD, B. (1982). Determinants of the word-frequency effect in recognition memory. *Memory & Cognition*, 10, 115-124.
- EYSENCK, M. W. (1979). Depth, elaboration, and distinctiveness. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing in human memory (pp. 89-118). Hillsdale, NJ: Erlbaum.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1-67.
- GLANZER, M., & ADAMS, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, 13, 8-20.
- GLANZER, M., & ADAMS, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology:* Learning, Memory, & Cognition, 16, 5-16.
- GLANZER, M., & BOWLES, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning & Memory*, 2, 21-31.
- GREEN, D. M., & SWETS, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley.
- GREGG, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), Recall and recognition (pp. 184-216). London: Wiley.
- KRUEGER, L. E. (1985). Effect of intermixed foveal and parafoveal presentation on same-different judgments: Evidence for a criterioninertia model. Perception & Psychophysics, 37, 266-271.
- LOCKHART, R. S., & MURDOCK, B. B., JR. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, 74, 100-109.
- Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition*, 10, 33-42.
- McCormack, P. D., & Swenson, A. L. (1972). Recognition memory

- for common and rare words. Journal of Experimental Psychology, 95, 72-77.
- McNicol, D. (1972). A primer of signal detection theory. London: George Allen & Unwin.
- McNicol, D., & Ryder, L. A. (1971). Sensitivity and response bias effects in the learning of familiar and unfamiliar associations by rote or with a mnemonic. *Journal of Experimental Psychology*, 90, 81-89.
- MURDOCK, B. B., JR. (1965). Signal-detection theory and short-term memory. *Journal of Experimental Psychology*, 70, 443-447.
- THE NATIONAL LANGUAGE RESEARCH INSTITUTE. (1962). Vocabulary and Chinese characters in ninety magazines of today: Vol. 1. General description and vocabulary frequency tables (Rep. No. 21). Tokyo: Author. (In Japanese)
- NEELY, J. H., SCHMIDT, S. R., & ROEDIGER, H. L., III. (1983). Inhibition from related primes and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9, 196-211.
- OGAWA, T., & INAMURA, Y. (1974). An analysis of word attributes: Imagery, concreteness, meaningfulness, and ease of learning for Japanese nouns. *Japanese Journal of Psychology*, **44**, 317-327. (In Japanese with English summary)
- PARKS, T. E. (1966). Signal-detectability theory of recognition-memory performance. *Psychological Review*, 73, 44-58.
- PETERSON, M. J., & McGEE, M. J. (1974). Effects of imagery instructions, imagery rating, and number of dictionary meanings upon recognition and recall. *Journal of Experimental Psychology*, **102**, 1007-1014.
- RAO, K. V., & PROCTOR, R. W. (1984). Study-phase processing and the word-frequency effect in recognition memory. *Journal of Ex*perimental Psychology: Learning, Memory, & Cognition, 10, 386-394.
- Schulman, A. I., & Lovelace, E. A. (1970). Recognition memory for words presented at a slow or rapid rate. *Psychonomic Science*, 21, 99-100.
- SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.
- Sperling, G., & Dosher, B. A. (1986). Strategy and optimization in human information processing. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of perception and human performance (Vol. 1). New York: Wiley.
- UNDERWOOD, B. J., & FREUND, J. S. (1970). Word frequency and short-term recognition memory. American Journal of Psychology, 83, 343-351.
- von Winterfeldt, D., & Edwards, W. (1982). Costs and payoffs in perceptual research. Psychological Bulletin, 91, 609-622.

# NOTE

1. Data from the pure-test condition could not be subjected to an ANOVA with word frequency and portion of test as independent variables, because portion of test and word frequency correlated to each other in that condition.

(Manuscript received May 1988; revision accepted for publication April 30, 1991.)