

Memorability judgments for high- and low-frequency words

ROBERT GUTTENTAG and DONNA CARROLL
University of North Carolina, Greensboro, North Carolina

Five experiments were conducted in order to examine subjects' judgments of the memorability of high- (HF) and low-frequency (LF) words in the context of a recognition memory task. In Experiment 1, the subjects were provided study/test experience with a list of HF and LF words prior to making memorability judgments for a new list of HF and LF items. The findings were consistent with previous evidence (Greene & Thapar, 1994; Wixted, 1992) suggesting that subjects are not explicitly aware of the greater recognition memorability of LF words. Experiments 2-5 embedded the memorability judgment task within the recognition test itself. In these experiments, the subjects consistently gave higher memorability ratings to LF items. The contrast between the pattern of results found when the subjects made their judgments at the time of list presentation (Experiment 1) and that when they made their judgments during the recognition test (Experiments 2-5) is consistent with recent evidence that even seemingly highly related metamnemonic judgments (e.g., ease of learning judgments vs. judgments of learning for the same items) may be based on very different factors if they occur at different points in the study/test cycle. The present findings are also consistent with the possibility that very rapid retrieval of memorability information for HF and LF words may affect recognition decisions and may contribute to the recognition memory word frequency effect.

The word frequency effect (WFE) in recognition memory refers to the common finding that memory performance tends to be more accurate, both in terms of hit rates and in terms of rates of false alarms, when subjects are tested with low-frequency (LF), rather than with high-frequency (HF) words (Gregg, 1976). The fact that the LF word advantage involves a higher response rate to LF than to HF targets but a lower response rate to LF than to HF distractors has been termed the *mirror effect* (Glanzer & Adams, 1985, 1990), and this phenomenon poses a difficult explanatory challenge for simple trace-strength or familiarity-based theories of recognition. More specifically, the mirror effect challenges such models to explain the finding that HF words seem, at the same time, to be both more familiar than LF words (in the case of distractors) and less familiar than LF words (in the case of targets; Glanzer, Adams, Iverson, & Kim, 1993; Hintzman, 1994). Accordingly, a number of modifications to simple trace-strength theories have been proposed, including Brown's (1976; Brown, Lewis, & Monk, 1977; Brown & Routh, 1970) proposal that word-frequency-related differences in subjective memorability contribute to the LF advantage for false alarms.

The memorability hypothesis proposes that subjects presented with a recognition task adjust their criterion for making a recognition judgment according to their judgment of the memorability of each item on the recognition

test. In order to explain the WFE, this view assumes that subjects correctly judge that LF target words are more memorable on a recognition test than are HF target words; consequently, subjects set a higher criterion for acceptance of LF words, thereby reducing the false alarm rate for LF distractors. In support of this theory, Brown et al. (1977) found that subjects were very unlikely to make false alarms to highly distinctive distractor words, such as their own names. Brown et al. argued that this effect reflected the subjects' awareness that, if their own name had actually been on the study list, they surely would have recollected it when performing the recognition test; accordingly, in the absence of clear recollection of the presentation of their names on the study list, subjects were able accurately to reject their names when they were presented on the recognition test as distractors.

Despite the apparent applicability of this hypothesis for explaining the distinctiveness effect described by Brown, the findings from a number of recent studies that attempted to evaluate the hypothesis have raised questions regarding the theory's relevance to most recognition contexts. Hintzman, Caulton, and Curran (1994), for example, found that the mirror effect was not attenuated by imposing time or capacity constraints on subjects at retrieval. Hintzman et al. interpreted this finding as precluding the memorability hypothesis, on the basis of the assumption that memorability judgments must require relatively slow and conscious reflection.

The findings from the Hintzman et al. (1994) study indicate that, if memorability judgments do affect patterns of recognition responding, the judgments must be made very rapidly and without placing large demands on the subject's limited attentional resources. Thus, although the

Correspondence concerning this article should be addressed to R. Guttentag, Department of Psychology, University of North Carolina, Greensboro, NC 27412 (e-mail: guttenta@goodall.uncg.edu).

—Accepted by previous editor, Geoffrey R. Loftus

Hintzman et al. (1994) findings do not refute the memorability hypothesis, the findings do place constraints on the mechanisms involved in making memorability judgments and using memorability information. More recently, Hintzman and Curran (1997) reported that word frequency information is, in fact, extracted very rapidly and early enough in the recognition process to make the memorability hypothesis a viable explanation for even the Hintzman et al. data. This finding is thus consistent with the suggestion made first by Brown et al. (1977) and more recently by Glanzer et al. (1993) that judgments of memorability may occur very rapidly and perhaps automatically.

The findings from a study by Wixted (1992), in contrast, challenge even the basic notion that subjects are capable of making accurate judgments of recognition memorability. In this study, the subjects rated HF and LF words (as well as rare words) for relative memorability. The subjects were instructed to imagine that the words had been encountered previously on a study list; the subjects' task was then to judge the relative memorability of these words, assuming that their memory would be tested with a recognition procedure. Wixted predicted that, if subjective memorability is taken into account during recognition decision making, subjects' estimates of memorability for HF and LF words should mirror the underlying distribution of correct recognition memory responses. Contrary to the predictions of the memorability judgment perspective, Wixted found that memorability ratings were not consistent with the standard recognition memory word frequency effect; on the contrary, the subjects in the Wixted study incorrectly judged HF words as significantly *more* memorable than LF words.

Recently, Greene and Thapar (1994) extended Wixted's (1992) findings in two important ways. First, Greene and Thapar examined memorability judgments, using a frequency of presentation task rather than a recognition memory task. Second, Greene and Thapar provided subjects with experience judging the frequency of presentation of HF and LF words immediately prior to the memorability judgment task. Greene and Thapar found that, despite the LF advantage that occurred on the frequency judgment task, the subjects nonetheless judged HF words as more memorable than LF words on the memorability judgment task.

One possible explanation for the Wixted (1992) findings is that subjects may simply have had difficulty accurately *imagining* a recognition task and therefore may have tended to rate words as though their memory were to be tested by recall. This explanation is more difficult to apply, however, to the findings from the Greene and Thapar (1994) study, in which subjects were provided experience with the criterial memory task immediately prior to making their memorability judgments.

More serious questions regarding the Wixted (1992) and Greene and Thapar (1994) findings are raised by recent research on the way in which subjects make differ-

ent kinds of metamemory judgments. In a study by Leonezio and Nelson (1990), for example, the subjects first made judgments about the ease with which they could learn each word item on a presented list (ease-of-learning judgments); then, the subjects studied each item and made judgments of how well each item had been learned (judgments of knowing). Finally, during a subsequent recall test, the subjects judged (for items they could not recall) the likelihood that they would recognize the item on a recognition test (feeling-of-knowing judgments). Leonezio and Nelson found that ease-of-learning judgments were much less accurate than either judgments of knowing or feeling-of-knowing judgments; moreover, even the ratings of judgments of knowing and feelings of knowing were not highly correlated with each other.

With the paradigms used both by Wixted (1992) and by Greene and Thapar (1994), subjects were required to make ease-of-learning judgments. The Leonezio and Nelson (1990) findings suggest that such judgments may not be based on the same factors as those that might be involved if subjects made memorability judgments at the time of the recognition test itself. Indeed, Glanzer and Kim (Glanzer et al., 1993; Kim & Glanzer, 1994) have specifically argued that recognition memorability judgments are derived, in part, from information extracted during each study trial itself, rather than being based on general metamnemonic knowledge of the ease of memorization of specific categories of words.

The present study was designed to investigate further the relationship between judged memorability and the pattern of WFEs in recognition memory. Five experiments are reported. The first experiment utilized a procedure similar to that used by Wixted (1992), but with the addition of Greene and Thapar's (1994) procedure of providing subjects with study/test experience with HF and LF words immediately prior to making memorability judgments on a new set of words. The remaining experiments, in contrast, assessed judgments of memorability online—that is, as subjects made their actual recognition judgments.

EXPERIMENT 1

The subjects in Experiment 1 were presented a study list of HF and LF words, followed by a standard yes–no recognition test. Immediately after completing the recognition test, the subjects were asked to rate the recognition memorability of words on a new list of HF and LF words.

Method

Subjects. Forty-two undergraduate students participated as subjects in the experiment in order to satisfy an introductory psychology course requirement.

Materials. A pool of 48 HF and 48 LF nouns was first compiled, using the Carroll, Davies, and Richman (1971) norms; the ranking for all HF words was <1,500, whereas the ranking for all LF words was >10,000 but <15,000.

The study list contained 48 words: 16 HF targets, 16 LF targets, 8 medium frequency recency items, and 8 medium frequency pri-

macy items. Three different study lists were constructed such that each word from the original pool of 96 words occurred on a single study list. The HF and LF items were arranged in random sequence on each study list, with the restriction that no more than 3 words of either frequency category occurred in sequence.

A different recognition test was generated to correspond to each study list. The three recognition tests each contained the 16 HF and 16 LF targets from the corresponding study list intermixed in random sequence with 16 HF and 16 LF distractors. Each word from the item pool served as a distractor item on one form of the recognition test.

After the recognition test, the remaining 16 HF and 16 LF words from the original word pool were presented on the memorability-rating task list. Words were arranged in random order on this list, with the restriction that no more than 3 words of either frequency category occurred in sequence. Each word from the item pool occurred on one version of the memorability-rating task list.

Procedure. The subjects were tested in three groups, with an approximately equivalent number of subjects in each group. Each group received a different version of the presentation list and the corresponding recognition test and memorability task. The subjects were informed that a list of words would be read aloud to them, followed by a memory test. The 48 study list items were then presented orally at a 2 sec per item rate.

For the recognition test, the subjects were presented 64 words (32 targets, 32 distractors) printed on two sheets of paper and were instructed to put a Y (for yes) next to each word they thought had been on the study list and an N (for no) next to each word they thought was a distractor. After the recognition test, the subjects were informed that they would receive another list of words printed on a sheet of paper and that their task was to rate the recognition memorability of each word. The subjects were instructed to imagine that each rating-list item had been read aloud in a manner similar to that used for presentation of the previous study list. For each word, the subjects were asked to rate how likely they thought it was that they would have recognized the item on a recognition test similar to the test they had just taken. The subjects rated each word on a 3-point scale, ranging from 1 ("think you might not recognize the item") to 3 ("very certain you would recognize the item").

Results and Discussion

Table 1 presents the hit and false alarm rates for HF and LF words on the recognition test. The pattern was consistent with the typical mirror effect: a higher hit rate for LF words but a higher rate of false alarms for HF words. Within-subjects analyses of variance (ANOVAs) revealed a significantly higher hit rate for LF than for HF words [$F(1,41) = 13.00, p < .01$]. The LF advantage for false alarms was marginally significant [$F(1,41) = 3.75, p < .10$].

In order to analyze the memorability ratings, the median rating for each subject was calculated for HF and LF words. Table 1 presents the median of the median ratings. A sign test for matched pairs (Hays, 1963) revealed that the difference between the ratings for HF and those for

LF words was not significant. Thus, unlike previous studies (Greene & Thapar, 1994; Wixted, 1992) that showed that subjects mistakenly believe HF words to be more memorable than LF words on an ease-of-learning task, the current experiment found that memorability ratings for HF and LF words were virtually identical. The reason for this discrepancy between the present findings and the pattern of results found in earlier studies is not clear; apparently, the present procedure produced some effect of the prior recognition test experience on the memorability judgments. However, it is also clear that the present findings do not support a memorability judgment explanation for the effect of word frequency on false alarms; the subjects in Experiment 1 exhibited the typical LF advantage with respect to responses to distractors on the recognition test but judged LF words as no more memorable than HF words when making judgments of memorability.

One way in which the present procedure differs from what might occur during an actual recognition test is that the subjects in Experiment 1 made their memorability judgments on a new set of words after experiencing a recognition test rather than as part of the recognition process itself. Experiment 2 was designed to measure judgments of memorability on-line—that is, as part of the recognition test itself.

EXPERIMENT 2

Method

Subjects. Fifty-one undergraduate students participated as subjects in the experiment in order to satisfy an introductory psychology course requirement.

Materials and Design. The same input lists and corresponding recognition tests as those used in Experiment 1 were used in Experiment 2.

Procedure. The distinguishing feature of Experiment 2 was that the subjects were not given a separate word list for the memorability-rating task. Rather, the subjects were asked to rate the memorability of rejected recognition test items (words they had judged not to have been on the presentation list), using the 3-point scale that was used in Experiment 1. The subjects were asked to rate, for each item that they did not recognize, the likelihood that they would have recognized that item if it really had been presented on the study list. In all other respects, the procedure used in Experiment 2 was identical to that used for the study list and recognition test components of Experiment 1.

Results and Discussion

Table 2 presents the mean hit and false alarm rates from the recognition task. The hit and false alarm data shown in Table 2 are consistent with the findings from Experiment 1 and conform to the predictions of the mirror effect. A within-subjects ANOVA revealed a significantly higher hit rate for LF words [$F(1,50) = 39.53, p < .01$], and a significantly higher rate of false alarms for HF words [$F(1,50) = 36.36, p < .01$].

Table 2 also presents the median memorability ratings for correct rejections. A sign test for matched pairs revealed that the subjects rated the LF words as significantly

Table 1
Mean Hit and False Alarm Rates and
Median Memorability Ratings in Experiment 1

Frequency	Hits	False Alarms	Memorability Rating
High	.67	.16	2.07
Low	.79	.11	2.11

Table 2
Mean Hit and False Alarm Rates and
Median Memorability Ratings in Experiment 2

Frequency	Hits	False Alarms	Memorability Rating
High	.72	.22	2.00
Low	.86	.10	2.25

more memorable than the HF words ($z = 3.00, p < .01$). This pattern is the opposite of the results found by Wixted (1992) and by Greene and Thapar (1994) but parallels the pattern of the WFE for false alarms. Thus, the present findings provide evidence that subjects may be sensitive to the greater memorability of LF as compared with HF words when the assessment of memorability occurs within the context of making recognition judgments.

EXPERIMENT 3

Given that the findings from Experiment 2 presented the first direct evidence for higher memorability ratings for LF than for HF words, it was considered prudent to replicate the finding with only minor changes in procedure. Experiment 3, accordingly, differed from Experiment 2 in only one respect: whereas in Experiment 2 a 3-point memorability rating scale was used, Experiment 3 utilized a 5-point scale, in order to make the measurement scale potentially more sensitive to item effects.

Method

Subjects. Fifty-four undergraduates participated as subjects in the experiment in order to satisfy an introductory psychology course requirement.

Materials. The presentation lists and corresponding recognition tests used in this experiment were the same as those used in Experiment 2.

Procedure. Experiment 3 was identical to Experiment 2, except that the subjects were given an expanded 5-point scale for rating the memorability of rejected test items (1 = "probably would not recognize" to 5 = "absolutely sure you would recognize"). In addition, whereas it was possible in Experiment 2 that some subjects may have made recognition judgments for several items prior to rating the memorability of any of the items, the subjects in Experiment 3 were explicitly instructed to rate the memorability of each nonrecognized item immediately after making the judgment that the item had not been on the study list.

Results and Discussion

Table 3 presents the mean hit and false alarm rates and median memorability ratings as a function of word frequency. As was found in the previous two experiments, the effects of word frequency on hit and false alarm rates conformed to the predictions of the mirror effect. A within-subjects ANOVA revealed that the subjects were significantly more accurate with LF than with HF targets [$F(1,53) = 24.33, p < .01$]; false alarm rates were higher for HF than for LF distractors [$F(1,53) = 18.43, p < .01$].

Table 3 also reveals that, as was found in Experiment 2, memorability ratings were significantly affected by word

frequency. A sign test revealed that the subjects rated LF correct rejections as significantly more memorable than HF correct rejections ($z = 3.57, p < .01$). Thus, the present findings provide further evidence that subjects are sensitive, while performing a recognition test, to the effects of word frequency on item memorability.

One possible explanation for this finding is that subjects may become explicitly aware of WFEs as a result of their experience with words of the two frequency classes during the early part of the recognition test. If this explanation is correct, it would suggest that the effects of word frequency on judgments of memorability should be attenuated early in the recognition test—that is, prior to the time at which subjects have had the opportunity to accumulate sufficient evidence of WFEs on recognition. Experiment 4 was designed to explore this possibility.

EXPERIMENT 4

Experiment 4 was essentially identical to Experiment 3, except that the recognition test was designed to permit a comparison of the magnitude of the memorability judgment WFE during different stages of the recognition test.

Method

Subjects. Twenty-nine undergraduate subjects participated as subjects in the experiment in order to satisfy an introductory psychology course requirement.

Materials and Design. A single presentation list of 48 words was constructed by randomly selecting 16 HF and 16 LF words from the original word pool. The presentation list also contained 8 medium frequency primacy items and 8 medium frequency recency items. The HF and LF words were arranged randomly on the study list, with the restriction that no more than 3 words of the same frequency category occurred in succession.

A 64-item recognition test was constructed such that each 16-item quartile contained four HF and four LF targets and four HF and four LF distractors. Items were randomly arranged within quartiles, with the restriction that no more than 3 items of the same type or frequency occurred in sequence. This initial version of the test was then used to generate four versions of the recognition test by counterbalancing quartiles of the test list across presentation quartiles on the test.

Procedure. The subjects were tested in a single group session; approximately one fourth of the subjects received each of the four versions of the recognition test. In all other respects, the procedure was identical to that in Experiment 3.

Results and Discussion

The general pattern of hits and false alarms was consistent with the mirror effect (see Table 4). A within-subjects ANOVA revealed a significantly higher hit rate

Table 3
Mean Hit and False Alarm Rates and
Median Memorability Ratings in Experiment 3

Frequency	Hits	False Alarms	Memorability Rating
High	.77	.14	3.00
Low	.88	.07	3.30

Table 4
Mean Hit and False Alarm Rates in Experiment 4

Frequency	Hits	False Alarms
High	.76	.18
Low	.91	.14

for LF than for HF targets [$F(1,28) = 25.57, p < .01$], but the difference in false alarm rates was not significant in this experiment [$F(1, 28) = 1.50, p > .20$].

Table 5 presents the median memorability ratings for HF and LF correct rejections as a function of test quartile. Inspection of Table 5 reveals that the magnitude of the effect of word frequency on memorability ratings was quite consistent across the four quartiles of the test, and sign tests revealed that the subjects rated LF items as significantly more memorable than HF items during each of the four quartiles of the test ($z = 2.40, 2.00, 3.20, \text{ and } 1.86$, for each of the four quartiles, respectively; all $ps < .05$).

Thus, consistent with the findings from Experiments 2 and 3, the subjects in Experiment 4 gave higher memorability ratings to LF than to HF distractors. Most notably, this effect emerged within the initial quartile of the recognition test, suggesting that the effect is dependent on, at most, very limited recognition test experience with the specific categories of materials used in the experiment.

EXPERIMENT 5

One feature common to Experiments 2, 3, and 4 was that the subjects made judgments of memorability only for rejected items; the analyses were based solely on their judgments for correctly rejected distractors. This feature of the procedure raises the possibility that the apparent effects of word frequency on memorability judgments may, in fact, have resulted from some form of item selection bias. In other words, it may be the case that, if judgments had been collected for all list items rather than just for rejected items, the distribution of judgments would have been similar for HF and LF words, rather than exhibiting higher overall ratings for LF words.

A second feature common to these experiments was that, for each item on the recognition test, the subjects first made their judgment of recognition and then (for rejected items) made their judgment of memorability. However, if memorability judgments are involved in making judgments of recognition, subjects should judge LF words as more memorable than HF words, even if they are asked to make their memorability judgment for an item prior to making their recognition judgment.

Experiment 5 was designed to address each of these potentially limiting features of the prior experiments. The subjects in Experiment 5 made judgments of memorability for all recognition test items, including both items judged to have been targets and items judged to have been distractors. Experiment 5 also manipulated the order in which judgments of recognition and judgments of memorability were made during the test trial; half the subjects

were instructed to judge an item's memorability prior to making their recognition judgment for that item, whereas the remaining subjects were instructed to make their recognition judgment for an item prior to making their judgment of that item's memorability.

Method

Subjects. Forty-seven undergraduate students participated as subjects in the experiment in order to satisfy an introductory psychology course requirement.

Materials and Design. Two 40-target-item presentation lists were constructed by selecting 20 HF and 20 LF words for each list from the original word pool. The presentation lists also contained 2 HF and 2 LF primacy items and 2 HF and 2 LF recency items. The target words were arranged randomly on each study list, with the restriction that no more than 3 words of the same frequency category occurred in succession.

An 80-item recognition test/memorability judgment test was constructed such that each 20-item quartile contained five HF and five LF targets and five HF and five LF distractors. The test was composed of the 40 target words from each of the two study lists; in this way, items that were targets for subjects receiving study list 1 were distractors for subjects receiving study list 2.

This test was presented in the form of a four-page booklet. At its top, each page contained a description of the 5-point memorability scale that was used in Experiments 3 and 4. The 20 test items on each page were printed in a single column beneath the information about the memorability scale. Next to each item was printed a Y and an N (for making recognition decisions) and the numbers 1 through 5 (for making memorability judgments). Approximately half the subjects ($n = 22$) were instructed to make memorability judgments prior to recognition judgments for any particular items; for these subjects the numbers for the memorability judgments were printed to the left of the letters Y and N. The position of the letters and numbers was reversed for the subjects ($n = 25$) instructed to make their recognition judgments prior to their memorability judgments.

Procedure. The subjects were tested in four groups. Subjects in two of the groups were presented study list 1; the remaining subjects were presented study list 2. For each study list, the subjects in one of the groups were instructed to make their memorability judgment for an item prior to making their recognition judgment for that item; the subjects in the other group receiving each study list were instructed to make their recognition judgment for an item prior to making their memorability judgment.

The method of list presentation was identical to that in the prior experiments. For the recognition/memorability test, the subjects were instructed to make both kinds of judgments for each item and to be sure to make the judgments for each item in the order specified.

Results and Discussion

The general pattern of hits and false alarms was again consistent with the mirror effect (see Table 6). A within-subjects ANOVA revealed a significantly higher hit rate for LF than for HF targets [$F(1,45) = 23.53, p < .01$], and a significantly higher rate of false alarms for HF than for

Table 5
Median Memorability Ratings by Quartile
of the Recognition Test in Experiment 4

Frequency	Quartile			
	Q1	Q2	Q3	Q4
High	2.88	2.88	3.00	3.00
Low	3.50	3.10	3.62	3.25

Table 6
Mean Hit and False Alarm Rates in Experiment 5

Frequency	Hits	False Alarms
High	.69	.25
Low	.80	.10

LF distractors [$F(1,45) = 40.20, p < .01$]. There were no significant effects of the order of task presentation.

In order to examine the judgments of memorability, the median rating was calculated for each subject for each of the four types of recognition test response items: hits, correct rejections, incorrect rejections, and false alarms. The median ratings are presented in Table 7 as a function of the order of task presentation (memorability rating followed by recognition vs. recognition followed by rating). It should be noted that the number of responses that contributed to the median ratings was much higher for the two categories of correct responses (hits and correct rejections) than for the two categories of incorrect responses (incorrect rejections and false alarms).

Overall, it may be seen that there were no consistent effects of the order in which the tasks were performed. It must be noted, however, that, although the subjects in the *rate first* groups were required to mark their memorability rating for an item prior to marking their recognition judgment, there is no way to know if the subjects actually made the rating prior to the recognition decision. That is, it is possible that even the subjects in the *rate first* groups may have covertly judged the old/new status of test items prior to making their memorability judgments. Indeed, our own experience with the task suggests that it is very difficult to avoid making a recognition judgment for each item as soon as it is read.

Not surprisingly, the subjects gave higher ratings to items that they judged as having been on the list (hits and false alarms) than to items that they judged as having not been on the list (incorrect rejections and correct rejections). It may also be seen that the subjects tended to give higher ratings to correctly judged items (hits and correct rejections) than to items on which they made a response error (incorrect rejections and false alarms): for HF items, $z = 4.07, p < .05$; for LF items, $z = 3.86, p < .01$. This effect presumably reflects the sensitivity of the subjects to the fact that some items were less memorable than others and, consequently, more difficult to classify correctly as old or new.

The primary focus of Experiment 5 was on the effects of word frequency on the memorability judgments. Table 7 reveals that, on average, the subjects rated LF items as more memorable than HF items for seven of the eight comparisons presented in the table. The one exception involved the rating of false alarms for subjects who rated memorability prior to making their recognition judgment. It should be noted that false alarms represented the least frequent category of responses; accordingly, the data in this cell are based on a relatively small number of responses by each subject. Moreover, the predicted pat-

tern of results emerged even for false alarms for subjects who made their ratings after making their recognition judgments.

Collapsing across task order and response category, the overall median rating given to target items was 3.17 for HF items and 3.90 for LF items; the overall median rating given to distractors was 3.00 for HF items, and 3.30 for LF items. Sign tests revealed that subjects rated LF items as significantly more memorable than HF items, both for targets (hits and incorrect rejections; $z = 4.34, p < .01$) and for distractors (correct rejections and false alarms; $z = 3.45, p < .01$). These findings indicate that the tendency for subjects to judge LF words as more memorable than HF words on a recognition test is not limited to correctly rejected items but rather reflects a more general effect of word frequency on memorability judgments during recognition test performance.

GENERAL DISCUSSION

Despite the long history of research on the WFE in recognition memory (see Gregg, 1976, for a review of research on the WFE extending back to 1961), the phenomenon continues to present an explanatory challenge for models of recognition memory, and no general consensus has yet emerged regarding the nature of the processes responsible for the LF word advantage (Gillund & Shiffrin, 1984; Glanzer & Adams, 1985, 1990; Glanzer, Adams, & Iverson, 1991; Glanzer et al., 1993; Hintzman, 1994; Hintzman et al., 1994; Hintzman & Curran, 1997; Jacoby, 1991; Mandler, 1980).

One proposal, first presented by Brown (1976; Brown et al., 1977; Brown & Routh, 1970), invokes the concept of memorability to help explain the WFE. According to this view, subjects are sensitive to the actual memorability of different categories of list items and adjust their response criteria accordingly. A similar rescaling mechanism is central to a number of more recent models as well (see, e.g., Gillund & Shiffrin, 1984; Glanzer & Adams, 1990; Glanzer et al., 1993). The viability, however, of any kind of memorability judgment mechanism is called into question by Wixted's (1992) and Greene and Thapar's (1994) direct assessment of judgments of memorability for HF and LF words. Contrary to the memorability judgment hypothesis for the WFE, these stud-

Table 7
Median Memorability Ratings as a
Function of Task Order and Word Frequency
for Each Type of Test Response Item in Experiment 5

Task Order	Word Frequency	Type of Item			
		Targets		Distractors	
		Hits	Incorrect Rejections	Correct Rejections	False Alarms
Rating/Recognition	High	3.90	2.38	2.90	3.75
	Low	4.50	3.17	3.48	3.50
Recognition/Rating	High	3.80	2.50	2.96	3.17
	Low	4.59	3.00	3.27	3.50

ies found that subjects incorrectly judged HF words as more memorable than LF words. These findings suggest either that memorability judgments play no role at all in judgments of recognition or else that they operate at a completely implicit level and that subjects have no conscious access to the products of the mechanisms involved.

The present study provides further evidence that, when subjects are instructed to judge the memorability of a set of HF and LF words, they do not judge LF words as more memorable, even if they have just completed a recognition test with similar materials. However, the present study also found that, if subjects make their memorability judgments while performing the recognition test itself, they do tend to give higher memorability ratings to LF than to HF test items. Moreover, this effect of word frequency on rated memorability emerged within the initial quartile of the recognition test and occurred even when the subjects were instructed to make their memorability judgments prior to their judgments of recognition.

These findings suggest an effect of word frequency on judgments of memorability that is related to processes operating at the level of individual item recognition rather than one involving more general metamnemonic knowledge of the recognition memorability of entire classes of recognition test items. An extension of Brown's (1976; Brown et al., 1977) memorability proposal, the attention/likelihood model (Glanzer & Adams, 1990; Glanzer et al., 1993), offers a more precise explanation of how memorability information may be obtained and used during the recognition experience. According to attention/likelihood theory, during study disproportionate attention is paid to certain classes of items relative to others (e.g., more to LF words than to HF words). As a consequence of this attentional advantage, more lexical features are marked for well-attended items; indeed, recent evidence indicates that more contextual details associated with item presentation are encoded as well (Bornstein & LeCompte, 1995; Guttentag & Carroll, 1994). The theory further assumes that subjects register information pertaining to attentional and marking variation and use this information for setting criterion levels during the recognition test. Thus, a LF distractor may be easily rejected for containing fewer marked lexical features and less associated contextual detail for item presentation than would be expected had the item been presented during study. A HF distractor is less easily rejected, because the difference in the number of marked features associated with distractors and targets is smaller.

The contrast between the present findings and the pattern of results observed by Wixted (1992) suggests that, under the conditions of assessment utilized in Wixted's paradigm, subjects were unable to make valid judgments of memorability, even though they are able to make such judgments when actually performing a recognition task. The pattern of results observed by Wixted, however, is entirely consistent with other research on judgments of ease of learning. This research has found that subjects perform such tasks by assessing the immediate familiarity

of the items and that, to the extent that immediate familiarity and delayed recall and recognition may be affected by different variables (as they are by word frequency when retention is assessed by recognition), immediate judgments of ease of learning are very inaccurate (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Dunlosky & Nelson, 1992; Leonesio & Nelson, 1990; Nelson & Dunlosky, 1992). Thus, the present findings provide further evidence that even seemingly highly related metamnemonic judgments (in the present case, judgments of memorability made for items on a study list vs. judgments of memorability for items on a recognition test) may be based on very different factors if they occur at different points in the study/test cycle.

Although the present findings suggest that subjects are sensitive to item differences in memorability when performing recognition tests, the findings present little direct evidence that such judgments are actually used when making recognition judgments. Hintzman et al.'s (1994) finding of a recognition WFE even under conditions of very rapid responding suggests that, if memorability information is used during the process of recognition, the processes involved must occur within a brief time window following test item presentation.

Hintzman et al. (1994) argued that their findings preclude a slow reflective process in which subjects consciously generate memorability judgments and then reflect, for each item, on the relationship between their feeling of familiarity for the item and their memorability judgment for that item. It is possible, however, that memorability information may be extracted quite rapidly, utilizing a mechanism such as that described in Glanzer and Adams's (1990; Glanzer et al., 1993) attention/likelihood model. This possibility is supported by Hintzman and Curran's (1997) recent finding that word-frequency-relevant information is extracted prior to the minimum time required for accurate judgments of recognition. If this is the case, it would suggest that the slower judgments of memorability assessed in the present study involved access to the products of the more rapid and automatic processes involved in actually utilizing memorability information during recognition test performance.

REFERENCES

- BEGG, I., DUFT, S., LALONDE, P., MELNICK, R., & SANVITO, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory & Language*, *28*, 610-632.
- BORNSTEIN, B. H., & LECOMPTÉ, D. C. (1995, November). *Stimulus characteristic effects on source monitoring*. Paper presented at the Annual Meeting of the Psychonomic Society, Los Angeles.
- BROWN, J. (1976). An analysis of recognition and recall and of problems in their comparison. In J. Brown (Ed.), *Recall and recognition* (pp. 1-35). New York: Wiley.
- BROWN, J., LEWIS, V. J., & MONK, A. F. (1977). Memorability, word frequency, and negative recognition. *Quarterly Journal of Experimental Psychology*, *29*, 461-473.
- BROWN, J., & ROUTH, D. A. (1970). Recognition assessed by d' and by non-parametric alternative (the A-Index) as a function of the number of choices. *Quarterly Journal of Experimental Psychology*, *22*, 707-719.

- CARROLL, J. B., DAVIES, P., & RICHMAN, B. (1971). *The American Heritage word frequency book*. New York: American Heritage.
- DUNLOSKY, J., & NELSON, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, **20**, 374-380.
- 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., ADAMS, J. K., & IVERSON, G. (1991). Forgetting and the mirror effect in recognition memory: Concentrating of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 846-858.
- GLANZER, M., ADAMS, J. K., IVERSON, G., & KIM, K. (1993). The regularities of recognition memory. *Psychological Review*, **100**, 546-567.
- GREENE, R. L., & THAPAR, A. (1994). Mirror effect in frequency discrimination. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 946-952.
- GREGG, V. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183-216). London: Wiley.
- GUTTENTAG, R. E., & CARROLL, D. (1994). Identifying the basis for the word frequency effect in recognition memory. *Memory*, **2**, 255-273.
- HAYS, W. L. (1963). *Statistics*. New York: Holt, Rinehart & Winston.
- HINTZMAN, D. L. (1994). On explaining the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 201-205.
- HINTZMAN, D. L., CAULTON, D. A., & CURRAN, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 275-289.
- HINTZMAN, D. L., & CURRAN, T. (1997). Comparing retrieval dynamics in recognition memory and lexical decision. *Journal of Experimental Psychology: General*, **126**, 228-247.
- JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.
- KIM, K., & GLANZER, M. (1994). Attention/likelihood theory: Reply to Hintzman (1994). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 206-208.
- LEONESIO, R. J., & NELSON, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 464-470.
- MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.
- NELSON, T. O., & DUNLOSKY, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science*, **3**, 317-318.
- WIXTED, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 681-690.

(Manuscript received April 23, 1996;
revision accepted for publication April 5, 1997.)