Memory & Cognition
1985, 13 (1), 8-20

The mirror effect in recognition memory

MURRAY GLANZER and JOHN K. ADAMS
New York University, New York, New York

The mirror effect in recognition memory refers to the fact that, with several different classes of stimuli, performance on new items from each class mirrors (is correlated with) performance on the corresponding classes of old items. Classes of stimuli that are accurately recognized as old when old are also accurately recognized as new when new; those that are poorly recognized as old when old are also poorly recognized as new when new. The statement above is shown not to be a tautology. A survey demonstrates that the effect holds for several types of variables (ways to classify stimuli)—word frequency, concreteness, meaningfulness, and others. The survey includes a total of 80 findings. The theoretical implications of the effect are considered.

This paper will present evidence that there is a general, systematic effect in recognition memory, called here the mirror effect. The effect refers to the situation in which a subject faces stimuli of different types, for example, high and low normative frequency words. In such situations, the type of stimulus that is accurately recognized as old when old is also accurately recognized as new when new. The type that is poorly recognized as old when old is also poorly recognized as new when new. Performance on new items *mirrors* performance on old items.

The effect will be shown to hold for all types of recognition tests. It refers basically to the ordering of the underlying theoretical distributions used to account for recognition performance. When the mirror effect holds, the order of the distributions representing new classes of stimuli mirrors in reverse order the distributions representing old classes. The mirror effect, if it holds, gives clear, corresponding patterns at the level of data. In the case of yes/no recognition data, the mirror effect is seen when, for different classes of stimuli, the false-alarm rates mirror in reverse order the hit rates.

It will be shown that the mirror effect is not a tautology. Several different relations can exist between the underlying theoretical distributions and their corresponding data. On theoretical grounds, these other relations would seem more likely than the one that actually holds. In fact, on the basis of single sets of data, some investigators have argued for one of these other relations.

The discussion of the mirror effect here will be in the framework of signal detection or decision theory. This is done for convenience of exposition, since that framework is familiar and is the basis of most discussions of recognition memory. The empirical relation on which this paper focuses is independent of any particular theory. It places, however, a very strong constraint on any theory of recognition memory.

In the framework of signal detection theory, each type of stimulus is represented by a pair of underlying theoretical distributions. If only one type of stimulus is being studied, then there is only a single pair—a new (N) and an old (O) distribution, as shown in the top panel of Figure 1. The old distribution is placed higher on a familiarity dimension. The distributions depicted are simply illustrative, and their form only indicates variability in familiarity values.

If there are two classes of stimuli, then two pairs of distributions have to be considered. The four distributions may be arranged in one of three main patterns if the subject responds differentially to the two classes, A and B. (Here, the subject's recognition performance is better with A than with B.) The differential response may be due to the new distributions' being separated although the old are not (Panel 2 of Figure 1). The greater distance between the new and old distributions of class A stimuli translates into the greater discriminability of that set.

Another pattern is shown in Panel 3 of Figure 1. Here, the greater discriminability of class A is due to differences solely in the old distributions. The new distributions are equally unfamiliar.

The pattern shown in Panel 4 of Figure 1 represents the mirror effect. The new distributions mirror in reverse order the old distributions. It will be argued here that this pattern holds generally for recognition memory involving different stimulus classes.

On a priori grounds, the mirror-effect pattern seems the least likely of the three patterns. Exposure or study of the stimuli results in a *reversal* of the order of their familiarity values when new. If the mirror effect can be shown to hold, any theory of recognition memory has to include a mechanism for this reversal.

A simple learning-theory view might be that the differences between class A and class B exist initially in the new distributions. Study moves the values for items in both classes up on the familiarity scale. If the learning operator preserved the order of the classes, the full order would be AN, BN, AO, BO in sequence. This is the opposite of a mirror effect. If the distances between the old

We thank Jean-Claude Falmagne, Alina Hsu, Micha Razel, and Gay Snodgrass for their helpful comments on the paper. The authors' mailing address is: Department of Psychology, New York University, 6 Washington Place, New York, NY 10003.

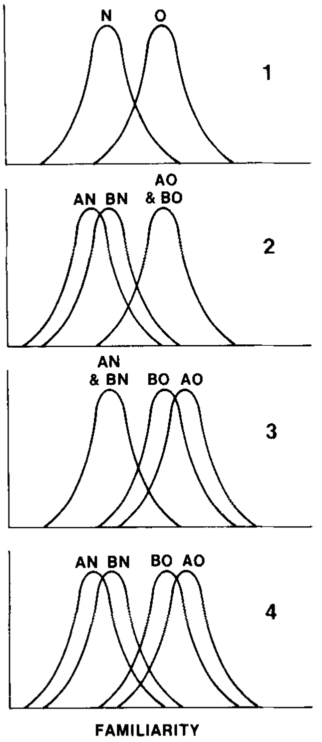

Figure 1. Possible arrangements of distributions underlying recognition. N = new, O = old. A and B indicate two different classes of stimuli.

distributions were reduced because of limits on the effect of the learning operator, then the pattern in Panel 2 of Figure 1 would be approximated. Murdock (1974) cited data on word frequency (see Table 2), and Morris and Reid (1974) cited data on imageability in support of that pattern.

Another simple view might be that noise is noise, that new items should be indistinguishable. Different classes may, however, differ in the ease with which they can be learned. This view leads to the pattern seen in Panel 3 of Figure 1.

The data that support the prevalence of the mirror effect, the pattern in Panel 4, will now be presented. Then its theoretical implications will be discussed.

The simple cases for the evaluation of the mirror effect are those in which the classes of stimuli are a within-subjects effect. That means that the same subject is responding to the different classes of stimuli. If the classes of stimuli are a between-subjects effect, it can be argued that assignment of positions of distributions on a single familiarity dimension is unclear. It can also be argued that different groups, using different criteria, make the assignment of positions complicated. In order to avoid such complexities, the focus will be on data from within-subjects designs. It will be pointed out, however, that the mirror effect appears clearly in data from between-subjects designs.

BACKGROUND

The mirror effect was demonstrated in a study by Glanzer and Bowles (1976), who used two-alternative forced choice in the recognition of high-frequency and low-frequency words. The effect was explicitly noted and empirically supported by Brown, Lewis, and Monk (1977). They also offered a theory for the high level of confidence with which new items from a well-recognized set of stimuli are rejected. This theory will be discussed later. Morris (1978) argued for a mirror effect in the specific case of word frequency.

In the Glanzer and Bowles (1976) study, the subjects, in a recognition test, had to choose the old item pairs in which both items were of high frequency, both were of low frequency, and pairs in which one item was of high and one was of low frequency. The data are summarized in Table 1. Examination of the proportions of choices reveals that the pattern in Panel 4 of Figure 1 holds. Low-frequency words correspond to A and high-frequency words to B in Figure 1. The picture implies that $P(LO, LN) > P(HO, LN), P(LO, HN) > P(HO, HN)$. The inequalities are obtained by taking the distances between the relevant underlying distributions. P(LO, LN) represents the proportion of choices of low-frequency old items over low-frequency new items. The comma between P(HO, LN) and P(LO, HN) indicates that the relation of those two proportions is indeterminate for the assumed array. Either $P(HO, LN) > P(LO, HN)$ or $P(LO, HN) > P(HO, LN)$, or their equality will satisfy the mirror pattern.

Other patterns of data would have been obtained if other theoretical arrays held. For example, if Panel 3 of Figure 1 held, the pattern of choices

$$P(LO, LN) = P(LO, HN) > P(HO, LN) = P(HO, HN)$$

would be seen, or, more loosely,

Table 1
Results for Three Forced-Choice Recognition Studies

	Choice			
Study	P(HO, HN)	P(LO, HN)	P(HO, LN)	P(LO, LN)
Bowles & Glanzer (1983)	.77	.82	.83	.88
Bowles & Poon (1982)	.74	.81	.80	.84
Glanzer & Bowles (1976)	.75	.80	.83	.89

Note—H = high frequency, L = low frequency, O = old, N = new. The order of terms within the parentheses for each column heading indicates the types of choice considered. P(LO, LN) is the mean proportion of choices of LO over LN.

$$P(LO, LN), P(LO, HN) > P(HO, LN), P(HO, HN),$$

with the difference between P(LO, LN) and P(LO, HN) and the difference between P(HO, LN) and P(HO, HN) being smaller than those marked by the inequality. This is clearly not the case in the Glanzer and Bowles (1976) data.

Another theoretical possibility bases the discrimination on initial differences in the new items. This possibility is represented in Panel 2 of Figure 1. It implies that, in forced choice,

$$P(LO, LN) = P(HO, LN) > P(LO, HN) = P(HO, HN),$$

or, more loosely,

$$P(LO, LN), P(HO, LN) > P(LO, HN), P(HO, HN),$$

with the difference between P(LO, LN) and P(HO, LN) and the difference between P(LO, HN) and P(HO, HN) being relatively small. This pattern is also contradicted by the Glanzer and Bowles (1976) data.

As noted earlier, the pattern that does fit that set of data is the mirror pattern in Panel 4 of Figure 1. The appropriateness of the mirror pattern is reinforced by results from two other types of forced choice given in that experiment. In those choices, labeled null choices, the subjects were given pairs in which the members were both old and differed only in frequency and pairs in which they were both new and differed only in frequency. According to the arrays showing the mirror effect in Panel 4 of Figure 1, the subjects should choose LO over HO and HN over LN. The subjects do just that. P(LO, HO) was .64; P(HN, LN) was .65.

With these data and other strongly concurring data from studies by Bowles and Glanzer (1983), Bowles and Poon (1982) (see Table 1), and Brown et al. (1977), it was decided to survey available recognition studies to determine whether the mirror effect holds generally for (1) word frequency and (2) other stimulus variables.

Two main criteria were set up to determine whether the study could be included in the survey: The study gave data from a within-subjects design, and it presented sufficient data to determine presence or absence of the mirror effect.

(1) *Within subjects design:* The reason for this criterion was given earlier. The same subject has to be tested on the several levels of the factor, for example, word frequency. This permits the argument that the derived distributions such as those in Figure 1 are arrayed along a single axis on which each distribution's relative position is specified. If the data for high-frequency words and low-frequency words come from different subjects, then the relative positions of the underlying distributions from different classes are indeterminate in the case of yes/no data. In the case of rating-scale data, it could be argued that the different subjects are using different systems for the assignment of their ratings. Since there is a large body of relevant within-subjects data that does not involve such problems, only those data will be considered fully here. Related data from between-subjects designs will, however, be noted briefly.

(2) *Sufficiency of data:* Information sufficient for the placement of both the old and the new distributions must be presented in the report of the study. One procedure usually is not reported with such information, the multiple-choice recognition test (e.g., Kinsbourne & George, 1974). In some cases, an investigator carries out a procedure that is often reported with the needed information, but does not report it. For example, the investigator may use the yes/no procedure, but does not report hits and false alarms, but only a single composite score such as hits minus false alarms (Gorman, 1961).

Certain other exclusions were made. A few studies were excluded because the investigators did not find any indication of an experimental effect of the variable of interest, particularly when such an effect is generally found. Also excluded was a study in which the subjects were trained on the stimuli before the start of the experiment. These exclusions will be noted in the appropriate section below.

There are two other general exclusions in the survey reported below. Recognition studies using the Sternberg paradigm were not examined. In that paradigm, the dependent variable of interest is latency. The procedure is designed to produce minimal error rates. This means that the data of interest here, such as hit rates and false-alarm rates, would show ceiling and floor effects. Recognition studies using adults as subjects were the focus of interest. Studies using children and infants were not examined.

There are four types of recognition study: yes/no, rating scale, two-alternative forced choice, and multiple choice. The yes/no data give a simple basis for the evaluation of the mirror effect. If both the hit rates and the false-alarm rates, or their complements, are reported, then the hit rates for the classes of stimuli should be in reverse order from the false-alarm rates for those classes. In the case of low-frequency and high-frequency words, the following would hold:

$$\text{Hits}(L) > \text{Hits}(H) \text{ and } FA(H) > FA(L),$$

where L represents low-frequency and H high-frequency words, and FA = false alarm. The rating-scale data also give a simple basis for the evaluation of the mirror effect. For example, if the scale goes from 0 (for "sure that is new") to 7 (for "sure that is old"), then the ratings for the old items of different classes should be in reverse order from the ratings for the corresponding classes of new items. In the case of high-frequency and low-frequency words, the following would hold:

$$\text{Rating } (LN) < \text{Rating } (HN) < \text{Rating } (HO) < \text{Rating } (LO).$$

Forced choice will refer here to the case of two-alternative forced choice. Forced-choice data, as indicated earlier, show the mirror effect for low- and high-frequency words if the following pattern of choice holds:

$$P(LO, LN) > P(HO, LN), P(LO, HN) > P(HO, HN).$$

To obtain information on the underlying distributions' placements, it is necessary that all four pairs be presented to the subjects.

Multiple choice will refer here to cases in which the test items consist of three or more alternatives. Data from this procedure usually are reported in such a way that the underlying distributions cannot be specified. In a few cases, reports do indicate information that permits determination of the presence or absence of the mirror effect. These will be included below.

PLAN OF SURVEY

A survey was made of recognition studies that satisfied the criteria listed earlier: within-subjects design and sufficiency of data. Except for a few special cases noted below, no study was knowingly excluded. Most of the studies fall into four groups reflecting the major stimulus characteristics that affect recognition: normative word frequency, concreteness, meaningfulness, and pictures versus words. In the course of the survey, many other stimulus variables were found that will be grouped together in a set labeled "miscellaneous." This set includes factors such as common versus proper nouns and orthographic distinctiveness. The results will be presented in the form of summary tables.

In many studies, data from a single experiment were reported for several subgroups, for example, for different age groups or for several levels of another experimental variable. The data in almost all such cases were combined into a single set of means, and that set of means was evaluated and reported here. The degree to which the summary means represent the underlying levels or groups will, however, be noted.

Combining of data was not done for studies that present a series of experiments, each labeled differently. In those cases, each experiment is evaluated and reported separately here. In all but five cases, data from a single experiment appear only once in the survey. In those five cases (Brown et al., 1977; Groninger, 1976; Mandler, Goodman, & Wilkes-Gibbs, 1982; Rao, 1983; Rao & Proctor, 1984), the experimenters used several different classes of stimuli, which are reported in the appropriate tables. The results involved can be eliminated from consideration without any weakening of the argument that the mirror effect is a general effect.

The evaluations of the effect are generally carried out simply on the basis of whether or not the pattern corresponding to the mirror effect is present in the study. Significance of differences, for example, between false-alarm rates, is not taken into account. In many of the studies, this information is not presented. Moreover, since the concern here is with the prevalence of an overall pattern, its frequency across the relevant studies will be tabulated. That frequency will then be evaluated statistically.

Normative Word Frequency

A strong stimulus effect on recognition is that of normative word frequency. Twenty-four experiments in 19 studies were found that met the criteria. The experiments are listed in Table 2. Table 2 includes the following column headings and information:

(1) *Method:* The type of method used—FC = two-alternative forced choice; MC = multiple choice; YN = yes/no; RS = rating scale.

(2) *N:* The total number of subjects that furnished data on the variable of interest. This number is presented to give some indication of the stability of the results.

(3) *Data sets:* Many experiments report data for several different levels or groups. These are of interest as internal replications. For example, there are four different groups in the Bowles and Glanzer (1983) experiment—a retroactive group, a proactive group, and two control groups. The number 4 is therefore listed.

(4) *Positive sets:* The number of distinct data sets within an experiment that show the mirror pattern is listed in this column. These are the numbers of positive internal replications. In Bowles and Glanzer (1983), all four groups show the pattern. A 4 is therefore entered in that column. This is another indication of the stability of the results.

(5) *M:* This column indicates whether the combined data for the experiment, the overall means across the data sets, exhibit the mirror pattern. If they do, a "+" is entered; if not, a "−" is entered. In the Mandler et al. (1982) Experiment 2, one data set, or internal replication, showed the mirror effect, whereas the other did not. The combined means for all the data do, however, show the effect.

In Table 2, 23 of 24 results show the mirror effect. The only deviant result is Shepard's (1967). To evaluate the findings statistically, it is necessary first to specify the probability of a positive result's occurring by chance. In the case of the yes/no procedure, it is the following: Given that either $\mathrm{Hits(L)} > \mathrm{Hits(H)}$ or that $\mathrm{FA(H)} > \mathrm{FA(L)}$, what is the probability that the other holds? That probability by chance is .50.[1] Similarly, for the rating-scale procedure, given that either $\mathrm{Rating(LO)} > \mathrm{Rating(HO)}$ or $\mathrm{Rating(HN)} > \mathrm{Rating(LN)}$, the probability that the other holds is .50.

For forced choice, the mirror pattern is demonstrated by the following relations:

$$\mathrm{P((LO,\ LN)} > \mathrm{P(LO,\ HN)},\ \mathrm{P(HO,\ LN)} > \mathrm{P(HO,\ HN)}.$$

The following inequalities represent all the possible relations that could be generated by the three different theoretical arrays in Figure 1 (arrays that could produce the word-frequency effect):

$$\mathrm{P(LO,\ LN)} > \mathrm{P(LO,\ HN)} > \mathrm{P(HO,\ LN)} > \mathrm{P(HO,\ HN)} \quad (1)$$

$$\mathrm{P(LO,\ LN)} > \mathrm{P(HO,\ LN)} > \mathrm{P(LO,\ HN)} > \mathrm{P(HO,\ HN)} \quad (2)$$

$$\mathrm{P(LO,\ HN)} > \mathrm{P(LO,\ LN)} > \mathrm{P(HO,\ HN)} > \mathrm{P(HO,\ LN)} \quad (3)$$

$$\mathrm{P(HO,\ LN)} > \mathrm{P(LO,\ LN)} > \mathrm{P(HO,\ HN)} > \mathrm{P(LO,\ HN)}. \quad (4)$$

As noted earlier, either (1) or (2) would be generated in the case of the mirror effect. The exact placement of the

Table 2
Occurrence of the Mirror Effect (M = +) in Studies of Word Frequency

Study	Method	N	Data Sets	Positive Sets	M
Balota & Neely (1980)[1]	YN	136	4		+
Bowles & Glanzer (1983)	FC	80	4	4	+
Bowles & Poon (1982)	FC	44	2	2	+
Brown, Lewis, & Monk (1977)[2], Exp. 3a,3b	RS	42	2	2	+
Clark (1981)[3] Exp. 1	YN	96	4	4	+
Dorfman & Glanzer (1984)[4]	YN	80	4	4	+
Dorfman, Glanzer, & Kaufman (1984)[5]	YN	54	4	3	+
Glanzer & Bowles (1976)	FC	48	4	4	+
Grace (cited in Murdock, 1974)[6]	YN		1	1	+
Lee, Tzeng, Garro, & Hung (1978)[7]					
Exp. 1	YN	80	4	4	+
Exp. 2	YN	80	5	5	+
Mandler, Goodman, & Wilkes-Gibbs (1982)[8]					
Exp. 1	YN	32	2	2	+
Exp. 2	YN	48	2	1	+
McCormack & Swenson (1972)[9]	YN	50	2	2	+
Morris (1978)	YN	84	1	1	+
Poon & Fozard (1980)[10]	YN	57	3	3	+
Rao (1983)[11]					
Exp. 1	YN	72	1	1	+
Exp. 2	YN	36	1	1	+
Rao & Proctor (1984)					
Exp. 1	YN	128	3	3	+
Exp. 2	YN	64	2	2	+
Schwartz & Rouse (1961)[12]					
Exp. 5	MC	94	1	1	+
Exp. 9	MC	98	1	1	+
Shepard (1967), Exp. 1	FC	17	1	0	−
Wilhite (1981)	YN	40	1	1	+

[1] *Balota and Neely describe the order of false-alarm and hit rates but do not present numerical data. The four internal replications were generated by two expectancy conditions times two training conditions. The order reported in the text on page 581 may, however, refer only to the combined data. The number of positive sets cannot, therefore, be listed.* [2] *Brown, Lewis, and Monk use high- and low-frequency first names.* [3] *Data are from four groups—young; old normal; old mildly memory-impaired; and old moderately memory-impaired. All show the same pattern.* [4] *Data are based on four groups defined by speed versus accuracy instructions and two list composition conditions.* [5] *Data are from four groups—old and young subjects, each group under speed or accuracy instructions.* [6] *Grace's data are from an unpublished continuous recognition study reported by Murdock (1974, p. 67). Murdock points out that the difference in hit rate is not statistically significant but that the difference in false-alarm rate is. Murdock uses those data to argue for the pattern in Panel 2 of Figure 1. The N in the study is not reported.* [7] *Hit and false-alarm rates derived from reported d's and βs.* [8] *Study included another level of words (VL) that were not in Kučera and Francis and that were unknown to the experimenters. These are, of course, effectively nonwords and will be considered separately in contrast to words in a later section (Table 4). The hits and false alarms are read from Figure 1 of that study.* [9] *McCormack and Swenson present data for cases in which comparisons can be made both within subjects and across different groups of subjects. All the data follow the same pattern. The N of 50 in the table represents the two groups for which within-subject comparisons can be made. These data, however, although from within subjects, are from pure lists, not mixed lists as in the other studies listed. It could be argued that the subjects moving from list to list make the relative placement of the theoretical distributions indeterminate, as in a between-subjects design. If that argument is accepted, then the McCormack and Swenson data should be excluded from the count. This would, of course, have little effect on the overall evaluation of the prevalence of the effect.* [10] *Data are from a continuous recognition study with seven lags. Three age groups participated. The data were collapsed over lags and age groups. The data were read from their Figure 1.* [11] *Data are from experiments in which subjects estimated the frequency of presentation of words. Rao derived hits and false alarms from those data. The 0 versus 1 estimates are used here. Since the test used was not a standard recognition test, the inclusion of these data might be questioned. An analysis of the differences between the two tasks is found in Proctor (1977).* [12] *Data are from two experiments that used a five-alternative multiple-choice test. The investigators correlated word frequency with the choice of the target words and with the choice of the distractor words. The correlation was negative in the first case and positive in the second case.*

two middle distributions determines the relative order of the middle terms in the inequalities.

If the picture in Panel 2 of Figure 1 holds and the requirement of strict equality for the old distributions is relaxed, then the near equality of LO and HO will be satisfied by inequalities (2) and (4). If the picture in Panel 3 holds and again the requirement of strict equality for the new distributions is relaxed, then the near equality of HN and LN will generate choices that satisfy inequalities (1) and (3). Therefore, of a set of four inequalities that are possible when there is a word-frequency effect, only two show the mirror effect. If either inequality (1) or (2) holds for a set of data, it will be considered a positive result, with a probability of .50.

Ordinarily, multiple-choice data such as those of Schwartz and Rouse (1961) are not reported with sufficient information to place the underlying new and old distributions. In that study, however, the investigators used correlations to analyze responses to both distractors and targets. The details are given in Footnote 11 of Table 2. The probability of obtaining the pattern they obtained, one that conforms to the mirror effect, surely occurs by chance with a probability lower than .5. The two positive cases the study generates will, however, be assigned a probability of .5 to keep the overall analysis simple.

The probability of 23 of 24 cases being positive when the probability of a positive result is .5 can be tested using the binomial distribution with $p = .5$ and $n = 24$. The probability of 23 or more positive results by chance is

$$P = \binom{24}{23}p^{23}q^{1} + \binom{24}{24}p^{24}q^{0} = .0000015.$$

That is a one-tailed probability. The two-tailed probability is .000003.

The mirror effect clearly holds for the case of normative word frequency. It is evident, moreover, whatever procedure is used—forced choice, yes/no, multiple choice, or rating scale.

The question about its presence in between-subjects data was raised earlier. The study by McCormack and Swenson (1972) also presents between-subjects data. Those data show a mirror effect ($N = 100$) that parallels the within-subjects data.

GENERALITY OF THE MIRROR EFFECT

The next question concerns the generality of the mirror effect. Is it restricted to the word-frequency variable, or does it hold for other variables? To determine generality, the mirror effect was examined in the studies of three classes of stimulus variables—word concreteness, pictures versus words, and meaningfulness—and a set of miscellaneous variables. The three named variables have been shown to affect recognition. If the mirror effect holds across all stimulus variables, then a basic characteristic of recognition memory is established.

Concreteness or Imageability

One classic stimulus effect on recognition memory is word concreteness. Studies of the variable have been grouped together here with studies of imageability measures. These two measures are, of course, highly correlated: Gilhooly and Logie (1980), $r = .78$; Paivio, Yuille, and Madigan (1968), $r = .83$; Rubin (1980), $r = .88$. One experiment that used imageability, Peterson and McGee (1974, Experiment 3), was excluded from the listing because it found no imageability effect. A study by Kuiper and Paivio (1977) was also excluded. They used two different types of distractors—synonymous and new. The new distractors gave the mirror effect. The synonymous distractors did not. A clear decision on how to count this case was not possible.

Of the nine studies listed, eight show the mirror effect (see Table 3). Applying the binomial again, the probability of that number or more of positive findings is .020. Here, and in the sections that follow, the one-tailed probability will be used, since the direction of the effect is specified.

There is also a between-subjects study of imageability effects (Paivio & Csapo, 1969). The overall data show the mirror effect.

Meaningfulness or Familiarity

Another variable that has an effect on recognition memory is the meaningfulness of the items presented. A broad definition of meaningfulness was adopted in order to obtain a sufficient sample of studies. Included were studies involving comparisons between words and nonsense syllables, between various approximations to English, between familiar and unfamiliar words, and between personally familiar and unfamiliar first names.

Among the 13 experiments listed in Table 4, 2 had three rather than two levels of the experimental variable. McNulty (1965) used letter strings at three levels of approximation to English. DaPolito, Barker, and Wiant (1972) used words, CVCs and CCCs. In the McNulty (1965) study, the results for the false alarms exactly mirror the results for the hits. The probability of the chance occurrence of that order is one in six, or .167. In DaPolito et al., the hits show the order words > CVC > CCC, but the false alarms show the order CVC < words < CCC. The probability of obtaining a matching of orders this close is .50. That probability is obtained by listing all six possible orders of the false alarms and determining how close each is to the order of the hits. When this is done, three of the six are as close as or closer than the order given above. The data will be considered here as a positive case.

Data from a study by Hunt and Elliot (1980, Experiment 6) were excluded. Their meaningfulness measure, number of elicited associates, had no effect on recognition. The hits for the low- and high-meaningfulness words and also the false-alarm rates are almost identical. That study, however, is included in Table 6 for another variable, orthographic distinctiveness.

Table 3
Occurrence of the Mirror Effect (M = +) in Studies of Concreteness and Imageability

Study	Method	N	Data Sets	Positive Sets	M
Bruning, Holzbauer, & Kimberlin (1975)[1]	YN	60	3	2	+
Groninger (1976)	YN	67	1	1	+
Jones & Winograd (1975)	YN	24	1	1	+
Moeser (1974)[2]					
Exp. 2	YN	10	1	1	+
Exp. 3	YN	10	1	1	+
Morris & Reid (1974)[3]					
Exp. 1	YN	50	2	2	+
Exp. 2	YN	52	2	2	+
Peterson & McGee (1974)[4],Exp. 1	YN	64	4	1	−
Winograd, Cohen, & Barresi (1976)[5], Exp. 1	YN	17	2	2	+

[1] Data are collapsed across three age groups. The recognition test was given after recall tests on the same items. Subjects were given imagery instructions. [2] Moeser compared concrete and abstract sentences. The distractors were sentences that had been changed in wording or in meaning. Both sets of distractors yield the mirror pattern. [3] Data are collapsed across two experimental conditions (imagery instruction and control). High- and low-imagery words were used, matched for frequency. [4] Data on imageability from Experiment 3 of Peterson and McGee were not included because no significant imageability effect was found. If those data were included, they would give another positive result. [5] Data are from an experiment with bilingual subjects and words in two languages. Several different groupings of data and several types of d's were computed for one of the groups. All of the results give the mirror pattern. One subset of data in their Table 2 presents equal hit rates for concrete and abstract words. Calculations based on Table 1, however, indicate that this equality is probably due to rounding. In any case, any combination of results across same- and other-language conditions and the two groups gives the mirror pattern.

Table 4
Occurrence of the Mirror Effect (M = +) in Studies of Meaningfulness or Familiarity

Study	Method	N	Data Sets	Positive Sets	M
Brown, Lewis, & Monk (1977)[1], Exp. 3a,3b					
personal familiarity of first names	RS	7	4	4	+
DaPolito, Barker, & Wiant (1972)[2]					
words, CVCs, CCCs	YN	60	2	1	+
Gordon & Clark (1974)[3]					
words vs. nonwords	YN	44	8	7	+
Groninger (1976)					
words vs. nonwords	YN	67	1	1	−
McNulty (1965)[4]					
3 levels of approximation to English, 8-letter strings	YN	36	2	2	+
Mandler, Goodman, & Wilkes-Gibbs (1982)					
known vs. unknown words					
Exp. 1	YN	32	2	0	−
Exp. 2	YN	48	2	0	−
Rao (1983)[5], Exp. 2					
words vs. nonwords	YN	36	1	0	−
Rao & Proctor (1984)[6]					
known vs. unknown words					
Exp. 1	YN	128	3	1	+
Exp. 2	YN	64	2	2	+
Schulman (1976)					
familiarity of rare words, Exp. 2	FC	29	1	1	+
familiarity of common names, Exp. 3	FC	27	1	1	+
Seamon & Murray (1976)					
rated meaningfulness of words, Exp. 2	RS	30	3	2	+

[1] Data are collapsed over two subsidiary experiments and two frequency levels. Only 7 of 42 subjects, however, furnished items of both high and low personal familiarity. The data on responses to the old items are not reported for Experiments 1 and 2. [2] Data are combined across two presentation rates, 1 and .1 sec. The .1-sec results do not give the mirror effect. The data have also been combined over a variety of context conditions, five for old items and four for new items. [3] Hits and false alarms are derived from reported d's and likelihood ratios for four groups after one and two study trials. Data are within subjects but between lists. See Footnote 9, Table 2. [4] Data are combined for two groups with different distractors in the test. [5] See Footnote 11, Table 2. [6] There were two groups of known words, high and low frequency. The unknown words show the mirror effect with respect to each.

The 13 studies are listed in Table 4. Nine of the experiments showed the mirror effect. If the McNulty (1965) positive result is assigned the chance probability of .5, then it is possible to use the binomial again to evaluate the set of results. The probability that 9 or more of 13 results would be positive is .133 by chance.

A study by Martin and Melton (1970) used CCCs and CVCs at three levels of meaningfulness in a between-subjects design. The data ($N = 120$) show an exact mirror effect.

Pictures Versus Words

Seven studies, giving eight experiments, were found that compared recognition for pictures and words and that satisfied the criteria. The studies are listed in Table 5. Six of the eight experiments show the mirror pattern. The probability of six or more positive results by chance is, using the binomial again, .145.

Two experiments were excluded from the survey. Both, if included, would give negative results. In one, by Snodgrass and McClure (1975), both targets and distractors were studied by the subjects before the first trial. The data from such an experiment are really second-trial data. Since, with the mirror effect, study reverses the position of the new distributions, it is impossible to develop a clear statement about where the distributions should stand on later trials. In the other study, Experiment 1 of Juola, Taylor, and Young (1974), the subjects first studied a word list and then were tested on pictures corresponding to the words or on the words themselves. For our purposes, a simple comparison of pictures versus words is not possible.

In addition to the seven studies listed, there are four studies in which a between-subjects design was used: Jenkins, Neale, and Deno (1967), Paivio and Csapo (1969), and Park, Puglisi, and Sovacool (1983) all used a yes/no procedure; Rowe and Rogers (1975) used a rating-scale procedure. All four show a mirror effect.

Miscellaneous Other Variables

There are a number of studies that fall outside the categories considered up to this point. They are summarized in Table 6, with the experimental variable indicated.

Of 26 relevant experiments, 17 show the mirror effect. The probability of chance occurrence of 17 or more positive results is .084.

Two experiments were excluded from the listing because they did not meet the requirement of showing statistical significance on the variable of interest. These were Experiments 1 and 2 of Glanzer and Ehrenreich (1979). If they had been included, they would have furnished two positive cases. Also excluded were several results on sex or race of photographed faces (Cross, Cross, & Daly, 1971; Going & Read, 1974; Malpass & Kravitz, 1969). The studies involved also varied the sex or race of the subjects. It is not clear whether the data should be partitioned according to the sex or race of the faces or the correspondence of each of those to the subjects' sex or race (e.g., own sex vs. other sex). If both partitions are considered, there are eight sets of findings. Five of the eight show the mirror effect.

Of the 9 negative results in Table 6, 4 occur in studies comparing recognition for normal, as opposed to transformed, text—the Kolers (1975) experiments and the Graf (1982) experiment. There are two possible reasons for the departure of the data obtained in this type of experiment from the mirror effect. One that will not

Table 5
Occurrence of the Mirror Effect ($M = +$) in Studies of Pictures Versus Words

Study	Method	N	Data Sets	Positive Sets	M
Bloom (1971)	YN	60	2	1	+
Gehring, Toglia, & Kimble (1976)[1], Exp. 2	YN	100	3	1	–
Juola, Taylor, & Young (1974), Exp. 2	YN	20	1	0	–
Rissenberg & Glanzer (1984)[2]	YN	44	3	3	+
Snodgrass & Burns (1978)[3]	RS	8	6	6	+
Snodgrass, Volvovitz, & Walfish (1972)[4]	YN	16	6	5	+
Snodgrass, Wasser, Finkelstein, & Goldberg (1974)[5]					
Exp. 1	YN	40	1	1	+
Exp. 3	FC	99	1	1	+

[1] *Data are collapsed across three delay intervals. The mirror pattern is present at immediate testing but not after delay (1 and 3 months).* [2] *Data are collapsed over three groups—young; old normal; and old memory-impaired. All three groups show the mirror effect.* [3] *Ratings were obtained for the items over six successive sessions.* [4] *Data are collapsed over two groups, with each group tested under three different list conditions. Group A had pictures versus words. Group B had pictures and words together versus words alone. Since the addition of words to the pictures did not have any effect, they are included here as a simple picture condition. Omission of this group would not change the conclusion recorded in the table.* [5] *In Experiment 1, hits were obtained from their $P\bar{W}$ and $\bar{P}W$ conditions, and false alarms, from their $\bar{P}\bar{W}$ condition. In Experiment 3, the forced-choice pattern was obtained from the P-P and W-W comparisons and the independent P-W and W-P comparisons.*

Table 6

Occurrence of the Mirror Effect (M = +) in Studies of Miscellaneous Other Variables

Study	Method	N	Data Sets	Positive Sets	M
Durso & O'Sullivan (1983)					
common vs. proper nouns, Exp. 2	YN	36	1	0	–
Fisher & Craik (1980)[1]					
sentence complexity					
Exp. 1	YN	32	2	2	+
Exp. 2	YN	32	2	1	+
Glanzer & Ehrenreich (1979)					
truth of statements, Exp. 3	YN	24	2	1	+
Going & Read (1974)					
uniqueness of faces	YN	80	4	4	+
Graf (1982)[2]					
normal vs. transformed text	YN	32	2	1	–
Groninger (1976)					
emotionality of words	YN	67	1	1	–
Hockley (1982)[3]					
nouns vs. nonnouns					
Exp. 2	YN	4	1	1	+
Exp. 3	YN	4	1	1	+
Hunt & Elliott (1980)					
orthographic distinctiveness, Exp. 6	YN	29	2	0	–
Kolers (1974)[4]					
normal vs. transformed text	YN	12	1	1	+
Kolers (1975)					
normal vs. transformed text					
Exp. 1	YN	12	1	0	–
Exp. 2	YN	8	1	0	–
Exp. 3	YN	12	1	0	–
Kolers & Ostry (1974)					
normal vs. transformed text	YN	42	7	5	+
Light, Kayra-Stuart, & Hollander (1979)[5]					
typicality of faces					
Exp. 1	RS	41	1	1	+
Exp. 2	YN	64	2	1	+
Exp. 3	YN	60	2	1	+
Exp. 4	YN	72	3	2	+
Ortony, Turner, & Antos (1983)					
pleasantness of sentences	YN	12	4	1	–
Peterson & McGee (1974)					
number of dictionary meanings, Exp. 3	YN	192	8	3	+
Rao (1983)[6]					
pronounceability of nonwords, Exp. 3	YN	32	1	0	–
Rubin (1983)[7]					
associative frequency of words, Exp. 3	MC	75	1	1	+
Wiseman & Neisser (1974)					
ambiguous figures: faces vs. nonfaces					
Exp. 1	YN	15	1	1	+
Exp. 2	YN	21	1	1	+
Zechmeister (1972)					
orthographic distinctiveness	FC	64	1	1	+

[1] *The Experiment 1 data show an interaction of sentence complexity with test conditions. Both test conditions, however, have a mirror effect. In Experiment 2, sentence complexity also interacts with another condition. Only one of two sets of data shows the mirror effect. The overall means, however, show the effect.*

[2] *The data show an interaction of transformation with sentence type. Meaningful sentences show the mirror effect. Anomalous sentences do not.*

[3] *False alarms are computed from presented d's and hit rates.*

[4] *False alarms are computed from presented d' (new) and hit rates. This uses the "old, same form" response for both hits and false alarms. This class of responses was used for the other Kolers papers listed below.*

[5] *In Experiment 1, typicality ratings were correlated with familiarity ratings of old items (positive) and also with familiarity ratings of new items (negative).*

[6] *See Footnote 11, Table 2.*

[7] *Rubin analyzed and reported the differential response to the distractor alternatives so that the equivalent of false-alarm-rate information is obtained. Associative frequency is the frequency with which a word is elicited as an associate. Normative word frequency is controlled for.*

be argued strongly is that the procedure used in the Kolers experiments is more complicated than the usual recognition procedure. The subject gives one of three different responses on the test: old, same form; old, different form; or new. The more complex discrimination required may interact with the factors that produce the pattern of responses. This argument is, however, weakened by the fact that such a three-response test in the study by Winograd, Cohen, and Barresi (1976) on concreteness did not remove the effect.

A more important factor may be changes in processing that occur between study and test. As the subjects become practiced in carrying out the reading of transformed text, they may process the items differently. The new and old items may not belong to the same class in the sense that new and old high-frequency items belong to the same class. This argument would not, however, apply to Kolers's (1975) Experiments 1 and 2. At this point, therefore, there is no basis for handling the transformed-text studies differently. They are, for the present, included with the other studies.

OVERVIEW

At this point, five sets of data have been examined. The type of variable and the probability of the mirror effect by chance are as follows: word frequency = .000003; concreteness or imageability = .020; meaningfulness or familiarity = .133; pictures versus words = .145; and miscellaneous = .084.

To check on the issue of the generality of the mirror effect, the probabilities of the four cases other than word frequency should be combined. Using Edgington's (1972) additive method,

$$\text{combined p} = \frac{(\Sigma p)^n}{n!} = \frac{(.382)^4}{24} = .0009.$$

There could of course be objection to the groupings of the cases. It is possible, then, to take all the relevant results as one group, excluding again the word-frequency results—the 9 in Table 3, 13 in Table 4, 8 in Table 5, and 26 miscellaneous results in Table 6. This gives a total of 56 results, of which 40 are positive. The probability of the chance occurrence of that frequency or higher, using the binomial evaluation, is .0009. This excludes the statistically significant results for the mirror effect with word frequency, since the interest here is in the generality of the effect across other variables.

The mirror effect therefore characterizes not only recognition when word frequency is a variable, but also recognition in general. Any theory of recognition memory must therefore cope with the effect.

Theory

One explanation of the mirror effect might be that the effect arises from the subjects' using two different criteria. For example, it might be possible in Panel 2 of Figure 1 to place separate criteria for A and B items so as to produce the appearance of a mirror effect. The problem with this approach, aside from its complexity, is that it cannot account for the fact that the mirror effect is found not only for yes/no data, but also for two-alternative forced-choice and multiple-choice data and rating-scale data. The Brown et al. (1977) proposal discussed next does, however, make use of the criterion-shift idea in combination with other mechanisms.

The main purpose of this paper has been to present evidence for the mirror effect. The next question that arises concerns its theoretical implications. One strong implication is negative. Any theory that assumes a simple linear transformation of the underlying variables is ruled out.

Two theories concerned with the effect will be outlined now. The first, which we will label a two-stage, cognitive theory, was presented by Brown et al. (1977). They note that Groninger (1976) anticipated elements of the theory.

The two-stage theory assumes the following process during a recognition test, such as a rating scale. In the first stage, the subject examines a test item and decides that he has or has not seen it before. In the case of hits, high-memorability items will, on the average, get a strong positive response. Low-memorability items will, on the average, get a weak positive response. If the subject decides "no" on an item, the second stage is entered. The subject evaluates the item's memorability. A high-memorability evaluation leads to a strong negative response (as though a subject said to him- or herself, "This is highly memorable. If I had seen it, I surely would remember it. Therefore, I am sure that it is a new item."). On the basis of the same type of covert reasoning, a low-memorability evaluation leads to a weak negative response. The theory therefore predicts that misses for high-memorability items will give rise to strong negative responses and that misses for low-memorability items will give rise to weak negative responses. The two-stage theory also implies that the mirror effect will hold only for subjects who either implicitly or explicitly know which variables affect memorability of an item. This is not implied by the next theory.

Another theory concerned with the effect we will label an excitation, multiple-observation theory. Its basic assumptions are (1) that different classes of stimuli excite the examination of different numbers of features of the stimuli and (2) that the subject, in evaluating a stimulus on a recognition test, responds both to the number of marked features and the total number of features examined. The efficiency of recognition in this model is determined by the number of features of a stimulus that the subject examines. If many features have been examined on the study trial, then the subject has stored a large amount of information to recognize items of this class as old on the test trial. If equally many features are examined on the test trial, that greater amount of information will be reflected in the subject's response to these old items. Hit rates for this class will be high.

New items appear for the first time on the test. New items of the class that evokes examination of a large number of features will display to the subject a large number of unmarked features. The subject's decision about new items, based on a larger amount of information, should again be more efficient. False-alarm rates for this class will be low.

This theoretical approach, using number of stimulus features, is closely related to theories of the effect of multiple observations on recognition performance. Examples of the theoretical structures possible may be found, therefore, in the signal detection literature (Green & Swets, 1974). A simple example of a possible theory will clarify the approach. Assume the following:

(1) There are two or more classes of stimuli—A, B, C, etc.

(2) Each stimulus is composed of a set of features. (No assertion is made that any stimulus has more features than any other stimulus.)

(3) Before being examined, the features of each stimulus are already marked (with a familiarity marker), with probability p_n. This is the source of noise in the system.

(4) Stimuli of class A, when presented, elicit the examination of more features than stimuli of class B. The number of features examined is indicated by n_a and n_b, with $n_a > n_b$.

(5) On a study trial, the subject marks each of the features examined with a fixed probability p_o. (No assertion is made that one class is more efficiently marked than the other.)

(6) On a test trial, the subject examines the same set of features for old stimuli as on the study trial. (Some variability in the set of features selected may be considered in a more fully developed model.)

(7) On a test trial, the subject evaluates the information obtained from the observation, taking account of the likelihood ratio implied by that information. (The subject evaluates the likelihood of obtaining x marked items out of y examined items given the information concerning the underlying distributions.)

An example will help convey the way in which this preliminary theory works. Assume that the numbers of features examined are $n_a = 6$ for class A and $n_b = 2$ for class B. Assume that for both classes, $p_o = .80$ and $p_n = .10$. For those parameters, the underlying distributions of scores would be the binomials shown in Figure 2. The two distributions have been aligned according to their likelihood ratios, λ, the basis for decision assumed in the model. It is necessary to consider these ratios or some equivalent.[2] The distributions cannot be aligned on the basis of absolute number of marked features. Clearly, two marked features out of two examined should mean something different from two marked features out of six examined.

In a rating-scale test, the ratings would be determined by the likelihood ratio for each item tested. The ratings would therefore show the mirror effect.

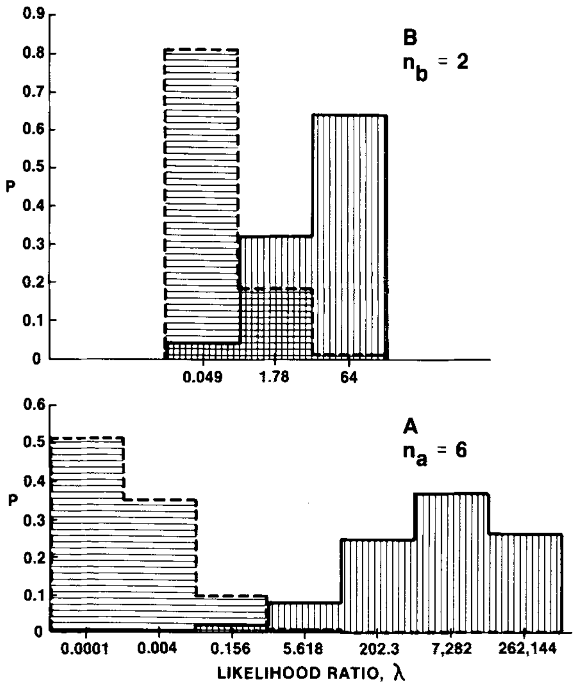

Figure 2. Distributions with feature sets of 6 and 2, $P_o = .80$, $P_n = .10$. Likelihood ratios, λ, are indicated on the abscissas.

In a yes/no test, a single criterion would be imposed on the two distributions, say, $\lambda \geq 1$, as the basis for deciding whether a stimulus is old or new. That criterion placement would give the following data.

For class B, with two features examined, the hit rate = .96 and the false-alarm rate = .19. For class A, with six features examined, the hit rate = .98 and the false-alarm rate = .02.

Note that the only difference between the two classes of stimuli assumed here is the number of features examined. The parameters p_o and p_n are the same for both classes. This was done deliberately in order to show that the mirror effect could be produced simply, on the basis of a single factor, the number of features examined.

The two theories outlined above differ in several ways. The two-stage theory derives the mirror effect solely from the subject's performance on the test. The multiple-observation theory derives the effect from processing both during study and test. The two-stage theory requires that the subject have knowledge of the memorability of different classes of stimuli. The multiple-observation theory does not require that knowledge. The two-stage theory handles the mirror effect only for rating-scale data. The multiple-observation theory handles the effect for all types of recognition data.

Both of the theories outlined above fall within the class of signal detection theories. The mirror effect can of course be considered within the context of other general theoretical frameworks. One is high-threshold theory. The mirror effect is a necessary consequence of some versions of high-threshold theory. For example, the version that yields the equations

$$P(\text{hit}) = p + (1\text{-}p)(g)$$

$$P(\text{FA}) = (1\text{-}p)(g)$$

produces a mirror effect. With g (the probability of guessing "yes") constant, the two sets of stimuli, one with a high p, the other with a low p, will produce a high hit rate and a low false-alarm rate for the more discriminable stimuli and the reverse for the less discriminable. High-threshold theory has problems, however, in coping with recognition data. Those problems have been pointed out in detail (Murdock, 1974).

All of the data reviewed in this paper involved stimulus variables, such as word frequency or meaningfulness. There are other classes of variables that affect recognition memory but that were not included. One is the class of experimenter-imposed variables, such as repetition or spacing of repetitions. Data on this class of variables usually do not meet the sufficiency criterion. For example, only one false-alarm rate seems possible, although two or more experimental conditions are used. Another class is that of subject-state variables, such as age. The cross-sectional data available for that variable do not meet the within-subjects criterion. In the future, it will be possible to consider whether the equivalent of the mirror effect can be demonstrated for these other classes.

In summary, a strong regularity of recognition memory has been demonstrated. Some of the theoretical implications of that regularity have also been indicated.

REFERENCES

BALOTA, D. A., & NEELY, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 576-587.

BLOOM, S. W. (1971). *Recognition memory for pictures and their word labels*. Unpublished doctoral dissertation, University of Rochester, Rochester, NY.

BOWLES, N. L., & GLANZER, M.(1983). An analysis of interference in recognition memory. *Memory & Cognition*, **11**, 307-315.

BOWLES, N. L. & POON, L. W. (1982). An analysis of the effect of aging on recognition memory. *Journal of Gerontology*, **37**, 212-219.

BROWN, J., LEWIS, V. J. & MONK, A. F. (1977). Memorability, word frequency and negative recognition. *Quarterly Journal of Experimental Psychology*, **29**, 461-473.

BRUNING, R. H., HOLZBAUER, I., & KIMBERLIN, C. (1975). Age, word imagery, and delay interval: Effects on short-term and long-term retention. *Journal of Gerontology*, **30**, 312-318.

CLARK, E. O. (1981). *Semantic and episodic memory impairment in old age*. Unpublished doctoral dissertation, New York University.

CROSS, J. F., CROSS, J., & DALY, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics*, **10**, 393-396.

DAPOLITO, F., BARKER, D., WIANT, J. (1972). The effects of contextual changes on component recognition. *American Journal of Psychology*, **85**, 431-440.

DORFMAN, D., & GLANZER, M. (1984). *List composition and speed-accuracy effects on lexical decision*. Unpublished manuscript, New York University.

DORFMAN, D., GLANZER, M., & KAUFMAN, J.(1984). *Aging effects on recognition memory: Encoding and speed-accuracy trade-off controlled*. Unpublished manuscript, New York University.

DURSO, F. T., & O'SULLIVAN, C. S. (1983)). Naming and remembering common nouns and pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 497-510.

EDGINGTON, E. S. (1972). An additive method for combining probability values from independent experiments. *Journal of Psychology*, **80**, 351-365.

FISHER, R. P., & CRAIK, F. I .M.(1980). The effects of elaboration on recognition memory. *Memory & Cognition*, **8**, 400-404.

GEHRING, R. E., TOGLIA, M. P. & KIMBLE, G. A. (1976). Recognition memory for words and pictures at short and long retention intervals. *Memory & Cognition*, **4**, 256-260.

GILHOOLY, K. J., & LOGIE, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, **12**, 395-427.

GLANZER, M., & BOWLES, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, **2**, 21-31.

GLANZER, M., & EHRENREICH, S. L. (1979). Memory for truths and falsehoods. *Memory & Cognition*, **7**, 13-18.

GOING, M., & READ, J. D. (1974). Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Perceptual and Motor Skills*, **39**, 109-110.

GORDON, S. K., & CLARK, W. C. (1974). Adult age differences in word and nonsense syllable recognition memory and response criterion. *Journal of Gerontology*, **29**, 659-665.

GORMAN, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, **61**, 23-29.

GRAF, P. (1982). The memorial consequences of generation and transformation. *Journal of Verbal Learning and Verbal Behavior*, **21**, 539-548.

GREEN, D. M., & SWETS, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Krieger.

GRONINGER, L. D. (1976). Predicting recognition during storage: The capacity of the memory system to evaluate itself. *Bulletin of the Psychonomic Society*, **7**, 425-428.

HOCKLEY, W. E. (1982). Retrieval processes in continuous recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **8**, 497-512.

HUNT, R. R., & ELLIOTT, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, **109**, 49-74.

JENKINS, J. R., NEALE, D. C., & DENO, S. L. (1967). Differential memory for picture and word stimuli. *Journal of Educational Psychology*, **58**, 303-307.

JONES, S., & WINOGRAD, E. (1975). Word imagery in recognition memory. *Bulletin of the Psychonomic Society*, **6**, 632-634.

JUOLA, J. F., TAYLOR, G. A., & YOUNG, M. E. (1974). Stimulus encoding and decision processes in recognition memory. *Journal of Experimental Psychology*, **102**, 1108-1115.

KINSBOURNE, M., & GEORGE, J. (1974). The mechanism of the word-frequency effect on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, **13**, 63-69.

KOLERS, P. A. (1974). Two kinds of recognition. *Canadian Journal of Psychology*, **28**, 51-61.

KOLERS, P. A. (1975). Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory*, **1**, 689-701.

KOLERS, P. A., & OSTRY, D. J. (1974). Time course of loss of information regarding pattern analyzing operations. *Journal of Verbal Learning and Verbal Behavior*, **13**, 599-612.

KUIPER, N. A., & PAIVIO, A. (1977). Incidental recognition memory for concrete and abstract sentences equated for comprehensibility. *Bulletin of the Psychonomic Society*, **9**, 247-249.

LEE, A. T., TZENG, O. J. L., GARRO, L. C., & HUNG, D. L. (1978). Sensory modality and the word-frequency effect. *Memory & Cognition*, **6**, 306-311.

LIGHT, L. L., KAYRA-STUART, F., & HOLLANDER, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, **5**, 212-228.

MALPASS, R. S., & KRAVITZ, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, **13**, 330-334.

MANDLER, G., GOODMAN, G. O., & WILKES-GIBBS, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition*, **10**, 33-42.

Martin, E., & Melton, A. W. (1970). Meaningfulness and trigram recognition. *Journal of Verbal Learning and Verbal Behavior*, **9**, 126-135.

McCormack, P. D., & Swenson, A. L. (1972). Recognition memory for common and rare words. *Journal of Experimental Psychology*, **95**, 72-77.

McNulty, J. A. (1965). An analysis of recall and recognition processes in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, **4**, 430-436.

Moeser, S. D. (1974). Memory for meaning and wording in concrete and abstract sentences. *Journal of Verbal Learning and Verbal Behavior*, **13**, 682-697.

Morris, P. E. (1978). Frequency and imagery in word recognition: Further evidence for an attribute model. *British Journal of Psychology*, 69, 69-75.

Morris, P. E., & Reid, R. L. (1974). Imagery and recognition. *British Journal of Psychology*, **65**, 7-12.

Murdock, B. B., Jr. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.

Ortony, A., Turner, T. J., & Antos, S. J. (1983). A puzzle about affect and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 725-729.

Paivio, A., & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental Psychology*, **80**, 279-285.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, **76** (1, Pt. 2).

Park, D. C., Puglisi, J. T., & Sovacool, M. (1983). Memory for pictures, words, and spatial location in older adults: Evidence for pictorial superiority. *Journal of Gerontology*, **38**, 582-588.

Peterson, M. J., & McGee, S. H. (1974). Effects of imagery instructions, imagery ratings, and number of dictionary meanings upon recognition and recall. *Journal of Experimental Psychology*,**102**, 1007-1014.

Poon, L. W., & Fozard, J. L. (1980). Age and word frequency effects in continuous recognition memory. *Journal of Gerontology*, **35**, 77-86.

Proctor, R. W. (1977). The relationship of frequency judgments to recognition: Facilitation of recognition and comparison to recognition-confidence judgments. *Journal of Experimental Psychology: Human Learning and Memory*, **3**, 679-689.

Rao, K. V. (1983). Word frequency effect in situational frequency estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 73-81.

Rao, K. V., & Proctor, R. W. (1984). Study-phase processing and the word frequency effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 386-394.

Rissenberg, M., & Glanzer, M. (1984). [Picture superiority in recognition]. Unpublished data.

Rowe, E. J., & Rogers, T. B. (1975). Effects of concurrent auditory shadowing on free recall and recognition of pictures and words. *Journal of Experimental Psychology: Human Learning and Memory*, **1**, 415-422.

Rubin, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior*, **19**, 736-755.

Rubin, D. C. (1983). Associative asymmetry, availability, and retrieval. *Memory & Cognition*, **11**, 83-92.

Schulman, A. I. (1976). Memory for rare words previously rated for familiarity. *Journal of Experimental Psychology: Human Learning and Memory*, **2**, 301-307.

Schwartz, F., & Rouse, R. O. (1961). The activation and recovery of associations. *Psychological Issues*, **3**(1, Whole No. 9).

Seamon, J. G., & Murray, P. (1976). Depth of processing in recall and recognition memory: Differential effects of stimulus meaningfulness and serial position. *Journal of Experimental Psychology: Human Learning and Memory*, **2**, 680-687.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, **6**, 156-163.

Snodgrass, J. G., & Burns, P. M. (1978). The effect of repeated tests on recognition memory for pictures and words. *Bulletin of the Psychonomic Society*, **11**, 263-266.

Snodgrass, J. G., & McClure, P. (1975). Storage and retrieval properties of dual codes for pictures and words in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, **1**, 521-529.

Snodgrass, J. G., Volvovitz, R., & Walfish, E. R. (1972). Recognition memory for words, pictures, and words + pictures. *Psychonomic Science*, **27**, 345-347.

Snodgrass, J. G., Wasser, B., Finkelstein, M., & Goldberg, L. B. (1974). On the fate of visual and verbal memory codes for pictures and words: Evidence for a dual coding mechanism in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, **13**, 27-37.

Wilhite, S. C. (1981). Word-frequency cueing effects: Recognition and encoding interference factors. *American Journal of Psychology*, **94**, 323-337.

Winograd, E., Cohen, C., & Barresi, J. (1976). Memory for concrete and abstract words in bilingual speakers. *Memory & Cognition*, **4**, 323-329.

Wiseman, S., & Neisser, U. (1974). Perceptual organization as a determinant of visual recognition memory. *American Journal of Psychology*, **87**, 675-681.

Zechmeister, E. B. (1972). Orthographic distinctiveness as a variable in word recognition. *American Journal of Psychology*, **85**, 425-430.

NOTES

1. This assumes no particular correlation between the old and new familiarity distributions across item types. If there were a positive correlation, the probability would be less than .50.

2. Another possibility would be for the subjects to respond on the basis of the proportion of marked elements.

(Manuscript received May 9, 1984;
revision accepted for publication November 14, 1984.)