

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/13677155>

Response distribution as an explanation of the mirror effect

Article in *Journal of Experimental Psychology Learning Memory and Cognition* · June 1998

DOI: 10.1037/0278-7393.24.3.633 · Source: PubMed

CITATIONS

16

READS

15

3 authors, including:



Murray Glanzer

New York University

75 PUBLICATIONS 4,248 CITATIONS

[SEE PROFILE](#)



John kofi Adams

Accra Institute of Technology

8 PUBLICATIONS 1,073 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



free recall [View project](#)

All content following this page was uploaded by **Murray Glanzer** on 26 March 2014.

The user has requested enhancement of the downloaded file.

Response Distribution as an Explanation of the Mirror Effect

Murray Glanzer and Kisok Kim
New York University

John K. Adams
Montclair State University

Response distribution has recently been proposed as an explanation of the mirror effect in recognition memory. According to the proposal, participants presented with distinctive sets of items (e.g., low- and high-frequency words) vary their responses to give an equal number of positive responses (e.g., the sum of hits and false alarms) to each set. Four experiments tested this proposal. Two experiments showed that the mirror effect is present in the absence of distinctive sets of items. Two experiments showed that the mirror effect is present in the absence of response equalization. Wherever the response distribution hypothesis can be tested, it fails.

The mirror effect (Glanzer, Adams, Iverson, & Kim, 1993) refers to an order found in recognition test responses when two classes of items have been studied (e.g., words and nonwords or low- and high-frequency words). In the case of a yes–no recognition test with low- and high-frequency words, the mirror effect is seen in the following order:

$$P(LN) < P(HN) < P(HO) < P(LO), \quad (1)$$

where P is the proportion of “yes” responses, L is low-frequency words, H is high-frequency words, N is new, and O is old. The first two terms are, therefore, proportions of false alarms and the last two are proportions of hits for the two item classes, L and H. The order shows an interaction of word frequency and old versus new: For old words, low frequency gives more “yes” responses than high frequency, but for new words the reverse holds.

In the case of two-alternative forced choice, the mirror effect is seen in the following order for standard pairs, the test pairs usually presented in forced choice:

$$P(HO, HN) < P(LO, HN), P(HO, LN) < P(LO, LN), \quad (2)$$

where P is the proportion of choices of the first argument over the second argument. The letters L, H, O, and N indicate the same states as those defined above.

In addition to the standard pairs just referred to, the forced choices used by Glanzer et al. (1993) included null choices in which pairs of old items, LO and HO, are presented and pairs of new items, HN and LN, are presented. These pairs

do not have a correct response but help define the positions of the underlying distributions that generate the entire set of participants’ responses. When the mirror effect holds, the null choices conform to the following inequality:

$$P(HN, LN), P(LO, HO) > .50. \quad (3)$$

The order of the arguments is important, and should be noted carefully. Here, LO is preferred to HO but HN is preferred to LN.

Both the order of “yes–no” test responses in Inequality 1 and the orders of forced-choice responses in Inequalities 2 and 3 can be derived by assuming that responses are based on the underlying distributions shown in Figure 1. In that figure, the positions of the new distributions, LN and HN, mirror the positions of the old distributions, HO and LO. Hence, the name *mirror effect*.

Several explanations of the mirror effect have been offered. One that is considered later is attention-likelihood theory. That theory generates underlying distributions such as those in Figure 1. It therefore predicts the order of responses in yes–no tests and the order of responses in forced-choice tests including the null choices (Glanzer, Adams, & Iverson, 1991; Glanzer et al., 1993; Kim & Glanzer, 1993). In forced-choice tests, the theory predicts that $P(HO, LN)$ and $P(LO, HN)$, referred to as mixed choices, will be approximately equal. (See Inequality 2.) It also predicts that the null choices will be approximately equal (see Inequality 3) and will change together. For example, if experimental conditions produce a decrease in one null choice, the other will also decrease. Such a decrease in null choices accompanying a decrease in the standard choices has been referred to as *concentering*. It has been demonstrated by Glanzer et al. (1991, 1993).

Recently Greene (1996) has offered an alternative explanation of the mirror effect as seen in both yes–no and forced-choice data. That explanation does not assume underlying distributions such as those in Figure 1. Instead it derives Inequalities 1, 2, and 3 on the basis of a response distribution hypothesis. According to the hypothesis, participants attempt to equalize the number of choices (e.g., “yes” responses in a yes–no test) of each of the item classes (e.g.,

Murray Glanzer and Kisok Kim, Department of Psychology, New York University; John K. Adams, Department of Psychology, Montclair State University. Kisok Kim is now at the Department of Psychology, St. Joseph’s College.

The research reported in this article was carried out under National Science Foundation Grant SBR 940 9560. We thank Andy Hilford for his comments on this article and for his assistance in its preparation.

Correspondence concerning this article should be addressed to Murray Glanzer, Department of Psychology, New York University, 6 Washington Place, New York, New York 10003. Electronic mail may be sent to mg@xp.psych.nyu.edu.

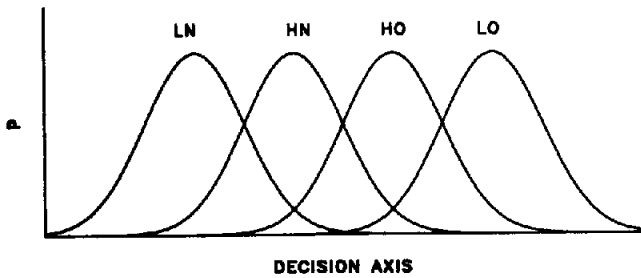


Figure 1. Arrangement of underlying old and new distributions showing the mirror effect. The figure is schematic, showing distributions with equal variance. Variance is generally not equal, according to either data or theory. p = probability; L = low frequency; H = high frequency; N = new; O = old.

low-frequency and high-frequency words). How response equalization produces the mirror pattern is given in detail shortly. First, we summarize the data that inspired the hypothesis.

Greene's (1996) Data

Greene (1996) reported a series of experiments in which the item classes used to produce the mirror effect were either words and nonwords or low-frequency and high-frequency words. In addition to using the usual recognition test in which the participant is asked whether an item was old or new, Greene also used other types of memory tasks. In one variant task, the participant was asked whether a test item was in the first or second half of the study list. The variant tasks give results similar to those of the usual recognition task and, to keep this presentation simple, are not discussed separately here.

Using the usual recognition task in which the participants discriminate new from old items, Greene (1996) showed that instructing the participants not to guess during a yes–no test disrupts the mirror effect. The data from his Experiment 8 with words and nonwords and from his Experiment 9 with high-frequency and low-frequency words are shown in Table 1. The mirror effect appears in the control condition of both experiments, when standard yes–no instructions are

Table 1
Data From Greene's (1996) Experiments 8 and 9

Condition	Mean proportion of "yes" response			
	$P(wN)$	$P(nN)$	$P(nO)$	$P(wO)$
Experiment 8				
Control	.20	.48	.58	.88
No guess	.19	.14	.27	.87
Condition	$P(LN)$	$P(HN)$	$P(HO)$	$P(LO)$
Experiment 9				
Control	.11	.25	.71	.83
No guess	.08	.09	.57	.78

Note. w = word; n = nonword; L = low-frequency word; H = high-frequency word; N = new; O = old.

given. In Experiment 8 with words and nonwords, $P(wN) < P(nN) < P(nO) < P(wO)$. Here w stands for words, n for nonwords, N for new, and O for old. In Experiment 9 with low-frequency and high-frequency words, $P(LN) < P(HN) < P(HO) < P(LO)$. The terms here are defined as in Inequality 1. When, however, instructions not to guess are given, the order of $P(wN)$ and $P(nN)$ reverses in Experiment 8 and the separation between $P(LN)$ and $P(HN)$ becomes minimal in Experiment 9. Greene also showed in Experiment 5 that in a forced-choice test with one of his variant tasks, a delay produces declines in both null pairs, part of the pattern referred to as *concentering*. This result is simply a replication with the variant task of part of the results of Glanzer et al. (1991) on *concentering*.

The Response Distribution Hypothesis

After demonstrating the effect of the no-guess instructions, Greene (1996) proposed a response distribution hypothesis to explain the mirror effect inequalities in both yes–no tests and forced-choice tests—Inequalities 1, 2, and 3—and *concentering* in forced-choice tests. He refers in the quotation that follows to words versus nonwords. The discussion applies as well to low-frequency versus high-frequency words or to any other variable that affects recognition memory. The following quotation summarizes the proposal (Greene, 1996, pp. 691–692):

First, consider yes–no tests. Participants go into a yes–no test . . . expecting roughly equal numbers of positive and negative items. They therefore try to give roughly equal numbers of positive and negative responses. To explain the mirror effect, one must only add that *when they are faced with two easily discriminable sets of items, they try to give equal numbers of positive responses to both sets* [italics added]. In the experiments here, participants could easily discriminate between words and nonwords. They assumed that a nonword was as likely to be a target as a word and therefore tried to give as many positive responses to nonwords as to words. To do this, participants choose to give positive responses as guesses more often to nonwords than words. Because more of the positive responses are guesses for nonwords than for words, there are more false alarms to nonwords than to words. The same argument applies when other stimulus comparisons are used to study the mirror effect. . . . The critical claim is merely that participants try to equate the number of positive responses they give to each class of stimulus. . . . This account can be extended in a straightforward way to performance on null trials in a forced-choice test. When participants choose between two options, both of which should get a positive response, they remember the word more often than the nonword. Thus, they choose the word more often on such pairs. To equate the number of times they pick words and nonwords, they tend to select nonwords more often than words when both options should get negative responses. This leads to the reversal of choices typically found on null trials.¹ [Footnote added]

¹ Greene referred here to the fact that in null trials for words versus nonwords, $P(wO, nO) \approx 1 - P(wN, nN)$, and for low-frequency versus high-frequency words, $P(LO, HO) \approx 1 - P(LN, HN)$. We avoided the subtraction from 1 by reversing the arguments for the new item null choice giving the following equivalent equations: $P(wO, nO) \approx P(nN, wN)$ and $P(LO, HO) \approx P(HN, LN)$.

The response distribution hypothesis asserts that in a yes-no recognition test the participants attempt to match the total number of "yes" responses (hits plus false alarms) they make to each stimulus class. Because a better recognized class generates more "yes" responses to old items, the participants lower the number of "yes" responses to new items of that class, thereby lowering their false-alarm rate. Because a less well-recognized class generates fewer "yes" responses to old items, the participants raise the number of "yes" responses to new items (thus raising their false-alarm rate) to make the total number of "yes" responses to the two classes equal. Therefore, the mirror effect, as it occurs in yes-no recognition tasks, would be produced (see Inequality 1). Note that the account places major emphasis in this case on the false-alarm rate, the response to new items.

A similar explanation is used for the mirror pattern found in forced-choice tests and, in particular, in forced null choices. The details of this explanation are presented later in Experiment 4.

To see how the hypothesis is applied, we return to the control condition data in Table 1. In Greene's (1996) Experiment 8, participants had a total of 60 words (30 old and 30 new) and 60 nonwords (30 old and 30 new) on the test. Translating the proportions for the control condition in Table 1 into the mean number of "yes" responses gives the following: for wN, 6.0; for nN, 14.4; for nO, 17.4; for wO, 26.4. The sum of wN and wO is 32.4. The sum of nN and nO is 31.8. As the hypothesis implies, the total number of "yes" responses (hits plus false alarms) is about equal in the two item classes. The same calculation applied to the control condition of Experiment 9 gives the following: for LN, 3.3; for HN, 7.5; for HO, 21.3; for LO, 24.9. The sum of LN and LO is 28.2. The sum of HO and HN is 28.8. Again, as the hypothesis implies, the number of "yes" responses is equal for the two item classes. These sums are shown in Table 2.

Greene's (1996) belief is that response equalization by the participant produces the mirror effect. We argue that the mirror effect produces the response equalization seen in Table 2. That the latter is so is shown later in Experiment 3. In that experiment, the mirror effect and response equalization were put in opposition. The mirror effect remained and response equalization disappeared.

We note also that the disruption of the mirror effect by

guessing instructions does not imply that the response distribution hypothesis is true. The disruption merely poses a general problem that any theory of the mirror effect has to deal with. We deal with it later in terms of attention-likelihood theory.

In summary, Greene (1996) has devised a procedure that disrupts the standard mirror order. He has also shown, as has been shown before (Glanzer et al., 1991), changes in null choices with forgetting. Although these two types of data are related by Greene to the response distribution hypothesis, neither offers a critical test of the hypothesis and neither is uniquely related to the hypothesis. The demonstration of changes in null choices is simply another demonstration of such changes, predicted by attention-likelihood theory. (Glanzer et al., 1991). The disruption of the mirror effect by instructions not to guess can be explained without the response distribution hypothesis. This is shown later. First, we tested the response distribution hypothesis directly. To do this, we started with a statement of the two factors that the hypothesis asserts are necessary for response distribution (and, therefore, the mirror effect).

The Two Factors Postulated by the Response Distribution Hypothesis

The response distribution hypothesis claims that the mirror effect occurs as a result of two factors, a condition presented by the experiment and a response to that condition by the participant. (See the italicized phrase in the earlier quotation.) The two factors are the following. (a) Distinct classes of items: In the quotation cited (Greene, 1996, p. 691), these are referred to as "two easily discriminable sets of items." This is the eliciting condition. (b) Equalization of responses: The participants adjust their responses (in Greene's formulation, guessing rates) differently for the distinct classes of items so that the total number of "yes" responses for one class equals that for the other.

We tested the hypothesis first in two experiments that examined what happens when the participants are not offered distinct classes of items. We then tested it in two experiments that examined directly whether participants, given two distinct classes of items, adjust their test responses to produce equal numbers of positive responses to those classes of items.

The Mirror Effect Without Distinct Classes of Items

The experiments reported in the literature that show the mirror effect usually have distinct, separated classes of items such as low- and high-frequency words with no middle-frequency words (Glanzer & Adams, 1985).² Greene's (1996) experiments had such distinct, separated classes. In his Experiment 9, there were high-frequency words with a frequency of 100 or more per million and low-frequency

Table 2
Mean Number of "Yes" Responses (Hits Plus False Alarms): Greene's (1996) Control Groups From Experiments 8 and 9

Item class	Mean number of "yes" responses
Experiment 8	
Nonwords	31.8
Words	32.4
Experiment 9	
High-frequency words	28.8
Low-frequency words	28.2

² Mirror effects with a continuous variable are, however, found in two experiments, Experiments 5 and 9 of Schwartz and Rouse (1961). The mirror effect appeared in the correlational analysis of the data. Details are given in Glanzer and Adams (1985).

words with a frequency of 5 or fewer per million. There was a frequency gap of 95. The two sets were clearly distinct. In his Experiment 8, there were words and nonwords. Again, the two sets were clearly distinct.

In the two experiments that follow, the participants do not have two distinct sets of items. If the mirror effect appears when the participants do not have distinct sets of items on which to base their assumed response equalization, the hypothesis fails. We eliminated distinct sets by using a continuous range of variables to generate a mirror effect. We imposed an arbitrary median split on this continuous range to define (for us, not the participant) two classes. There was no separation, no boundary between upper and lower values of the range. No basis was offered to the participant for distinct sets for which responses could be equalized. In the first experiment, the continuous variable was word frequency. The entire range between low and high frequency was used. In the second experiment, the continuous variable was concreteness rating.

Experiment 1: Continuous Word Frequency

In this experiment, word frequency was varied in a special way. The participants did not see distinct sets of high- and low-frequency words separated by a large gap in the frequencies, as is typical in experiments on word frequency. They saw words selected from a continuous range of frequencies. The word frequency variable was defined by the experimenter in scoring the responses by a median split in this continuous distribution of frequencies.

Method

The participants were first given a lexical-decision list (half words and half nonwords), in which they were asked to decide whether each item was a word or nonword. Then they were given a recognition test list consisting of "old" words, words that had appeared in the lexical-decision list, and new words. The participants were asked to decide whether each item was new or old. The main variable, word frequency, was continuous in the study lists. Distinct sets of high-frequency versus low-frequency words were not presented to participants. There was a secondary variable, test list composition. The proportion of old words in the test list was varied across three groups of participants. This variable was, however, irrelevant to our present concerns.

Materials. A pool of 512 nouns that ranged in natural logarithm frequency³ from 2.4 to 6.3 with a median of 3.3 was used (Kučera & Francis, 1967). The distribution of frequencies was continuous across its range, as shown in Figure 2. This set of words was divided into a high- and low-frequency set on the basis of a median split. Nonwords for the lexical-decision task were pronounceable, orthographically legal strings. The lexical-decision list and recognition test list were drawn at random and were ordered at random independently for each participant.

Procedure. The participants' lexical-decision lists consisted of 256 words and 256 nonwords. Words and nonwords were presented one by one on a computer monitor in random order, and the participants pressed a key labeled *word* if they thought the item was a word or a key labeled *nonword* otherwise. The task was self-paced. This procedure was used in all three experiments that follow. The lexical-decision task was followed by a recognition test list of 320 words. Each test word was presented one by one on a

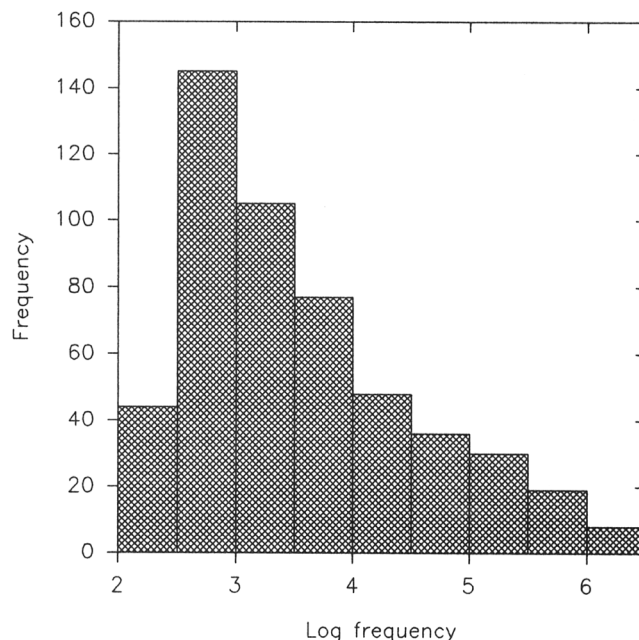


Figure 2. Distribution of natural logarithm normative word frequencies for the set of words used in Experiment 1. The abscissa refers to the frequency of words in the Kučera and Francis (1967) norms. The ordinate refers to the frequency of occurrence of each word class in the word pool of the experiment.

computer monitor in random order, and the participants pressed a key labeled *yes* if they thought the word was old, or a key labeled *no* if they thought it was new. The recognition test was also self-paced.

As noted earlier, a secondary variable, list composition, was present. The proportion of old items, words seen in the lexical-decision task, was varied across groups of participants. For one third of the participants, the test list consisted of .80 old words, for one third .50 old words, and for one third .20 old words. In all test lists, half of the words, both new and old, were high frequency, and half were low frequency. For example, in the .80 test condition, 256 words were old, 128 high frequency and 128 low frequency, and 64 words were new, 32 high frequency and 32 low frequency. Participants were not informed about the composition of their test lists.

Participants. Thirty undergraduate students from an introductory psychology course participated to fulfill a course requirement. They were assigned at random to the three experimental groups of 10 each. All participants spoke English from the age of 6 years or earlier. This description of the participants holds for Experiments 2, 3, and 4.

Results

The results are shown in Table 3. The mean proportions (bottom row of the table) showed the mirror pattern of Inequality 1: $P(LN) < P(HN) < P(HO) < P(LO)$. The overall differences between these four means were statisti-

³ Carroll (1968) has shown that the word frequency distribution is lognormal.

Table 3
Mean Proportions of "Yes" Responses for the Three Test Conditions of Experiment 1 With Word Frequency as the Variable

Proportion of old items in the test list	Mean proportions of "yes" responses			
	<i>P</i> (LN)	<i>P</i> (HN)	<i>P</i> (HO)	<i>P</i> (LO)
.80	.18	.18	.51	.56
.50	.29	.34	.55	.59
.20	.13	.19	.42	.46
<i>M</i>	.20	.23	.49	.54

Note. L = low-frequency word; H = high-frequency word; N = new; O = old.

cally significant, $F(3, 81) = 108.92, p < .001, MSE = 0.0084$.

The effect of list composition was statistically significant, $F(2, 27) = 3.42, p < .05, MSE = 0.0605$. The effect was, however, only a difference in bias for the three groups. All three groups were conservative: All tended to give more "no" than "yes" responses. The .20 group showed a very strong conservative bias—the mean proportion of "yes" responses summed across all four response categories is .30. The .50 group showed the least conservative bias, with the mean proportion of "yes" responses equal to .44. This effect did not, moreover, interact with any other variable. Most important, it did not interact with the mirror pattern of means, $F(6, 81) = 1.28, p > .25, MSE = 0.0084$, and is not relevant to the present discussion.

The pattern of inequalities is the same pattern found in 22 studies of word frequency (Glanzer & Adams, 1985, 1990). We decomposed the pattern into three orthogonal comparisons. One, of lesser interest, tested for overall effect of old versus new words: $[P(LO) + P(HO)] - [P(LN) + P(HN)]$. The other two addressed the specific differences that characterize the mirror effect: $P(HN) - P(LN)$ and $P(LO) - P(HO)$. Evaluation of the comparisons finds the following: For $[P(LO) + P(HO)] - [P(LN) + P(HN)]$, $t(27) = 11.68, p < .001, SE = 0.0256$; for $P(HN) - P(LN)$, $t(27) = 2.47, p < .05$; and for $P(LO) - P(HO)$, $t(27) = 3.22, p < .01$, with $SE = 0.0139$ for the last two tests.

The significant difference, $P(HN) - P(LN)$, is particularly damaging for the hypothesis. The hypothesis claims that participants equalize "yes" responses by adjusting positive responses to new items, adjusting false-alarm rates. See the sentence in the quotation from Greene (1996) given earlier: "Because more of the positive responses are guesses for nonwords than for words, there are more false alarms to nonwords than to words" (p. 691). If, however, distinct classes of times are absent, there is, according to response distribution hypothesis, neither the basis nor the need to equalize responses. There is, therefore, no reason for false-alarm rates to differ.

We have demonstrated a mirror effect without distinct classes of items. A condition for response distribution to produce the mirror effect has been eliminated, but the mirror effect remains. We now repeat the demonstration using another variable.

Experiment 2: Continuous Concreteness

This experiment is similar to Experiment 1. A continuous variable was used to generate the mirror effect. Distinct sets were not presented to the participant. Instead of word frequency, however, concreteness was the variable. This variable was again defined by the experimenter by a median split in a continuous distribution.

Method

The method was similar to that of Experiment 1—a lexical-decision task followed by a yes–no recognition test.

Procedure. The participant's lexical-decision list consisted of 200 words and 100 nonwords. The yes–no test list that followed consisted of the 200 lexical-decision words (old) plus 200 new words. The procedure on both the lexical-decision task and the recognition test was the same as in Experiment 1. The main lists were preceded by a brief practice session—lexical decision with 10 words and 10 nonwords followed by 20 words in a yes–no recognition test. At the end of both the practice and main lexical-decision task, and also at the end of both the practice and main recognition test, the number of correct responses was displayed on the screen.

Participants. Thirty-six undergraduate students from an introductory psychology course participated to fulfill a course requirement.

Materials. A pool of 400 words was drawn from Paivio, Yuille, and Madigan (1968). All of the selected words were high-frequency words with a frequency count of 50 or higher by the Thorndike and Lorge (1944) count. The words varied continuously in concreteness rating from 2.66 to 6.71, with a median of 4.69. The distribution is continuous across its range, as shown in Figure 3.

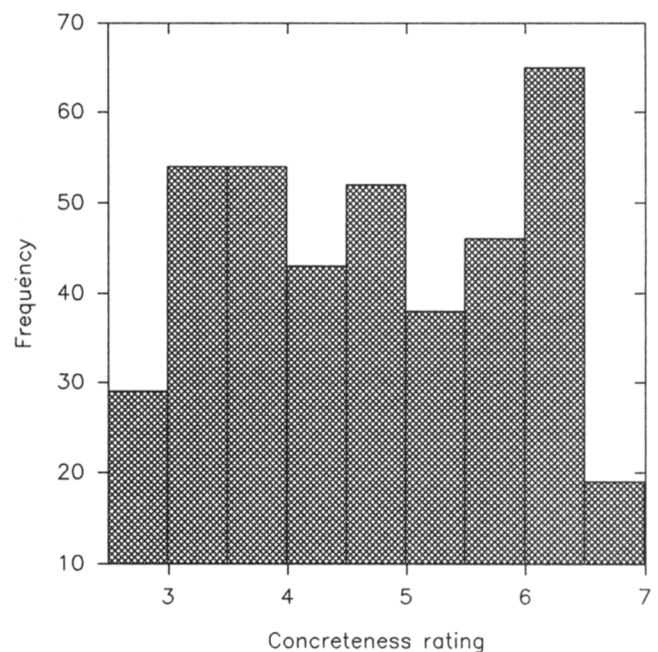


Figure 3. Distribution of concreteness ratings for the set of words used in Experiment 2. The ordinate refers to the frequency of occurrence of each concreteness rating in the word pool of the experiment.

The pool was split at the median into two sets defined as concrete and abstract. The concrete set had a mean frequency of 80; the abstract set, 77. An additional 40 words were selected from Paivio et al. (1968) and used for the practice lists and for the main study list filler items. As in Experiment 1, nonwords were pronounceable, orthographically legal strings. Study and test list items were randomly selected from the concrete and abstract set as defined by the median split and were randomly ordered for each participant independently. Each participant, therefore, had a unique study and test list.

Results and Discussion

The results are shown in Table 4. The mean proportions show the mirror pattern of Inequality 1: $P(\text{CN}) < P(\text{AN}) < P(\text{AO}) < P(\text{CO})$ —where C is concrete and A is abstract. This is the same pattern found in studies with concreteness (or imageability) as a variable: three experiments reported by Glanzer and Adams (1990) and eight of nine experiments summarized by Glanzer and Adams (1985). The overall differences between the means were statistically significant, $F(3, 105) = 134.62, p < .001, \text{MSE} = 0.0073$. Orthogonal comparisons corresponding to those carried out for Experiment 1 were carried out here. A test of the overall effect of old versus new words [$P(\text{CO}) + P(\text{AO})$] – [$P(\text{CN}) + P(\text{AN})$] yielded $t(35) = 12.78, p < .001, SE = 0.0222$. Tests of the differences specific to the mirror pattern found for $P(\text{AN}) - P(\text{CN})$, $t(35) = 5.87, p < .001$. However, for $P(\text{CO}) - P(\text{AO})$, the difference was not statistically significant, $t(35) = 1.57, p > .10$. For both tests, $SE = 0.0094$.

The small difference between CO and AO may be due to the participants' greater tendency to say "yes" to abstract words as compared with concrete words. This bias increased the difference, $P(\text{AN}) - P(\text{CN})$, but decreased the difference, $P(\text{CO}) - P(\text{AO})$.⁴

The evidence, however, from this experiment goes against the response distribution hypothesis. The overall mirror pattern appeared despite the absence of distinct stimulus classes. The difference, $P(\text{AN}) - P(\text{CN})$, was statistically significant. That alone contradicts the hypothesis. The fact that the difference, $P(\text{CO}) - P(\text{AO})$, although in the right direction, was not statistically significant does not protect the hypothesis. The focus of the hypothesis is, as noted earlier, on effects on new words, in this case the difference, $P(\text{AN}) - P(\text{CN})$.

An argument in defense of the response distribution hypothesis has been made with respect to both Experiments 1 and 2—that despite the continuity of the variables presented, participants distinguish some subsets of the items

(e.g., the very high- and very low-frequency words) and carry out response equalization on those subsets. The argument is implausible. No reason is given why a random sequence of words that vary continuously in a number of characteristics will cause participants to define two subsets. There is, moreover, no evidence that such a process occurs. The argument, furthermore, takes a key term of the hypothesis, "easily discriminable sets of items," which has a clear meaning, and replaces it with reference to a covert, unobservable process. The argument reformulates the hypothesis, replacing a testable condition with an untestable condition.

In summary, mirror effects were found in two experiments in which, because the variables were continuous, the participants were not "faced with two easily discriminable sets of items." It has been argued in defense of the hypothesis that the participants, despite the continuity, construct discriminable sets of items and do the hypothesized equalizing of positive responses. That argument is implausible and reformulates the theory, making a major assertion of the hypothesis untestable. There is little to be gained by debating further the plausibility of the argument or the reasonableness of the reformulation involved. We now examine instead whether the second assertion of the hypothesis holds, whether participants equalize responses when they clearly do have distinct classes of items.

The Mirror Effect Without Response Equalization

The next two experiments tested for the presence of response equalization as asserted by the hypothesis, namely that participants "faced with two easily discriminable sets of items . . . try to give equal numbers of positive responses to both sets." In these experiments participants were, by design, faced with two easily discriminable sets of items. One experiment used a yes-no test. The other experiment used a forced-choice test. The conditions were arranged so that response equalization and the mirror pattern were put into opposition. The participants could either equalize responses or show the standard mirror pattern of responses, but not both. In both experiments the participants' responses showed the mirror pattern. They did not show response equalization.

Experiment 3: Yes-No Test With Imbalance in Frequency

In this experiment there were two groups of participants. Both groups received the same type of study list, but the composition of the test list was varied. In the test list, Group 1 (90 HO) had more high-frequency than low-frequency old words; Group 2 (90 LO) had more low-frequency than high-frequency old words. There were also complementary imbalances in the new test words so that in total, the participants saw the same number of low-frequency (old

Table 4
Mean Proportions of "Yes" Responses for Experiment 2
With Concreteness Rating as the Variable

Mean proportions of "yes" responses			
$P(\text{CN})$	$P(\text{AN})$	$P(\text{AO})$	$P(\text{CO})$
.34	.39	.64	.66

Note. C = concrete word; A = abstract word; N = new; O = old.

⁴ The mean proportion of "yes" responses summed across new and old items was .50 for concrete words (no bias) but .52 for abstract words. The difference between the two was statistically significant, $t(35) = 2.39, p < .025, SE = 0.0085$.

plus new) and high-frequency (old plus new) words. The numbers of test items in each test item class in each group are shown in Table 5. Group 1 (90 HO) with a majority of high-frequency old words and a minority of low-frequency old words in the test should, according to the hypothesis, decrease the number of "yes" responses to high-frequency new words and raise the number of "yes" responses to low-frequency new words (false alarms in both cases). Group 2 (90 LO) with a majority of low-frequency old words and a minority of high-frequency old words in the test should decrease the number of "yes" responses to low-frequency new words and raise the number of "yes" responses to high-frequency new words. The sums of "yes" responses (hits plus false alarms) to both frequency classes in each group should, according to the hypothesis, be equal. Furthermore, the response distribution predicted on the basis of the hypothesis should result in distortion of the mirror effect. As we shall see, the hypothesized response equalization did not occur. Neither did the predicted distortion of the mirror effect.

Method

Procedure. The participants carried out lexical decision on a list consisting of 180 words (half low and half high frequency) and 180 nonwords plus 10 filler words both at the start and end of the list. The words in the lexical-decision list furnished the old words in the test list. Group 1 (90 HO) received a test list that had 90 high-frequency old words (HO) and 30 high-frequency new words (HN) and 30 low-frequency old words (LO) and 90 low-frequency new words (LN). Note that on the test, Group 1 (90 HO) had a total of 120 low-frequency words and a total of 120 high-frequency words. Group 2 (90 LO) received a complementary pattern of item classes in its test list with 30 HO, 90 HN, 90 LO, and 30 LN. Note again that on the test, Group 2 (90 LO) had a total of 120 low-frequency words and a total of 120 high-frequency words. The session started with a short, practice lexical-decision list consisting of six words and six nonwords and a short practice recognition test list of six old and six new words. The procedure for the lexical-decision and recognition tests was the same as that of Experiment 2.

Participants. Forty undergraduate students from an introductory psychology course participated to fulfill a course requirement. They were assigned at random to the two groups of 20 each.

Materials. There were 180 high-frequency words (Thorndike & Lorge, 1944; count of 40 or higher) and 180 low-frequency words (count of 1 to 9) randomly selected from Gilhooly and Logie (1980). An additional set of 32 words was selected for filler and practice items. Study and test list items were randomly selected and

Table 5
Design of Experiment 3: The Numbers in Each Test Item Class for the Two Groups

Group	Numbers of test items			
	LN	HN	HO	LO
1 (90 HO)	90	30	90	30
2 (90 LO)	30	90	30	90

Note. L = low-frequency word; H = high-frequency word; N = new; O = old.

Table 6
Mean Numbers of "Yes" Responses in the Four Test Item Classes for the Two Groups in Experiment 3

Group	Mean numbers of "yes" responses			
	LN	HN	HO	LO
1 (90 HO)	21.0	10.2	57.7	21.8
2 (90 LO)	8.0	29.7	21.8	70.5

Note. L = low-frequency word; H = high-frequency word; N = new; O = old.

randomly ordered for each participant independently so that each participant's lists were unique.

Results

The mean numbers of positive responses to all test conditions in both groups of participants are summarized in Table 6. Note that these are numbers, not proportions. These numbers cannot be compared directly because they are based on different numbers of items, 30 or 90. They are presented to display the base from which the values for two critical tables, Table 7 and Table 8, were computed. Table 7 permits the simplest, most direct test of the response distribution hypothesis. Table 8 permits a test of the hypothesis in the context of the mirror pattern.

Mean number of "yes" responses. According to the hypothesis, participants try to give an equal number of "yes" responses (hits plus false alarms) to each distinct item class (high-frequency and low-frequency words). Therefore, the hypothesis predicts that those sums should all be about the same, approximately 60 (half of 120 total test words—old plus new—in each frequency class). In other words, for the mean number of "yes" responses the hypothesis predicts no effect of group and no interaction of group with frequency class.

We obtained the mean number of "yes" responses for each frequency class by summing LO and LN, and summing HO and HN in Table 6 for Group 1 and doing the same for Group 2. Thus, the numbers could be compared. They were all based on 120 items.

The mean numbers of "yes" responses to each class in each group obtained in this way and the predicted numbers (in brackets) are shown in Table 7. According to the response distribution hypothesis, all of the means should be approxi-

Table 7
Mean Sum of "Yes" Responses (Hits Plus False Alarms) in Each Frequency Class for the Two Groups of Experiment 3, Response Distribution Predictions in Brackets

Group	Mean sum of "yes" responses	
	High frequency	Low frequency
1 (90 HO)	67.9 [60]	42.8 [60]
2 (90 LO)	51.5 [60]	78.5 [60]

Note. Each mean consists of LO + LN or HO + HN in Table 6. L = low-frequency word; H = high-frequency word; O = old; N = new.

Table 8
Mean Proportions of "Yes" Responses for the Two Groups of Experiment 3

Group	Mean proportions of "yes" responses			
	<i>P</i> (LN)	<i>P</i> (HN)	<i>P</i> (HO)	<i>P</i> (LO)
1 (90 HO)	.23	.34	.64	.73
2 (90 LO)	.27	.33	.73	.78
<i>M</i>	.25	.34	.68	.76

Note. L = low-frequency word; H = high-frequency word; N = new; O = old.

mately the same and equal to 60, half of the test items in each class. It is obvious that the means were not the same and not equal to 60. Analysis of variance of the means found a statistically significant effect of group, $F(1, 38) = 4.39$, $p < .05$, $MSE = 422.25$, with Group 2 (90 LO) giving more "yes" responses than Group 1 (90 HO). More important is the statistically significant interaction of group with frequency, $F(1, 38) = 100.43$, $p < .001$, $MSE = 135.40$. The interaction effect made it unnecessary to test whether any one of the obtained proportions in the table departs significantly from the theoretical 60. We can, however, test whether the smallest departure from the theoretical value of 60, that of Group 1 high frequency with a mean of 67.9, was statistically significant. A separate test of that mean gave $t(38) = 3.04$, $p < .01$, $SE = 2.6019$.

A simple nonparametric test also indicated that the participants were not equalizing responses. If they were, then they should be giving half of their positive responses to each word frequency class. In Group 1 (90 HO), 19 out of 20 participants gave more "yes" responses to high-frequency test words than to low-frequency test words. With a binomial test ($p = .50$, $n = 20$), the probability of one or fewer positive responses is .00002. In Group 2 (90 LO), 19 out of 20 participants gave more "yes" responses to low-frequency test words than to high-frequency test words. With the same binomial test, the result is of course the same. The participants were, therefore, not equalizing their responses. To see what the participants were doing, and also to check on a possible objection to the preceding analysis, we looked at the data in another way.

Proportion of "yes" responses. An objection to the preceding analysis could be that the participants were really trying to equalize responses but that the imbalances we introduced could not be overcome. If the participants were equalizing their responses to any extent, then the imbalances introduced in this experiment should distort the standard mirror pattern. To see whether there was evidence of such distortion, we recast the data in terms of the proportion of "yes" responses to each separate test category LN, HN, HO, and LO, as in Experiment 1. This is the form in which the mirror pattern is ordinarily evaluated.

We returned to Table 6 and divided the number of "yes" responses in each cell by the number of test items in that cell. These are the numbers presented in the corresponding cells of Table 5. This operation gave the proportion of "yes" responses in the cells in Table 8.

The complete set of proportions of "yes" responses is seen in that table. Both groups show the mirror pattern as given in Inequality 1:

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

The overall differences between the means were statistically significant, $F(3, 114) = 183.60$, $p < .001$, $MSE = 0.0137$. Evaluation of the differences between the means that define the mirror pattern found for $P(\text{LO}) - P(\text{HO})$, $t(38) = 4.74$, $p < .001$, and for $P(\text{HN}) - P(\text{LN})$, $t(38) = 5.60$, $p < .001$, with $SE = 0.0152$ for both tests. No other effects were statistically significant: neither experimental group (list structure), $F(1, 38) = 1.21$, $p > .25$, $MSE = 0.0599$, nor the interaction of list structure with any other variable. Most important, there was no interaction of experimental group (list structure) with the order of the four means: $F(3, 114) = 1.158$, $p > .30$, $MSE = 0.0137$.

Comparisons of the proportions of "yes" responses in each response category (LO, HO, HN, and LN) furnished further evidence against the response distribution hypothesis. If the participants were trying to equalize positive responses, the two groups should have produced different patterns of response. Group 1 (90 HO), with three times as many HO items as LO items, should have decreased $P(\text{HO})$ and $P(\text{HN})$ and raised $P(\text{LO})$ and $P(\text{LN})$. Group 2 (90 LO), with three times as many LO items as HO items, should have decreased $P(\text{LO})$ and $P(\text{LN})$ and raised $P(\text{HO})$ and $P(\text{HN})$. This means that the response distribution hypothesis predicts that the following inequalities should hold: $P(\text{LN})$ in Group 1 $>$ $P(\text{LN})$ in Group 2, $P(\text{HN})$ in Group 1 $<$ $P(\text{HN})$ in Group 2, $P(\text{HO})$ in Group 1 $<$ $P(\text{HO})$ in Group 2, and $P(\text{LO})$ in Group 1 $>$ $P(\text{LO})$ in Group 2. Examination of the four pairs of mean proportions in Table 8 finds that only one of the four inequalities, that for HO, was satisfied. The other three were reversed. The fact that the first two inequalities, those for LN and HN, were reversed is particularly damaging to the response distribution hypothesis. As noted earlier, the hypothesis relies primarily on response adjustments to new items.

Discussion

Given the structure of the test lists in this experiment, if there were response distribution, the mirror pattern should have been disrupted. For example, $P(\text{LN})$ and $P(\text{HN})$ should have been reversed in Group 2 (90 LO). Instead, the mirror pattern occurred in the same form in both groups and overrode any tendency toward equalizing response distribution, if such a tendency were to exist. In fact, the variation of total number of "yes" responses seen in Table 7, the violation of response equalization, occurred because the mirror pattern seen in Table 8 required it. Response distribution was clearly not producing the mirror pattern. The distribution of responses seen here and also in Table 2 was produced by the mirror effect, not the reverse as Greene (1996) had argued for the data of Table 2.

In summary, the response distribution hypothesis failed three tests in this experiment. (a) According to the hypoth-

Table 9
Mean Proportions of Choice in Glanzer et al. (1991, Experiment 1)

Condition	Null choice		Standard choice			
	$P(\text{HN}, \text{LN})$	$P(\text{LO}, \text{HO})$	$P(\text{HO}, \text{HN})$	$P(\text{LO}, \text{HN})$	$P(\text{HO}, \text{LN})$	$P(\text{LO}, \text{LN})$
Immediate	.65	.60	.74	.78	.80	.83
Delay	.57	.55	.56	.62	.63	.64

Note. The first argument within the parentheses is the preferred choice; for example, $P(\text{HN}, \text{LN})$ is the proportion of choice of HN over LN. L = low-frequency word; H = high-frequency word; N = new; O = old. From "Forgetting and the Mirror Effect in Recognition Memory: Concentrating of Underlying Distributions," by M. Glanzer, J.K. Adams, and G. Iverson, 1991, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, p. 87. Copyright 1991 by the American Psychological Association.

esis, the participants should distribute the number of positive responses to produce equalization. They did not (see Table 7). (b) According to the hypothesis, the participants should show a pattern of distortions of the proportions of response item classes—the four inequalities just noted. They did not (see Table 8). (c) According to the hypothesis, the mirror effect should be disrupted. It was not (see Table 8).

Experiment 4: Forced Choice With Imbalance in Choice Conditions

In the preceding experiment, participants did not show response equalization in a yes–no recognition test that produced a mirror effect. In this experiment participants were given another opportunity to demonstrate response equalization. The test here was two-alternative forced choice.

To set the stage for the examination, we first consider an example of the forced-choice pattern that is evidence of the mirror effect. The data are those reported in Table 1 of Glanzer et al. (1991). They show a pattern that has been replicated many times (Glanzer et al., 1991, 1993; Kim & Glanzer, 1993). We use the data to clarify the structure of Experiment 4.

The data in Table 9 show the mirror pattern for forced choice. For standard choices, as in Inequality 2 cited earlier, $P(\text{HO}, \text{HN}) < P(\text{LO}, \text{HN})$, $P(\text{HO}, \text{LN}) < P(\text{LO}, \text{LN})$. For the null choices, as in Inequality 3, $P(\text{HN}, \text{LN})$, $P(\text{LO}, \text{HO}) > .5$. The data in Table 9 also show centering, the collapsing of all underlying distributions toward the midpoint as indicated by the decline of all proportions of choice including the null choices in the delay condition. According to the response distribution hypothesis, the order in the first set of inequalities, that for the standard choice conditions, occurs because LO is very well remembered and HO poorly remembered. This places $P(\text{LO}, \text{LN})$ highest and $P(\text{HO}, \text{HN})$ lowest. We note that response equalization is not set in motion by these two types of choice because whichever alternative is chosen with HO, HN, the participant is choosing H. Similarly, whichever alternative is chosen with LO, LN, the participant is choosing L. The participant does not have a surplus of H or L choices on the basis of those two types of comparisons because ordinarily there is an equal number of such comparisons. Response distribution comes into play with the mixed comparisons LO, HN and HO, LN. Here, according to the hypothesis, if there were no response equalization, the participant would choose a high proportion of

LO in $P(\text{LO}, \text{HN})$ and a low proportion of HO in $P(\text{HO}, \text{LN})$. However, according to the hypothesis, response distribution enters to equalize the number of choices of L and H by biasing the choices toward H and against L, thereby lowering $P(\text{LO}, \text{HN})$ and raising $P(\text{HO}, \text{LN})$ to a middle position in the inequality.

The same logic is applied to the null choices. $P(\text{LO}, \text{HO})$ is greater than .50 because LO is better remembered than HO. Response adjustment then enters to raise $P(\text{HN}, \text{LN})$ so that it equals to $P(\text{LO}, \text{HO})$. Thus, L and H are chosen an equal number of times across both choices. (Note the order of the arguments in the two.) Moreover, if a condition is introduced that lowers $P(\text{LO}, \text{HO})$, such as delay, then the participants will do less response adjustment on HN, LN and centering will be produced.

In the present experiment, we eliminated some of the conditions in Table 9 to test the hypothesis. Two groups were tested. For both groups, the null condition, HN, LN, was retained but LO, HO was omitted. This eliminated the pressure, assumed by the response distribution hypothesis, to adjust HN, LN on the basis of choices of LO, HO. For Group 1 (LO, HN), moreover, the mixed condition LO, HN was kept but HO, LN was omitted. That eliminated the hypothesized pressure from HO, LN to lower $P(\text{LO}, \text{HN})$. For Group 2 (HO, LN), by contrast, the mixed condition HO, LN was kept but LO, HN was omitted. That eliminated the hypothesized pressure from LO, HN to raise $P(\text{HO}, \text{LN})$. On the basis of the hypothesis, the two pure comparison conditions HO, HN and LO, LN were exempt from response adjustment. Any choice in the LO, LN condition resulted in a choice of L. Any choice in the HO, HN condition resulted in a choice of H. The only conditions that were available for response adjustment were the now-isolated mixed conditions and the single null condition.

The response distribution hypothesis does not predict the quantities that the mixed conditions (LO, HN and HO, LN) should take in our reduced experimental design.⁵ The null condition HN, LN is, however, completely available according to the hypothesis, for response adjustment to equalize the number of positive responses to H and L. We can, therefore,

⁵ It does make the qualitative prediction that $P(\text{HO}, \text{LN}) < P(\text{LO}, \text{HN})$. See Greene (1996, p. 691). That prediction is clearly wrong. See the mixed comparisons in Table 11.

predict unambiguously the values the null comparison HN, LN should take if the hypothesis were correct. In the case of Group 1 (LO, HN), with only mixed comparison LO, HN present, the null comparison HN, LN should be adjusted to equal the choice of LO, HN (to counterbalance any preponderance of choice of L). $P(\text{HN, LN})$ should equal $P(\text{LO, HN})$. In the case of Group 2 (HO, LN), with only mixed comparison HO, LN present, the null comparison HN, LN should be adjusted to give the complement of the choices of the mixed comparison. $P(\text{HN, LN})$ should equal $1 - P(\text{HO, LN})$.

Method

Procedure. Before the main study and test lists, the participants were given a practice study and test list. The practice study list consisted of 10 words and 5 nonwords in a lexical-decision task followed by a forced-choice recognition test of 10 word pairs. The words used in the practice lists were middle-frequency words. Following the practice lists, the participants were given a lexical-decision task consisting of 180 words (half high frequency and half low frequency) and 60 nonwords. This list started with five filler items and ended with five filler items.

The recognition test consisted of four types of comparison conditions. Both groups had the pure comparison conditions HO, HN and LO, LN and the null condition HN, LN. Group 1 (LO, HN) had in addition the mixed comparison condition LO, HN to produce a preponderance of "yes" responses to low-frequency words. Group 2 (HO, LN) had instead the mixed comparison condition HO, LN to produce a preponderance of "yes" responses to high-frequency words. The design is summarized in Table 10.

According to Greene (1996), the reason that $P(\text{HN, LN})$ is greater than .50 and changes with $P(\text{LO, HO})$ is that $P(\text{LO, HO})$ is greater than .50 and by its size determines the size of $P(\text{HN, LN})$. Now $P(\text{LO, HN})$ in Group 1 (LO, HN) and $P(\text{HO, LN})$ in Group 2 (HO, LN) will play that determining role.

Each pair was presented on a computer monitor, with the two words arranged vertically, one above the other. There were two response keys on the keyboard, one labeled with an arrow pointing up, the other, below it, with an arrow pointing down. Participants pressed the key with the upward arrow to choose the upper word as the one seen earlier, and the key with the downward arrow to choose the lower word. The word pair remained on the screen until the participant responded. At the end of the lexical-decision task and also at the end of the recognition test, the number of correct responses was displayed on the screen.

Participants. Forty-two undergraduate students participated to fulfill a course requirement. They were assigned at random to the two groups of 21 each.

Materials. Sets of 180 high-frequency words and 180 low-frequency words were selected from Kučera and Francis (1967).

Table 10
Design of Experiment 4

Group	Null choice	Standard choice		
1 (LO, HN)	HN, LN	HO, HN	LO, HN	LO, LN
2 (HO, LN)	HN, LN	HO, HN	HO, LN	LO, LN

Note. The two groups' tests differed only in the composition of the mixed choice. The conditions that distinguish between the two experimental groups are in bold. L = low-frequency word; H = high-frequency word; O = old; N = new.

Table 11

Mean Proportions of Choice in Experiment 4, Response Distribution Predictions for Null Choices in Brackets

Group	Null choice: $P(\text{HN, LN})$	Standard choice		
		$P(\text{HO, HN})$	$P(\text{mixed})$	$P(\text{LO, LN})$
1 (LO, HN)	.60 [.80]	.73	.80	.83
2 (HO, LN)	.61 [.12]	.75	.88	.87
M	.61	.74	.84	.85

Note. $P(\text{mixed})$ is $P(\text{LO, HN})$ in Group 1; $P(\text{mixed})$ is $P(\text{HO, LN})$ in Group 2. The first argument within the parentheses is the preferred choice. L = low-frequency word; H = high-frequency word; N = new; O = old.

High-frequency words were those that occurred 40 or more times per million words, and low-frequency words occurred 8 or fewer times per million. Word length ranged from four to nine letters. The two groups of words were matched on word length and concreteness value according to the Paivio et al. (1968) norms. An additional 30 middle-frequency words were selected as practice study and test items and as filler items.

For each participant the high- and low-frequency word sets were sorted at random to form the four groups of 36 test pairs each (HN, LN; HO, HN; LO, HN or HO, LN; and LO, LN). The study list consisted of the old words in the test pairs. All groups of test words were matched in terms of word frequency, concreteness, and word length. Each participant's study and test list was drawn independently at random from the word sets, assigned independently at random to word pairs, and ordered independently at random. Each participant had, therefore, a unique set of lists.

Results

The results are shown in Table 11. The overall means (bottom row of the table) for the two experimental conditions show the mirror pattern. In the standard choices, $P(\text{HO, HN}) < P(\text{mixed}) < P(\text{LO, LN})$, despite a reversal of $P(\text{HO, LN})$ and $P(\text{LO, LN})$ in Group 2. The null choice, $P(\text{HN, LN})$, was greater than .50 in both groups. The pattern of results corresponds to that of experiments involving forced choice with all six comparison conditions. See, for example, the means in Table 9, which give a typical pattern. Analysis of variance was carried out. It found, as expected, that choice conditions differed, $F(3, 120) = 54.68, p < .001, MSE = 0.0098$. There was a borderline difference between the two groups, $F(1, 40) = 3.70, p = .06, MSE = 0.0158$. Most important, however, there was no interaction between the groups and the choice conditions ($F < 1$). The response distribution hypothesis predicts an interaction. See the predicted values for $P(\text{HN, LN})$ in Table 11.

For Group 1 (LO, HN), the response distribution hypothesis predicts that $P(\text{HN, LN}) = P(\text{LO, HN})$, the mixed choice = .80, to equalize responses. For Group 2 (HO, LN), it predicts that $P(\text{HN, LN}) = 1 - P(\text{HO, LN}) = .12$, to equalize responses. A test of the difference between the obtained and predicted values for $P(\text{HN, LN})$ for Group 1 (LO, HN) found a statistically significant difference, $t(120) = 6.55, p < .001$; the same test for Group 2 (HO, LN) gave $t(120) = 16.00, p < .001$. For both tests, $SE = .0306$.

Discussion

The data contradict the response distribution hypothesis. The participants showed no evidence of adjustment of responses to high- and low-frequency words to equalize the number of choices of each. The values of the null choice, $P(\text{HN}, \text{LN})$, were nowhere near those predicted by the hypothesis. The values of the null choice, $P(\text{HN}, \text{LN})$, were the same in both experimental groups, contrary to the predictions on the basis of the hypothesis.

General Discussion

The response distribution hypothesis states that two factors produce the mirror effect in both yes-no and forced-choice tests. One is the presence of distinct stimulus classes. The other is participants equalizing the number of positive responses they make to each class.

Experiments 1 and 2 tested whether mirror effects appear in the absence of distinct stimulus classes. Contrary to the hypothesis, they did. Experiment 3, using yes-no data, and Experiment 4, using forced-choice data, tested whether the participants equalize the number of positive responses they make to stimulus classes. Contrary to the hypothesis, they did not. In fact, the mirror pattern appeared even though participants had to give unequal numbers of positive responses to the different stimulus classes to produce the pattern. In summary, wherever the response distribution hypothesis can be brought to test, it is proved wrong.

We are left then with the question of what Greene (1996) demonstrated. He showed that instructions to avoid guessing disturb the mirror effect. He did not demonstrate response distribution. The preceding experiments rule out that hypothesized process and its relation to the mirror effect. He did show that it is possible, with two distinct sets of items, to have participants use different criteria for the two sets. We now demonstrate that criterion shift in his data.

First, we applied a standard signal-detection analysis to the data of Greene's (1996) two experiments on the effect of no-guessing instructions, Experiments 8 and 9. Examination of the data of those experiments in Table 1 reveals a simple pattern. Participants given no-guessing instructions changed

Table 13

Data From Greene's (1996) Experiment 8 Fitted by Attention-Likelihood Theory

Group	$P(\text{wN})$	$P(\text{nN})$	$P(\text{nO})$	$P(\text{wO})$
Obtained				
Control	.20	.48	.58	.88
No guess	.19	.14	.27	.87
Fitted				
Control	.17	.45	.64	.87
No guess	.17	.12	.25	.87

Note. The parameters are as follows: $p(\text{new}) = .110$, $N = 1,000$ (preset), $n(\text{nonword}) = 31$, $n(\text{word}) = 88$, $\beta(\text{control}) = 0$ (preset), $\beta(\text{nonword}) = 0.499$, $\beta(\text{word}) = 0.064$, $r^2 = .993$. w = word; n = nonword; N = new; O = old.

their criteria for the two item classes. Table 12 presents the signal-detection d' s (the accuracy or sensitivity measures) and betas (the bias or criterion placement measures) for the conditions in both experiments. In both experiments, the main effect of the instruction was to produce an increase in beta. Participants became more conservative. They became more so with the inferior item class, nonwords (n) in Experiment 8 and high-frequency words (H) in Experiment 9. The changes in d' were relatively minor. There was no change in d' for the superior item classes—word (w) in Experiment 8 and low-frequency word (L) in Experiment 9. There was a slight increase in d' for the inferior items classes, nonwords (n) and high-frequency words (H). The latter increase may be the result of more deliberate evaluation of those classes of items with no-guessing instructions.

The no-guessing instruction, therefore, simply causes participants to change their decision criteria differentially for the item classes presented. This aspect of the performance is easily handled by any theory of recognition memory that incorporates the mechanisms of signal-detection theory. We show that that is so by fitting attention-likelihood theory (Glanzer et al., 1993) to Greene's (1996) data. There are two reasons to use attention-likelihood theory here. It has been designed to handle the mirror effect. It also incorporates the mechanisms of signal-detection theory.

To fit attention-likelihood theory to the data, we permitted the criteria in the instruction condition to move. We fitted the eight observed means in each experiment with five free parameters. In the case of Experiment 8, there were five parameters free to vary: two parameters setting stimulus sampling rates, $n(\text{nonword})$, $n(\text{word})$; a noise parameter $p(\text{new})$; and two criteria, $\beta(\text{nonword})$ and $\beta(\text{word})$ for the instruction condition. Two parameters were preset: N , the total number of features in a word, was fixed at 1,000 on the basis of past work, and the criterion for the control condition was fixed at 0. In the case of Experiment 9, the stimulus sampling rates were set by $n(\text{low})$ and $n(\text{high})$. The criteria for the instruction condition were set by $\beta(\text{low})$ and $\beta(\text{high})$. The results of the fitting allowing five parameters to vary are shown in Tables 13 and 14. The fit accounted for 99.3% of the variance in the means of

Table 12
 d' s and Betas for Greene's (1996) Experiments 8 and 9

Group	d'		β	
	n	w	n	w
Experiment 8				
Control	0.25	2.02	0.98	0.72
No guess	0.47	2.00	1.49	0.78
Group	H	L	H	L
Experiment 9				
Control	1.23	2.18	1.08	1.35
No guess	1.52	2.18	2.42	1.99

Note. n = nonword; w = word; H = high-frequency word; L = low-frequency word.

Table 14
Data From Greene's (1996) Experiment 9 Fitted by
Attention-Likelihood Theory

Group	P(LN)	P(HN)	P(HO)	P(LO)
Obtained				
Control	.11	.25	.71	.83
No guess	.08	.09	.57	.78
Fitted				
Control	.14	.22	.70	.84
No guess	.08	.13	.57	.76

Note. Parameters are as follows: $p(\text{new}) = .108$, $N = 1,000$ (preset), $n(\text{high}) = 62$, $n(\text{word}) = 87$, $\beta(\text{control}) = 0$ (preset), $\beta(\text{high}) = 0.760$, $\beta(\text{low}) = 0.324$, $r^2 = .996$. L = low-frequency word; H = high-frequency word; N = new; O = old.

Experiment 8 and 99.6% of the variance in the means of Experiment 9. As can be seen, the shift in criteria produced the departure from the mirror order found in the instruction condition of Experiment 8 and the reduction in the difference between $P(\text{LN})$ and $P(\text{HN})$.

The focus of this article is not, of course, on a comparison of the response distribution hypothesis with attention-likelihood theory. The focus is on whether the response distribution hypothesis explains the mirror effect. Attention-likelihood theory is brought in here only to show that an alternative approach involving standard signal detection ideas can explain Greene's (1996) data.

In summary, eliminating the conditions that the response distribution hypothesis asserts produce the mirror effect finds no support for the hypothesis. The mirror effect occurs without distinct classes of items. It occurs without equalization of responses. Moreover, response equalization does not produce the mirror effect. The reverse is true. When there is no imbalance of classes of items, the mirror effect produces those cases that appear to show response equalization. When there is imbalance in classes of items (e.g., more high-frequency old words), the mirror effect remains and response equalization disappears. The demonstration of the effect of no-guessing instructions is simply another demonstration that participants can adjust bias in a recognition test.

References

- Carroll, J. B. (1968). Word-frequency studies and the lognormal distribution. In E. M. Zale (Ed.), *Proceedings of the Conference on Language and Language Behavior* (pp. 213-235). New York: Appleton-Century-Crofts.
- 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., & 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, and 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, and Cognition*, 17, 81-93.
- Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, 100, 546-567.
- Greene, R. L. (1996). Mirror effect in order and associative information: Role of response strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 687-695.
- Kim, K., & Glanzer, M. (1993). Speed versus accuracy instructions, study time, and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 638-652.
- Kučera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 76(1, Pt. 2).
- Schwartz, F., & Rouse, R. O. (1961). The activation and recovery of associations. *Psychological Issues*, 3(1, Whole No. 9).
- Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teacher's College, Columbia University Bureau of Publications.

Received June 8, 1995

Revision received September 8, 1997

Accepted September 8, 1997 ■