Journal of Experimental Psychology:
Learning, Memory, and Cognition
1996, Vol. 22, No. 3, 687–695

0278-7393/96/$3.00

Mirror Effect in Order and Associative Information: Role of Response Strategies

Robert L. Greene
Case Western Reserve University

Classes of stimuli that are relatively easy to classify in recognition as old when old also appear to be relatively easy to classify as new when new. Five experiments are reported that extend this mirror effect to tests requiring retention of order and associative information. These experiments tested memory for words and nonwords; across experiments, the mirror effect was found on both yes–no and forced-choice tests. In addition, a new account of the mirror effect is proposed, one which suggests that this pattern results from participants' attempts to distribute responses equally across stimulus classes on tests. Support from this account came from additional experiments in which the mirror effect was eliminated when participants were asked to refrain from using positive responses as guesses.

Numerous studies have been performed in which recognition memory has been tested for study lists containing two distinct sets of stimuli. The typical finding is that if one class is easier to classify as old when old than the other, then that class will also be easier to classify as new when new. On a yes–no recognition test, this pattern is shown when the stimulus class that has the higher hit rate also has a lower false-alarm rate than the other class. On a forced-choice recognition test, the equivalent pattern would be shown when participants are most accurate on test pairs containing a target and a distractor from the same class, least accurate on test pairs containing a target and a distractor from the other class, and intermediate in accuracy on the pairs containing one stimulus each from the two classes. Results such as these have been called the *mirror effect* and have been found in many recognition studies, such as comparisons of words with nonwords and low-frequency words with high-frequency words (Glanzer & Adams, 1985, 1990).

The generality of the mirror effect across stimuli in recognition has been clearly demonstrated. The purpose of Experiments 1–5 was to demonstrate its generality across memory tests. It has been shown that the mirror effect can be found in frequency discrimination (Greene & Thapar, 1994). Frequency discrimination, like recognition, is a test of item information, and evidence suggests that frequency judgments and recognition are performed using similar processes (Hintzman & Curran, 1994). The first five experiments reported here extend the mirror effect to measures of order and associative information. All current accounts of the mirror effect have been intended to explain its occurrence on recognition tests. If frequency discrimination is performed using the same processes as recognition, such theories may be able to be extended there as well. It is less clear whether such theories could be extended to tests requiring order and associative information. In Experiments 6–9, I tested a new response-distribution account that can be easily extended to all memory tests.

The choice of stimuli was influenced by the desire to have two classes that could be expected to differ as greatly as possible. The first five experiments tested memory for words and nonwords (unpronounceable strings of consonants). The other comparison often used in the mirror-effect literature (high-frequency words vs. low-frequency words) was not studied here in Experiments 1–5 because the previous literature makes it unclear whether those two classes differ on the measures used here. For example, Gregg (1976) claimed that word frequency has no effect on tests requiring order information once response availability is equated. As for associative information, the literature offers a mixed picture, but word frequency at best has only a small effect that is not always found (Clark, 1992; Clark & Burchett, 1994; Clark & Shiffrin, 1992; Gronlund & Ratcliff, 1989; Hockley, 1991, 1994). The word–nonword comparison was used here because it seemed likely to lead to large differences in performance on all tasks. The focus of interest then becomes the pattern that these differences exhibit.

Experiment 1

This experiment determined whether a mirror effect could be found on a yes–no measure of order memory. The measure used might be called a *list-half specific recognition test.* Participants saw lists containing words and nonwords. After each list, they were given a test in which they had to give a positive response if a test stimulus came from a specified half of the list and a negative response if it came from the other half.

Method

Participants. Twenty-four students from introductory psychology classes participated individually to fulfill a course requirement.

Stimuli. The stimuli were those used by Greene and Thapar (1994, Experiments 1–3). The critical words were 60 nouns with a frequency of 1–5 occurrences per million words (Thorndike & Lorge, 1944) selected from the Paivio, Yuille, and Madigan (1968) norms. The

I would like to thank Bryce Burba, Julie Cho, Sharon Evans, and Laura Wellett for their help in testing participants. Tim Curran, Murray Glanzer, and John Wixted made many useful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Robert L. Greene, Department of Psychology, Case Western Reserve University, Cleveland, Ohio 44106. Electronic mail may be sent via Internet to rlg2@po.cwru.edu.

critical nonwords were 60 strings of consonants, judged to have no obvious interpretation. The nonwords were matched with the words for number of letters and number of repeated letters. From this set of 120 stimuli, the computer randomly selected for each participant 48 words and 48 nonwords and distributed these equally across four lists.

Procedure. Participants received four lists of words and nonwords, each followed immediately by a test. At the beginning of the experiment, participants were told the nature of the lists and tests that they would receive. Participants saw four lists containing 28 items; these were shown 1 item at a time on a computer screen at a rate of 1 every 5 s. Each list consisted of 12 critical words, 12 critical nonwords intermixed with the words, and 4 untested buffer stimuli (2 words and 2 nonwords) occupying the first two and last two serial positions. Immediately after the last item on each list, the test on that list began. Participants were shown 1 list item at a time; all stimuli used as test items had been included on the list. On two of the lists, the computer asked participants to give a positive response if the item came from the first half of the list and a negative response if it did not; on the other lists, participants were to give a positive response if the item had been shown in the second half. Participants did not know which half would require a positive response until the test began for a specific list, and there had been no demarcation of list halves during presentation. After responding by pressing keys to 12 words and 12 nonwords, the next list immediately began, until participants had seen four list–test sequences.

Results

Preliminary analyses showed that the form of the test question (asking either about the first or second half) had no effect on performance. This was not surprising, because logically these two forms can be answered with the same information. Also, the serial position of the four lists had no effect. Data were collapsed across these variables in later analyses, which were conducted on the number of hits and the number of false alarms. A .05 significance level was used for analyses in all experiments reported here.

Hits and false alarms were defined here in an analogous way to conventional recognition tests: A hit was a correct positive response, and a false alarm was an incorrect positive response. A mirror effect would be present if words had higher hit rates and lower false-alarm rates than nonwords. The proportions of hits were .80 for words and .61 for nonwords. The proportions of false alarms were .28 for words and .41 for nonwords. There were significant differences in the number of hits, $F(1, 23) = 38.10$, $MSE = 6.26$, and in the number of false alarms, $F(1, 23) = 21.51$, $MSE = 5.45$. Discussion of the implications of these results is deferred until after Experiment 5.

Experiment 2

This experiment was conducted to determine whether a mirror effect could be found in a different measure of order information, a forced-choice recency-discrimination test on words and nonwords.

Method

Participants. Twenty students from introductory psychology classes participated individually to fulfill a course requirement.

Procedure. The same stimuli were used as in Experiment 1, but they were now used in five lists. Participants saw five lists of 28 items comprising 12 words, 12 nonwords, and 4 untested buffers at the first two and last two positions. Immediately after each list, participants were given a test in which they were shown pairs of items from the list. They were asked to press a key to indicate the item in each test pair that had been shown more recently on the list. The computer randomly paired list items together on the list with the constraints that at least 3 items separate the input positions of the stimuli on a test pair and that the test contain equal numbers of pair types (word–word, nonword–nonword, recent word paired with older nonword, recent nonword paired with older word). Left–right position for the two alternatives was determined randomly by the computer on each test trial. After 12 test pairs were responded to, the next list began until five list–test sequences had been done.

Results

A mirror effect is present in a forced-choice test when performance is best when the test pair contains two items from one class, worst when the test trial contains two items from one class, and intermediate on mixed test pairs. This pattern was found here. The proportions correct were .75 on pure-word trials, .56 on pure-nonword trials, .67 on mixed trials where the more recent item was a word, and .66 on mixed trials where the more recent item was a nonword. The overall effect of pair type was significant, $F(3, 57) = 7.45$, $MSE = 3.68$. The pure-word pairs were significantly easier than the pure nonword pairs, $F(1, 19) = 17.89$, $MSE = 4.54$, than the mixed pairs where the word was the correct choice, $F(1, 19) = 4.55$, $MSE = 2.91$, and than the mixed pairs where the nonword was the correct choice, $F(1, 19) = 8.05$, $MSE = 1.94$. The pure-nonword pairs were significantly more difficult than the word-correct mixed pairs, $F(1, 19) = 6.77$, $MSE = 4.27$, and the nonword-correct mixed pairs, $F(1, 19) = 4.66$, $MSE = 5.50$. The two kinds of mixed pairs did not differ significantly, $F(1, 19) = 0.03$, $MSE = 2.94$.[1]

Experiment 3

Hockley (1994) has reported a mirror effect on tests requiring associative information in comparisons of nouns with other words and of concrete and abstract words. In Experiment 3, I attempted to generalize this claim by seeking a mirror effect in a yes–no test requiring associative information with words and nonwords as stimuli. Participants saw lists of pairs and were given tests in which they had to give positive responses to intact pairs and negative responses to pairs of rearranged (mismatched) items.

Method

Participants. Twelve students from introductory psychology classes participated individually to fulfill a course requirement.

Stimuli. An additional 60 words and 60 nonwords with the same characteristics as those used in the first two experiments were added to the previously used stimuli. For each participant, the computer randomly made word pairs and nonword pairs. Four lists were randomly constructed by the computer for each participant; each list

[1] In these (and all other) pairwise comparisons, the error terms used do not come from the omnibus significance test but are based only on the specific two conditions being compared.

contained 12 word pairs, 12 nonword pairs, and 4 untested buffer pairs occupying the first two and last two serial positions.

Procedure. Before the experiment began, participants were informed of the nature of the lists and tests they would be receiving. Each list was presented on a computer screen at a rate of one pair every 5 s. Immediately after the list, participants were given a test in which they were shown pairs of list items. Participants were to give a positive response if the two stimuli had been paired on the list and a negative response if they had belonged to different pairs. All list and test pairs consisted of two words or two nonwords. Stimuli on rearranged (distractor) pairs were combined randomly by the computer for each participant with the constraints that words could only be paired with words and nonwords with nonwords and that stimuli in a rearranged pair never came from adjacent input positions.

Results

A mirror effect was found, with word pairs having higher hit rates and lower false-alarm rates than nonword pairs. The proportions of hits were .86 for word pairs and .53 for nonword pairs. An analysis performed on the number of hits found a significant difference between words and nonwords, $F(1, 11) = 33.00$, $MSE = 11.64$. The proportions of false alarms were .17 for word pairs and .39 for nonword pairs. There was a significant difference in number of false alarms between words and nonwords, $F(1, 11) = 41.23$, $MSE = 4.01$.

Experiment 4

This experiment determined whether a mirror effect could be found on a forced-choice test of associative information. Participants were shown a list of word pairs and nonword pairs. They were then given a forced-choice associative recognition test in which they had to discriminate between intact and rearranged pairs. A mirror effect would be present if accuracy was highest when participants chose between two word pairs, lowest when participants chose between two nonword pairs, and intermediate on test trials containing one word pair and one nonword pair.

Method

Participants. Twenty-two students from introductory psychology classes participated individually to fulfill a course requirement.

Procedure. The stimuli were the same, and the same method of list presentation was followed as in Experiment 3, with participants seeing four lists composed of 28 pairs. Immediately after each list, participants were given a test consisting of 12 trials. On each test trial, they were shown two pairs. One pair consisted of two stimuli that had been shown together on the list. The other pair was made of items that had been presented in different (and nonadjacent) pairs. Participants were asked to indicate the intact pair by pressing a key on the keyboard. Equal numbers of test trials involved discriminating between two word pairs, two nonword pairs, an intact word pair and a rearranged nonword pair, and an intact nonword pair and a rearranged word pair.

Results

A mirror effect was found here. The proportions correct were .91 on trials containing two word pairs, .63 on trials containing two nonword pairs, .83 on mixed trials where the intact pair contained words, and .79 on mixed trials where the intact pair contained nonwords. The main effect of pair type was significant, $F(3, 63) = 12.84$, $MSE = 3.36$. Pure-word trials were significantly easier than pure nonword trials, $F(1, 21) = 52.56$, $MSE = 2.30$, than mixed trials where the word pair was the correct choice, $F(1, 21) = 5.80$, $MSE = 1.57$, and than mixed trials where the nonword pair was the correct choice, $F(1, 21) = 7.56$, $MSE = 2.89$. Pure-nonword trials were significantly more difficult than mixed pairs where the word pair was intact, $F(1, 21) = 24.08$, $MSE = 2.65$, and than mixed trials where the nonword pair was intact, $F(1, 21) = 8.51$, $MSE = 4.71$. The two kinds of mixed trials did not differ significantly, $F(1, 21) = 0.46$, $MSE = 6.04$.

Experiment 5

Glanzer, Adams, and Iverson (1991; see also Glanzer & Bowles, 1976) proposed that null test trials could be used as a converging operation to demonstrate mirror effects on forced-choice recognition tests. A null test trial is one on which there is no one correct answer. Although participants are told that all test trials involve one old alternative and one new alternative, at least some trials contain two old alternatives or two new alternatives. All null trials contain one alternative from each of the two stimulus classes. A mirror effect is demonstrated when participants tend to choose items from one stimulus class on null trials containing two old alternatives and items from the other class on null trials containing two new alternatives.

Experiment 5 tested performance on null trials in a test requiring associative information. After seeing a list composed of word pairs and nonword pairs, participants were given a forced-choice test in which they were asked to discriminate between intact and rearranged pairs. In reality, all of the test trials contained either two intact pairs or two rearranged pairs. Every test trial involved a choice between a word pair and a nonword pair. A mirror effect would be shown if participants picked the word pair when both pairs were intact and the nonword pair when both were rearranged.

This experiment was also designed to demonstrate the phenomenon of *concentering* on null trials. Glanzer et al. (1991) reported concentering in recognition of high- and low-frequency words. Some null trials were included on their test. When participants were tested immediately, they tended to choose the low-frequency word on test pairs containing two old stimuli and the high-frequency word on test pairs containing two new stimuli. When the test was delayed for a week, both the proportion of low-frequency words chosen on old–old trials and the proportion of high-frequency words chosen on new–new trials moved closer to the random level of .50. In Experiment 5, half of the participants were tested immediately and half were tested 5 days after list presentation. Concentering would be present if the proportion of word pairs chosen on old–old trials and the proportion of nonword pairs chosen on new–new trials were closer to .50 on the delayed test than on the immediate test.

Method

Participants. Sixty students from an advanced psychology course participated as unpaid volunteers.

Procedure. Unlike in the previous experiments, each participant saw a single 52-pair list (24 critical word pairs, 24 critical nonword pairs, and 4 buffer pairs occupying the first two and last two positions); the stimuli were chosen randomly by the computer separately for each participant. Then, a single test consisting of 24 trials was given. This test was given immediately after presentation of the study list for 30 participants. The remaining 30 participants were asked to return to the laboratory 5 days later and were given the test then. On each trial, a pair of words and a pair of nonwords were given. Participants were asked to choose the 1 intact pair on each trial. In reality, every trial involved a choice between 2 intact pairs or 2 rearranged pairs. After the experiment, participants were debriefed, and all said they were surprised when told that null pairs had been used.

Results

A mirror effect was obtained, with participants choosing the word pair on null trials that contained two intact pairs and the nonword pair on null trials that contained two rearranged pairs. Of the 12 trials containing two intact pairs, participants who were given the immediate test chose the word pair on a mean of 8.97 trials (75%, SD = 2.58), which differed significantly from the chance level of six trials (50%), $t(29)$ = 6.30. When the test trials contained two rearranged pairs, the mean number of word pairs chosen by immediate-test participants was 2.33 (19%, SD = 2.32), which differed significantly from the chance level of six trials, $t(29)$ = 8.64. For participants given the delayed test, the mean numbers of word pairs chosen were 7.03 (59%, SD = 1.50) on trials containing intact pairs and 5.40 (45%, SD = 1.59) on trials containing rearranged pairs; each of these differed significantly from the chance level of six trials, $t(29)$ = 3.78, and $t(29)$ = 2.07.

Concentering is found when the participants' responses fall closer to 50% after a delay than on an immediate test. This was shown here. On trials with two intact pairs, the number of word pairs chosen was significantly higher on the immediate test than on the delayed test, $F(1, 58)$ = 12.61, MSE = 4.45; on trials with two rearranged pairs, the number of word pairs chosen was significantly lower on the immediate test than on the delayed test, $F(1, 58)$ = 35.59, MSE = 3.96.

It is impossible to test null trials on order tests, such as the recency-discrimination test used in Experiment 2. I could not guarantee that two stimuli would be encoded by the participant at exactly the same time, so the question of which of two stimuli had been seen more recently may always have a correct answer. However, Experiment 5 shows that a mirror effect and concentering can be found in associative information using null pairs.

Discussion of Experiments 1–5

A mirror effect was found on yes–no and forced-choice tests requiring both order and associative information. These results show that the mirror effect generalizes beyond the recognition test usually employed to study it.

Implications for Previous Accounts

Several kinds of accounts have been proposed for the mirror effect. One kind (Glanzer & Bowles, 1976; Hintzman, 1988; Hoshino, 1991; Wixted, 1992) assumes that all responses are determined by the strength of a signal to a memory probe and attributes the mirror effect to differences in signal strength; items from one class lead to stronger signals than the other when they are new, but there is a reversal of their relative strengths when they are old. This account has been proposed to explain mirror effects in recognition, and it is not clear how it could be developed to explain the presence of mirror effects on tasks that involve discriminating between old items, such as frequency discrimination (Greene & Thapar, 1994) or the order tasks studied in Experiments 1 and 2. Consider the recency-discrimination data in Experiment 2: The fact that pure-nonword pairs were more difficult than mixed pairs where the word was the correct choice could be explained by assuming that words yield stronger signals than nonwords and that participants choose the stronger stimulus as more recent. However, that assumption is inconsistent with the finding that pure-word pairs were easier than mixed pairs where the word was the correct option and with the finding that accuracy on mixed pairs was not significantly affected by whether the word or the nonword was the more recent stimulus. It should also be noted that studies of recency discrimination have generally not supported accounts that rely on concepts such as signal strength (Flexser & Bower, 1974; Hacker, 1980; McElree & Dosher, 1993).

A second sort of account (Brown, Lewis, & Monk, 1977) attributes the mirror effect to a deliberate analysis of the memorability of each stimulus. It is assumed that participants take into account the fact that some items are inherently more memorable than others. This explanation of the mirror effect in recognition is inconsistent with evidence that mirror effects can be found even when participants have inaccurate information about the memorability of items (Greene & Thapar, 1994; Wixted, 1992) or when such memorability evaluations should be impaired by having participants respond quickly or while they are performing a secondary task (Hintzman, Caulton, & Curran, 1994). For that matter, it is unclear how one could even apply this sort of account here. Would such memorability assessments help participants decide on which half of a list an item had been presented (Experiment 1) or which of two stimuli had been shown more recently when they might have been separated by only a few items on the list (Experiment 2)?

A third approach suggests that mirror effects reflect rescaling of signal strengths, possibly in an unconscious fashion (Gillund & Shiffrin, 1984; Glanzer & Adams, 1990; Hintzman, 1994). These accounts are able to predict mirror effects in recognition, and it is a matter of taste whether one believes that the complex mathematical rescalings proposed could be carried out by participants (e.g., transforming signal strengths into log likelihood ratios; Glanzer & Adams, 1990). Versions of this approach that claim that these rescalings occur at time of test have an additional difficulty in explaining the presence of mirror effects when participants have to respond rapidly (Hintzman et al., 1994). With respect to the data reported here, one can imagine how these models might be extended to the associative tasks used in Experiments 3–5; for example, participants could unitize the members of a pair and then perform a recognition process comparable to that carried out on individual items. It is less clear how this approach could explain performance on the order tasks studied in Experiments

1 and 2. Even if signal strength carries information about relative position, could participants perform rescalings to determine the log likelihoods that particular items occurred in the first half of a list or in the second half (Experiment 1)? Even this sort of information would not be decisive in the recency-discrimination task in Experiment 2, where the two members of a test pair may have been presented on the same half of the study list. The generality of the mirror effect across tasks presents problems for accounts that attribute it to complex processes specific to recognition.

A Response-Distribution Account

The fact that mirror effects appear on a number of memory tests raises the possibility that they might not reflect any specific memory processes. An account is proposed here that suggests that this effect is actually rather uninteresting. That is, it does not reveal any particular properties of the memory system, but rather it results from the response strategies that participants tend to adopt in testing situations.

First consider yes–no tests. Participants go into a yes–no test (whether in the laboratory or in the classroom) 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. 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 to 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. Such a process that does not depend on memory may be executed quickly on any particular trial, leading to mirror effects even when participants answer quickly (Hintzman et al., 1994). Participants do not necessarily even have to realize that they are remembering one class of stimuli more than the other; they might just realize that they are giving positive responses more often to one class than the other without knowing whether they are correct.

The response-distribution account has been phrased in terms of distributing guesses. This wording was chosen to remain as free of assumptions as possible. The critical claim is merely that participants try to equate the number of positive responses they give to each class of stimulus. How participants do this is left unspecified and may vary between tests. On a recognition test, this may be done by choosing separate decision criteria for each class. In fact, the response-distribution account is essentially consistent with recent proposals by Hirshman (1995), who argued that changes in criterion placement may underlie the mirror effect in recognition memory. Different mechanisms may be involved on tests, such as those requiring memory for order, which seem less dependent on pure familiarity.

It should be kept in mind that this account is not trying to explain why one class of stimuli is more memorable than another. After all, the reason why words are more memorable than nonwords is probably not exactly the same reason why low-frequency words are more memorable than high-frequency words on some tests or why concrete words are more memorable than abstract words. The reason why one class of stimuli leads to more hits in recognition than another may not be the reason why it also leads to more hits on order and associative tasks. Perhaps participants recognize items on the basis of signal strength and determine an item's position through retrieval of contextual information. None of that is addressed by this account. Rather, this account is only trying to explain why false-alarm rates go in the opposite direction of hit rates, and it explains this by claiming that this pattern merely reflects participants' attempts to give roughly equal numbers of positive responses to both classes of stimuli. This is why this account can rightly be described as being uninteresting; it assumes that the mirror effect can be explained without making any claims about the properties that make one class of stimuli more memorable than another. This proposal says that the goal of making equal numbers of positive responses to both categories leads participants to make positive responses to items they otherwise would not respond positively to; this simple principle could presumably be grafted onto any model of memory functioning.

Extension to forced-choice tests. This account can be easily extended to forced-choice tests. A forced-choice test is one where participants are asked to give exactly one positive response to every pair of test items, and they will try to give positive responses to items from each stimulus class about equally often. Consider an experiment comparing words and nonwords. Participants are able to remember the words more often. This leads to pure-word test trials being easier than pure-nonword test trials. The critical finding to explain is that mixed test trials are intermediate in accuracy. This can be explained by realizing that participants will be biased to choose nonwords on mixed trials so that the total number of positive responses to words and to nonwords will be equal.

For example, let us say that participants are able to remember word targets correctly 70% of the time they occur and nonword targets 30% of the time. When participants are unable to remember either of the options on a test pair, they guess. On pure-word pairs, this would lead to an overall proportion correct of .85 (that is, participants are correct by memory .70 of the time, and they are correct half of the .30 of the time that they guess). On pure-nonword test trials, participants would have an overall proportion correct of .65 (correct by memory .30 of the time and correct half of the .70 of the time they guess). On mixed test trials, participants would be correct by memory .70 of the time when the word is the target and .30 of the time that the nonword is the target. However, when participants resort to guessing, they will do so in a way such that nonwords are given positive responses as often as words. In this case, participants would give 70% of their guesses on mixed trials to nonwords and 30% to words. Participants would have a proportion correct of .79 on mixed trials where the word is the correct option because they are correct by memory .70 of time and they guess .30 of the time, choosing the word only .30 of the time they guess [.70 + (.30)(.30) = .79]. Participants would also have a proportion correct of .79 on mixed trials where the nonword is the correct option because they are correct by memory .30 of the time and guess .70 of the time, choosing the nonword .70 of the time they guess [.30 + (.70)(.70) = .79]. This is a demonstration that the mirror effect could result if participants use their guesses on mixed trials to equate the numbers of times they choose stimuli from each class.

Extension to null trials. 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.

This could also explain the phenomenon of concentering on null trials, the tendency for the proportion of choices to approach .50 after a delay. According to the account offered here, there is nothing mysterious about this process. In Experiment 5, the delay reduced participants' ability to recognize the word pairs on trials containing two intact pairs; participants would then guess more often on these trials, thereby increasing the number of times they would select the nonword pairs. As the number of times the nonword pair is selected on intact–intact trials increases, the number of times the word pair is selected on trials with two rearranged pairs increases because there will be less need to balance the number of word and nonword pairs chosen. The result is that all of the proportions head toward .50. If delay were lengthened until forgetting was complete, all the proportions would be at .50. An analogous explanation would explain concentering on recognition tests (e.g., Glanzer et al., 1991).

When is a mirror effect not found? A mirror effect will be found whenever participants can easily discriminate between two different classes of stimuli that differ in their memorability. For participants to decide to give positive guesses more to one kind of stimuli than to the other, they must realize that two discrete classes of stimuli are being shown and that they are giving unequal numbers of positive responses to the two classes. Because words and unpronounceable nonwords are easily discriminated and differ greatly in their memorability, a mirror effect should be found with these stimuli on any yes–no or forced-choice test. Other comparisons may not be obvious to some participants and thus fail to exhibit mirror effects. For example, comparisons of abstract and concrete words lead to mirror effects only inconsistently unless participants encode items by rating them for concreteness (Hintzman et al., 1994); this simultaneously increases the recognizability of concrete words and makes participants more aware of the distinction between concrete and abstract words. In the absence of this kind of encoding task, many participants might not realize that two distinct classes of stimuli are being used and would therefore not try to equate the number of positive responses given to abstract and concrete words. Similarly, a lexical-decision encoding task enhances the mirror effect in recognition for low- and high-frequency words (Hoshino, 1991), perhaps because it makes more salient the fact that two different classes of word are being used.

One way to test this account is to ask participants to avoid guessing and to give a positive response only when they believe that it is the correct answer. This was done in Experiments 6–9. In Experiment 6, a yes–no test for order memory identical to that used in Experiment 1 was used. In Experiment 7, a yes–no test for associative information identical to that used in Experiment 3 was used. Recognition was tested in Experiments 8 and 9.

Experiment 6

Method

Participants. Twenty-four students from introductory psychology classes participated individually to fulfill a course requirement.

Procedure. The stimuli and procedure of presentation were identical to those used in Experiment 1, with participants going through four study–test cycles. Each test consisted of the list-half specific recognition test used in Experiment 1. The only new aspect was that before the first test was administered, participants were asked to avoid using positive responses as sheer guesses and to give a positive response only when they believed that yes was the correct answer.

Results

The hit rates were .71 for words and .37 for nonwords. The false-alarm rates were .19 for words and .14 for nonwords. A mirror effect was not present as it was no longer the case that nonwords had a higher false-alarm rate than words. The difference in hits was significant, $F(1, 23) = 88.47$, $MSE = 9.33$, but the difference in false alarms was not, $F(1, 23) = 2.37$, $MSE = 9.56$.

Experiment 7

Method

Participants. Twelve students from introductory psychology classes participated to fulfill a course requirement.

Procedure. The stimuli and procedure were identical to those used in Experiment 3, with a yes–no associative recognition test again being given. As with Experiment 6, participants were asked to avoid giving positive responses as sheer guesses and to answer yes only when they believed that this was the correct answer.

Results

No mirror effect was found. Hit rates were .84 for words and .33 for nonwords. False-alarm rates were .15 for words and .13 for nonwords. As in Experiment 6, the difference in hits was significant, $F(1, 11) = 66.88$, $MSE = 14.39$, but the difference in false alarms was not, $F(1, 11) = 0.13$, $MSE = 9.56$.

Experiments 6 and 7 were conducted approximately 7 months after Experiments 1 and 3. Between-experiment comparisons should be done cautiously. Still, the general pattern between experiments is reasonable. Instructions to avoid giving positive responses as guesses had only small effects on hits and false alarms for words, presumably because participants rarely gave positive responses as guesses to these stimuli. However, hits and false alarms dropped dramatically for nonwords when the no-guessing instructions were used in Experiments 6 and 7.

Experiment 8

Experiments 6 and 7 found that the mirror effect on tests of order and associative information was eliminated when participants were asked not to give positive responses as guesses. Although the focus of this study has not been on mirror effects in recognition, this finding would be of greater interest if it could be shown on that task as well. In Experiment 8, two groups of participants each studied a single list of words and nonwords and were then given a yes–no recognition test. One group was not given any instruction about guessing; the second group was asked to avoid using positive responses as guesses.

Method

Participants. Twenty-four students from introductory psychology classes participated to fulfill a course requirement.

Procedure. All participants were asked to study a list for a recognition test and were then given a single list containing 30 words and 30 nonwords, as well as 10 untested buffers, divided equally between words and nonwords and placed in the first five and last five serial positions. The items were shown on a computer screen 1 at a time at a rate of 1 every 5 s. Immediately after the list was done, participants were given a yes–no recognition test with 60 items from the list, along with 30 new words and 30 new nonwords. Participants were asked for recognition judgments. Half of the participants were asked to avoid giving positive responses as guesses and to answer yes only when they thought that this was the correct answer; the other participants received no instructions about guessing.

The words and nonwords used all came from previous experiments. For each participant, the computer randomly determined which stimuli were to be used as targets and which as distractors and assigned items to input list positions and to test positions.

Results

In the condition in which participants received no specific instructions regarding guessing, a mirror effect was obtained. There were more hits to words (.88) than to nonwords (.58), $F(1, 11) = 27.49$, $MSE = 10.95$, and there were more false alarms to nonwords (.48) than to words (.20), $F(1, 11) = 19.51$, $MSE = 14.01$. In the condition in which participants were asked to avoid using positive responses as guesses, no mirror effect was found. There were still more hits to words (.87) than to nonwords (.27), $F(1, 11) = 83.53$, $MSE = 14.76$, but there was no longer a significant difference in false alarms between words (.19) and nonwords (.14), $F(1, 11) = 2.39$, $MSE = 3.92$.

Supplemental analyses were conducted to compare the groups. When numbers of hits were analyzed, there were significant effects of group, $F(1, 22) = 8.40$, $MSE = 20.54$, and type of item (word or nonword), $F(1, 22) = 107.05$, $MSE = 12.85$, and a significant interaction between these effects, $F(1, 22) = 12.27$, $MSE = 12.85$. Similarly, an analysis of the numbers of false alarms revealed significant effects of group, $F(1, 22) = 16.17$, $MSE = 13.94$, and type of item, $F(1, 22) = 10.12$, $MSE = 8.97$, as well as a significant interaction, $F(1, 22) = 21.44$, $MSE = 8.97$. The significant interactions in these two analyses reflected the fact that the groups differed more in their responses to nonwords than to words. Instructions to avoid using positive responses as guesses had little effect on responses to words because participants rarely gave positive responses as guesses to words; however, these instructions greatly lowered both hits and false alarms to nonwords.

A related demonstration may also be mentioned here. An additional 12 participants (recruited from an upper-level psychology course) received a list of 20 words and 20 nonwords. They then received a recognition test containing 20 old words, 20 old nonwords, 40 new word distractors, and 40 new nonword distractors. Participants were informed at the beginning of the test that there were twice as many distractors as old items on the test. This can be expected to reduce participants' willingness to use positive responses as guesses. Indeed, in this sample, there was a significant difference in hit rates between words (.80) and nonwords (.34), $F(1, 11) = 92.39$, $MSE = 5.56$, but the false-alarm rates of words (.10) and nonwords (.10) did not differ significantly ($F < 1.0$).

Experiment 9

All of the experiments discussed so far have used words and nonwords as stimuli. The response-distribution account may work particularly well with these stimuli because they are so easily discriminable and differ so greatly in memorability. Still, many previous investigations have used more subtle categories of stimuli, such as high- and low-frequency words or concrete and abstract words. Clearly, participants would be able to discriminate between words belonging to these categories. After all, concreteness classifications are directly based on participants' ratings (such as those of Paivio et al., 1968), and the frequency-estimation literature has consistently shown that participants are quite accurate in determining the relative frequencies of words in everyday life (e.g., Howes, 1954; Shapiro, 1969). Still, it is not obvious that participants would spontaneously notice when an experiment contains stimuli from two distinct categories based on frequency or concreteness or that they would monitor how often they respond positively to each category. Therefore, it appeared worthwhile to demonstrate that results like those obtained in Experiment 8 could be found in an experiment examining recognition of low- and high-frequency words.

Method

Participants. Twenty-four students from introductory psychology classes participated to fulfill a course requirement.

Materials. This experiment compared low- and high-frequency words. The low-frequency words were those used in the previous experiments here, with frequencies of 1–5 occurrences per million words (Thorndike & Lorge, 1944). Sixty high-frequency words (frequency of 100–150 occurrences per million; Thorndike & Lorge, 1944) were chosen so that these words would be equivalent to the 60 low-frequency words on concreteness (5.08; Paivio et al., 1968), length (5.2 letters), and syllables per word (2.1).

Design and procedure. In all other respects, the method was identical to that of Experiment 8. Participants saw a list containing 60 critical items and 10 buffers and then received a recognition test on 120 items. Half of the participants received no particular instruction about guessing strategies, and the other participants were asked to avoid using positive responses as guesses.

Results

When no instructions about guessing were given, participants displayed a mirror effect. In this group, there were significantly more hits to low-frequency words (.83) than to high-frequency words (.71), $F(1, 11) = 5.52$, $MSE = 13.95$, and significantly more false alarms to high-frequency words (.25) than to low-frequency words (.11), $F(1, 11) = 16.53$, $MSE = 6.56$. In the group in which participants were asked to avoid using positive responses as guesses, the mirror effect was eliminated; there were still significantly more hits to low-frequency words (.78) than to high-frequency words (.57), $F(1, 11) = 15.73$, $MSE = 15.30$, but the difference in false alarms between high-frequency (.09) and low-frequency (.08) words did not approach significance, $F(1, 11) = 0.30$, $MSE = 5.05$. These results are comparable to those found in Experiment 8, in which words and nonwords were compared.

The groups were compared in supplementary analyses. When hit rates were analyzed, the only significant effect was type of word (low- or high-frequency), $F(1, 22) = 20.17$, $MSE = 22.69$. When false alarms were analyzed, there were significant effects of group, $F(1, 22) = 6.95$, $MSE = 15.11$, and type of word, $F(1, 22) = 11.67$, $MSE = 5.80$. There was also a significant interaction between these variables, $F(1, 22) = 7.27$, $MSE = 5.80$, indicating that guessing instructions had a greater impact on high-frequency words than on low-frequency words.

Conclusion

The generality of a phenomenon may sometimes be seen as evidence for its importance. However, the generality of the mirror effect may be evidence for its unimportance in evaluating theories of memory. The fact that mirror effects may be found on tests of order and associative information presents difficulties for accounts that attribute it to processes that are not easily extended beyond recognition. Moreover, it raises the possibility that this phenomenon may not reflect specific processes in memory but may instead be due to general response strategies that can be applied to a wide variety of tasks and presumably added to a wide variety of models.

Of course, it is possible that there is no one explanation for all occurrences of mirror effects. The response-distribution account proposed here may apply in some situations but not in others. Particularly critical will be whether mirror effects can be demonstrated in situations where no participants are able to discriminate between the two stimulus categories being used. Such a demonstration would be clear evidence that the response-distribution account could not be a complete account of mirror effects. However, the account proposed here does lead to the conclusion that some (and possibly all) occurrences of mirror effects may be due to seemingly uninteresting response strategies.

References

Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency, and negative recognition. *Quarterly Journal of Experimental Psychology, 29,* 461–473.

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition, 20,* 231–243.

Clark, S. E., & Burchett, R. E. R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition, 22,* 55–62.

Clark, S. E., & Shiffrin, R. M. (1992). Cuing effects and associative information in recognition memory. *Memory & Cognition, 20,* 580–598.

Flexser, A. J., & Bower, G. H. (1974). How frequency affects recency judgments: A model of recency discrimination. *Journal of Experimental Psychology, 103,* 706–716.

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

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16.

Glanzer, M., Adams, J. K., & Iverson, G. (1991). Forgetting and the mirror effect in recognition memory: Concentering of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 81–93.

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.

Greene, R. L., & Thapar, A. (1994). Mirror effect in frequency discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 946–952.

Gregg, V. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183–216). New York: Wiley.

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative recognition: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 846–858.

Hacker, M. J. (1980). Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 651–675.

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

Hintzman, D. L. (1994). On explaining the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 201–205.

Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 275–289.

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

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 302–313.

Hockley, W. E. (1991). Recognition memory for item and associative information: A comparison of forgetting rates. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 227–248). Hillsdale, NJ: Erlbaum.

Hockley, W. E. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition, 22,* 713–722.

Hoshino, Y. (1991). A bias in favor of the positive response to high-frequency words in recognition memory. *Memory & Cognition, 19,* 607–616.

Howes, D. (1954). On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology, 48,* 106–112.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General, 122,* 291–315.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement, 76,* 1–25.

Shapiro, B. J. (1969). The subjective estimation of relative word frequency. *Journal of Verbal Learning and Verbal Behavior, 8,* 248–251.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words.* New York: Columbia University Press.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 681–690.

Received December 1, 1994
Revision received March 29, 1995
Accepted April 19, 1995 ■