

# What information underlies correct rejections in short-term recognition memory?

E. E. JOHNS and D. J. K. MEWHORT  
Queen's University, Kingston, Ontario, Canada

We propose that correct rejections are based on information that contradicts the study set, rather than on insufficient familiarity. Using two-dimensional stimuli, we varied the featural overlap between lures and the study set so that one feature of the lure had occurred during study and the second feature of the lure had not occurred. Familiarity varied with the number of times the studied feature had occurred, whereas detectability of the extralist feature varied with the number of studied alternatives on the same dimension. Lures increased in difficulty with familiarity only when familiarity of the studied feature was confounded with the number of alternatives to the extralist feature. Hence, the difficulty in rejecting the lure was controlled by the number of alternatives to the extralist feature, not by the familiarity of the lure. Current theory requires additional representation assumptions and a new comparison process to accommodate these data.

In typical experiments on recognition from short-term memory, subjects study a list of items and then indicate whether a test item had occurred in the study list. Although experiments normally include equal numbers of test items that had occurred in the study list (*positive* items) and those that had not occurred in the study list (*negative* items, *lures*, or *distractors*), accounts of recognition have focused on how list items are recognized. As a result, *no* is treated as a default response to be used whenever the evidence is insufficient to justify saying *yes*.

Early scanning models (e.g., Murdock, 1971; Sternberg, 1969, 1975) proposed that subjects search the study set, saying *yes* if they find the probe and *no* otherwise. More recent models have perpetuated the no-as-default idea. They base decision on a mechanism derived from signal detection theory (Green & Swets, 1966) operating on a unidimensional value (called echo intensity by Hintzman [1984], dot-product of the probe and memory vectors by Murdock [1982, 1983], strength by McElree & Doshier [1989], familiarity by Raaijmakers & Shiffrin [1981], log odds for positive by Shiffrin & Steyvers [1997], etc.); we shall refer to the value as familiarity. A test item elicits a *yes* response if the familiarity value lies above an arbitrary criterion and *no* otherwise (see Clark & Gronlund, 1996; Neath, 1998, pp. 210–216). Hence, any factor that promotes a *yes* response must inhibit a *no* response, and vice versa.

Because theory has focused on identification of targets, there is relatively little work on lure rejection per se. We do know that lures are increasingly difficult to reject as their similarity to studied items increases. Familiarity models place the locus of similarity effects in the retrieval, not the encoding, stage: They explain the effect by assuming that the probe's similarity to the study set directly increases its calculated familiarity. The closer the calculated familiarity lies to the criterion, the more difficult it should be to reject.

Mewhort and Johns (2000) proposed an alternative view—namely, that *no* is a deliberate choice that is given when an aspect of the probe contradicts the memorized items. Contradictory evidence need not be perfectly correlated with the evidence that underlies a *yes* decision. In this article, we adopt their methods to further the argument.

Two aspects of the experimental technique are pertinent. First, the stimuli comprise a small number of features, each with a finite set of values (low-dimensionality stimuli). We vary the similarity of the study and the test items by manipulating their featural overlap. Second, we use subspan lists and discard trials on which the subject categorizes his response as *unsure*. These methods ensure that we are testing retrieval in a situation with accurate encoding (see Sternberg, 1969). We measure reaction time (RT) and assume that RT increases with decision difficulty (cf. Murdock, 1985).

Using two-dimensional stimuli, Mewhort and Johns (2000) compared performance on two types of distractors. Each distractor shared one feature with each of two study items, so that they were equivalent in summed similarity to the study list; that is, we held familiarity constant. Performance was better when the distractor overlapped two study items by the same feature than when the distractor overlapped two study items by two different features. For example, a red star that overlapped a red triangle and a

---

The research was supported by a grant to the second author from the Natural Science and Engineering Research Council of Canada. Some of the data were reported in a paper read at the 39th Annual Meeting of the Psychonomic Society, Dallas, November, 1998. Correspondence concerning this article should be addressed to D. J. K. Mewhort, Department of Psychology, Queen's University, Kingston, ON, K7L 3N6 Canada (e-mail: doug@ebbinghaus.psy.queensu.ca or beth@vip.psy.queensu.ca).

—Accepted by previous editorial team

red donut was easier to reject than a red star that overlapped a red triangle and a yellow star, although the familiarity value would be the same for both cases. We argued that subjects use the feature that did not occur in the study set (the extralist feature) as evidence for a *no* decision. Conversely, when we held the number of extralist features constant and varied the number of occurrences in the study set of the studied feature, performance did not vary. We argued that a *no* response occurs if there is enough negative evidence, regardless of the amount of positive evidence.

The data thus far show an effect of the number of extralist features when the standard measure of familiarity is held constant: clear evidence of a role for contradiction in recognition decisions. The failure to find an effect of familiarity is less conclusive. Small numerical differences in the predicted direction did occur, and the conclusion requires asserting the null hypothesis. Experiment 1 in the present paper, therefore, provides a new test of familiarity when the number of extralist features is held constant.

### EXPERIMENT 1

The experiment used a varied-set Sternberg (1969) procedure with set size constant at two items. The stimuli were chosen from a pool of 512 visual items, created from the factorial combination of eight shapes, eight colors, and eight fill patterns. The shapes and fill patterns are illustrated in Figure 1; eight colors were used: red, blue, green, purple, pink, brown, gray, and yellow.

On each trial, the two study items were selected so as to have the same value on one, and only one, dimension; that is, the study items had the same shape or the same color or were filled in with the same pattern. Using uppercase letters to represent the values on the three dimensions, the study items had the structure AAA and ABB.

In the experiment, we compared two types of negative probes; the probes were defined by the way they overlapped the studied items. We label the negative conditions with three digits representing the number of times each dimension had occurred in the study set. For convenience, the first digit refers to the dimension on which the study items took a common value, rather than to a particular dimension. A Condition 0:1:1 probe had a new value on the dimension shared by the study items, combined with the values of one of the study items for the other two dimensions (i.e., an XAA or XBB probe). A Condition 2:1:0 probe had the common value on the dimension shared by the study items, combined with a value from one of the study items on the second dimension and a new value on the third dimension (i.e., an AAX, ABX, AXA, or AXB probe). Our definitions treat each trial as an independent unit; hence, a new value is one that had not yet occurred on the trial, not one that had not occurred earlier in the experiment.

Figure 2 provides two examples that illustrate the relation between the features of the study and the test items.

The first study set is composed of two crosses: One is blue and filled with a solid pattern; the other is red and filled with a striped pattern. Shape is the dimension in common for the two items. Sample negative probes for the study set are shown to the right of the study set. The red striped heart illustrates the 0:1:1 condition. It shares color (red) and pattern (striped) with the right-hand study item, but it has a new value (heart) on the common dimension. The green striped cross (on the right) illustrates the 2:1:0 condition; it shares shape (cross) with both studied items and pattern (striped) with the right-hand study item, but the green color is unstudied. In the second study set, the study items share color; one is a red striped cross, and the other is a red solid moon. For the second study set, the green striped cross serves as a Condition 0:1:1 probe. It shares shape (cross) and pattern (striped) with the left-hand study item, but it has a new value (green) on the common dimension. The red striped heart serves as a Condition 2:1:0 probe. It shares color (red) with both study items and pattern (striped) with the left-hand study item, but its shape (heart) was not studied.

The two negative conditions define the similarity manipulation. By common measures of similarity—the number of items overlapped (one for the 0:1:1 condition and two for the 2:1:0 condition) or the number of times that its features had occurred in the study set (two for the 0:1:1 condition and three for the 2:1:0 condition)—the 2:1:0 negative is more like the study set than is the 0:1:1 negative and, hence, should be more difficult to reject.

### Method

**Subjects.** The subjects were 7 volunteers who participated in return for payment of \$5. All had normal or corrected-to-normal vision and none was color blind.

**Apparatus.** The experiment was controlled by an MS-DOS PC with VGA color graphics. The subjects responded by pushing micro-switches attached to the games port of the computer; three buttons were used for *start*, *yes*, and *no*. The subjects used the thumb or index finger of the nondominant hand on the *start* button and the index and middle fingers of the dominant hand for the *yes* and *no* response buttons, respectively. Millisecond timing was achieved, using a version of Heathcote's (1988) timing routines.

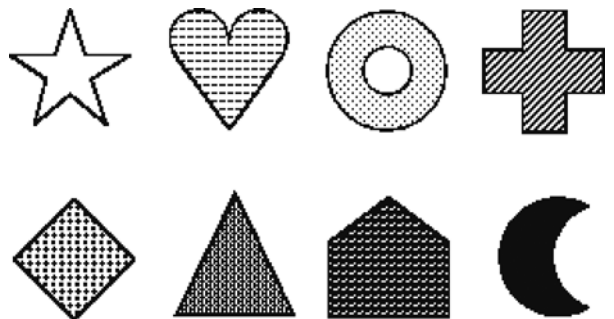


Figure 1. Sample stimuli illustrating the eight fill patterns and the eight shapes. The samples shown are a hollow star, a lined heart, a dotted donut, a striped cross, a checkered diamond, a lacy triangle, a wavy house, and a solid moon.

**Stimuli.** Each shape was drawn within a square measuring 120 × 120 pixels on a screen measuring 640 pixels wide × 480 pixels high; each shape had a maximal extent of about 4.5 cm horizontally and vertically.

The colors were chosen, subjectively, to be discriminable without varying appreciably in luminance. By their common verbal labels, with RGB values in parentheses, the colors used were gray (*RGB*), blue (*GbB*), purple (*RbB*), pink (*rRB*), red (*rR*), brown (*Rg*), yellow (*rRG*), and green (*gGb*), where *r* indicates the presence of low-intensity red, *R* indicates high-intensity red, and so forth. Stimuli appeared on a black background.

**Procedure.** The subjects were tested individually, in 1-h sessions. The session began with a preview of the stimuli. First, the eight shapes were shown together. The subject confirmed that the shapes were discriminable by naming them; the experimenter suggested alternatives for awkward names (e.g., *ring*, *wheel*, or *donut* instead of *lifesaver*). Next, the eight colors were displayed, and the subject confirmed the detectability of the colors by giving them names. Lastly, the eight patterns were displayed and named.

Next, the subject was tested on a long series of trials. Each trial began with the message "When ready, press **START**" written in white in the center of the screen. After the start button had been pressed, the two study items were displayed, one at a time, each centered on the screen. The second study item was followed by a mask 120 pixels square. Each pixel was painted in one of the eight colors chosen at random. The study items and the mask each appeared for 1,750 msec; a blank screen between displays lasted for 250 msec. The test probe appeared next; it remained until a response button was pressed. The message "Are you sure?" followed in white at the center of the screen, remaining until a response button was pressed. The confidence judgment was followed by the "When ready, press **START**" message to begin the next trial.

The subjects were fully informed as to the sequence of events on each trial. They were told that the names they had used during preview gave them the option of naming the study items but that they should adopt whatever study technique they thought would work. They were advised that the mask signaled the end of the memory set and the approach of the test item and that they should not close their eyes. The subjects were asked to try for high accuracy. They were also informed that RT was being recorded for the recognition decisions, so they should respond as quickly as they could without making errors.

Trials were organized into five blocks: a block of 20 warm-up trials, followed by four blocks of 48 experimental trials each. The 20 warm-up trials were randomly selected from a fifth experimental block; warm-up data are not included in any analysis. At the end of each block, the number of errors on that block was displayed. Trials within each block were presented in random order.

**Study and test conditions.** Each experimental block included 24 positive tests and 24 negative tests in random order. The 24 positive trials included 8 trials in which the common feature between the two study items was color, 8 in which it was shape, and 8 in which it was pattern. Within each of the three types of trials, both serial positions were tested equally often.

The 24 negative trials also subdivided into three sets of eight trials, depending on which feature was shared by the two study items. Each set of eight trials comprised four tests in the 0:1:1 condition and four tests in the 2:1:0 condition. In the 0:1:1 condition, the studied features were drawn from the item in each serial position equally often. In the 2:1:0 condition, there were two tests that had a new value on each of the two unshared dimensions; for example, if the study items shared shape, two 2:1:0 probes would have the new value on color, and the other two would have the new value on pattern.

## Results and Discussion

Table 1 summarizes performance on positives and the two types of negative probes; the subscripts on the con-

dition labels will be explained below. We report means and standard deviations (based on the subject means) for five different measures: the percentage of correct responses, the mean RT for correct responses, the percentage of responses called *sure*, the percentage of *sure* responses that were correct, and the mean RT for correct *sure* responses.

Accuracy was sufficiently high, 95.9%, to justify analyses of RT data. The correct trials, however, included several cases in which the subject indicated that he or she was not sure. We cannot assume correct encoding on *not sure* trials or on errors. Accordingly, our preferred measure is RT for correct *sure* trials. We report the data on a variety of measures, however, to document that, in all the experiments, the measures agree: The slowest RTs are associated with the least accurate conditions.

In Table 1, all five measures suggest that it was easier to reject a 0:1:1 probe than a 2:1:0 probe. Correct *sure* responses were 153 msec faster for 0:1:1 negative probes than for 2:1:0 negatives [ $F(1,6) = 11.90, p < .05$ ].

The results support the role of familiarity in recognition: By standard measures of familiarity—by counting the number of times the features of the probe occurred in the study set (three vs. two) or the number of items in the study set that overlap the probe (two vs. one)—the 2:1:0 probe is more similar to the study set than is the 0:1:1 probe. Hence, the 2:1:0 probe should be more difficult to reject than the 0:1:1 probe, and it was.

Familiarity is not the only way to explain the difference, however. Even though the 0:1:1 and the 2:1:0 probes each included one and only one extralist feature, we contend that the contradictory information was not equally apparent in the two conditions. In the 0:1:1 condition, the extralist feature occurred on the dimension with a single value in the study set, whereas in the 2:1:0 condition, the extralist feature occurred on a dimension with two different values in the study set. We suggest that it is easier to detect a contradiction when an extralist feature contrasts with a single value in the study set than when it contrasts with a set of values.

In the examples in Figure 2, the extralist feature in the red striped heart is the heart shape. When the red striped

**Table 1**  
Summary of Performance in Experiment 1

Measure	Probe Type		
	Positive	0 <sub>1</sub> :1:1	2:1:0 <sub>2</sub>
Mean correct trials (%)			
<i>M</i>	94.8	97.3	96.7
<i>SD</i>	4.5	2.0	2.4
Mean reaction time for correct trials (msec)			
<i>M</i>	1,006	942	1,111
<i>SD</i>	207	182	254
Mean trials classified as <i>sure</i> (%)			
<i>M</i>	97.3	98.8	97.6
<i>SD</i>	2.3	1.1	3.1
Mean correct trials ( <i>sure</i> only) (%)			
<i>M</i>	96.0	98.5	97.9
<i>SD</i>	4.6	1.6	2.1
Mean reaction time for correct <i>sure</i> trials (msec)			
<i>M</i>	992	942	1,095
<i>SD</i>	205	182	227

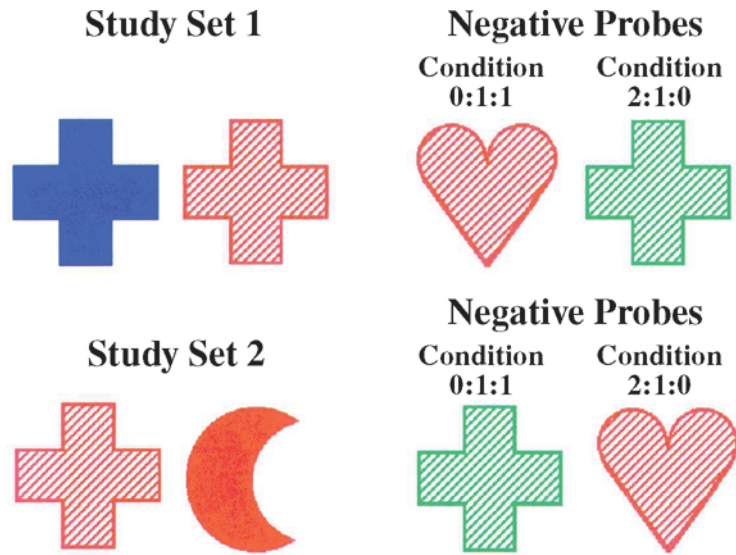
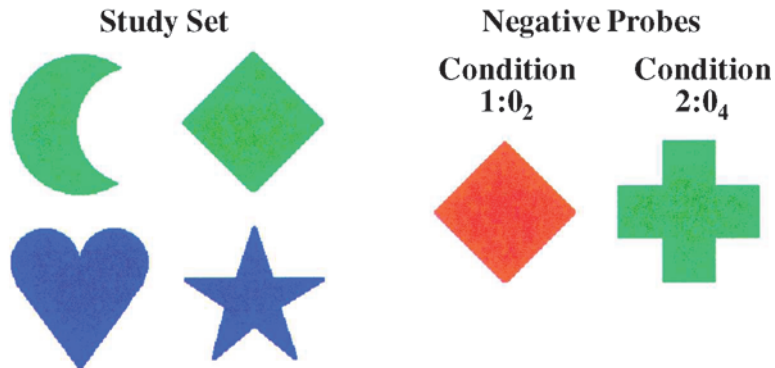


Figure 2. Sample study sets for Experiment 1, with illustrations of negative probes for Conditions 0:1:1 and 2:1:0.

### Unidimensional Organization



### Bidimensional Organization

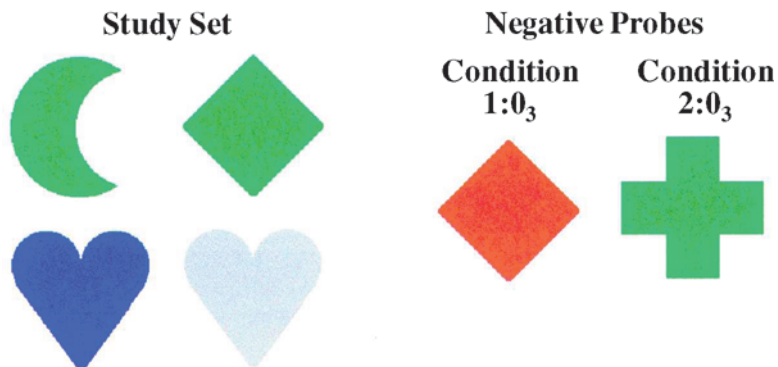


Figure 3. Sample study sets for Experiment 2 illustrating uni- and bidimensional organization and negative probes for Conditions 1:0 and 2:0.

heart is the 0:1:1 probe (Study Set 1), the heart shape mismatches the cross shape, whereas when it is the 2:1:0 probe (Study Set 2), the heart shape mismatches the cross and moon shapes. Our position is that the advantage for Condition 0:1:1 may occur because it is easier to see that a heart is not in a study set composed of crosses than it is to see that a heart is not in a study set composed of a cross and a moon.

By this analysis, the 0:1:1 and 2:1:0 conditions of Experiment 1 differ on two perfectly confounded factors: the number of study presentations of the studied feature and the number of alternatives to the extralist feature. To make the confound clear, we will indicate the number of studied alternatives to the extralist feature in each negative condition with a subscript within the condition label; for Experiment 1, the negative conditions were  $0_1:1:1$  and  $2:1:0_2$ . Familiarity theories maintain that decision difficulty reflects the familiarity of the studied feature within the probe. We suggest that decision difficulty reflects the number of alternatives to the extralist feature.

The contradiction hypothesis claims that correct rejections depend on the detectability of contradictory evidence. The number of extralist features has been shown to be one factor that affects detectability. The number of alternatives to the extralist feature is proposed as a second factor: An extralist feature should be easier to detect the fewer the studied values on that dimension.

## EXPERIMENT 2

The second experiment was designed to separate the factors that were confounded in the first—namely, the familiarity of the probe and the number of studied alternatives to the extralist feature contained in the probe.

On each trial of Experiment 2, the subjects studied four objects defined by color and shape; the four items consisted of two pairs of items. Items within a pair shared either the same color or the same shape. The negative probes always contained one familiar (i.e., studied) feature and one extralist feature. We varied familiarity by choosing a familiar feature that had occurred either once or twice in the study set. Thus, our two basic negative conditions were 1:0 and 2:0 (the order of the digits is not significant).

We varied the number of alternatives to the extralist feature by manipulating the organization of the study set. Under unidimensional organization, each pair of related items shared a value on a single organizing dimension. Four different values were used on the nonorganizing dimension; the structure of the study set was AA, AB, BC, BD. Thus, if two colors were used, four different shapes were used; if two shapes were used, four different colors were used. Under bidimensional organization, the pairs of related items shared values on different dimensions; that is, there were two items that were one color and two that were one shape (AA, AB, BC, and CC). With this arrangement, there are three feature values on each of the two dimensions.

Figure 3 shows examples of the two organizations. The unidimensional organization is shown on the top, where the four study items are arranged into pairs by color, green

and blue, respectively. The bidimensional organization is shown below where the four study items are arranged into pairs, one defined by color (two green objects) and one defined by shape (two hearts).

Now consider the structure of negative probes under unidimensional organization. The 1:0 probe combines a studied feature from the nonorganizing dimension with an extralist feature on the organizing dimension (i.e., XA, XB, XC, and XD). The extralist feature mismatches the two values from the organizing dimension, so the complete designation of the condition is  $1:0_2$ . The 2:0 probe combines a studied feature from the organizing dimension with an extralist feature on the nonorganizing dimension (i.e., AX and BX). The extralist feature mismatches the four values from the nonorganizing dimension, so the complete designation of the condition is  $2:0_4$ .

In the example in Figure 3, the  $1:0_2$  probe consists of a red diamond. The diamond shape occurred once in the study set. Red did not occur, and there are two alternatives to red in the study set—namely, green and blue. The  $2:0_4$  probe consists of a green cross. Green occurred twice in the study set. Cross did not occur in the study set and has four alternatives in the study set (moon, diamond, heart, and star).

Next, consider the structure of the same negative probes under bidimensional organization. Neither dimension can be called organizing or nonorganizing, and there are three studied values on each dimension. Thus, if one combines an extralist feature with a studied feature that has been seen once, to make Condition 1:0 (i.e., XA, XB, CX, or DX), there will be three studied alternatives to the extralist feature. Similarly, if one combines an extralist feature with a studied feature that has been seen twice, to make Condition 2:0 (i.e., AX or XC), there will be three studied alternatives to the extralist feature. The conditions under bidimensional organization are completely designated as  $1:0_3$  and  $2:0_3$ .

From Figure 3, under bidimensional organization, the  $1:0_3$  probe consists of a red diamond. The diamond shape occurred once in the study set. Red did not occur, and there are three alternatives to red in the study set—namely, green, blue, and gray. The  $2:0_3$  probe consists of a green cross. Green occurred twice in the study set. Cross did not occur in the study set and has three alternatives in the study set (moon, diamond, and heart).

The logic of the experiment is now evident. Like the design of Experiment 1, the unidimensional condition perfectly confounds the number of study presentations of the familiar feature with the number of alternatives to the extralist feature. Both the familiarity and the contradiction accounts anticipate that Condition  $1:0_2$  should be easier than Condition  $2:0_4$ . In the bidimensional condition, however, the familiarity of the studied feature changes, whereas the detectability of the extralist feature remains constant. If the difference between Conditions  $1:0_2$  and  $2:0_4$  depends on the familiarity of the studied feature, the same difference should occur between Conditions  $1:0_3$  and  $2:0_3$ . If it depends on the detectability of the extralist fea-

ture, there should be no difference between Conditions 1:0<sub>3</sub> and 2:0<sub>3</sub>. That is, the familiarity view predicts a main effect for Conditions 1:0 and 2:0, whereas the contradiction view predicts an interaction of that factor with the organization of the study set.

## Method

**Subjects.** Twenty students from an introductory psychology course participated in the experiment in return for bonus credit in the course. All the subjects had normal or corrected-to-normal vision; none was color-blind. Ten subjects were assigned randomly to each of two groups defined by the organization of the study sets used, either bidimensional or unidimensional.

**Apparatus and Stimuli.** The apparatus was largely the same as that in Experiment 1, but a different set of response buttons were used. Half the subjects in each group used the index finger of the dominant hand on the *yes* button and the index finger of the nondominant hand on the *no* button, with either thumb on the *start* button; the remaining subjects used the opposite arrangement.

The stimuli were constructed by combining the eight shapes and eight colors used in Experiment 1. Fill pattern was not varied; all the shapes were solidly colored. Each shape was two thirds of its size in Experiment 1.

**Procedure.** The procedure was similar to that in Experiment 1, except that the study items were displayed simultaneously. We used a simultaneous display to encourage the subjects to capitalize on the relations among the features within the set. After the *start* button was pressed, the four items appeared for 8 sec. Each of the study items was centered at one corner of an imaginary square measuring 7.1 × 7.1 cm, leaving a blank central square that measured 3.8 × 3.8 cm. Items with the same color or shape were aligned horizontally on half of the trials and vertically on the other half. The study set was followed by a blank screen for 250 msec; the blank screen was followed by a square mask for 2 sec. The mask covered the entire area in which study and test items appeared (10.7 × 10.7 cm), with each pixel painted in one of the eight colors chosen at random.

The test probe and request for a confidence rating were the same as those in Experiment 1, but we added feedback. If the recognition response was incorrect, a quick sequence of descending tones (a "groan") was given; no feedback was given for correct responses.

The subjects were fully informed as to the sequence of events on each trial. Their instructions were the same as those in Experiment 1.

**Study and test conditions.** Each subject completed 16 warm-up trials, followed by four blocks, with 32 experimental trials per block. Within each experimental block, there were equal numbers of positive and negative trials within each condition; color and shape were assigned equally often as the shared feature for the positive probes

and as the familiar feature for the negative probes. The trials were presented in a random order. Across the 128 experimental trials, each possible configuration of the display was used equally often, and the display position for positive items and for the familiar feature of negative items was drawn equally often from each position within each configuration.

## Results and Discussion

The results are summarized in Table 2. The various measures agree as to the pattern of results, and, as in Experiment 1, we report analyses of RTs for correct *sure* responses. RTs for positives were equivalent under unidimensional and bidimensional organization ( $F < 1$ ). For negatives, the interaction of familiarity (the basic 1:0/2:0 comparison) with organization (uni- or bidimensional) was highly reliable [ $F(1,18) = 17.83, p < .001$ ]. The interaction is illustrated in Figure 4. Separate contrasts confirmed that, with unidimensional study sets, the subjects rejected 1:0<sub>2</sub> probes more quickly than 2:0<sub>4</sub> probes [ $F(1,9) = 29.61, p < .001$ ]. With bidimensional study sets, by contrast, there was no difference between the 1:0<sub>3</sub> and the 2:0<sub>3</sub> probes ( $F < 1$ ).

The results are unambiguous. With unidimensional organization, we replicated the similarity effect found in Experiment 1: Lures increased in difficulty with increased similarity to the studied items. Because the organization confounded familiarity of the studied feature with number of studied alternatives to the unstudied feature, the mechanism underlying the similarity effect cannot be determined in this condition. With bidimensional organization, where familiarity of the studied feature varied but the number of alternatives to the unstudied feature was held constant, no similarity effect was obtained. Hence, the obtained similarity effect depends on the number of alternatives to the extralist feature, not on conventional familiarity.

Mewhort and Johns (2000) also found no effect of familiarity. They held the number of extralist features constant and varied the familiarity of the studied feature (the 1:0 vs. 2:0 comparison in their Experiment 2 and the 1:0/0 vs. 2:0/0 comparison in their Experiment 4). They acknowledged, however, that the numerical differences between the conditions lay in the direction predicted by fa-

**Table 2**  
Summary of Performance in Experiment 2

Probe Type	Measure									
	Correct Trials (%)		Reaction Time for Correct Trials (msec)		Trials Classified as <i>Sure</i> (%)		Correct Trials ( <i>Sure</i> Only) (%)		Reaction Time for Correct <i>Sure</i> Trials (msec)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Unidimensional Organization										
Positive	88.9	4.9	1,746	494	88.4	8.8	91.6	4.5	1,615	398
1:0 <sub>2</sub>	97.8	3.3	1,251	224	97.5	3.6	98.7	2.4	1,252	238
2:0 <sub>4</sub>	97.5	2.5	1,590	341	94.7	6.8	98.7	1.7	1,549	345
Bidimensional Organization										
Positive	92.4	4.2	1,604	552	94.2	5.9	93.8	3.8	1,516	401
1:0 <sub>3</sub>	97.8	3.3	1,491	457	97.2	5.0	98.7	3.0	1,481	427
2:0 <sub>3</sub>	96.6	3.4	1,456	399	97.2	5.0	97.7	2.2	1,440	379

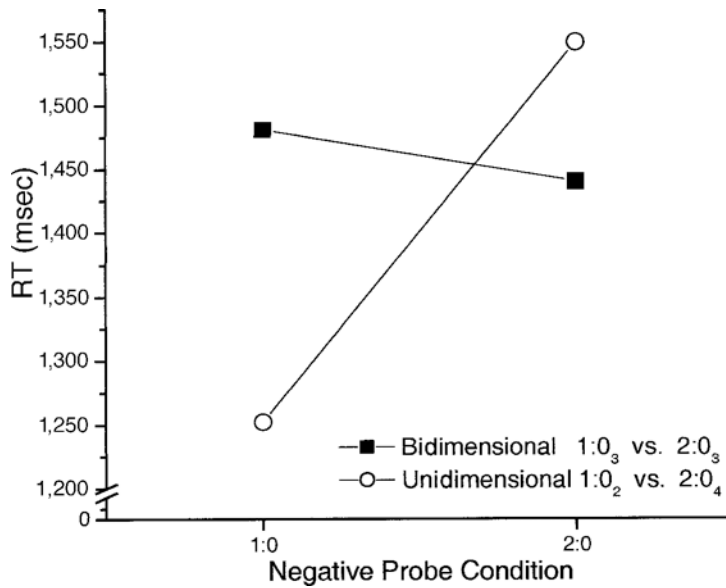


Figure 4. Mean reaction time (RT) for correct *sure* responses as a function of the organization of the study set (uni- or bidimensional) and negative probe condition (Condition 1:0 and 2:0) in Experiment 2.

miliarity theory and that the conclusion required us to affirm the null hypothesis. The present experiment provides stronger evidence against familiarity theory, because the evidence hinges on an interaction, not on a null effect. Furthermore, in the unconfounded bidimensional condition, the 2:0 probes were numerically faster than the 1:0 probes.

In our view, the detectability of contradictory information is the critical psychological variable. The manipulations of number of extralist features and of alternatives to an extralist feature both map onto detectability. It is worth noting that neither the number of extralist features nor the number of alternatives to the extralist feature has a direct parallel for *yes* responses. By definition, extralist information affects only negative responses.

Because our experiments favor a contradiction view over the familiarity view, it is important to establish that they do not represent a special case lying outside the domain of traditional theory. Our last two experiments address this issue.

### EXPERIMENT 3

Our experiments might lie outside general theory if the experimental manipulations induced idiosyncratic processing strategies on the part of the subjects. Our experiments are unusual in that the stimuli are easily decomposed into features and the relation between study and test items is explicit and limited. Subjects might, therefore, adopt a feature-by-feature processing strategy, one that could not be applied to traditional experiments using verbal stimuli. In Experiment 3, we tested whether subjects were using a directed-search strategy, a retrieval mechanism that has some empirical antecedents.

Directed search was explored in the heyday of scanning theory. Scanning was originally proposed for cases in which the study set lay below memory span. For longer lists, it was thought that subjects might still scan but restrict the scan to a relevant subset of the list (e.g., Clifton & Gutschera, 1971; Homa, 1973; Johns, 1985; Kaminsky & DeRosa, 1972; Naus, Glucksberg, & Ornstein, 1972). For example, suppose that a study list was composed of items from semantic categories. Subjects might categorize the test probe and then scan the list of categories represented in the study set. The subject would proceed to scan individual items only if the probe's category label were found in the list of categories. The scenario predicts an additive effect of number of categories and number of items in the relevant subcategory. Although the studies cited gave modest support for directed search, the evidence never coalesced into a convincing theory.

Nevertheless, a variant of the directed-search idea offers a possible explanation for the results of the experiments reported here. According to directed search, subjects should examine the organizing dimension first, checking the non-organizing dimension only if a match is found on the organizing dimension. By this theory, in Experiment 1, the subjects should first check a probe for a match to the shared feature in the study set, proceeding to check other features only if the shared feature appears in the probe. A 0<sub>1</sub>:1:1 probe should be rejected after the first check, but a 2:1:0<sub>2</sub> probe will require further processing; thus, the 2:1:0<sub>2</sub> probe is predicted to take longer to reject.

Experiment 2 presents a more complicated situation. Under unidimensional organization, it is quite clear what the organizing dimension is: It is the dimension that takes two values in the study set (e.g., color in Figure 3). With unidimensional organization, when a 1:0<sub>2</sub> probe is pro-

cessed, no match will be found on the organizing dimension, and a *no* response can be issued without checking the other dimension. For a 2:0<sub>4</sub> probe, by contrast, a match will be found on the organizing dimension, and the second dimension must be examined for a match. Hence, the 2:0<sub>4</sub> probe should be slower than the 1:0<sub>2</sub> probe, as was obtained with unidimensional organization in Experiment 2.

With bidimensional organization, by contrast, both dimensions take three values in the study set, and the dimensions are equally likely to be used as the organizing dimension. Subjects might choose to organize a set on one or the other dimension because of the order of stimulus presentation, because of their personal biases, or simply at random. Indeed, with a four-item set, they might choose not to organize the set. There is no way of knowing which dimension will be checked first. For both the 1:0<sub>3</sub> and the 2:0<sub>3</sub> probes, the first-checked dimension might yield a mismatch, allowing a rapid *no* response, or it might yield a match, requiring processing of the other dimension. Thus, under bidimensional organization, the directed-search strategy may benefit the 1:0<sub>3</sub> and the 2:0<sub>3</sub> probes equally; there will be no difference between them, just as obtained in Experiment 2.

Experiment 3 was designed to rule out an account based on the directed-search idea. The subjects studied two objects defined in three dimensions. One feature was shared among the items; that is, as in Experiment 1, the study set had the structure AAA and ABB. There were four types of negative probe: 0:0:0, 0:1:0, 0:1:1, and 2:0:0. To simplify the text, for the time being, we have omitted the subscripts denoting the number of alternatives to the extralist feature. As in Experiment 1, the first digit refers to the dimension

on which the study items took a common value. A Condition 0:0:0 probe had new values on all three dimensions (i.e., XXX). A Condition 0:1:0 probe had a new value on the dimension shared by the study items and on one other dimension, combined with the value of one of the study items for the third dimension (i.e., XAX, XXA, XBX, or XXB). A Condition 0:1:1 probe had a new value on the dimension shared by the study items, combined with the values of one of the study items for the other two dimensions (i.e., XAB or XBA). A Condition 2:0:0 probe had the value shared by the study items on the common dimension, combined with new values on the other two dimensions (i.e., AXX).

Figure 5 shows a sample study set with a possible probe for each negative condition. The two study items share color (blue) but have different values on shape (cross and moon) and on pattern, (solid and striped). The Condition 0:0:0 probe has new values on all three dimensions (red lacy heart). The Condition 0:1:0 probe has new values on two dimensions (red and lacy) but shares its moon shape with the second study item. The Condition 0:1:1 probe has a new value on color (red) but shares pattern (solid) with the first study item and shape (moon) with the second item. Hence, the first three conditions all include a feature (color) that mismatches the feature shared by the study items. The Condition 2:0:0 probe includes the value of the shared feature (blue) but has new values on pattern and shape (lacy heart).

The directed-search idea claims that subjects check one dimension first and exit immediately if it is contradicted. In keeping with the familiarity advantage offered for the unidimensional organization, the feature common

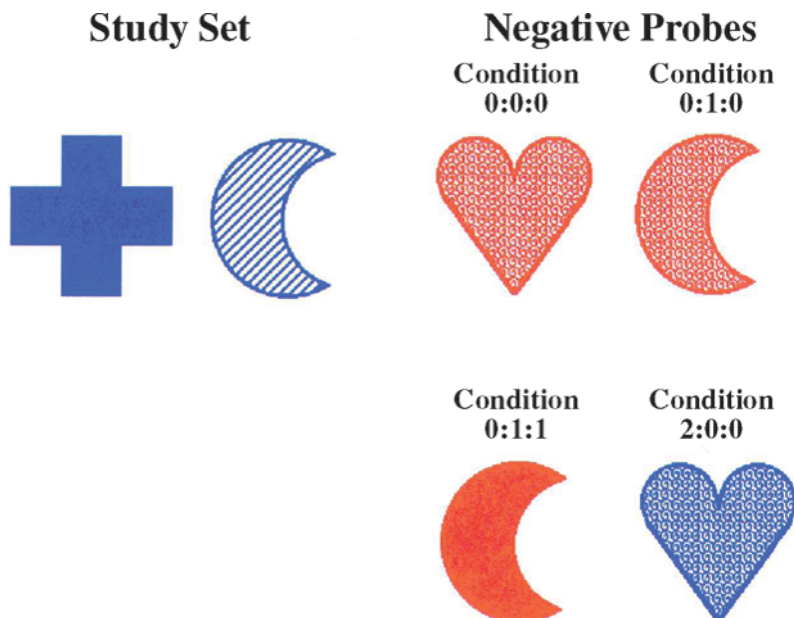


Figure 5. A sample study set for Experiment 3 with illustrations of negative probes for Conditions 0:0:0, 0:1:0, 0:1:1, and 2:0:0.



to both study items should be the dimension considered first. Three of the negative probe conditions (0:0:0, 0:1:0, and 0:1:1) all contradict the shared feature. Hence, all three conditions should be rejected in equivalent time. Importantly, all should be rejected more quickly than the 2:0:0 condition.

Alternately, if negatives are rejected on the basis of extralist features, it should be easier to detect an extralist feature as the number of extralist features in the probe increases (Mewhort & Johns, 2000). Conditions 0:0:0, 0:1:0, and 0:1:1 include three, two, and one extralist feature, respectively. Hence, Conditions 0:0:0, 0:1:0, and 0:1:1 should become increasingly difficult as the number of extralist features decreases.

## Method

**Subjects.** Fourteen students from the same source as that in Experiment 2 participated in the experiment.

**Apparatus and Stimuli.** The apparatus was the same as that in Experiment 2. The stimuli were the eight shapes, colors, and fill patterns used in Experiment 1.

**Procedure.** The procedure was basically the same as that in Experiment 1. Trials were organized into three blocks: a block of 16 warm-up trials, followed by two blocks of 96 experimental trials each. No feedback was given within the block, but at the end of each block, the number of errors on that block was displayed. Trials within each block were presented in random order.

**Study and test conditions.** Each experimental block included 96 tests, 48 positive tests and 48 negative tests. The 48 positive trials included 16 trials in which the common feature between the two study items was color, 16 in which it was shape, and 16 in which it was pattern. Within each of the three types of trial, both serial positions were tested equally often. The 48 negative trials were also balanced with respect to the dimension shared by the positive items. Within each of the three types of trials, all possible probes in each condition (0:0:0, 0:1:0, 0:1:1, and 2:0:0) were used equally often.

## Results and Discussion

Table 3 summarizes performance. As is evident in Table 3, RT for correct *sure* responses increased linearly across the 0:0:0, 0:1:0, and 0:1:1 conditions; that is, the difference between the 0:0:0 and 0:1:1 conditions was highly significant [ $F(1,13) = 17.66, p < .01$ ], and the 0:1:0 cell was equal to the average of the outer cells [ $F(1,13) = 1.61, .15 < p < .25$ ]. The difference between the 0:1:1 cell and the 2:0:0 cell was not reliable [ $F(1,13) = 1.26, .25 < p < .35$ ].

Because the 0:0:0, 0:1:0, and 0:1:1 negative probe conditions all included a feature that contradicted the shared feature, the directed-search idea predicts that they should be rejected as soon as that feature has been checked. The conditions differ from each other in terms of their overlap with the study set on other features, but the other features should not need to be processed, so the probes should be rejected in equivalent time. The data, however, showed very fast rejections in Condition 0:0:0, slower rejections in Condition 0:1:0, and slower still in Condition 0:1:1. Moreover, the directed-search idea predicts that the probes in the conditions that contradict the shared feature (i.e., Conditions 0:0:0, 0:1:0, and 0:1:1) should all be rejected more quickly than in the 2:0:0 condition. In fact, probes in the

**Table 3**  
**Summary of Performance in Experiment 3**

Measure	Probe Type				
	Positive	0:0:0	0:1:0	0:1:1	2:0:0
Correct trials (%)					
<i>M</i>	94.4	100	99.7	97.9	98.2
<i>SD</i>	4.1	0	1.1	2.7	4.5
Reaction time for correct trials (msec)					
<i>M</i>	1,026	825	868	959	931
<i>SD</i>	390	272	295	304	300
Trials classified as <i>sure</i> (%)					
<i>M</i>	91.7	99.4	99.4	97.3	96.1
<i>SD</i>	10.3	1.5	1.5	5.1	6.6
Correct trials <i>sure</i> only (%)					
<i>M</i>	96.5	100.0	100.0	99.4	100.0
<i>SD</i>	3.3	0.0	0.0	1.5	0.0
Reaction time for correct <i>sure</i> trials (msec)					
<i>M</i>	955	804	844	949	912
<i>SD</i>	306	276	260	305	302

2:0:0 condition were rejected more quickly than Condition 0:1:1 probes, although the difference was not reliable. By contradicting both predictions derived from it, the data deny the directed-search hypothesis.

Although it is impossible to reject categorically all possible accounts that depend on the featural structure of the stimuli, it is clear that directed search is not viable. The corpus of data that a feature-by-feature processing model would have to accommodate is now sufficiently complicated that we believe such a model would prove unsatisfying.

The data confirmed the prediction made on the basis of the detectability of the extralist feature. RT increased as the number of extralist features decreased across Conditions 0:0:0, 0:1:0, and 0:1:1. As Mewhort and Johns (2000) argued, the number of extralist features, like the number of alternatives to the extralist feature, affects the detectability of the extralist feature.

There is an interesting comparison for the present data with data from Experiments 3, 5, 6, and 7 in Mewhort and Johns (2000). Their stimuli involved two dimensions, and the critical finding was that Condition 2:0, which included an extralist feature, was much easier than Condition 1:1, which did not, even though the conditions were equivalent in familiarity. In the present experiment, however, the two conditions that seem most comparable to their 2:0 and 1:1 conditions, the 2:0:0 and 0:1:1 conditions, yielded equivalent RTs. In fact, when the two factors that affect detectability—the number of extralist features and the number of alternatives to the extralist feature—are taken into account, the contradiction position readily explains the apparent anomaly. In Experiment 3, the 2:0:0 probe included two extralist features, but both had two studied alternatives in the memory set and were, therefore, difficult to find (i.e., the complete designation of the condition is 2:0<sub>2</sub>:0<sub>2</sub>). By contrast, the 0:1:1 probe included only one extralist feature, but it was easy to find (i.e., the complete designation of the condition is 0<sub>1</sub>:1:1). The two factors—number of extralist features and the number of alternatives to the extralist feature—work against

each other to produce equivalent detectability in the two cells.

## EXPERIMENT 4

Whereas many recognition experiments use words, our work is distinguished by its use of constructed stimuli. Mewhort and Johns (2000) did use words in their Experiments 5–7. When they held familiarity constant, they found an effect of the number of extralist features, but they did not hold detectability of the extralist features constant and test for an effect of familiarity. In Experiment 4, accordingly, we provide a test of familiarity with words. We use the design of Experiment 2 to pit the familiarity view against the effect of the number of alternatives to the extralist feature.

Because the structure of words cannot be manipulated, we selected words to define our experimental conditions. The words began and ended with a limited number of bigrams. We treated the bigrams as features of the words, equivalent to the color and shape used in the preceding experiments. Unlike Experiment 2, the organization of the study set was manipulated within subjects.

### Method

**Subjects.** The subjects were 12 students from a 2nd-year cognitive psychology course, who participated in return for bonus credit in their course. All were fluent speakers of English and had normal or corrected-to-normal vision.

**Stimuli.** The stimuli were 360 four-letter words drawn from the MRC psycholinguistic database (Coltheart, 1981). Each word had a CVCC spelling pattern. Within the pool, there were 63 initial bigrams and 44 final bigrams. We used the initial and final bigrams as the two features that defined the stimuli, corresponding to color and shape in the preceding experiments. For our manipulations, two bigrams were defined to be different only if both letters of the first bigram mismatched both letters of the second bigram. The 360 words are listed in the Appendix.

**Design.** The experiment followed a  $2 \times 2$  within-subjects design. Table 4 provides examples of the study sets and probes.

The first factor was study set organization: Study sets comprised two pairs of words organized on the basis of either the same feature or different features. In sets with unidimensional organization, the word pairs shared either the initial or the final bigram. In sets with bidimensional organization, one word pair shared the initial bigram, and the other word pair shared the final bigram.

The second factor was type of negative probe. A 1:0 probe comprised a bigram (either initial or final) that had occurred in one study word, paired with a bigram that had not been presented. A 2:0 probe comprised a bigram (either initial or final) that had occurred in two study words, paired with a bigram that had not been presented. As in Experiment 2, the probes are fully designated as 1:0<sub>2</sub> and 2:0<sub>4</sub> in the unidimensional condition and as 1:0<sub>3</sub> and 2:0<sub>3</sub> in the bidimensional condition.

**Procedure.** The procedure was similar to that in Experiment 2, with the following changes prompted by the change in stimuli. (1) The subjects were informed that the stimuli would be common four-letter words but did not preview them. (2) The study words were presented sequentially, each for 1,250 msec, with 250 msec between displays. (3) A row of six question marks was presented between the study and the test items, instead of the colored mask used in Experiment 2. It was presented for 1,250 msec, with a 250-msec blank before the probe appeared. (4) A confidence judgment was not re-

**Table 4**  
**Examples of Study Sets and Probes Used in Experiment 4**

Organization	Probe Type		
	Positive	1:0 <sub>2</sub>	2:0 <sub>4</sub>
Unidimensional			
Example 1	path pact boss born	less	bond
Example 2	self gulf rack tick	send	wolf
	Probe Type		
	Positive	1:0 <sub>3</sub>	2:0 <sub>3</sub>
Bidimensional			
Example 1	path pact self gulf	send	wolf
Example 2	boss born rack tick	less	bond

quired. Pilot work suggested that subjects rarely use the *not-sure* response when dealing with word stimuli and that accuracy would be high. Hence, it would not be necessary to conditionalize on confidence. (5) Errors were signaled (by a groan) and were replaced within the block by an additional trial in the same condition. The subjects were informed that errors were replaced, information that likely encouraged them to respond as accurately as possible.

Each subject completed a block of 12 warm-up trials, followed by three experimental blocks of 64 trials each. As before, conditions and stimuli were fully counterbalanced and presented in random order.

### Results and Discussion

The results are summarized in Table 5. The mean RT for the correct positives in the unidimensional condition did not differ reliably from the mean for the bidimensional organization ( $F < 1$ ). As is documented in Table 5, with unidimensional study sets, subjects rejected 1:0<sub>2</sub> probes about 50 msec more quickly than 2:0<sub>4</sub> probes [ $F(1,11) = 15.47$ ,  $p < .01$ ]. With bidimensional study sets, however, the mean difference between the 1:0<sub>3</sub> and the 2:0<sub>3</sub> probes was less than 4 msec ( $F < 1$ ). The interaction of study set organization with familiarity was reliable [ $F(1,11) = 6.03$ ,  $p < .05$ ]. In short, the pattern of results was the same as that obtained in Experiment 2 with colored forms.

As in Experiment 2, the data provide clear confirmation that performance depends on the subject's use of contradiction and, in particular, on the detectability of the extralist feature. Moreover, the similarity of the results with words and with colored forms leaves little room for criticism based on the nature of the stimuli. One cannot dismiss the data from Experiment 2 on the grounds that colored forms fall outside of the domain of current models of recognition memory: Words yield the same pattern of results.

Generality across stimuli is no surprise. Historically, memory experiments have used letters, numbers, sentences, faces, line drawings of objects, Landau rings, and even sounds and odors as stimuli; the different stimuli do not normally require different theories. Accordingly, we doubt that the authors of current models would restrict their application to a subset of experiments that use verbal stimuli. Within the Sternberg (1969) paradigm, Cavanagh (1972) has documented a common pattern of results across stimuli. Experiment 4 confirms the fact that the number of alternatives to the extralist feature, like other processing variables, is pertinent to various stimuli.

**Table 5**  
**Summary of Performance in Experiment 4**

Measure	Probe Type		
	Positive	1:0 <sub>2</sub>	2:0 <sub>4</sub>
Unidimensional Organization			
Correct trials (%)			
<i>M</i>	97.0	99.3	97.1
<i>SD</i>	2.3	1.6	3.6
Reaction time for correct trials (msec)			
<i>M</i>	784	747	795
<i>SD</i>	187	175	177
	Probe Type		
	Positive	1:0 <sub>3</sub>	2:0 <sub>3</sub>
Bidimensional Organization			
Correct trials (%)			
<i>M</i>	96.8	98.4	98.3
<i>SD</i>	3.6	2.6	2.1
Reaction time for correct trials (msec)			
<i>M</i>	776	767	769
<i>SD</i>	199	134	148

## GENERAL DISCUSSION

### The Empirical Findings

Using stimuli defined by color, shape, and fill pattern, in the first experiment we demonstrated how a manipulation of familiarity can be confounded with the number of alternatives to the extralist feature, a factor that affects the detectability of the extralist information. In the second experiment, we used stimuli defined by color and shape to separate the confounded factors. By manipulating the organization of the study set, we showed that the number of alternatives to the extralist feature, not familiarity, is the operative factor. The third experiment ruled out an explanation based on a strategy of directed search. Finally, the fourth experiment replicated Experiment 2 with verbal stimuli.

Our procedures were taken deliberately in order to test feature-based models of recognition. Although such models allow for differential encoding, the predictions we have drawn for the 1:0 and 2:0 conditions are based on retrieval mechanisms. Current practice has largely de-emphasized the distinction between encoding and retrieval failure, but we maintain that a fair test of retrieval mechanisms presupposes accurate encoding. Hence, we used subspan lists and conditionalized on high confidence.

Our choice of stimuli was also deliberate. It is impossible to test predictions of feature-based models unless the dimensions of the stimuli can be identified. The nonverbal stimuli were constructed to have well-defined and manipulable components: color, shape, and pattern. The words were selected to satisfy the same requirement.

### How Current Theories Fail

The important finding is that performance in Conditions 1:0 and 2:0 was controlled by the number of alternatives to the extralist feature, not by the number of occurrences in the study list of the presented feature. Current models claim that the test item is compared against the memorized items and that a measure of familiarity is calculated.

In Minerva2 (Hintzman, 1984) and REM (Shiffrin & Steyvers, 1997), for example, the test item is compared, feature by feature, against each memorized item and the similarity of the item to the whole memory set is calculated by summing across items. Because a 2:0 probe shares a feature with two memorized items but a 1:0 probe shares a feature with only one memorized item, the 2:0 probe must yield a higher similarity value than a 1:0 probe, and it must always be more difficult to reject. In our data, however, the difference between the 1:0 and 2:0 conditions vanished when the number of alternatives to the extralist feature was held constant. The number of alternatives to the extralist feature maps onto the detectability of the extralist feature. Current models cannot explain why the difference between the 1:0 and 2:0 probes depends on the detectability of the extralist feature, not on calculated familiarity.

The literature provides no convincing evidence favoring a role for familiarity of the studied features. Because most studies use verbal stimuli and because we cannot define the featural composition of words, manipulations that are claimed to affect familiarity are often crude and/or vague. For example, researchers often use words drawn from semantic categories and assume that words from the same category are more similar than words drawn from different categories. The organization is clearly unidimensional; it ignores similarities between items on other dimensions (e.g., the similarity between earwig and earring). With so little information about the featural composition of the stimuli, one cannot make the fine distinction between familiarity of the studied features and detectability of the extralist feature that is at issue here.

One way to escape the implications of our data (and, thereby, to save current theory) would be to discount the results by treating them as a special case. Our experiments are distinguished by the low dimensionality of the stimuli; that is, our stimuli were defined by their values on a small number of dimensions. Moreover, our procedure ensured that the subjects understood the nature of the stimuli. In

the experiments using colored shapes, the dimensions and their values were pointed out explicitly. All of the study lists were organized according to the dimensions, and most subjects were able, at debriefing, to explain at least some of the rules by which study lists had been constructed. Finally, in each experiment, the features of the negative probes were related to the features of the study items in a limited number of identifiable ways.

Both because of the nature of the stimuli and because of the identifiable relation between the study and the test items, the subjects may have adopted a feature-processing strategy that could not occur with randomly selected, less contrived, stimuli.<sup>1</sup> We have no evidence to support the special-strategy idea, and several points argue against it.

One cannot argue that our experiments constitute a special case on the grounds that we did not obtain the standard effects of similarity: We did obtain the standard effects. Familiarity predicts that a negative probe will be more difficult to reject if its constituent features occurred repeatedly in the study set. The effect occurred in all four of the experiments reported here: (1) the 2:1:0<sub>2</sub> versus 0<sub>1</sub>:1:1 difference in Experiment 1, (2) the 2:0<sub>4</sub> versus 1:0<sub>2</sub> difference under unidimensional organization in Experiments 2 and 4, and (3) the 2:0<sub>2</sub>:0<sub>2</sub> versus 0<sub>1</sub>:1:0<sub>2</sub> difference in Experiment 3. The problem for familiarity theory is that, in all of these cases, the familiarity effect was confounded with the number of alternatives to the extralist feature. By demonstrating that the effect disappears under bidimensional organization, we have shown that familiarity is not the effective factor. Because previous demonstrations of similarity have relied on stimuli that cannot be analyzed in terms of constituent features, we suggest that traditional support for familiarity theory may reflect confounds with detectability of extralist information.

Secondly, Experiment 3 addressed the most obvious special case: a strategy based on directed search. Although a directed-search strategy has some empirical basis in the literature, we found no evidence for it.

Finally, the detectability of extralist features affects performance even when the features have been permitted to vary naturally. Mewhort and Johns (2000, Experiments 6 and 7) used four-letter words selected according to their first and last letters. Although the identity of the middle letters was not controlled, it was known, so we were able to demonstrate that RT decreased linearly as the number of extralist letters increased across the range 0 to 3. That is, an extralist feature effect was obtained for features that were not manipulated or limited in any manner.

In any case, a special-case interpretation of our data would not leave standard theory unscathed. A special-case view implies that our experiments define a situation in which current models cannot apply. If so, we have both limited the models' generalizability and forced them to propose a mechanism that can decide if and when the regular strategy does apply. In addition, although current models both depend on a feature-based representation of items and base decisions on comparisons of the features, the special-case argument implies that they do not apply when the features are defined and

can be manipulated. In the extreme, such models are untestable.

### New Directions for Theory

What modifications to theory do our results require? We will speculate here about implications for representation, for the comparison process, and for the decision process. Which and how many changes are necessary, of course, remains an open question.

**Representation.** In a previous paper, we suggested that the global models already calculate the information necessary to accommodate the effect investigated there: the number of extralist features (Mewhort & Johns, 2000). The composite memory in TODAM and the "echo content" in Minerva2 both constitute a feature-by-feature profile of the whole memory set; that profile could be examined for the presence of extralist features. The present data complicate the situation. It is not clear if and how the existing framework could be altered to allow a role for the number of alternatives to the extralist feature.

At a descriptive level, it seems reasonable that it is easier to decide whether a given value on a dimension occurred in the memory set the fewer the values on that dimension that did occur in the memory set. For example, a color would be most easily identified as an extralist feature if the studied colors were the same. With a composite memory, the more colors, the blurrier the representation of color, and the harder to determine whether a particular color were there.

The problem for theory is that an item cannot be represented by a flat vector of features. The representation must include some information concerning how the features relate to each other. Features that describe color are compared only against features that describe color. The matching operator knows which features describe color and distinguishes them from features that concern other dimensions, such as shape. Some sort of hierarchical structure is required.

One possible structure, which deviates minimally from existing notions of composites, assumes that our manipulated features are encoded in terms of subfeatures. Various colors are represented in terms of their values on a common set of subfeatures, whereas the shapes are represented in terms of their values on a different set of sub-features. The idea is comparable with the way in which the colors on a computer monitor are created by combinations of high- and low-intensity red, green, and blue. The features used to represent a color include its relevant subcolors, as well as features that define the particular combination of subcolors involved. Colors are similar to the extent that they share subfeatures; for example, purple would share some subfeatures with red and other subfeatures with blue.

When a color is encoded, the value of each of its defining features and subfeatures is incremented. The larger the number of studied colors, the greater the number of subfeatures that would be encoded, and the more likely it would be that the subfeatures of the distractor's color would be encoded. In the most difficult case, all of the probe's subfeatures may have been encoded, so that the only extra-

list features are the features representing the combination of subfeatures.

**Comparison.** Mewhort and Johns (2000) proposed that recognition decisions might be based on an examination of the composite feature profile. A *no* decision would be issued if enough extralist (unmatched) features occurred in the probe item, and a *yes* decision would be issued if all the probe's features had occurred in the study set. A similarity scale based on the presence or absence of features differs fundamentally from current theory; it is based mathematically on set theory, not on distance between points in an assumed Euclidean space (cf. Tversky, 1977).

A feature evaluation comparison process operating on a representation with features and subfeatures could accommodate the effect of the number of alternatives to the extralist feature. The greater the number of studied values on the extralist feature's dimension, the larger the number of the probe's subfeatures that have been encoded. Thus, the difficulty of finding the extralist feature will increase with the number of studied alternative values to the extralist feature. Suppose, for example, that purple were the extralist feature in a probe. If all the study items had colors that did not share any subfeatures with purple, it would be easy to find enough extralist features to support a *no* decision. Alternatively, if the study items included red and blue items, colors that share subfeatures with purple, it would be more difficult to find the requisite number of extralist features. The more studied colors there are, the higher the probability of an overlap with the probe's subfeatures. The explanation uses the feature comparison process operating on features and subfeatures.

**Decision.** The number of alternatives to the extralist feature is, by definition, a factor that affects only negative decisions. If different factors affect positive and negative decisions, *yes* and *no* responses may depend on different types of evidence. Established dual-accumulator models (e.g., Anderson, 1973) have proposed that when a test probe is processed, two distinct kinds of evidence are calculated: the similarity of the probe to the memory set, and the similarity of the probe to the set of known or possible distractor items. Exceeding a criterion on the first kind of evidence would lead to a *yes* response; exceeding a criterion on the second would lead to a *no* response. REM (Shiffrin & Steyvers, 1997) also contrasts similarity to positive items with similarity to the set of possible distractor items but uses a Bayesian decision mechanism instead of the dual-accumulator mechanism.

Our view adopts the dual-accumulator notion but proposes a different kind of negative evidence: Both positive and negative evidence are based on comparing the test probe against the memory set. Confirmatory evidence favors a *yes* response, and contradictory evidence favors *no*. Positive and negative evidence would be related, but not perfectly correlated.

One would then need a decision rule, the simplest possibility being that there are two criteria, one for a *yes* and one for a *no* response. As soon as one evidence accumulator reaches criterion, the associated response is issued.

Our bias is toward such a dual-accumulator theory, but we intend to establish an empirical motivation for this direction before pursuing it in detail.

## REFERENCES

- ANDERSON, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review*, **80**, 417-438.
- CAVANAGH, J. P. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review*, **79**, 525-530.
- CLARK, S. E., & GRÖNLUND, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, **3**, 37-60.
- CLIFTON, C., & GUTSCHERA, K. D. (1971). Hierarchical search of two-digit numbers in a recognition memory task. *Journal of Verbal Learning & Verbal Behavior*, **10**, 528-541.
- COLTHEART, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-505.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- HEATHCOTE, A. (1988). Screen control and timing routines for the IBM microcomputer family using a high-level language. *Behavior Research Methods, Instruments, & Computers*, **20**, 289-297.
- HINTZMAN, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, **16**, 96-101.
- HOMA, D. (1973). Organization and long-term memory search. *Memory & Cognition*, **1**, 369-379.
- JOHNS, E. E. (1985). Effects of list organization on item recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 605-620.
- KAMINSKY, C. A., & DEROSA, D. V. (1972). Influence of retrieval cues and set organization on short-term recognition memory. *Journal of Experimental Psychology*, **96**, 449-454.
- MCLELREE, B., & DOSHER, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, **118**, 346-373.
- MEWHORT, D. J. K., & JOHNS, E. E. (2000). The extralist-feature effect: Evidence against item matching in short-term recognition memory. *Journal of Experimental Psychology: General*, **129**, 262-284.
- MURDOCK, B. B., JR. (1971). A parallel-processing model for scanning. *Perception & Psychophysics*, **10**, 289-291.
- MURDOCK, B. B., JR. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89**, 609-626.
- MURDOCK, B. B., JR. (1983). A distributed memory model for serial-order information. *Psychological Review*, **90**, 316-338.
- MURDOCK, B. B., JR. (1985). An analysis of the strength-latency relationship. *Memory & Cognition*, **13**, 511-521.
- NAUS, M. J., GLUCKSBERG, S., & ORNSTEIN, P. A. (1972). Taxonomic word categories and memory search. *Cognitive Psychology*, **3**, 643-654.
- NEATH, I. (1998). *Human memory: An introduction to research, data, and theory*. Pacific Grove, CA: Brooks/Cole.
- RAAJMAKERS, J. G. W., & SHIFFRIN, R. M. (1981). Search of associative memory. *Psychological Review*, **88**, 93-134.
- SHIFFRIN, R. M., & STEYVERS, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145-166.
- STERNBERG, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, **57**, 421-457.
- STERNBERG, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, **27**, 1-32.
- TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.

## NOTE

1. We thank R. Shiffrin for expressing this view. We also thank A. Heathcote for articulating the notion of limited dimensionality and for debating the implications with us.

**APPENDIX**  
**Words Used in Experiment 4**

back	bald	ball	band	bang	bank	bark	barn	bass	bath	bell	belt	bend
bent	best	bill	bind	bird	bold	bolt	bomb	bond	born	boss	both	bowl
buck	buff	bulb	bulk	bull	bump	bunk	bunt	burn	burr	bush	bust	butt
calf	call	calm	camp	card	cart	cash	cask	cast	cell	cent	cold	colt
comb	cord	cork	corn	cost	cuff	cult	curb	curl	curt	damp	dark	darn
dash	dawn	deck	dell	dent	desk	dill	dirt	dish	disk	dock	doll	down
duck	duct	dull	dumb	dump	dusk	dust	fact	fall	farm	fast	fawn	fell
felt	fern	fill	film	find	fish	fist	fold	folk	fond	fork	form	fort
full	fund	fuss	gang	gash	gasp	germ	gift	gill	gilt	girl	gist	gold
golf	gown	gulf	gull	gulp	gush	gust	hack	half	hall	halt	hand	hang
hard	hark	harm	harp	hash	hawk	helm	help	herb	herd	high	hill	hilt
hind	hint	hiss	hold	horn	host	howl	hulk	hull	hump	hung	hunk	hunt
hurl	hurt	hush	lack	lamb	lamp	land	lard	lash	lass	last	lawn	
left	lend	lens	lent	less	lest	lewd	lick	lift	lilt	limb	limp	link
lint	list	lock	loft	long	lord	loss	lost	luck	lull	lump	lung	lurk
lush	lust	mall	mark	mart	mash	mask	mass	mast	math	melt	mend	mesh
mess	mild	milk	mill	mind	mink	mint	miss	mist	mock	mold	monk	moss
most	moth	much	must	pack	pact	palm	park	part	pass	past	path	pawn
peck	perk	pert	pest	pick	pill	pimp	pink	pint	poll	pomp	pond	pork
port	post	puff	pull	pulp	pump	push	putt	rack	raft	ramp	rang	rank
rash	rent	rest	rich	rift	ring	rink	risk	rock	roll	romp	rump	rung
runt	rush	rust	sack	salt	sand	sang	sank	sash	self	sell	send	sent
sick	sigh	sign	silk	sill	sing	sink	sock	soft	sold	song	sort	such
suck	suds	sung	sunk	surf	tack	tact	talk	tall	tank	tart	task	tell
tend	tent	term	test	text	tick	till	tilt	tint	told	toll	tomb	torn
toss	town	tuck	turn	walk	wall	wand	want	ward	warm	warn	warp	wart
wash	wasp	weld	well	went	west	wick	wild	will	wilt	wind	wing	wink
wish	wisp	with	wolf	womb	word	work	worm	worn				

(Manuscript received November 24, 2000;  
revision accepted for publication July 16, 2001.)