The relation between recognition memory and classification learning

JANET METCALFE

University of British Columbia, Vancouver, British Columbia, Canada

and

RONALD P. FISHER

Florida International University, Miami, Florida

Two experiments investigated the relation between recognition memory and classification learning. The subjects were instructed that they would see a series of random-dot patterns and later would be asked to classify or to recognize the patterns. Following study, the subjects performed a classification task, a recognition-memory task, or both. It was found that classification-learning instructions were superior to recognition-memory instructions for the classification task, but that there was little or no effect of instructions on the recognition task. When subjects performed both recognition and classification tasks, there was no relation between saying "old" to a probe and correctly classifying it, except with old exemplars, and then only when the initial instructions had been to expect a recognition-memory test. Overall, the data show that classification and recognition can be experimentally separated. In addition, classification is often statistically independent of recognizing that items are old. These observed relations provide some constraints for the further development of models of classification learning and recognition memory.

A number of theorists have proposed that people may classify or categorize by retrieving from memory particular events that they have previously experienced (e.g., Brooks, 1978; Medin & Schaffer, 1978). Reber and Allen (1978) illustrated this exemplar view of classification as follows:

> A novel four-legged beast is perceived as a dog not because it fits with the viewer's abstract feature system for dog but rather because it reminds him of some specific critter he has met before which was identified as a dog. (p 192)

Hintzman and Ludlum (1980) showed how a model that stores only presented instances or exemplars can account for the differential forgetting of new prototypical and old exemplar test items, as is found in the empirical data (Homa, Cross, Cornell, Goldman, & Schwartz, 1973; Posner & Keele, 1970; Strange, Keeney, Kessel, & Jenkins, 1970). This finding had previously been thought to provide evidence for the special status of concepts that were distinct from the presented exemplars. Hintzman (1983) further elaborated this exemplar model to account for a wide range of memory and classification data. Jacoby

and Brooks (1984) also argued that specific instances may account for much of the classification-learning data, and Kahneman and Miller (in press) noted that normative judgments may be attributable to retrieval of specific episodes or instances. The exemplar-based idea of classification learning stimulates the question addressed in the present article, of the relation between classification and recognition memory for specific instances.

Models that propose that classification is based on memory for specific instances suggest that there should be a relation between recognition of items as old and classification of the items, since both judgments presumably use the same information. Note that the predictions are not in terms of positive, negative, or zero correlations between recognition and classification accuracy. Rather, subjects should be better at classifying items they think are old than those they think are new, because on average the "believe old" items are more likely to be in or like items in memory than the "believe new" items. Smith and Medin (1981, p. 164) summarized the workings of exemplar models of classification as follows: "typical exemplars dominate concept representations, and a test item is categorized as a member of a target concept if and only if the former retrieves a criterial number of the latter's exemplars." They are explicit about the relation that is predicted between classification and recognition of items as old:

> If a new pattern B3 [which is similar to old item B1] is presented, it should be correctly classified because it would most likely access the representation associated with B1.

Note further that on a new-old recognition test B3 would very likely be recognized falsely *as old* for the same reason. (p. 245, our emphasis)

Thus, the more likely an item is to be recognized as old, the better it should be classified.

Despite the pleasing parsimony of single-system exemplar models, other research provides evidence that two memory systems with different characteristics may exist. These two systems have been called episodic and semantic memory (Tulving, 1972, 1983), memory with and without awareness (Eich, 1984; Jacoby & Witherspoon, 1982), locale and taxon (Jacobs & Nadel, 1985; O'Keefe & Nadel, 1978), and explicit and implicit memory (Graf & Schacter, 1985a, 1985b; Schacter, 1985; Schacter & Graf, 1985). The explicit-locale-episodic system is usually characterized as requiring that the individual be aware that he or she is remembering a particular event. It is also thought to be highly context sensitive (O'Keefe & Nadel, 1978) and interference prone (Schacter, 1985). Standard memory tasks such as free recall, cued recall, and recognition are usually thought to be explicit-locale-episodic tests (although the implicit system may also be used under certain circumstances; see Mandler, 1980). The implicit-taxon-semantic system does not require conscious awareness that an event occurred. Perceptual fluency, rather than awareness, seems to be one hallmark of this system. It may be characterized by unidirectional stimulus-response associations (Jacobs & Nadel, 1985; O'Keefe & Nadel, 1978). Tasks such as fragment completion, priming, word identification, anagram solving, and lexical decision are considered to be implicit-memory tasks. Both systems are vulnerable to new information, although perhaps differentially vulnerable. Several lines of evidence suggest that the two types of memory are separable, are partially independent, and handle information in different ways. According to Schacter and Graf (1985), amnesics are often impaired on explicit-memory tasks, such as recall and recognition, and yet spared on fragment completion, priming, and other implicit tasks (Cohen & Squire, 1980; Graf, Squire, & Mandler, 1984; Kinsbourne & Wood, 1982; Milner, Corkin, & Teuber, 1968; Moscovitch, 1982; Wood, Ebert, & Kinsbourne, 1982). Normals also have shown a pattern of results indicating that performance on tasks usually ascribed to one system (e.g., fragment completion task) may be independent of performance on tasks usually ascribed to the other system (e.g., recognition memory task; see Tulving, Schacter, & Stark, 1982). Fisher and Chandler (1985) showed that training on an episodic task transfers to an episodic but not to a semantic task, whereas training on a semantic task transfers to a semantic but not to an episodic task.

It is plausible to speculate that the implicit-taxon-semantic system is mainly responsible for categorization. Distributed perceptron models and unidirectional stimulus-response matrix models, such as those of McClelland and Rumelhart (1985) and Knapp and Anderson (1984),

respectively, can do a good job of handling the data from classification-learning experiments. The McClelland and Rumelhart (1985) model can also handle priming effects, and both models are capable of redintegration, so they have the potential to do other implicit-memory tasks, such as fragment completion. These models seem ideally suited to handling much of the data that have been attributed to the taxon or implicit-memory system. (The models have not been applied to episodic-memory data, so the limit of application is not known.) On the other hand, episodic-distributed models, such as CHARM (Eich, 1982), which is highly context sensitive, can account for much data from such tasks as free recall (Metcalfe & Murdock, 1981), cued recall, recognition (Eich, 1982, 1985), and serial recall (Murdock, 1983). Interestingly, the CHARM model requires a semantic-memory pattern-recognition system in addition to the episodic system. Interesting, also, is the fact that although the episodic CHARM model can do some categorization, it does not mirror in detail the results of experiments done on prototype formation (see Eich, 1982). It may be that this failure of the episodic model is informative. Perhaps classification is mainly done in the taxon or semantic system. O'Keefe and Nadel (1978, p. 100) argued that concepts and categories are represented in the taxon system. Children seem to be able to form concepts before the hippocampus—the anatomical structure necessary for explicit-episodic memories—is fully functioning. Jacobs and Nadel (1985, Footnote 3) also noted "the central role of all taxon systems in prototype or concept formation." The idea that there may be two memory systems—one implicit and mainly responsible for concept formation, and one explicit and mainly responsible for recognition memory—provides a contraposition to the exemplar-only view of classification learning. The relation between recognition memory and classification learning is of considerable interest from the dual-memory perspective. This view suggests that these two tasks might be dissociable.

Reber and Allen (1978) conducted a study on the learning of synthetic grammars in which they asked subjects to introspect about information and decision processes that were relevant to how they were able to tell whether a particular probe letter string was or was not grammatical. Of special interest in the present context are those introspections indicating that subjects classified letter strings as grammatical because they judged the strings to be old (i.e., because they recognized the string, or thought they did). Reber and Allen noted that these kinds of response deserve special treatment, since they represent the primary evidence for the idea that classification is done on the basis of specific instances. They expected that if subjects classified primarily by referencing back to the stimuli from the learning session, then they would find a large number of items on which the subjects justified their classification responses by citing an item as being an "old" item or one that "reminds me of...." In contrast to expectations, Reber and Allen found that these kinds of introspection were infrequent. In addition, subjects were often

wrong about whether an item was, in fact, old. However, when they did classify an item as grammatical because it reminded them of an old exemplar, they were usually right in the grammaticality decision. To our knowledge, these introspective reports comprise the only empirical evidence that has direct bearing on the relation between saying "old" to an item in recognition memory and correctly classifying that item. The primary goal of the present study was to further investigate the relation between recognition memory and classification, in an objective form.

The other aspect of the relation between classification and recognition that is of interest is the goodness of performance on the recognition and classification tasks as a function of the instructions given to subjects. Although models of classification learning make no predictions on this issue, the results could be of interest from a practical or pedagogical standpoint. Accurate recognition may be inversely related to good classification performance, or it might be the case that the compatibility between instructions and test (Bransford, Franks, Morris, & Stein, 1979) is an important factor in performance. Luria's (1968) mnemonist, while exhibiting spectacular recognition abilities, was almost totally unable to conceptualize and classify. Reber and Allen (1978) found that under paired-associate learning instructions, subjects were better able to remember strings of letters than under incidental learning instructions, but they were less able to determine the grammaticality of the strings. Medin and Smith (1981) found that subjects were best at classifying when they were simply told to expect a classification task; performance deteriorated with more explicit instructions, such as to memorize a rule plus exceptions or to try to form a general impression and use this impression to classify. Reber (1976) also found that instructions to try to find and use a rule hurt classification performance. However, since this strategy also impaired recognition performance, the search for an obscure rule may have been simply a distracting activity. Although many different strategies have been investigated, there are no previous studies in which subjects were told to expect a recognition test or a classification test and then were given either a classification test or a recognition test or both. Such a study is of interest because it should tap the usual encoding that is done for one task or the other. In the experiments that follow, we manipulate the encoding by means of our instructions about the task that will be given at time of test.

EXPERIMENT 1

Method

Materials. The materials were 36 random-dot patterns, each of which consisted of six black dots (.5 cm in diameter) randomly placed on an 8.5×11 in. white background. There were 10 patterns in each of three categories, arbitrarily labeled A, B, and C. Each category contained one pattern that served as the prototype, eight patterns that were small distortions of the prototype, and one pattern that was a large distortion of the prototype. The small dis-

tortions were constructed by moving each of the six dots in the prototype .5 in. in one of eight compass-point directions N, NE, E, SE, S, SW, W, NW. Each dot was moved in a different direction. The large distortions were constructed in a similar fashion, but by moving each of the six dots 1 in. in one of the eight directions.

Procedure and design. Subjects heard one of two sets of instructions before studying a list of random-dot patterns. They were told to study the patterns in anticipation of either a recognition test or a classification test. The instructions were read verbatim from one of the following texts:

(A) *Recognition instructions* In this experiment, I am going to show you some pictures that you may think of as constellations of stars if you wish. I will show you three different lists of 6 patterns each for a total of 18 patterns The first 6 patterns will be called List A patterns, the second List B, and the third List C [For half of the subjects, the order was C, B, A] Your task is to try to remember each of the 18 patterns so that you can recognize them later I am later going to present some of these same 18 patterns mixed in with other very similar patterns that you have not seen, and I will ask you to say whether a given pattern is old or new

(B) *Classification instructions.* In this experiment, I am going to show you some pictures that you may think of as constellations of stars if you wish. I will show you three different lists of 6 patterns each for a total of 18 patterns. The first 6 patterns will be called List A patterns, the second List B, and the third List C Your task is to study the patterns so that later you can correctly classify them as A, B, or C. I am later going to present you with some of the patterns mixed in with some new patterns that you have not seen, and I will ask you to sort all of the patterns into Type A, B, or C

The stimuli were arranged in three six-page booklets (lists), with each booklet containing six exemplars from one category. The six exemplars were chosen randomly from among the eight small distortions of each prototype. Subjects examined each pattern for 10 sec, and then on the experimenter's signal ("turn") turned to the next page of the booklet to examine the next pattern. After the subjects examined the last page of each booklet, there was a brief pause during which time the experimenter reminded the subjects which list they had just finished and which was to be examined next

Immediately after the learning phase, the subjects performed either a recognition test, a classification test, or both. The same sequence of 24 randomly ordered test patterns was used in both the recognition and the classification tests The test booklets included 8 patterns from each category, the prototype, four small distortions that had been presented during acquisition (old exemplars), two small distortions that had not been presented before (new exemplars), and the large distortion.

For the recognition test, subjects judged whether each test pattern had been seen before (old) or not (new). The subjects gave a 6-point confidence judgment (1 = *guess*, 6 = *positive*) after each decision. For the classification test, subjects judged whether each of the test patterns was an A, B, or C pattern, and gave a 6-point confidence judgment. Testing was subject-paced, and the subjects were asked to respond to every probe, even if unsure.

Subjects. The subjects were 211 UCLA undergraduates who participated to fulfill a course requirement. The subjects were run in small groups of 5 to 16. One hundred seven subjects received the recognition instructions. Of these, 51 performed the recognition test and 56 were given the classification test, as their first, or only, test. Fifty-one of the subjects who were given the recognition instructions performed both the recognition and classification tests, 27 in the order recognition followed by classification, and 24 in the reverse order. One hundred four subjects received the classification instructions. Of these, 50 performed the recognition test and 54 were given the classification test as their first, or only, test. Forty-eight subjects who were given the classification instructions performed both tests, 22 in the order recognition followed by classification, and 26 in the reverse order.

Results and Discussion

A criterion of $p < .05$ was chosen for considering an effect to be significant. If an effect is mentioned, but did not reach the .05 level, the p value of the trend is specifically stated.

Recognition test. In this section, we examine the results of only the first test. The recognition data were analyzed in terms of the proportion of items that subjects thought were old. These data are presented in Figure 1, top panel. As the figure shows, the probability of calling an item old did not depend upon instructions [$F(1,99) = 1.52$, $MSe = .06$]. There was no interaction between instructions and the type of probe [$F(3,297) = .38$, $MSe = .04$]. However, there was considerable discrimination among the different probe types [$F(3,297) = 181.49$, $MSe = .04$]. Subjects in both instructional conditions thought that the nonpresented prototypes were old most of the time, followed by the old exemplars, the new exemplars, and finally by the large distortions, which were rarely called old. All comparisons among probe types were significant by Duncan's multiple-range test. Subjects discriminated among the probes, and thus, presumably were making contact with stored memorial information. However, there was no indication that contact with the memorial information was affected by the instructions.

Four other analyses were conducted on the recognition data. First, the data were analyzed in terms of proportion correct, collapsed over List A, B, or C. Again, there was no effect of instructions on correct recognition [$F(1,99) = 1.09$, $MSe = .05$]. There was no interaction between instructions and the type of probe [$F(3,297) = .53$, $MSe = .04$]. However, there was a significant effect of type of probe [$F(3,297) = 152.55$, $MSe = .04$]. The probability of correct recognition was best for the large distortions and the old exemplars. The new exemplars were about at chance, and the new prototypes tended to be wrongly recognized as being old. Bearing in mind that subjects were only correct in calling the old exemplars old, the mean values for this effect are redundant with those given in Figure 1.

Second, the data were collapsed over probe type and analyzed by lists, to see if input order was important. There was no main effect of instructions, of list (ABC vs. CBA), or of order (first, second, or third list presented) (Fs < 1). None of the interactions was significant except that between list and order [$F(2,192) = 12.18$, $MSe = 1.19$]. This interaction can be interpreted as indicating that List A was easier than List C. This analysis was fairly uninteresting since there was no hint of an interaction with instructions. The overall mean proportion correct was .57 for prototype instructions and .58 for recognition instructions.

Third, the data, collapsed over probe type, were analyzed on the basis of the number of items that were called old. Again, there were no interactions with instructions. There was an effect of order, showing that items in the first list were more likely to be called old than the third than the second [$F(2,192) = 47.04$, $MSe = 2.10$]. There was also an interaction between list and order that can

be interpreted as indicating that List C items (whether first or last) were more likely to be called old than A items than B items [$F(2,192) = 19.09$, $MSe 2.10$]. Since these materials effects did not interact with instructions, they will not be discussed further.

Fourth, the overall mean confidence ratings, collapsed over probe type and order, were compared with those ratings on just the correct items. There was no effect of instructions, and no interactions with instructions. The only significant effect revealed by the analysis showed that people were slightly more confident on correct responses (3.92) than they were overall (3.83) [$F(1,96) = 14.99$, $MSe = .045$]. The lack of an effect of instructions on confidence ratings (as shown above with proportion correct) suggests that this lack of an effect for instructions in recognition is not due just to an insensitive measure of proportion correct.

Classification test. The bottom panel of Figure 1 presents the unconditional classification data. As can be seen, the group given classification instructions classified better than did the group given recognition instructions [$F(1,108) = 13.85$, $MSe = .12$]. This effect is in contrast to the failure to find a difference with instructions in the recognition-memory task. There was an effect of probe type [$F(2,324) = 34.17$, $MSe = .05$]. Goodness of classification decreased for prototypes, old exemplars, new exemplars, and large distortions. There was also a significant interaction between instructions and the type of probe [$F(3,324) = 4.04$, $MSe = .05$]. This interaction is attributable to the large distortions, which were classified about equally poorly in both instructional conditions. Duncan's multiple-range tests showed that there were significant differences between instructions for each of the prototype, old exemplar, and new exemplar probes. The differences among these probes were not significant under classification instructions, although all three gave significantly better results than did the large distortion probe. Under recognition instructions, there was a significant difference between the prototype and new exemplar probes, but none of the other comparisons among prototype, old exemplars, and new exemplars were significant. All three of these probes produced significantly better results than did the large distortion probes. There was no significant difference between the large distortion probes depending upon instructions. The superiority of the classification instructions is thus shown with all probes except with the large distortion probe. We did not take similarity judgments on the materials. However, the fact that under both instructional conditions classification performance was well above chance, even on nonpresented probes, indicates that the within-category similarity was considerably higher than the between-category similarity.

These findings suggest that contact with episodic memory, as evidenced by the recognition scores, was the same in the two instructional conditions. There was neither a main effect of instructions, nor an interaction between instructions and probe type, when the dependent measure was the probability of calling an item old. Although the instructional manipulation did not affect contact with epi-
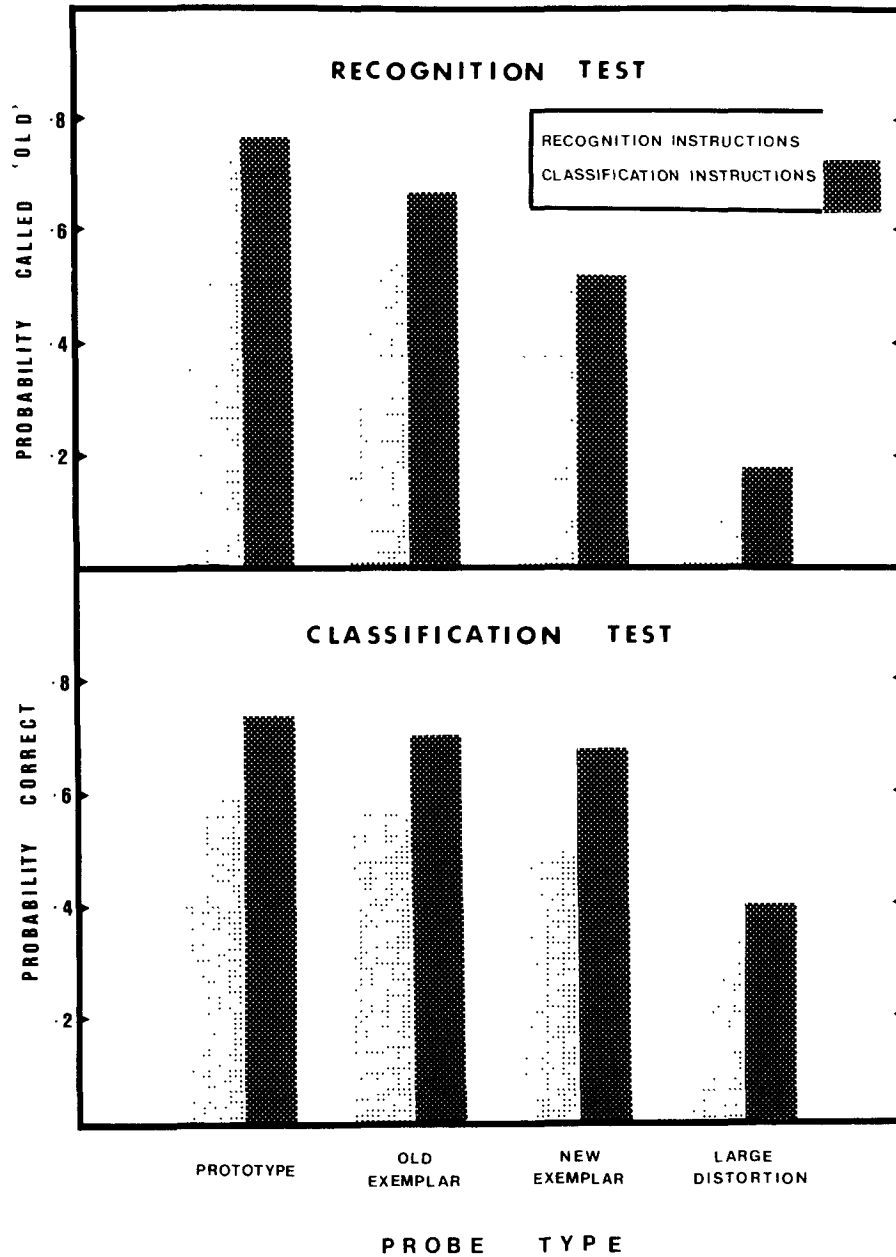
Figure 1. Experiment 1, recognition and classification. The top panel shows the probability that new prototypes, old exemplars, new exemplars, and new distortions were called old in recognition. The bottom panel shows the probability of correct classification of these probes.

sodic memory as shown in the recognition task, there was a large effect in the classification task. Similarly, there was no instructions × probe type interaction in the recognition task, but there was a reliable interaction in the classification task. Classification learning and recognition memory thus appear to be separable experimentally.

To further investigate the relation between recognition memory and classification learning, we looked at the conditional probabilities of classification given recognition

as old or new, for those subjects who where given both the recognition and classification tests. If both recognition and classification judgments are based on the same information, and if the processes are related, we should find an item-by-item dependence relation between recognition memory and classification. Specifically, if people call an item old, they should be more likely to classify that item correctly than if they call an item new. This finding would corroborate the data on subject introspections

found by Reber and Allen (1978). If recognition and classification were handled by different memory systems, we might expect little or no dependence.

Conditional analyses. The difference between proportion of correct classification of those items judged to be old versus items judged to be new in recognition was calculated for each subject. As can be seen from Table 1, under classification-learning instructions, it did not matter for classification performance whether an item was judged to be old or new. In fact, the items judged as new were actually slightly more likely to be classified correctly than were the items judged to be old. With recognition instructions, there was a tendency for classification performance to depend upon recognizing an item as old. A *t* test showed that there was a reliable difference between the two instructional conditions. There was no dependence between recognition and classification with classification instructions, whereas dependence was found with recognition instructions [$t(27) = 2.13$].

EXPERIMENT 2

Experiment 2 was similar to Experiment 1 in overall design. We were primarily interested in further investigating the finding from Experiment 1, that classification performance was better with classification instructions than with recognition instructions, and more importantly, that classification performance can be independent of saying "old" in a recognition-memory task. To obtain more data on the dependency relation, all subjects were tested for both recognition and classification. Subjects were tested for recognition of a particular probe and then were tested immediately for classification of that probe. Finally, we increased the number of old and new exemplar probes to provide enough observations to calculate the conditional probabilities separately for both probes.

Method

Materials. The stimuli were similar to those used in Experiment 1. Three new prototypes were constructed, one for each category (A, B, and C). From each prototype, we constructed 12 exemplars that were small distortions and 1 exemplar that was a large distortion.

Procedure. The instructions and learning phases were identical to those of Experiment 1. Subjects examined six small distortions of the prototype in each category, one at a time, at a 10-sec rate. Immediately after the learning phase, subjects were given a 42-page booklet containing 14 patterns from each of the three categories: the prototype, the six small distortions that had been presented earlier (old exemplars), six small distortions that had not been presented

Table 1
Simple Probability of Classification and Differences in Conditional Probabilities, Experiment 1

Instructions	Old Exemplars	
	P(Correct Classification/Said Old) −P(Correct Classification/Said New)	P(Correct Classification)
Classification	−.04	.66
Recognition	.12	.50

earlier (new exemplars), and one large distortion. In all there were 18 old test patterns and 24 new test patterns. The 42 patterns were ordered randomly. All subjects made both a recognition and a classification decision about each test pattern before turning to the next pattern. Subjects were instructed to decide whether a given pattern was old or new (plus give a confidence judgment: 1 = *guess*, 3 = *positive*) and then judge whether the pattern was a Type A, B, or C (and give a 3-point confidence judgment).

Subjects. The subjects were 100 UCLA undergraduates who participated to fulfill a course requirement. There were 51 subjects in the recognition-instructions group and 49 in the classification-instructions group.

Results

Recognition test. When the data were analyzed in terms of the probability of calling an item old, there was an effect of instructions [$F(1,98) = 4.81, MSe = .05$]. Probes were more likely to be called old with classification than with recognition instructions. However, there was no trend at all toward an interaction between instructions and type of probe [$F(3,294) = .40, MSe = .04$]. An interaction had been found on the classification task in Experiment 1 and was also found in this experiment, as will be discussed shortly. The main effect of instructions, then, seems to be due to a response bias during recognition rather than to differential remembering. As had been the case in Experiment 1, there was a main effect of type of probe [$F(3,294) = 81.02, MSe = .04$]. Duncan's multiple-range tests showed that prototypes were called old significantly more than were old exemplars. The difference between old exemplars and new exemplars was significant, as was the difference between new exemplars and the large distortions, which were rarely called old. The data, plotted in terms of the probability of calling the various probes old, are presented in the top panel of Figure 2.

It was thought that the confidence ratings might provide a more sensitive measure than did the simple probabilities of calling items old, and might show evidence for an interaction between probe type and instructional condition. Thus, the mean confidence ratings for the called-old items were calculated for each subject who had at least one "old" response for each probe type. There were 61 such subjects. There was no effect of instructions ($F < 1$). There was an effect of probe type such that confidence was higher for the prototype (2.31) and old exemplars (2.32) than for the new exemplars (2.12) and new distortions (2.18) [$F(3,177) = 3.77, MSe = .16$]. However, there was no interaction between probe type and instructions ($F < 1$). Thus, the confidence analysis supports the idea of a simple response bias rather than memory difference between instructional conditions.

There was no effect of instructions on correct recognition [$F(1,98) = 1.34, MSe = .04$], nor was there an interaction between instructions and type of probe [$F(3,294) = 1.75, MSe = .04$]. There was a significant effect of type of probe [$F(3,294) = 71.82, MSe = .04$]. These results were consistent with those found in Experiment 1.

When the data were collapsed across probe type and analyzed for correct recognition by lists, there was also
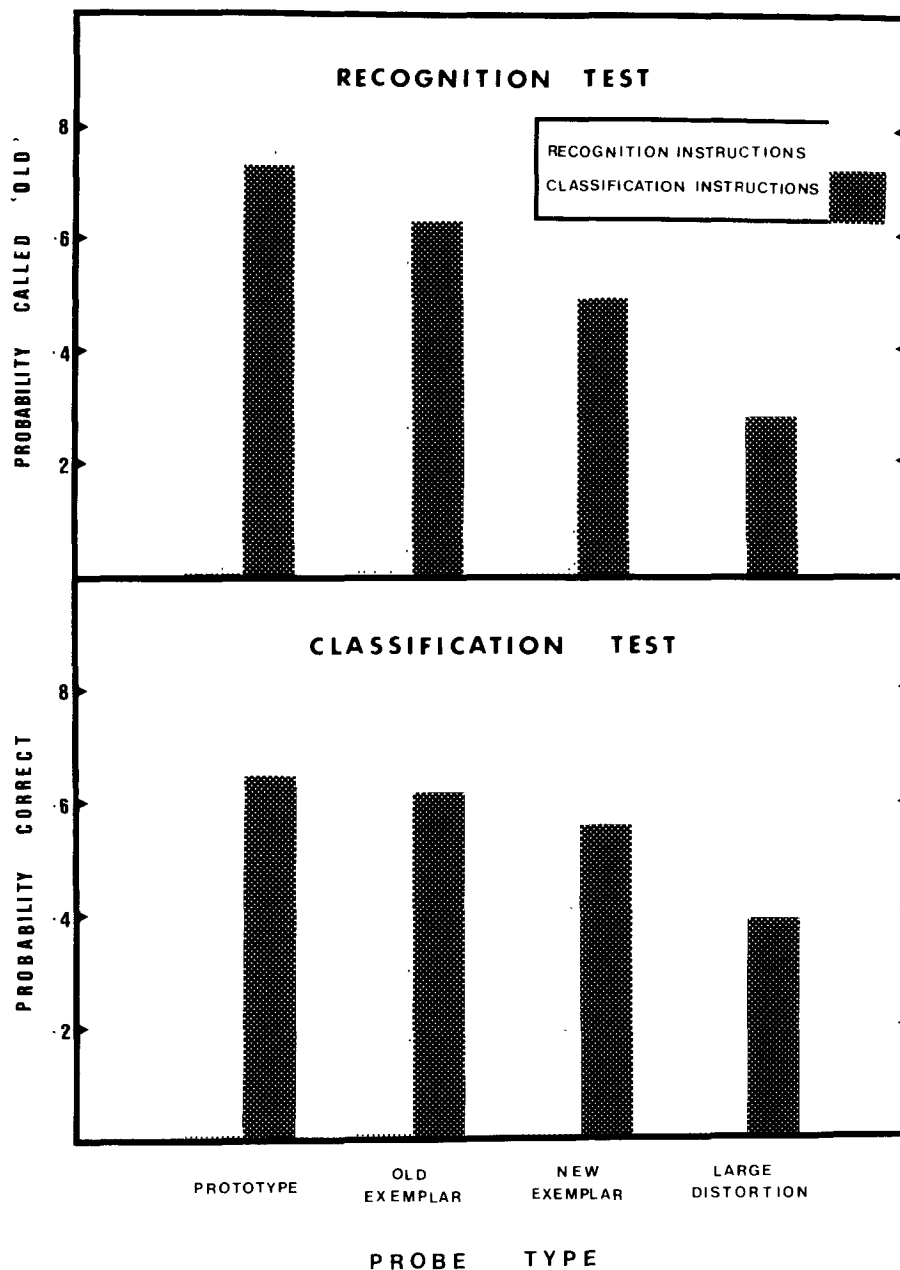
Figure 2. Experiment 2, recognition and classification. The top panel shows the probability that new prototypes, old exemplars, new exemplars, and new distortions were called old in recognition. The bottom panel shows the probability of correct classification of these probes.

no effect of instructions. The means were .56 and .55 for recognition and classification instructions, respectively. There were no significant interactions with instructions. The only significant effect that this analysis revealed was a list × order interaction, which indicated that recognition was best on List C, then List A, and finally List B. When the data, collapsed over probe type, were analyzed for the number of items called old, there was an effect of instructions such that more items (7.78 or 56%) were

called old with classification instructions than with recognition instructions (7.05 or 50%). This is the same response bias reported above. There were no significant interactions with instructions. There was an effect of order such that items from List B were most likely to be called old.

The analysis of confidence ratings on correct responding collapsed over probe type showed no effects at all except one of list order. Subjects given presentation order

ABC had higher confidence (2.22) than those given the CBA order (2.06) [$F(1,95) = 6.00, MSe = .02$]. We have no explanation for this effect.

Overall, the recognition results showed that subjects discriminated among probe types in both instructional conditions and tended to call more items old, regardless of the correctness of this judgment, when given classification as opposed to recognition instructions. This latter effect seems to be a response bias, since there was no interaction between instructions and probe type (as was found in the classification results described below) on any dependent measure.

Classification test. The classification results are presented in the bottom panel of Figure 2. As had been the case in Experiment 1, there was a significant main effect of instructions, such that performance was better with classification instructions than with recognition instructions [$F(1,98) = 14.54, MSe = .07$]. There was a main effect of probe type [$F(3,294) = 10.39, MSe = .05$]. The order of goodness of classification was prototypes > old exemplars > new exemplars > large distortions. Duncan's tests showed that prototypes were significantly better classified than new exemplars, and that prototypes, old exemplars, and new exemplars were all better classified than large distortions. There was also an interaction between the type of probe and instructions [$F(3,294) = 4.11, MSe = .05$]. As had been the case in Experiment 1, the interaction was attributable mainly to the large distortions, which were not differentially classified depending upon instructions. Duncan's tests showed that the prototypes, old exemplars, and new exemplars were all better classified in the classification- than in the recognition-instructions condition. In addition, the post hoc tests showed that the differences among probes given above for the main effect apply only to the classification-instructions condition. There were no significant differences among probe types with recognition instructions. It may be noted, however, that subjects were well above the guessing probability of .33, even under recognition instructions, and even for the large distortions probes which were classified the worst [$t(50) = 2.07$].

Conditional analysis. Table 2 shows the (simple) probabilities of correctly classifying an item and the difference in classification probability given that an item was recognized as old as compared to new. Separate difference scores were calculated for the old and the new exemplars for each subject, and an ANOVA was conducted on these difference scores. There was neither a main effect of instructions, nor a main effect of type of probe ($Fs < 1$). However, the interaction between instructions and probe type was significant [$F(1,98) = 4.06, MSe = .06$]. A Duncan's multiple-range test showed that this interaction was attributable to the fact that there was a large difference with the old exemplar probes when the instructions had been to expect recognition. In this treatment combination, subjects were more likely to correctly classify an item that they had recognized as old than one they thought was new. This dependence between recognition

Table 2
Simple Probabilities of Classification and Differences in Conditional Probabilities, Experiment 2

Instructions	P(Correct Classification/Said Old) −P(Correct Classification/Said New)	P(Correct Classification)
Old Exemplars		
Classification	.04	.623
Recognition	15	.476
New Exemplars		
Classification	.09	.556
Recognition	.05	.435

and classification was not found in the other three treatment combinations, nor were the difference scores in the other three conditions reliably different from one another. Only the old exemplar probes in the recognition-instructions condition produced difference scores that were statistically greater than zero (Bruning & Kintz, 1968).

These results replicate those found in Experiment 1 in which only the old exemplar probes were examined. Dependence between recognition and classification was found, in both experiments, only when subjects expected a recognition test *and* when the actual old items were given as test stimuli. Otherwise, recognition and classification were statistically independent.

GENERAL DISCUSSION

The results of the two experiments presented here show that classification learning and recognition memory are experimentally separable and, under some conditions, are also statistically independent. The reason for claiming experimental separability is that an instructional manipulation improved classification performance without altering recognition performance (Experiment 1) or with very little effect on recognition (Experiment 2). The statistical independence in some treatment combinations does not appear to be artifactual because we also found dependence in one condition. Furthermore, the pattern of independence between recognition of items as old and classification under classification-learning instructions and some dependence under recognition-memory instructions, with the old exemplar probes, replicated over the two experiments. The dependence relation that is expected by a simple interpretation of exemplar-only models exists, but it is not a universal phenomenon. In particular, we were able to find dependence only when subjects encoded in a way that presumably would give priority to the specific exemplars—in preparation for a recognition test. When subjects encoded in order to be able to classify later, recognition and classification performance were independent.

These experiments do not show an inverse relation between classification learning and recognition memory. People who were told that they would get a recognition test were not better on recognition, as would be predicted by a compatibility hypothesis. Classification instructions,

however, did improve performance on the classification task. Our results are consistent with those found by Reber and Allen (1978), insofar as their episodic-memory instructions for paired-associate recall and ours for recognition memory both impaired classification performance. In addition, the one treatment combination in our experiments in which we found a relation between "said old" and correct classification corresponds to the episodic-memory-instructions condition of Reber and Allen's study, in which a dependency relation was most likely to occur.

The fact that a different pattern of dependencies, as well as a difference in simple probabilities, occurred depending upon instructions suggests that the instructional manipulation may have resulted in different strategies. It seems unlikely to us that the strategy used under classification instructions was one of trying to abstract rules. Both Medin and Smith's (1981) and Reber's (1976) experiments indicated that classification performance is worse, not better, under rule-based strategies than under recognition-memory-based strategies. The results seem most interpretable by the conjecture that there may be two systems that are differentially used if one is expecting a recognition as compared to a classification test. We have two sorts of evidence for a distinction. First, the instructions produced different effects on the recognition and classification tasks. Second, classification was independent of saying "old" in recognition in all treatment combinations except the one that gave explicit priority to specific exemplars. We acknowledge that the idea that two types of information develop (e.g., exemplar and prototype information, Posner & Keele, 1970; although other possibilities exist, Jacoby, 1984; Medin, Alton, & Murphy, 1984; and see Smith & Medin, 1981, for a review) might potentially handle the results presented in this article. It is even conceivable that some version of an exemplar-only model might produce these results. The measures taken to induce an exemplar-only model to yield the pattern found in these data could generate interesting new predictions. The dissociation between recognition and classification found here, however, seems to us to be most compatible with recent research pointing to a difference between semantic and episodic memory (Tulving, 1983), between locale and taxon systems (Jacobs & Nadel, 1985; O'Keefe & Nadel, 1978), or between explicit and implicit memory (Graf & Schacter, 1985a, 1985b; Schacter, 1985; Schacter & Graf, 1985).

REFERENCES

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing in human memory (pp. 331-354). Hillsdale, NJ: Erlbaum.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), Cognition and categorization (pp. 169-211). Hillsdale, NJ: Erlbaum

Bruning, J. L., & Kintz, B. L. (1968). Computational handbook of statistics. Glenview, IL: Scott, Foresman.

Cohen, N. J., & Squire, L. R. (1980). Presented learning and reten-
tion of pattern-analyzing skill in amnesia. Dissociation of "knowing how" and "knowing that." Science, 210, 207-209.

Eich, E. (1984). Memory for unattended events. Remembering with and without awareness. Memory & Cognition, 12, 105-111

Eich, J. M. (1982). A composite holographic associative recall model Psychological Review, 89, 627-661

Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. Psychological Review, 92, 1-38.

Fisher, R. P., & Chandler, C. C. (1985) The independence of theme-cued and speaker-cued recall. Manuscript submitted for publication.

Graf, P., & Schacter, D. L. (1985a) Implicit and explicit memory for new associations in normal and amnesic subjects Journal of Experimental Psychology: Learning, Memory, & Cognition, 11, 501-518

Graf, P. & Schacter, D. L. (1985b). Selective effects of practice and retroactive interference on implicit and explicit memory for new associations. Unpublished manuscript.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget Journal of Experimental Psychology. Learning, Memory, & Cognition, 10, 164-178.

Hintzman, D. L. (1983). "Schema abstraction" in a multiple-trace memory model Paper presented at the conference "The Priority of the Specific." Elora, Ontario, Canada

Hintzman, D. L., & Ludlum. G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. Memory & Cognition, 8, 378-382.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances determining the prototype. Journal of Experimental Psychology, 101, 116-122.

Jacobs, W. J., & Nadel, L. (1985). Stress-induced recovery of fears and phobias. Psychological Review, 92, 512-531.

Jacoby, L. L. (1984). Incidental versus intentional retrieval: Remembering and awareness as separate issues. In L. R. Squire & N. Butters (Eds.), Neuropsychology of memory (pp. 145-156). New York: Guilford Press.

Jacoby, L. L., & Brooks, L. (1984) Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 18). New York: Academic Press.

Jacoby, L. L., & Witherspoon, D. (1982). Remembering without awareness. Canadian Journal of Psychology, 36, 300-324.

Kahneman, D., & Miller, D. T. (in press). Norm theory: Comparing reality to its alternatives. Psychological Review.

Kinsbourne, M., & Wood, F. (1982). Theoretical considerations regarding the episodic-semantic memory distinction. In L. S Cermak (Ed.), Human memory and amnesia (pp. 195-218). Hillsdale, NJ: Erlbaum.

Knapp, A., & Anderson, J A (1984). A signal averaging model for concept formation. Journal of Experimental Psychology: Learning, Memory, & Cognition, 10, 616-637.

Luria, A. R (1968) The mind of a mnemonist. New York: Basic Books.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence Psychological Review, 87, 252-271

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. Journal of Experimental Psychology: General, 114, 159-188.

Medin, D. L., Alton, M. W., & Murphy, T. D (1984). Given versus induced category representations: Use of prototypes and exemplar information in classification. Journal of Experimental Psychology: Learning, Memory, & Cognition, 10, 333-352.

Medin, D. L., & Schaffer, M M. (1978) Context theory of classification learning. Psychological Review, 85, 207-238

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. Journal of Experimental Psychology Human Learning & Memory, 7, 241-253

Metcalfe, J., & Murdock, B. B., Jr (1981). An encoding and retrieval model of single-trial free recall. Journal of Verbal Learning & Verbal Behavior, 20, 161-189.

Milner, B., Corkin, S, & Teuber, H -L. (1968) Further analysis of the hippocampal amnesic syndrome 14 year follow-up study of H.M. Neuropsychologia, 6, 215-234.

Moscovitch, M. (1982). Multiple dissociations of function in amne-

sia. In L. S. Cermak (Ed.), *Human memory and amnesia* (pp. 337-370). Hillsdale, NJ: Erlbaum.

MURDOCK, B. B., JR. (1983). A distributed memory model for serial-order information. *Psychological Review*, **90**, 316-338.

O'KEEFE, J., & NADEL, L. (1978). *The hippocampus as a cognitive map*. London: Oxford University Press.

POSNER, M. O., & KEELE, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, **83**, 304-308.

REBER, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 88-94.

REBER, A. S., & ALLEN, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition*, **6**, 189-221.

SCHACTER, D. L. (1985). Priming of old and new knowledge in amnesic and normal subjects. *Annals of the New York Academy of Sciences*, **444**, 41-53.

SCHACTER, D. L., & GRAF, P. (1985). *Effects of elaborative processing on implicit and explicit memory for new associations*. Manuscript submitted for publication.

SMITH, E. E., & MEDIN, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

STRANGE, W., KEENEY, T., KESSEL, F. S., & JENKINS, J. J. (1970). Abstraction over time of prototypes from distortions of random dot patterns. *Journal of Experimental Psychology*, **83**, 508-510.

TULVING, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press.

TULVING, E (1983). *Elements of episodic memory*. New York: Oxford University Press.

TULVING, E., SCHACTER, D. L., & STARK, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 336-342.

WOOD, F., EBERT, V., & KINSBOURNE, M. (1982). The episodic semantic memory distinction in memory and amnesia: Clinical and experimental observations. In L. S. Cermak (Ed.), *Human memory and amnesia* (pp. 167-194). Hillsdale, NJ: Erlbaum.