

A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli

SAFA R. ZAKI and ROBERT M. NOSOFSKY
Indiana University, Bloomington, Indiana

In previous research (Reed, Squire, Patalano, Smith, & Jonides, 1999), amnesic patients performed at near normal levels in a categorization task involving stimuli with discrete features, but showed impaired recall of the features. These results were taken as evidence of the existence of separate memory systems for categorization and recall/recognition. The present research addresses a single-system account of this dissociation. In Experiment 1, results closely matching the dissociation pattern were obtained when normal participants' classification and recognition performance was tested either immediately or after a week delay. In addition, formal modeling of the data suggested that participants use only a few of the dimensions in the categorization task, whereas they must use many dimensions in the cued-recall task. In Experiment 2, we found that many participants could perform the categorization task without any exposure to the training sequence. These results suggest that different memory demands across the two tasks may be responsible for the observed dissociation—separate memory systems are not necessarily involved.

Research in cognitive neuroscience has recently brought to the forefront an intriguing dissociation between performance on categorization tasks and explicit memory tasks (Knowlton & Squire, 1993; Knowlton, Squire, & Gluck, 1994; Reber & Squire, 1999; Squire & Knowlton, 1995). Amnesic patients are shown to perform at near normal levels in categorization tasks, whereas they show a substantial deficit in explicit memory tasks such as recognition and recall. This dissociation in performance is interpreted as evidence that separate cognitive systems govern recognition/recall and categorization. A declarative memory system, housed in the medial temporal lobe and diencephalic brain structures, is responsible for recognition and recall, whereas a procedural memory system is responsible for more implicit types of memory, including categorical knowledge.

For example, in an influential article, Knowlton and Squire (1993) demonstrated this dissociation using the classic Posner and Keele (1968) dot-pattern paradigm. In the categorization task, participants trained on 40 high-level distortions of a prototype. Following this study phase, amnesic and control participants were tested on various new items that were generated from the same category prototype. The participants' task was to decide

whether each of the new patterns was a member of the previously presented category. In a recognition task, the participants studied five unrelated dot patterns eight times each and were then tested with five new and five old patterns. As one would expect, the amnesics performed substantially worse than controls on the recognition task. The pivotal result, however, was that the amnesics performed at near normal levels in the categorization task. This finding led Knowlton and Squire to argue that categorization and recognition were based on different memory systems. A declarative system that is responsible for recognition is damaged in amnesics, causing their performance decrement in that task, whereas the amnesics' procedural system is intact, thereby accounting for their unimpaired categorization performance.

However, a plausible single system explanation is also consistent with these findings. Following the exemplar-based approach (Medin & Schaffer, 1978; Nosofsky, 1986, 1988, 1991), Nosofsky and Zaki (1998) suggested that participants perform both recognition and categorization by comparing the test items to individual exemplars stored in memory. Naturally, the exemplar-based memories of the amnesics are impaired. Such impairment would significantly reduce recognition performance for the amnesics relative to the healthy controls because recognition requires the ability to make fine-grained discriminations between old versus new items. By contrast, the impaired memory for the individual items does not have a drastic impact on categorization performance because the categorization task does not require the observer to make fine-grained discriminations between items. Indeed, for purposes of categorization,

This research was supported by Grant PHS R01 MH48494 from the National Institute of Mental Health. We thank John Jonides and three anonymous reviewers for their helpful criticisms of an earlier version of this article. We also thank John Jonides for suggesting the idea behind Experiment 2 of this article. Correspondence should be addressed to S. Zaki, Department of Psychology, Indiana University, Bloomington, IN 47405 (e-mail: szaki@indiana.edu).

the observer should instead take advantage of whatever similarity exists between the novel transfer items and the old training exemplars. In a nutshell, because of the redundancy of information that is available in the categorization task, impaired memory does not substantially hurt categorization performance, so amnesics perform at near normal levels.

Nosofsky and Zaki (1998) provided two forms of evidence to support this single-system interpretation of Knowlton and Squire's (1993) results. The first involved indirect experimental evidence in the form of an "amnesia simulation." Participants in this experiment were given the Knowlton and Squire recognition or categorization test either immediately after training or after a 1-week delay. Delay of the testing session was introduced to test the effect of degrading participants' memory for the training items on categorization and recognition performance. Interestingly, the results of the delayed group were strikingly similar to those exhibited by the amnesic group in that there was a large decline in recognition performance but little decline in categorization performance. Although it is possible to reconcile these results with a multiple-systems account by positing different decay parameters for the declarative and procedural systems, this demonstration supports the plausibility of a single-system model.

More importantly, Nosofsky and Zaki (1998) also demonstrated that a formal exemplar model, the generalized context model (GCM), which incorporated this difference in memory sensitivity across the two groups, provided an excellent quantitative fit to the recognition and categorization data. Therefore, the formal modeling of the data provided converging evidence to support the idea that the categorization task contains more redundant information than the recognition task, and that different task requirements in categorization and recognition may be solely responsible for the observed dissociation. This notion is further supported by Palmeri and Flanery (1999), who demonstrated that participants could perform the categorization task at normal levels without *any* exposure to the sequence of training exemplars.

However, in a recent article, Reed, Squire, Patalano, Smith, and Jonides (1999) reported a new set of findings that they argued to be damaging to the single-system account of recognition and categorization. In this article, Reed et al. suggested that the dot-pattern paradigm might not be conducive to learning to categorize using a declarative system, because the dot patterns are difficult to memorize as distinct items. Thus, they suggested that a paradigm in which participants studied object-like stimuli that vary on discrete dimensions may be more likely to induce participants to use their declarative system for categorization. Therefore, if categorical knowledge depends on the explicit items in memory, then amnesics should not be able to perform the task at normal levels. According to these authors, a demonstration that amnesics performed well in a categorization task involving such object-like stimuli would

provide convincing evidence in favor of the multiple-systems view.

In Reed et al.'s (1999) study, amnesics and matched controls studied cartoon animals that varied on nine binary-valued dimensions. Following this training, the participants were tested on a new set of animals that either belonged to the category or did not. In addition, the participants were given an explicit memory test in which they were asked to recall the two possible values along each of the nine dimensions. As expected, amnesics showed significantly worse performance in the recall of the training items relative to the control group. Interestingly, however, they showed statistically equivalent classification performance. Reed et al. interpreted these results as being challenging to the single-system approach to categorization and recognition.

However, much like in the dot-pattern paradigm used by Knowlton and Squire (1993), the memory requirements of the two tasks used by Reed et al. (1999) may be quite different. To illustrate this point, consider Table 1, which shows the structure of the type of categorization training stimuli used by Reed et al. An examination of this structure reveals that there is a great deal of redundant information across dimensions in the categorization task, but not the recall task. For example, it turns out that if participants attend to only three of the nine dimensions, and always place a transfer item into the category if it matches the prototype on two of these three dimensions, they can achieve 96% correct on the categorization task. Participants can even do extraordinarily well and classify 87% of the stimuli correctly if they attend to any single one of the nine dimensions.¹ Indeed, these performance levels are considerably higher than the accuracy achieved by the participants in the Reed et al. study, who scored 74% correct in the categorization task. In contrast, participants can perform well in the recall task only to the extent that they retain information from all of the nine dimensions of the stimuli. Perhaps it is this difference in task requirements that allowed both amnesics and controls to perform reasonably well in the categorization task, but caused a difference to emerge in the results of the recall test.

The purpose of this article is to examine a single-system account of recognition/recall and categorization with respect to the results of the experiments reported by Reed et al. (1999). Specifically, the article will test the idea that different task requirements may have allowed the amnesics to perform at normal levels in the categorization task but not in the explicit memory task. Because we do not have access to amnesic participants, our methods, following those of Nosofsky and Zaki (1998), will involve the use of normal participants only.

EXPERIMENT 1

There were two main goals of this experiment. The first involved a manipulation similar to that used by Nosofsky and Zaki (1998), designed to test the idea that

Table 1
Binary Values of the Nine Dimensions of the Training Stimuli Used in Experiment 1

Body Shape	Feet	Legs	Tail	Head Ornament	Face	Neck Length	Neck Curvature	Body Markings
1	1	1	1	1	1	0	1	1
1	1	0	1	1	1	1	1	1
1	1	1	0	1	1	1	1	1
1	1	1	1	1	0	1	1	1
0	1	1	1	1	0	1	1	1
1	0	1	0	1	1	1	1	1
0	0	1	1	1	1	1	1	1
1	0	1	1	1	1	1	0	1
1	1	0	1	1	1	1	1	0
1	1	1	1	0	0	1	1	1
0	1	1	1	1	1	1	0	1
0	1	1	1	0	1	1	1	1
1	0	1	1	0	1	1	1	1
1	1	1	1	1	1	1	0	0
1	1	1	0	1	1	0	1	1
1	1	0	1	0	1	1	1	1
1	1	1	1	1	0	1	0	1
1	1	1	1	1	1	0	1	0
0	1	1	1	1	1	0	1	1
1	1	0	0	1	1	1	1	1

because of the redundancy of information in the categorization task, degraded memory for the training items would lead to a small decline in categorization performance but a large decline in explicit recognition performance. Our categorization task was the same as the one used by Reed et al. (1999). The stimuli were cartoon animals that varied on nine binary-valued dimensions. Participants trained on low-level distortions of a prototype and were then tested with new items of mixed distortion levels. Some of the items belonged to the previously presented category and some did not. Participants judged which of the items were category members. We substituted a recognition task for the cued-recall task used by Reed et al. to allow for more direct comparison with the categorization results and to allow for possible quantitative modeling of the data. This recognition task was modeled after the one used earlier by Knowlton and Squire (1993). In the study phase, the participants trained on five nonrelated cartoon animals. They were then given a recognition test that contained the five old items and five new randomly generated items. Following the instructions used by Knowlton and Squire, participants judged whether or not each item was exactly the same as one of the items presented during study.

The critical difference between the present experiment and the one described by Reed et al. (1999) was that, for half the participants, the testing session did not always occur immediately, but rather after 1-week delay. Although we did not expect to truly simulate amnesia in the participants, the delay manipulation was intended to degrade their memories for the training items. The hypothesis was that delay, just like amnesia, would affect categorization and recognition differentially because of the different memory loads required in these two tasks.

The second goal was to collect a set of data that would allow for testing the GCM. As will be described below,

the GCM includes certain parameters that measure the extent to which participants attend to individual dimensions of the objects. The main goal of the present experiment was to determine how many dimensions of the objects participants in this task were actually using for the purpose of categorization. That is, do participants rely on the redundancy of information in the categorization task and focus their attention on only a few dimensions, or do they distribute their attention to most or all of the dimensions, as suggested by Reed et al. (1999)? If participants are using only a few dimensions in the categorization task, that would imply different memory demands in the categorization and recognition/recall tasks and would therefore provide a potential alternative explanation of the Reed et al. findings.

Method

Participants. A total of 88 Indiana University undergraduate students participated in the study. Half of these were randomly assigned to the categorization task, and the other half to the recognition task. In addition, participants were randomly assigned to one of two subconditions in which they participated in the test session either immediately after training or following a 1-week delay. All participants completed the research in order to fulfill a class requirement. A bonus of \$15 was offered to the top 4 performers in the experiment to ensure motivated participants. Data from 2 participants in the categorization task were not used because they pressed only one key during the transfer phase.

Stimuli. For the categorization task, the stimuli consisted of 82 cartoon animals similar to those used by Reed et al. (1999). Figure 1 gives examples of the stimuli. The stimuli differed from each other along nine binary-valued dimensions. The dimensions were head type, face shape, head ornaments, neck curvature, body shape, body markings, tail shape, leg length, and foot shape. For the categorization task, the prototype used by Reed et al. was constructed and 20 unique study items were generated from this prototype. The study items were low distortions; that is, they differed from the prototype on one or two features. The transfer test consisted of 12 instances of the prototype, 24 new low distortions (items with seven

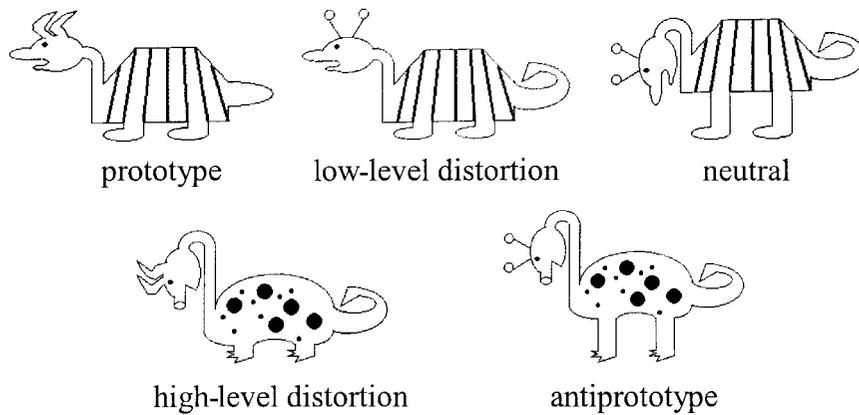


Figure 1. Examples of the stimuli used in Experiments 1 and 2. Prototypes contained nine prototypic features; low distortions contained seven or eight; neutral animals contained four or five; high distortions contained one or two; and the antiprototype contained no prototypic features.

or eight prototypic features), 24 neutral animals (items with four or five prototypic features), 24 high distortions (items with one or two prototypic features), and 12 instances of the antiprototype (no prototypic features). For half the participants, the direction of distortion was reversed in such a way that the prototype became the antiprototype, the low-level distortions became the high-level distortions, and so forth. The actual structure of the training exemplars used in the categorization task is the one listed in Table 1. For the recognition task, the stimuli consisted of 10 items that were a random subset of all possible combinations of the binary values of the nine features. Five of these were reserved for use in training and five were reserved for use as new items in the test.

Procedure. Participants were trained on either the categorization or recognition stimuli and were given a transfer test either immediately or after a 1-week delay. In the categorization training phase, the participants saw two blocks of the 20 low-distortion an-

imals on a computer screen for 5 sec each. They were instructed to simply attend as best they could. Following Reed et al. (1999), in the transfer test, participants were told that all of the animals that they had seen were members of the category "Peggle." They were instructed that they would see some new animals, half of which would be Peggles and half would not. Their goal was to decide whether a particular animal belonged to the Peggle category. In the transfer instructions, the participants were also told that the animals could vary on nine dimensions, the nine dimensions were listed,² and the participants were instructed to use as many of these dimensions as possible when making their categorization decisions in order to win one of the bonuses. They were then shown the 96 transfer items, one at a time, and were asked to indicate whether or not each animal belonged to the category. No feedback was provided.

In the recognition task, participants studied the five old items eight times each, with the order of stimuli randomized in each

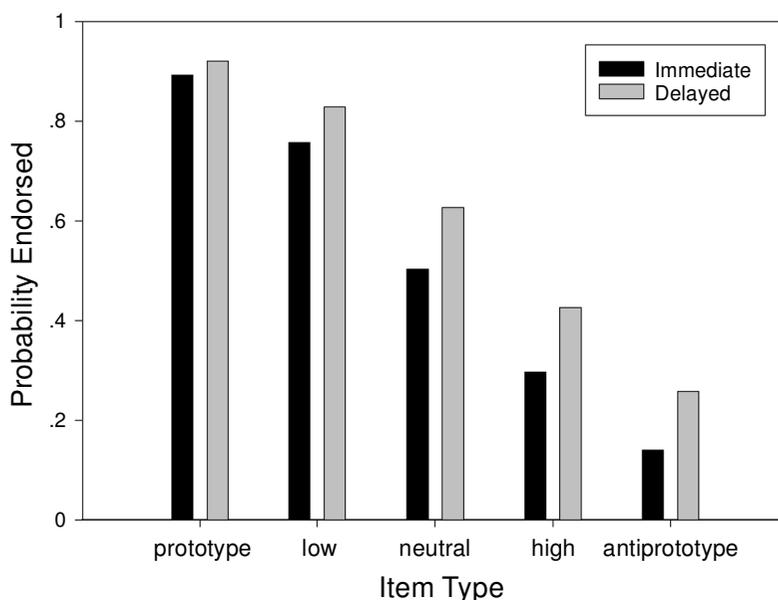


Figure 2. Endorsement of the different item types in the immediate and delayed conditions.

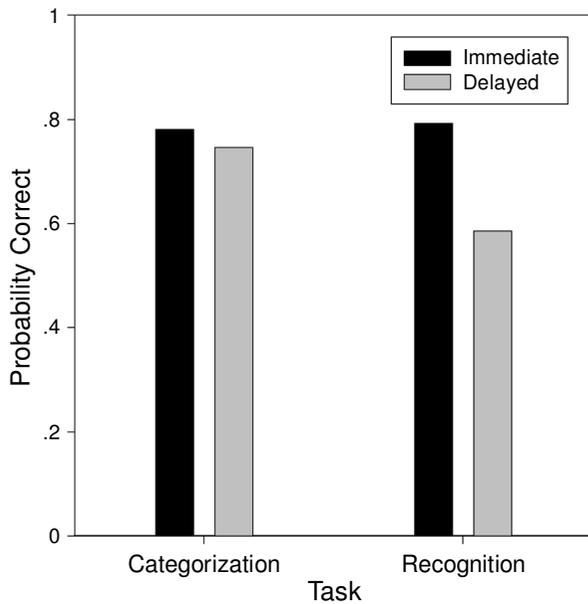


Figure 3. Categorization and recognition accuracy for the immediate and delayed conditions in Experiment 1.

block. Each item was displayed for 5 sec. Participants received the same training instructions as participants in the categorization group. In the test that followed, participants saw the five old items as well as five new items and were asked to indicate whether they thought a particular animal was exactly the same as one they had previously seen.

Results

The probability with which participants endorsed the animals as members of the category is shown in Figure 2. As expected, items that were closer to the prototype had higher endorsement rates in both the immediate and the delayed condition [$F(4,160) = 103.97, p < .001$]. In addition, there was a significant effect of delay, with participants tending to place more items into the category in the delayed test [$F(1,40) = 5.737, p < .05$]. There was no interaction between item type and time of test.

Overall categorization and recognition performance are displayed graphically in Figure 3. Following Reed et al. (1999), accuracy on the categorization task was calculated as follows. A response was scored as correct if the participant endorsed a prototype or low-level distortion. In addition, the response was scored as correct if the participant did not endorse a high-level distortion or antiprototype. The performance on the neutral animals was not used in the calculation of accuracy. Critically, categorization accuracy was statistically equivalent across the immediate (.78) and the delayed (.74) [$t(40) = 0.716, p > .10$] conditions. Furthermore, the numeric difference between the categorization accuracy of the groups (.04) was similar to the numeric difference found by Reed et al. between the control and amnesic participants (.74 vs. .69 when two of their most amnesic patients are eliminated because of the possible reversal of

category labels; see Reed et al. for details). On the other hand, overall performance in the recognition task showed a steep drop in accuracy in the delayed condition (.59) versus the immediate condition (.79) [$t(42) = 4.965, p < .001$]. The same pattern of results was reflected in a d' analysis indicating that the observed dissociation was not due to changes in criterion settings.³ That is, there was no significant difference in the d' values of the immediate group (mean $d' = 1.58$) versus the delayed group (mean $d' = 1.36$) in categorization [$t(40) = 0.728, p > .10$]. However, for the recognition condition there was a significant drop in d' in the delayed group (mean $d' = 0.53$) relative to the immediate group (mean $d' = 1.78$) [$t(42) = 4.80, p < .001$]. Clearly, the delay manipulation was successful in inducing memory impairments in the explicit recognition task.

The greater tendency of our delayed-condition participants to endorse items into a category was not observed in the amnesic group in Reed et al. (1999). Undoubtedly, our delayed group differs from the amnesic group in certain critical respects. However, the general pattern of results that is taken as evidence of separate memory systems in Reed et al. is replicated in this study. That is, our delay manipulation yielded a dissociation between recognition and categorization performance that was similar in form to the dissociation that Reed et al. observed for their normal and amnesic participants. Our interpretation is that because good performance in the old–new recognition task (or in Reed et al.'s cued-recall task) requires retention of information from multiple dimensions, a delay in testing (or an amnesic disorder) is particularly detrimental to performance in such a task. By contrast, because good performance in the categorization task requires retention of information from exceedingly few dimensions, the delay in testing (or an amnesic disorder) has little impact. This line of reasoning is indirect, however, and such an interpretation of the results is not forced. For example, it may be that separate memory systems govern recognition and categorization, and that the categorization system simply shows slower forgetting characteristics than does the recognition system. Therefore, we now turn to the more important goal of this study—namely, the formal modeling of the categorization data to test whether observers do indeed perform the task by focusing on only a small subset of the component dimensions of the stimuli.

THEORETICAL ANALYSIS

In this section we describe the details of the GCM and how it was applied to the data. According to the model, people represent categories by storing individual exemplars in memory. Test items act as retrieval cues for the stored exemplars. Specifically, exemplars from alternative categories are retrieved on the basis of how similar they are to the test item or probe. In a process-oriented version of the model (Nosofsky & Palmeri, 1997), the retrieved exemplars drive a random walk process that ac-

cumulates evidence toward alternative categories. Once sufficient information is retrieved, the observer initiates a response.

A crucial assumption in the GCM is that similarities among exemplars are influenced by a selective attention process in which observers differentially weight the individual dimensions that compose the objects. Objects that match along the attended dimensions are judged to be highly similar. The selective attention process is expected to play a crucial role in classification learning situations involving stimuli that vary along discrete, separable dimensions, such as the object-like stimuli used by Reed et al. (1999).

Formally, according to the model, exemplars are represented as points in a multidimensional space. The present stimuli varied along nine binary-valued separable dimensions. Therefore, the psychological distance between two stimuli is calculated by using the weighted city-block metric,

$$d_{i,j} = \sum_{m=1}^9 w_m |x_{i,m} - x_{j,m}|, \tag{1}$$

where $x_{i,m}$ is the value (either 0 or 1) of item i on dimension m , and w_m ($0 \leq w \leq 1, \sum w_m = 1$) is the attention weight given to dimension m . A geometric interpretation of the weights is that they “stretch” the space along the attended dimensions and “shrink” the space along the unattended dimensions (Nosofsky, 1986). The similarity between two stimuli is an exponential decay function of their distance:

$$s_{i,j} = e^{-cd_{i,j}}. \tag{2}$$

There is widespread empirical support for the assumption of an exponential relation between similarity and psychological distance (see Shepard, 1987, for a review and theoretical analysis). The c parameter in Equation 2 is a sensitivity parameter that reflects the overall discriminability of the items in the psychological space.

In the present single-category situation, the probability that an observer endorses an item as being a member of the category is given by

$$P(\text{Peggle} | i) = \frac{[\sum s(i, \text{peggle})]^\gamma}{[\sum s(i, \text{peggle})]^\gamma + k_c}, \tag{3}$$

where $\sum s(i, \text{peggle})$ is the summed similarity of item i to all previously presented exemplars from the Peggle category. The parameter k_c is a categorization criterion that represents the bias to endorse items as members of the category. The response-scaling parameter γ , introduced by Ashby and Maddox (1993), describes the extent to which participants use deterministic versus probabilistic response rules. When $\gamma = 1$, participants respond probabilistically by “matching” to the relative summed similarities of each category, whereas as γ grows greater than 1, participants respond more deterministically. In the random-walk version of the model developed by Nosof-

sky and Palmeri (1997), γ corresponds to the magnitude of the decision criteria in the random-walk process. Intuitively, it represents the amount of exemplar-based evidence that needs to be retrieved before an observer will initiate a response.

For the categorization data, the full version of the GCM has one overall sensitivity parameter, one response-scaling parameter, a criterion parameter, and eight freely varying attention weights (the nine weights are constrained to sum to 1). For simplicity in the modeling, however, we set the response scaling parameter at $\gamma = 1$. We fit the model to each individual participant’s data separately by using a maximum-likelihood criterion.⁴ Three sets of starting parameters were used to avoid local minimum problems, and the best fit from these three is the one reported here.

In Table 2 we report the mean observed and predicted categorization probabilities for the various types of stimuli in the immediate and delayed tests. Recall that the model was fitted to each individual participant’s data. The predicted values in Table 2 are averages computed across the individual participants’ fits. Inspection of Table 2 reveals that the predictions of the exemplar model are very similar to the observed findings. The model captures nicely the prototypicality effect, where items closer to the prototype have a higher endorsement rate. The model also does a good job of fitting the differences between the immediate and delayed conditions for reasons that we describe below.

In Table 3 we report the mean maximum likelihood parameters as well as the summary fits for the full version of the model. Again, the reported mean parameter values are averages computed across each individual participant’s best-fitting parameter values. Inspection of the mean parameter values indicates that, averaged across participants, there is little difference in the salience of the dimensions. Because particular features had particular

Table 2
Observed and Predicted Classification Probabilities in Experiment 1

	Observed	GCM-Full	GCM-3
Immediate			
Prototype	.89	.86	.83
Low	.76	.78	.75
Neutral	.50	.54	.53
High	.30	.26	.30
Antiprototype	.14	.12	.14
Accuracy	.78	.80	.77
Delayed			
Prototype	.92	.92	.90
Low	.83	.85	.82
Neutral	.63	.65	.64
High	.43	.39	.43
Antiprototype	.26	.22	.26
Accuracy	.75	.77	.77

Note—GCM-Full refers to the baseline version of the GCM with one sensitivity parameter, one criterion parameter, and nine attention weight parameters. GCM-3 refers to the version of the GCM in which nonzero weights were allowed on only three of the nine dimensions.

Table 3
Average Maximum-Likelihood Parameters and
Summary Fits of GCM With Nine Dimension Weights
to the Categorization Data

Parameter	Immediate	Delay
c	15.90	13.54
k	2.98	1.42
$w_{\text{body shape}}$.11	.23
w_{feet}	.12	.10
w_{legs}	.20	.10
w_{tail}	.05	.07
$w_{\text{head ornament}}$.11	.17
w_{face}	.08	.06
$w_{\text{neck length}}$.15	.09
$w_{\text{neck curvature}}$.10	.10
$w_{\text{body markings}}$.07	.08
$-\ln L$	30.40	33.15
SSE	9.28	9.90

Note— c , general sensitivity parameter; k , criterion parameter; w_m , attention weight given to dimension m ; $-\ln L$, negative value of log likelihood; SSE , sum of squared deviations between observed and predicted classification probabilities.

abstract stimulus values associated with them, these dimension weights must be interpreted with caution. However, it is not unusual for a relatively equal salience of dimensions to be reported in studies that utilize cartoon animal stimuli (e.g., Minda & Smith, 2001). In addition, the obtained values of the sensitivity parameter were also very similar across the immediate and delayed conditions. The main difference between the parameter values across the two conditions is in the criterion parameter (immediate = 2.98, delayed = 1.42). That is, the model, like participants in the experiment, places more items

into the category in the delayed condition. This parameter difference also allows the model to capture the slight decrease in accuracy in the delayed condition.

The key question, however, is how are *individual* observers distributing their attention across the nine dimensions? Are participants attending to only a subset of the dimensions that compose the objects? To gain some preliminary insight into this issue, we conducted an analysis in which, for each individual participant, we rank-ordered the attention weights from greatest to smallest and then computed the cumulative weight distribution for that participant. In Figure 4 we report the mean cumulative weight distribution averaged across all participants. (The cumulative distributions were essentially the same for the participants from the immediate and delayed groups.) The figure reveals that, on average, the most utilized dimension received about half of a given participant's attention, and that participants distributed over 90% of their attention to the top four utilized dimensions. This analysis provides preliminary evidence that in the present categorization task, participants were focusing attention on a relatively small subset of the dimensions that compose the objects.

To examine individual participants' performance more fully, we fitted various restricted versions of the model to each individual participant's data. In these restricted models, the attention weight parameters were constrained so that nonzero weights were allowed on only one, two, three, or four of the dimensions. For example, for the one-dimension version of the model, we assumed that participants attended to only a single dimension of the nine-dimension objects, and all remaining attention weights were set to zero. For each individual participant, the attended dimensions were assumed

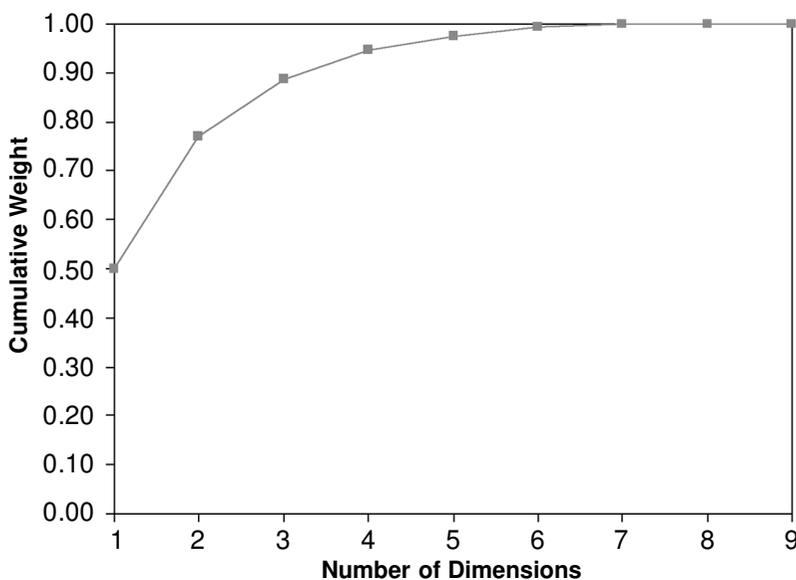


Figure 4. Mean cumulative dimension weights computed from the full version of the generalized context model.

to be different. That is, given the structure of the category, different participants may have chosen very different dimensions to which to attend with similar degrees of success in the categorization task. Therefore, for each participant, we rank-ordered the dimensions according to the magnitude of weights when the full model was fitted to the data, and entered them into the restricted model according to this ranking. For example, for each participant, a two-dimension weight model was fitted to the data. The two dimensions that entered into the fit for the given participant were those whose weights had the highest magnitude when the full version of the model was fitted to that participant's data.

A one-, two-, three-, and four-dimension weight model was fitted to each participant's data individually, as described above. We then used the method of likelihood ratio testing to determine whether each restricted model fit a participant's data significantly worse than the full version, which allowed all nine attention weights to vary freely.⁵ The results indicated that for 15 of the 42 participants in the categorization task, a model in which participants attended to only one of the nine dimensions fit the data as well as did the full model. For 29 out of the 42 participants, the two-dimension weight model was equivalent in fit to the nine-dimension weight model. For 40 out of the 42 participants, the three-dimension weight model was equivalent in fit, and for all 42 participants, there was no significant improvement in fit from the four-dimension weight model to the nine-dimension weight model.

To illustrate the adequacy, for example, of the three-dimension restricted version of the model, we plot in Figure 5 the mean observed category endorsement probabilities for all 72 objects in the categorization test against the endorsement probabilities predicted by the exemplar model. The top panel of Figure 5 shows the results for the full version of the exemplar model, and the bottom panel shows the results from the three-dimension restricted version. Clearly, the version of the model that assumes that participants attended to only three (or potentially fewer) dimensions in performing the task provides an excellent account of the pattern of categorization data. The summary results for the five main types of patterns are shown along with the results from the full model in Table 2.

At first glance, the results of these model fits appear to be at odds with the results of an analysis performed by Reed et al. (1999) on their data. In their article, Reed et al. raised the point that the averaged pattern of endorsement of the various stimulus types could have emerged if participants based their responses on a single dimension. They dismissed this possibility, however, on the grounds of the following analysis. Reed et al. calculated the endorsement probabilities for the most and least frequently endorsed feature for each participant in their experiment. They argued that participants could not be using a single-feature strategy because, on average,

participants endorsed the most frequently endorsed feature with a probability of only .773. In addition, because the least frequently endorsed feature was endorsed (.594) with higher probability than chance, the authors argued that "the participants had learned about the prototypical category in a broad sense by learning something about each of the nine features" (p. 415). A similar analysis conducted on the present data revealed the same pattern, in which the most frequently endorsed feature for each participant was endorsed with a probability well below 1 ($M = .760$), and the least endorsed feature was endorsed with a probability greater than .5 [$M = .573$, $t(41) = 4.23$, $p < .001$].

It turns out, however, that the low-dimension versions of the exemplar model, even the single-dimension version, predict this pattern of feature endorsement rates. For example, we obtained predictions from the single-dimension exemplar model by using the model to generate responses from 1,000 simulated participants. For each simulated participant, a single dimension was chosen randomly and all attention weight was placed on this dimension. The model was then used to predict the endorsement probabilities for each of the stimuli in the categorization transfer phase. Let p denote the predicted probability with which a given stimulus is endorsed. (Note that this predicted probability will differ depending on whether the attended feature is present or absent.) To simulate the response, a random number in the interval (0,1) was selected. If the random number was less or equal to p , then an endorsement response was chosen; if not, a nonendorsement response was chosen. Finally, we conducted the Reed et al. (1999) analysis on these simulated data.

The results are shown in Table 4.⁶ As can be seen, the single-dimensional model predicts that the most frequently endorsed feature is endorsed with a probability of .812, whereas the least frequently endorsed feature is endorsed with a probability of .563. The predicted endorsement probabilities for the remaining features decrease regularly from the most frequently endorsed feature to the least frequently endorsed one. This pattern of results is similar to the one observed by Reed et al. (1999).

The reason that the single-dimension model predicts that even the most frequently endorsed feature has an endorsement rate below 100% is that the exemplar model uses a probabilistic response rule (Equation 3). Thus, even if an observer attends to information from just a single dimension, the information from that dimension

Table 4
Predicted Endorsement Probabilities for the Nine Features
Rank-Ordered From Most Endorsed (1) to Least Endorsed (9)

Feature Ranking								
1	2	3	4	5	6	7	8	9
.812	.677	.653	.637	.625	.613	.600	.585	.563

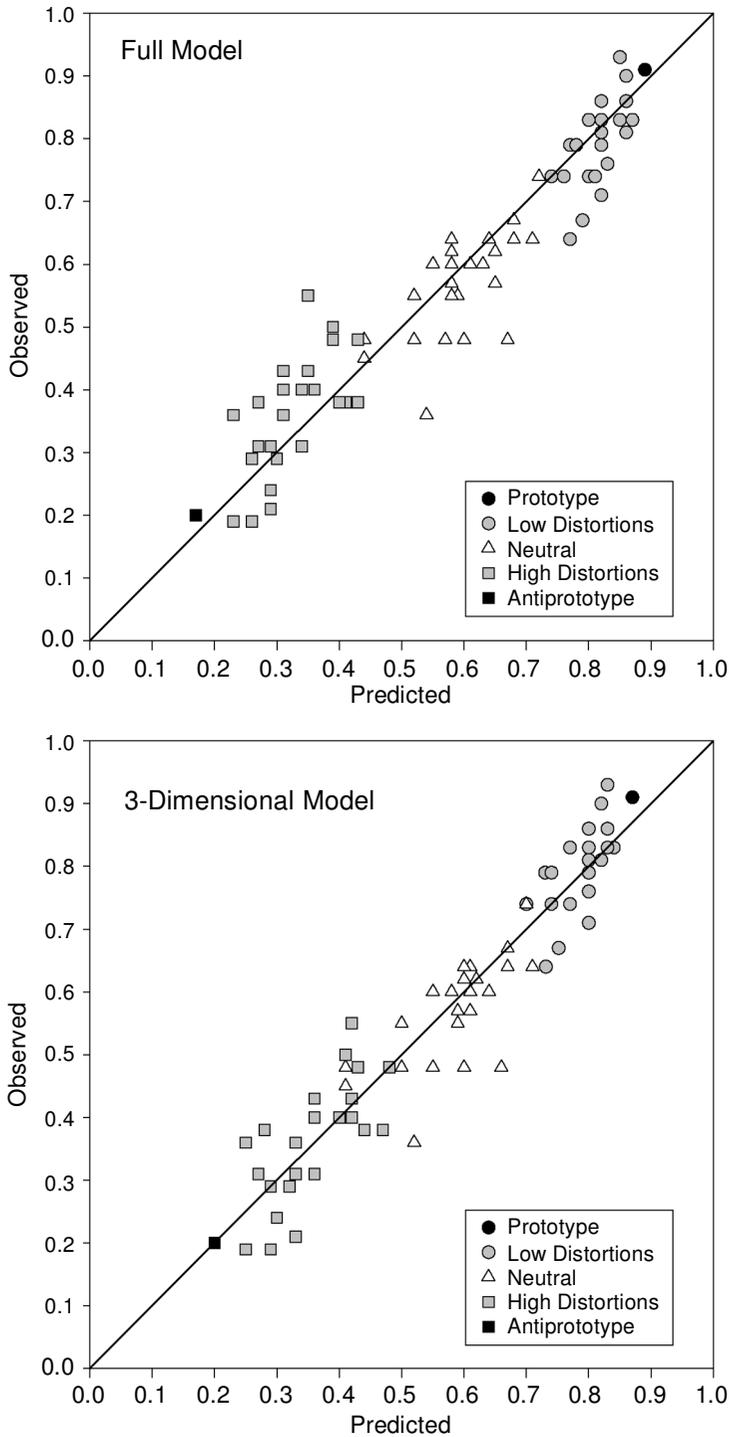


Figure 5. Scatter plots of observed against predicted classification endorsements. The top panel is for the full version of the exemplar model, and the bottom panel is for the three-dimensional version of the model.

does not result in a deterministic response. More interesting, the single-dimension exemplar model can predict that even the least frequently endorsed feature has an endorsement rate greater than chance (50%) because the

dimension values in the set of test stimuli used by Reed et al. (1999) are highly correlated. For example, if Dimension 1 of an object has a value that is consistent with the category, it is likely that the remaining dimensions

do as well, and vice versa. Thus, if an observer endorses a pattern because its Dimension 1 value points to the category, the observer is likely making a response that is consistent with the values on the remaining dimensions as well.

Although the results in Table 4 provide an example involving only a single-dimension version of the exemplar model, the same pattern is obtained when the two-, three-, and four-dimension versions are used to simulate the participants' responses. Thus, the single-feature analyses reported by Reed et al. (1999) provide converging evidence in favor of the low-dimension versions of the exemplar model rather than in opposition to them.

EXPERIMENT 2

To provide further evidence that separate memory systems mediate categorization and recall/recognition, Reed et al. (1999) reported data from a densely amnesic single subject, E.P., with extensive damage to the medial temporal lobe. The intriguing aspect of E.P.'s performance is that although E.P. showed zero ability to recall the dimensions, he performed at the same levels as controls in the categorization task. More precisely, in Experiment 1A of Reed et al.'s study, E.P. displayed an inverted pattern of responding in which he consistently endorsed nonmembers of the category and rejected category members. Thus, E.P. discriminated among the item types, but confused which end of the feature continuum represented category members versus nonmembers.

To address the concern regarding E.P.'s inverted response profile, in a second experiment Reed et al. (1999) tested E.P. with a new set of stimuli that included only a single category endpoint as well as neutral items. Reed et al. reasoned that E.P. would be able to perform this categorization task correctly because he would not need to remember which end of the feature continuum to endorse. In accord with this prediction, E.P. did show the correct pattern of endorsement in this experiment. Taken together, Reed et al. interpreted these results as strong support for the existence of separate memory systems for categorization and recognition/recall.

However, as we have already demonstrated, observers can perform these categorization tasks by attending to only a small subset of the dimensions of the objects and responding consistently on the basis of these dimension values. In the extreme, E.P. could categorize the test stimuli on the basis of a single dimension. This response strategy would yield performance levels as good or better than the ones that Reed et al. (1999) observed. Although Reed et al. considered this possibility, they rejected it on the grounds of the single-feature analysis that we described earlier in this article. However, as we have explained previously, the low-dimension versions of the exemplar model *predict* the results of the single-feature analyses conducted by Reed et al.

Given that E.P. could not recall any of the features of the training stimuli, one might question whether he would

have sufficient memory to use even a single-dimension value as a basis for classification during test. Our hypothesis, however, does not require any long-term memory of the dimension values experienced during training. All that is required is that E.P. continually recycle in his short-term working memory his use of a single-feature strategy at time of test. (Note that E.P. would need the same type of working memory in order to simply follow the experimental instructions.) Indeed, the inverted response profile that E.P. displayed in Reed et al.'s (1999) Experiment 1A is highly consistent with the idea that he used a low-dimension response strategy without memory for the feature values experienced during training.

It is true that E.P. reversed his pattern of endorsements in Reed et al.'s (1999) Experiment 2. However, these results involving E.P. and the direction of endorsements involve a sample size of 1. The reliability of the phenomenon is unknown. According to a single-feature hypothesis, there is a 50% chance that E.P. would go in either direction. Moreover, in the design of Reed et al.'s Experiment 2, the base rate of category features exceeded the base rate of noncategory features. Conceivably, this difference in base rates may increase the likelihood of observers making feature endorsements in the right direction, even in the absence of any long-term memory for the features experienced during training.

The purpose of the present Experiment 2 was to test the idea that even with no exposure to the original training instances, numerous participants would exhibit categorization behaviors that were similar to the behaviors displayed by E.P. in Reed et al.'s (1999) tasks. To achieve this goal, we employed a mock subliminal training procedure similar to the one used by Palmeri and Flanery (1999) in an earlier investigation of E.P.'s behavior. Participants were led to believe that they had been subliminally shown various members of a category, when in fact they had seen no study items. Following this phase, we tested the participants in a transfer test with the same design as that used in Experiment 2 of Reed et al. (1999). The participants' task was to decide for each test item whether it was a member of the category of previously displayed items. The key question was whether, even with no exposure to the training instances, participants would respond with a typicality gradient similar in form to those observed for E.P. in Reed et al.'s tasks. If E.P.'s performance falls within the range of performances of participants who have had no prior experience with the training items, this result would imply that E.P. may not have had any long-term implicit memory for the training items. It would also lead one to question the need to posit separate memory systems for categorization and recognition/recall.

Method

Participants. Thirty-seven Indiana University undergraduate and graduate students were paid \$5 for participating in the experiment. The students were unaware of the issues under investigation in this study.

Stimuli. The stimuli consisted of the same kind of cartoon animals used in Experiment 1. The test consisted of 60 items. There were 12 instances of the prototype, 24 low distortions, and 24 neutral animals.

Procedure. The experiment consisted of a mock subliminal training phase followed by a transfer test similar in form to the one used in Experiment 2 of Reed et al.'s study. On each trial in the mock subliminal phase, participants saw a fixation point followed by a blank screen. Participants were told that we were testing their subliminal perception of cartoon animals that were very briefly flashed on the screen. In the transfer phase, participants were given a test that was conceptually similar to the test used in the transfer phase of Reed et al.'s Experiment 2. The participants were instructed to guess whether they believed each animal was a Peggles or not. The instructions read in part as follows:

For each animal, simply guess whether you believe it is a PEGGLE or not. Obviously, you will have no conscious recollection of the cartoon animals, so you will need to rely on various forms of unconscious information to perform the task—gut instinct, a sense of familiarity, and so forth.

In addition, participants were informed that if they scored significantly above chance in deciding category membership, they would be paid a bonus of \$5. The order of test items was randomized for each participant.

Results

We divided the participants into three groups according to their overall accuracy in the transfer phase. Assuming random responding on each item, chance accuracy is 50% correct. Assuming binomial variability, on

the basis of 60 observations, the total probability of an observer scoring greater or equal to 39 correct responses (65% correct), or less than or equal to 21 correct responses (35% correct), is .027. Thus, we consider participants who scored at these levels to have performed significantly differently from chance, and we will refer to them as “learners.” Of the 37 participants in our study, there were 23 learners. (Assuming binomial variability, the probability of observing this many learners by chance alone is less than 3.54×10^{-27} .) Fourteen of the learners used the category labels correctly, whereas 9 used reversed category labels, generally endorsing only nonmembers of the category. Figure 6 gives the overall distribution of accuracy in the test, with the performance of the learners indicated by the solid bars. For ease of comparison, E.P.'s performance in Reed et al.'s (1999) Experiment 2 is indicated by the dotted line. The critical result is that E.P.'s accuracy level falls well within the range displayed by participants who had no prior exposure to the stimuli. Indeed, 54% of our participants displayed more extreme accuracy levels (i.e., levels differing more from 50%) than did E.P.⁷

Figure 7 shows the typicality gradient for the learners who used the reverse category labels, participants who were at chance, and learners who used the correct category labels. In addition, E.P.'s typicality gradient, reported in Reed et al.'s (1999) Experiment 2, is also plot-

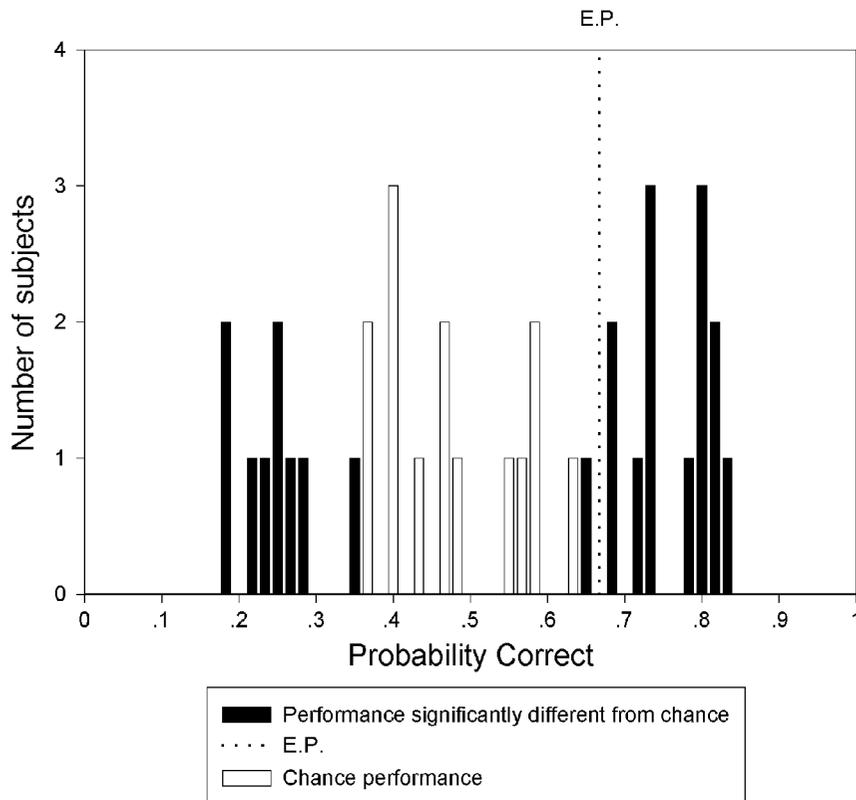


Figure 6. Histogram of probability correct for participants in Experiment 2. The dotted line indicates the performance of E.P. in Reed et al.'s (1999) Experiment 2.

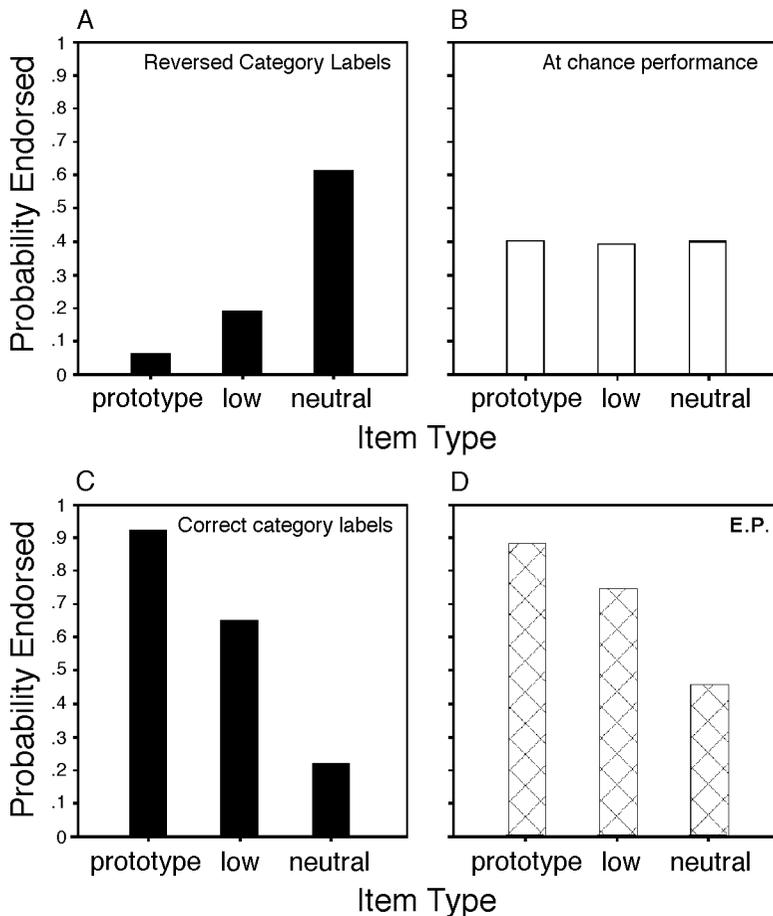


Figure 7. Typicality gradients observed for (A) participants in Experiment 2 who scored significantly below chance; (B) participants in Experiment 2 who performed at chance; (C) participants in Experiment 2 who scored significantly above chance; and (D) E.P. in Reed et al.'s (1999) Experiment 2.

ted. The typicality gradient for E.P. is quite similar to the gradient for those learners who used the correct category labels. In addition, the participants who consistently reversed the category labels showed a typicality gradient similar in form to the gradient that E.P. displayed in Reed et al.'s Experiment 1.

Although the result is not statistically significant, it is suggestive that of the 23 learners, 14 used the correct category labels and only 9 used the reversed ones. As we noted earlier, in the present design, the base rate of category features exceeds the base rate of noncategory ones. Perhaps observers are sensitive to this difference in base rates in choosing which feature values to endorse. To the extent that E.P.'s switch from an inverted typicality gradient in Experiment 1 to the correct direction in Experiment 2 is a reliable phenomenon, the feature base-rate difference could be playing some role.

Discussion

Experiment 2 was aimed at investigating classification behavior in Reed et al.'s (1999) design in a situation

in which the observers had no prior experience with the actual training instances (see Palmeri & Flanery, 1999). The results indicate that a sizable portion of our sample performed similarly to E.P. in terms of their accuracy and their typicality gradients.⁸ We conclude, therefore, that E.P.'s performance could very well be a result of recycling a low-dimension classification strategy in his short-term working memory at time of test, without memory for the original training instances. E.P.'s classification performance in Reed et al.'s design is therefore not diagnostic of the existence of multiple memory systems for categorization and recognition/recall.

GENERAL DISCUSSION

Summary

In summary, in a previous study involving object-like stimuli varying along discrete features, Reed et al. (1999) found that amnesic patients were significantly impaired in recalling the features, but learned to categorize the object-like stimuli as well as controls. They in-

terpreted this dissociation as supporting multiple-system models in which an explicit memory system is responsible for recall/recognition, whereas an implicit memory system governs categorization. The goal of the present article was to demonstrate the plausibility of a single-system approach to explaining the dissociation between categorization and explicit memory performance exhibited by amnesics and matched controls in Reed et al.'s study. Our key point is that good performance in Reed et al.'s categorization task can be achieved by retaining information from just a small subset of the dimensions that compose the objects. By contrast, adequate recall or recognition performance requires retention of information from most of the multiple dimensions. Thus, in Reed et al.'s task, a general impairment in the operating characteristics of a single memory system is more likely to be detrimental to recall or recognition than to categorization. There is no need to posit that separate memory systems govern the two types of tasks.

We provided several lines of evidence in support of the single-system hypothesis. The first line of evidence was indirect. Using Reed et al.'s (1999) stimuli and category structure, we measured normal participants' recognition and categorization performance in an immediate test and after a week's delay. Whereas the week's delay caused a major deficit in old–new recognition performance, categorization accuracy remained essentially intact. Although the amnesic group in Reed et al. and our delayed group are undoubtedly qualitatively different in various ways, the general pattern of results across the two experiments was identical. This pattern follows naturally from the idea that the old–new recognition task makes high demands on memory, whereas the categorization task does not. More important, we conducted detailed formal modeling analyses of the obtained categorization data. These analyses provided support for the hypothesis that participants did indeed use only a limited subset of the dimensions that composed the objects when making their classification decisions. Models that assumed attention to just three (or even fewer) dimensions provided good quantitative fits to the individual participant data. Furthermore, the quantitative fits of these low-dimension models could not be distinguished statistically from a model that allowed attention to all nine dimensions of the objects.

Previously, Nosofsky and Zaki (1998) accounted for dissociations between categorization and recognition in amnesics and normal controls by positing that amnesics had a lowered level of memory sensitivity—that is, a reduced ability to discriminate between distinct items in memory. The present account of categorization–recognition dissociations placed emphasis on the differential attentional requirements of the two tasks. The reason that different formal parameters (memory sensitivity vs. attention weights) were emphasized in these explanations is that the nature of the stimuli across the two tasks was vastly different. In the tasks considered by Nosofsky and Zaki, participants categorized and recognized dot-pattern stim-

uli. Such stimuli are not organized into highly discrete, separable dimensions, but rather are perceived as unitary wholes. The memory-sensitivity parameter governs the extent to which participants can discriminate among such holistic stimuli. By contrast, in the present study, the stimuli were composed of highly discrete, separable dimensions. Selective attention to individual dimensions plays a major role in the processing of such stimuli, and so the present analysis placed emphasis on the role of the attention weights instead. We believe it is plausible that in situations involving highly separable dimension stimuli, a major memory impairment would reduce an observer's ability to simultaneously retain information from multiple dimensions.

Evidence From a Densely Amnesic Patient

Data from a severely amnesic patient (E.P.) with extensive damage to the medial temporal lobe have been used to seriously challenge the predictions of a single-system approach (Squire & Knowlton, 1995; Squire & Zola, 1996). For example, using the dot-pattern paradigm, Squire and Knowlton demonstrated that E.P. could not discriminate between old and new exemplars in a recognition task but performed normally on a categorization task. Similarly, in Reed et al. (1999), although E.P. could not recall any of the features of the stimuli, he could discriminate between category members and non-members (although not always assign the correct labels to these two classes of stimuli).

However, even these striking dissociations between categorization and recognition/recall can be explained within a single-system framework if one considers differences in the requirements of the two tasks. First, consider the Squire and Knowlton (1995) results. Nosofsky and Zaki (1998) pointed out that in the dot-pattern paradigm of Knowlton and Squire (1993) there were frequent presentations of category members at the time of test. They reasoned that if E.P. had even a small amount of residual exemplar-based memory from the training phase, these frequent presentations of the category members could serve to reintegrate this knowledge. The same type of short-term memory that allowed E.P. to follow the instructions for the task could allow E.P. to recycle his knowledge of the category structure. In contrast, no such strategy would be successful for the recognition test, because the individual old patterns bear no resemblance to one another. Indeed, in an elegant demonstration, Palmeri and Flanery (1999) reported that participants could perform Squire and Knowlton's categorization task without *any* exposure to the training instances. Apparently, participants can use the similarities of the items presented during the testing sequence to infer the category structure. However, with no exposure to the recognition training items, Palmeri and Flanery's participants were at chance in the recognition task.

Analogously, in Experiment 2 of the present article, we used the same procedure as did Palmeri and Flanery (1999) and found that many participants discriminated

among category members and nonmembers at levels as high as did E.P. Furthermore, these individuals displayed typicality gradients for the test instances of the same form as the gradients observed in Reed et al.'s (1999) study. In other words, without any exposure to the original training instances, participants could perform well on the Reed et al. categorization task. (Like E.P., however, although these participants discriminated between category members and nonmembers, they did not always assign the correct category label to the two classes of stimuli.) These results bring into question whether E.P. did indeed have any long-term implicit memory for the original training instances. Therefore, we conclude that E.P.'s performance in Reed et al.'s tasks is not diagnostic of the existence of separate memory systems for categorization and recognition/recall.

Evidence for a Double Dissociation

Much of the strength of the separate-systems argument depends on data that form a double dissociation. Whereas amnesic patients show intact categorization and impaired recognition, Parkinson's disease (PD) patients have demonstrated the mirror dissociation with intact recognition but impaired categorization. These results were observed in a probabilistic classification paradigm (Knowlton, Mangels, & Squire, 1996; Knowlton et al., 1994) in which participants learned to classify stimuli that varied on four binary-valued dimensions into two categories. Each of the four dimensions of the stimuli was associated probabilistically with the category. Amnesic participants performed at normal levels during the first 50 trials of this task, although the control participants showed a clear advantage in later training. Critically, the PD patients, who displayed good memory on a recall task, were impaired during the early stages of training, although their performance was similar to that of the amnesic group in later stages of training.

Nosofsky and Zaki (1998) noted that a number of possible deficits in the PD group other than an impaired category representational system could give rise to these results. To illustrate the effects of one such possibility, Nosofsky and Zaki simulated a group that used a suboptimal probability matching strategy to perform the task. Whereas a lowered memory sensitivity parameter was sufficient to replicate the results for the amnesic group, a lowered response-scaling parameter was sufficient to account for the early training deficit demonstrated by the PD group.

There is now converging evidence that a factor other than an impaired categorization system may have been responsible for the PD patients' performance deficit in Knowlton et al.'s (1994; Knowlton et al., 1996) probabilistic classification learning task. In particular, recent studies have demonstrated that PD patients perform normally in other categorization tasks. For example, Reber and Squire (1999) showed that PD patients performed at normal levels in the Knowlton and Squire (1993) dot-pattern task as well as in an artificial grammar-learning

task. Therefore, although PD patients may show deficits in categorization, these deficits appear to be limited to particular paradigms. Therefore, it appears that the generality of the double dissociation involving PD patients may be limited. We acknowledge, however, the more complicated possibility that there are multiple category representation systems, only some of which are impaired by PD (e.g., Poldrack, Prabhakaran, Seger, & Gabrieli, 1999; Reber & Squire, 1999; Reber, Stark, & Squire, 1998; Smith, Patalano, & Jonides, 1998). Future research is needed to test these alternative possibilities.

Conclusions and Future Research

In general, we believe that to provide clearer evidence that categorization and recognition are governed by separate representational systems, investigators will need to vary parametrically the overall difficulty of the categorization tasks that are tested. If the categorization task is simply easier than the recognition or recall tasks, then even single-system models can predict an observed dissociation between categorization and explicit memory performance. Amnesics may perform well on some particular categorization task simply because it makes small demands on overall memory resources.

In an attempt to study performance of amnesic patients on a more complex categorization task, Filoteo, Maddox, and Davis (2001) trained amnesics and controls on a perceptual categorization task involving stimuli that consisted of horizontal and vertical lines of varying lengths connected in one corner. The category was defined by a complex quadratic rule. Filoteo et al. observed that accuracy for the last block of trials was comparable for controls (85%) and 2 amnesics (84%). These results appear to support the idea that amnesics are able to learn a complex nonlinear categorization rule. However, Filoteo et al. noted that a simple strategy could have been adopted by participants to determine category membership in this task. The category was constructed so that if the two lines of the stimulus were approximately equal (or looked like part of a square), the item generally belonged to Category 1. If the two lines were unequal (or looked like part of a rectangle), the stimulus generally belonged to Category 2. Filoteo et al. reported some analyses indicating that a highly simplified model based on shape failed to account fully for their data. However, Flanery, Palmeri, and Schaper (2001) demonstrated that if the same abstract category structure was mapped on another stimulus domain, one in which the two dimensions could not be easily integrated, participants performed dramatically worse. These results suggest that participants in the Filoteo et al. study may indeed have used a simple shape rule to complete the categorization task.

Clearly, further research is needed to test whether performance of amnesics might suffer in more difficult categorization tasks. To begin to address the possibility, Zaki, Nosofsky, Jessup, and Unverzagt (2001) have tested patients who were diagnosed as likely being at the

early stages of Alzheimer's disease. Their results indicated that although these patients showed the classic dissociation pattern observed in the Knowlton and Squire (1993) paradigm, a more challenging dot-pattern paradigm demonstrated a categorization deficit for this memory-impaired group. It would be interesting as well to test amnesic patients on more challenging categorization paradigms that involve object-like stimuli such as those used by Reed et al. (1999).

REFERENCES

- ASHBY, F. G., & MADDOX, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372-400.
- FILOTEU, J. V., MADDOX, W. T., & DAVIS, J. D. (2001). Quantitative modeling of category learning in amnesiac patients. *Journal of the International Neurological Society*, **7**, 1-19.
- FLANERY, M. A., PALMERI, T. J., & SCHAPER, B. L. (2001). Investigating dissociations between perceptual categorization and explicit memory. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 291-296). Mahwah, NJ: Erlbaum.
- KNOWLTON, B. J., MANGELS, J. A., & SQUIRE, L. R. (1996). A neostriatal habit learning system in humans. *Science*, **273**, 1399-1402.
- KNOWLTON, B. J., & SQUIRE, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, **262**, 1747-1749.
- KNOWLTON, B. J., SQUIRE, L. R., & GLUCK, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, **1**, 106-120.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.
- MINDA, J. P., & SMITH, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 775-799.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- NOSOFSKY, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 700-708.
- NOSOFSKY, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 3-27.
- NOSOFSKY, R. M., & PALMERI, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.
- NOSOFSKY, R. M., & ZAKI, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, **9**, 247-255.
- PALMERI, T. J., & FLANERY, M. A. (1999). Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science*, **10**, 526-530.
- POLDRACK, R. A., PRABHAKARAN, V., SEGER, C. A., & GABRIELI, J. D. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology*, **13**, 564-574.
- POSNER, M. I., & KEELE, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353-363.
- REBER, P. J., & SQUIRE, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behavioral Neuroscience*, **113**, 235-242.
- REBER, P. J., STARK, C. E., & SQUIRE, L. R. (1998). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory*, **5**, 420-428.
- REED, J. M., SQUIRE, L. R., PATALANO, A. L., SMITH, E. E., & JONIDES, J. (1999). Learning about categories that are defined by object-like stimuli despite impaired declarative memory. *Behavioral Neuroscience*, **113**, 411-419.
- SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.
- SMITH, E. E., PATALANO, A. L., & JONIDES, J. (1998). Alternative strategies of categorization. *Cognition*, **65**, 167-196.
- SQUIRE, L. R., & KNOWLTON, B. J. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Sciences*, **92**, 12470-12474.
- SQUIRE, L. R., & ZOLA, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences*, **93**, 13515-13522.
- ZAKI, S. R., NOSOFSKY, R. M., JESSUP, N., & UNVERZAGT, F. (2001). *Categorization and recognition performance of a memory-impaired group: Evidence for single-system models*. Manuscript in preparation.

NOTES

1. These estimates are based on calculations performed on a new set of stimuli constructed according to the specifications outlined in Reed et al. (1999). Minor differences in these estimates may exist across stimulus sets but would be exceedingly small.

2. An anonymous reviewer suggested that because the instructions listed the dimensions of the stimuli, we might have predisposed participants to use a small number of dimensions in the task. To address this possibility we tested 10 new participants in the task with instructions in which the dimensions were not listed. The general pattern of results did not differ from the one reported in this article.

3. We computed the d' values for individual participants by converting the proportions of 0 and 1 to $1/(2N)$ and $1 - 1/(2N)$, respectively (Macmillan & Creelman, 1991, p. 10).

4. Specifically, we searched for the parameters that maximized the log-likelihood function $\ln L = \sum \ln N_i! - \sum \ln f_{ij}! + \sum \sum \ln f_{ij} \ln p_{ij}$, where N_i is the number of times that stimulus i was presented, f_{ij} is the observed frequency with which stimulus i was classified into category j , and p_{ij} is the predicted probability of stimulus i being classified into category j . Note that in the present case, there are two "categories"—endorsed and nonendorsed.

5. The method of likelihood ratio testing is used to ascertain whether a restricted model that is nested within a full model fits the data significantly worse than the full model. Let $\ln L_F$ denote the log likelihood measure of the full model and $\ln L_R$ denote the log likelihood measure of the restricted model. Then the quantity $2(\ln L_F - \ln L_R)$ is distributed as a chi-square random variable with degrees of freedom equal to the number of parameters constrained in moving from the full model to the restricted one. If this quantity exceeds the critical value of chi square, then at least one of the extra parameters in the full model is warranted. Note that in the present application, the exact difference in the number of free parameters between the full model and a particular restricted version is difficult to determine. The reason is that the zero-weighted dimensions in the restricted model are not chosen a priori. Nevertheless, we assume that the difference in the number of dimension weights used in the two models being compared provides a good approximation to the degrees of freedom in the chi-square test.

6. The version of the model illustrated here assumed $c = 2$ and $\gamma = 3$. With $\gamma = 1$, there is too much noise in the decision process, and so the least frequently endorsed feature has an endorsement probability close to chance. The two- and three-dimension models, however, predict the qualitative pattern of results even with $\gamma = 1$.

7. One possibility raised by a reviewer is that although participants in the mock subliminal task had no memory for the unseen training items, they may have used their memory of the test items during the transfer task. However, because E.P. could not form any lasting memory of the test items, he would have access only to these items to the extent that he could keep them in his working memory. Although any information used in the transfer test must have been gleaned from the transfer test items, it does not appear that our participants benefited from continued exposure to the test items. For example, the learners performed similarly in the first 10 trials (mean absolute difference from .50 was .24) and the last 10 trials in the transfer test (mean absolute difference from

.50 was .22) [$t(22) = 0.71, p > .10$]. Therefore, even in the early stages of the transfer test, participants with no prior exposure to the stimuli achieved accuracy levels similar to E.P.'s overall performance.

8. The results of our Experiment 2 may not be very different from a similar one reported by Reed et al. (1999). These researchers also tested participants' ability to classify during the test phase without providing a study phase, although they did not use a mock subliminal training procedure (see Reed et al., Experiment 1B). The main result they reported was that, when averaging across the results from the 7 observers they tested, a flat typicality gradient was produced (see Reed et al., Figure 4). That is, on average, the observers endorsed each type of pattern with the

same probability. However, if some of the individual participants had produced an increasing typicality gradient, and others had produced an inverted decreasing one, then averaging across all observers would make the typicality gradient look flat. Indeed, Reed et al. (p. 416) acknowledged that analyses of their individual participant data suggested that 4 of their 7 participants had adopted a single-feature strategy and that these participants produced endorsement patterns similar to those from their Experiment 1A.

(Manuscript received December 10, 2000;
revision accepted for publication October 31, 2001.)