Memory & Cognition
2007, 35 (8), 2088-2096

A high-distortion enhancement effect in the prototype-learning paradigm: Dramatic effects of category learning during test

Safa R. Zaki
Williams College, Williamstown, Massachusetts

and

Robert M. Nosofsky
Indiana University, Bloomington, Indiana

Recent research suggests that exemplar models of classification are disconfirmed by the finding of extreme prototype-enhancement effects and steep typicality gradients in a version of the prototype-learning paradigm. We argue that these results are due to learning-during-transfer effects and not to the abstraction of a prototype from the training instances. In the standard version of the paradigm, observers are flooded with multiple presentations of the prototype and its low distortions during transfer. In a modified transfer condition, we instead present multiple instances of an arbitrary high distortion and low distortions of that high distortion. An extreme "high-distortion enhancement effect" is observed. Also, there is a flattening of the typicality gradient associated with the standard patterns (prototype, low distortions, and standard high distortions). The results provide dramatic evidence of the role of learning during transfer in this task and force a reevaluation of the dominant current interpretation of the steep typicality gradient.

One of the central issues in categorization research has been the debate over the nature of category representation. According to prototype models, people abstract the central tendency, or prototype, of a category, and use that abstraction as the basis for classifying new items (Homa & Vosburgh, 1976; Posner & Keele, 1968; S. K. Reed, 1972; J. D. Smith, Murray, & Minda, 1997). In contrast, exemplar models assume that people store particular exemplars of a category and base categorization decisions on similarity to the exemplars (Heit, 1994; Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986).

Prototype theory initially gained support due to results from the classic dot-pattern paradigm. In this paradigm, first introduced by Posner, Goldsmith, and Welton (1967), a prototype is created by randomly placing nine dots in a grid. Patterns at various levels of distortion are generated by displacing each of the dots according to a statistical-decision rule. Early on, researchers observed that, following training on a number of distorted patterns, endorsement levels of the previously unseen prototype were relatively high, often exceeding endorsement levels of the old study patterns (Posner & Keele, 1968; Strange, Keeney, Kessel, & Jenkins, 1970). This *prototype-enhancement effect* reinforced the idea that a prototype was abstracted during training.

However, exemplar models also predict a prototype-enhancement effect because of the similarity of the prototype to the numerous items stored in memory (Busemeyer, Dewey, & Medin, 1984; Hintzman, 1986; Shin & Nosofsky, 1992). In addition, exemplar models also make accurate predictions regarding generalization effects from particular old items (Homa, Sterling, & Trepel, 1981; Nosofsky & Zaki, 2002).

However, in recent years, prototype theorists have argued that exemplar theory has been disconfirmed because of the typicality gradient observed in a highly influential version of the dot-pattern paradigm (J. D. Smith, 2002; J. D. Smith & Minda, 2001, 2002). In this task, first used by Knowlton and Squire (1993), participants view various high-level distortions from a single prototype in the training phase (see Figure 1, top row). In the test phase, participants see new patterns and decide whether each belongs to the category. The test patterns include the previously unseen prototype, low-level distortions of the prototype, new high-level distortions, and random patterns (see Figure 1, middle row). The classic result is a typicality gradient in which the prototype is endorsed with the highest probability, followed by the low-level distortions, high-level distortions, and random patterns.

The exemplar model correctly predicts the ordering of these endorsements, because patterns closest to the center of the category have the greatest summed similarity to the old items. However, using measures of physical dot

S. R. Zaki, szaki@williams.edu

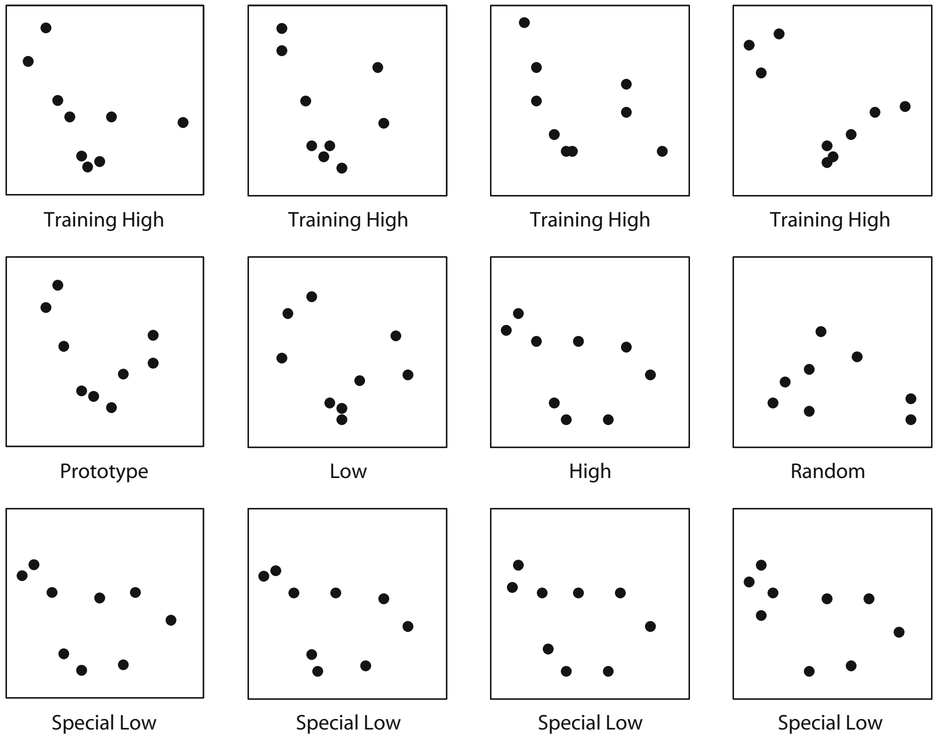

Figure 1. Examples of stimuli used in the dot-pattern category learning task. The top row displays examples of the training instances—that is, various high distortions of the prototype. The middle row shows examples of transfer patterns from the standard task—that is, the prototype, a low distortion of the prototype, a new high distortion of the prototype, and a random pattern. The bottom row shows examples of low distortions of the high distortion from the middle row. These special low distortions were presented in the transfer phase of the modified task.

distance as a method of computing similarity, J. D. Smith (2002) and J. D. Smith and Minda (2001, 2002) fitted quantitative versions of exemplar and prototype models to these data. They reported that the exemplar model was unable to capture the steepness of the typicality gradient, whereas the prototype model fit the data well; that is, the exemplar model systematically underpredicted the rate at which the prototype and low-level distortions were endorsed relative to the endorsement rates of the other patterns. Nosofsky and Zaki (1998) had previously acknowledged a similar result in formal analyses that made use of human similarity judgments, rather than measures of physical dot distance, among pairs of patterns.

Zaki and Nosofsky (2004) pointed to several reasons why a steeper typicality gradient than that predicted by the exemplar model might be observed in this particular paradigm. The most important factor involved learning that occurs during the transfer phase. In testing categorization models, researchers often make the simplifying assumption that the category representation established during the learning phase remains stable during transfer. This simplifying assumption is usually a reasonable one: In most category learning paradigms, there is an extensive training phase in which observers receive trial-by-trial corrective feedback, and training often continues until observers reach a performance criterion. By contrast, the transfer phase is generally much shorter and feedback is withheld on the transfer patterns.

Learning conditions are very different, however, in the Knowlton–Squire (1993) paradigm. In the training phase, observers are simply presented with 40 high distortions of the prototype, and often do not even know until time of test that they have just experienced members of a category. There is no trial-by-trial feedback during training, and no performance criterion to achieve. Furthermore, as explained below, the distribution of category members changes dramatically at time of transfer, when observers first need to try to discriminate between members and nonmembers. Moreover, there is clear evidence that observers can use forms of implicit feedback at time of transfer as a basis for learning. In particular, Palmeri and Flanery (1999, 2002) demonstrated that participants can perform well in the Knowlton–Squire task, even in the complete absence of a training phase, by learning the category at time of test. Taken together, these considerations suggest that the assumption of a stable category representation at time of transfer is likely to be false.

Zaki and Nosofsky (2004) argued that, in the standard version of the paradigm that includes a training phase, learning during transfer is likely to affect the steepness of the typicality gradient by boosting the endorsement rates of the prototype and low distortions. During training, observers experience only 40 high distortions of the prototype; they never experience the prototype or its low distortions. By contrast, in the transfer phase, there are 4 separate presentations of the prototype, 20 presentations of its low distortions, and 20 presentations of new high distortions. Thus, during transfer, observers are flooded with patterns that lie close to the center of the category, after having received no experience with such patterns during initial training. To the extent that observers update their category representations with new exemplars at time of transfer, the summed similarity for the prototype and low distortions will grow precipitously, so exemplar models will predict high endorsement rates for these patterns.

To test this explanation, Zaki and Nosofsky (2004) conducted experiments that manipulated the distribution of the patterns presented during transfer. In a "full set" condition, they replicated the standard Knowlton–Squire paradigm: During transfer, there were 4 presentations of the prototype, 20 low distortions, 20 new high distortions, and 40 random patterns. However, in a "subset" condition, there was only a single presentation of the prototype, 2 low distortions, 20 new high distortions, and 20 random patterns. In accord with the learning-during-transfer explanation, the observed typicality gradient was steeper in the standard condition than in the subset condition, supporting the hypothesis that the enhanced endorsement rate of the prototype and low distortions was due to their frequent presentations at time of test. Furthermore, Zaki and Nosofsky (2004) demonstrated that a version of the exemplar model that assumed that people augment their category representations with the patterns presented during transfer provided a good qualitative account of the results.

Despite these findings, J. D. Smith (2002) and J. D. Smith and Minda's (2001, 2002) research continues to exert a major influence in favor of prototype models in the category representation debate (e.g., Ashby & Maddox, 2005; Juslin, Jones, Olsson, & Winman, 2003; Ramsey, Langlois, & Marti, 2005; Rhodes & Jeffery, 2006; Storms, 2004). For example, in their recent *Annual Review* chapter on Human Category Learning, Ashby and Maddox devoted a complete section to the Knowlton–Squire prototype-learning paradigm and identified J. D. Smith (2002) and J. D. Smith and Minda's (2001, 2002) typicality-gradient evidence as being a critical test that shows the inadequacy of the exemplar view. Likewise, J. D. Smith (2005) argued recently that "exemplar theory has been shown to have serious flaws and to fail qualitatively at critical points" and cited the typicality gradient evidence to bolster this claim (p. 59). The work has even influenced the face perception literature, where the typicality gradient evidence is also cited as support for prototype representations over exemplar representations (Ramsey et al., 2005; Rhodes & Jeffery, 2006).

Given the major impact of the J. D. Smith (2002) and J. D. Smith and Minda (2001, 2002) articles, we felt it important to demonstrate the potential magnitude of the effect of learning during transfer on the typicality gradients in the Knowlton–Squire task. Although Zaki and Nosofsky's (2004) research suggested that the steepness of the gradients can be modified on the basis of the transfer test's composition, the effects were of a subtle quantitative form. By contrast, the goal of the present research is to provide a dramatic, qualitative demonstration of the role of these learning-during-transfer effects. This demonstration, we hope, will ultimately lead the field to reevaluate whether or not past results involving the steepness of the typicality gradient are, in fact, due to the abstraction of a prototype from the training instances.

In the present research, we tested a condition in which participants experienced the same training phase as in the Knowlton–Squire task; but this time during transfer, instead of flooding the observer with the prototype and its low-level distortions, we chose an arbitrary high-level distortion to play that role. That is, for each observer, one of the high-level distortions was chosen at random and was presented four times during the test phase. In addition, using the same statistical-distortion procedure as in the previous studies, we presented 20 low-level distortions of that high-level distortion (see Figure 1, bottom row). Thus, in the transfer phase of this condition, participants experienced a cluster of patterns surrounding one of the high-level distortions instead of experiencing the cluster surrounding the prototype. The key question was the extent to which, in this modified condition, observers endorsed as category members the high-level distortion and its surrounding cluster of low-level distortions. To reiterate: The dominant current interpretation of the extreme prototype enhancement effect in the standard version of the Knowlton–Squire paradigm is that observers abstract the prototype from the training instances. A strong "high-distortion enhancement effect" in the present modified paradigm would call into question this interpretation and would suggest instead that the steep typicality gradient is a reflection of learning during transfer.

EXPERIMENT

We compared the standard Knowlton–Squire dot-pattern task in one condition to a modified version of the task in another condition. As described in the Method section, the conditions differed only in the distribution of category members presented during transfer.

Method

Participants. One hundred ninety-nine Indiana University undergraduates participated to fulfill a course requirement. Of those, 50 participants were randomly assigned to the standard Knowlton and Squire (1993) condition and 149 to the modified transfer-test condition. We opted to test considerably more participants in the modified transfer condition in order to achieve reasonable sample sizes for the standard prototype and low-level distortions, which are presented with very low frequency in that condition.

Stimuli. The stimuli were nine-dot patterns (Posner et al., 1967; Posner & Keele, 1968). The training stimuli for both conditions consisted of the 40 high-level distortions used by Knowlton and Squire (1993). For the standard condition, the transfer stimuli consisted of 4 instances of the prototype, 20 low-level distortions, 20 new high-level distortions, and 40 random patterns; for the modified version, the transfer test included only 1 instance of the prototype and 2 low-level distortions, randomly selected for each participant. One of the high-level distortions was randomly chosen for each participant to be the center of a cluster of new patterns. That special high-level distortion was shown four times, and the remaining 19 were shown

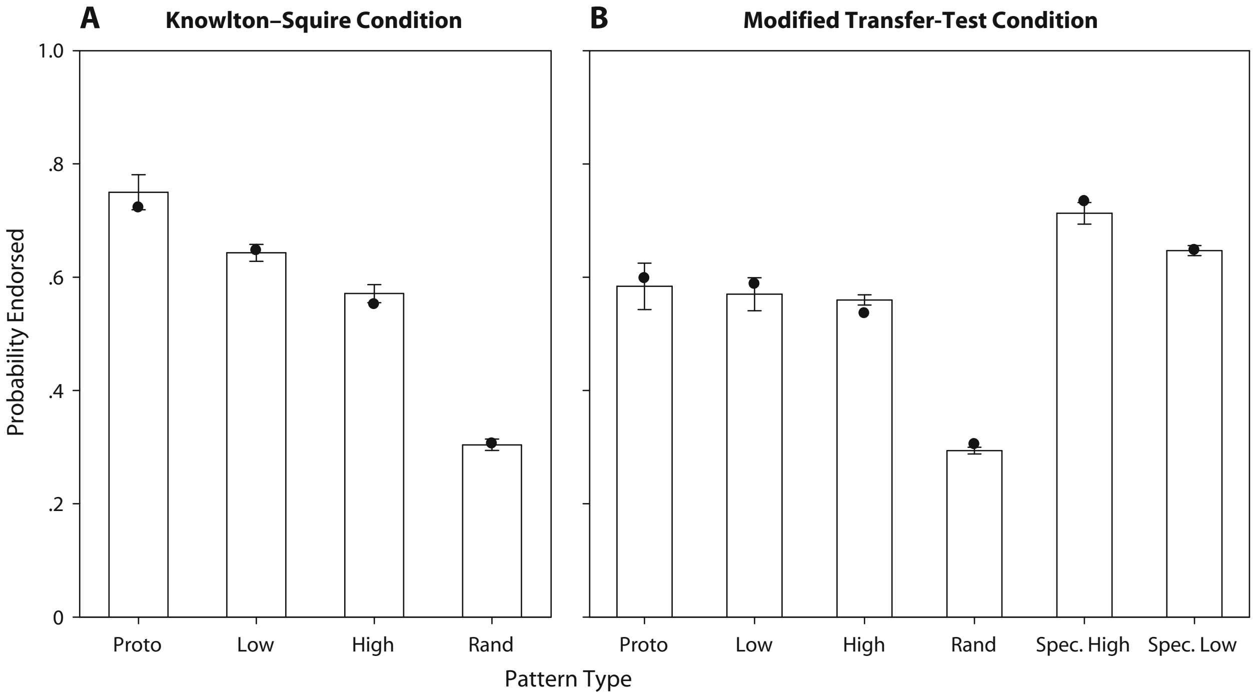

Figure 2. (A) the observed level of endorsement of the various patterns in the standard Knowlton and Squire (1993) condition; (B) The endorsement levels of the various patterns in the modified transfer condition. Proto, prototype; Low, low-level distortions; High, high-level distortions; Rand, random pattern; Spec. High, special high distortion chosen at random to be surrounded by a cluster of items in the transfer test; Spec. Low, low distortions of the Spec. High pattern. Open bars, observed data; solid dots, predictions from learning-during-transfer exemplar model. Error bars indicate ±1 *SEM*.

once. Twenty low-level distortions of the special high-level distortion were generated for each participant. Finally, the transfer test included the 40 random patterns from the Knowlton and Squire (1993) set. Examples of the stimuli are shown in Figure 1.

Procedure. Participants were tested individually in a computerized task. In the training phase, 40 high-level distortions were shown, one at a time, for 5 sec each. Following the procedure from previous studies, the participants' task was simply to observe. In the transfer phase, following Knowlton and Squire (1993), participants were told that all of the dot patterns in the first phase belonged to a category of patterns in the same sense that if they had seen pictures of dogs, each picture would be a member of the category of dogs. On each trial, one of the test patterns was displayed on the computer screen, and participants judged whether it was a category member. The display was response terminated, and no feedback was provided. All patterns appeared in a different random order for each participant.

Results

Figure 2A (open bars) shows the probability with which each stimulus type was endorsed in the standard condition of the Knowlton and Squire (1993) task. As in previous studies, an orderly typicality gradient was observed, with the prototypes being endorsed with the highest probability followed by the low-level distortions, high-level distortions, and random patterns. The steepness of the typicality gradient is similar to that reported by J. D. Smith (2002) and J. D. Smith and Minda (2001, 2002).

The results from the modified condition, shown by the open bars in Figure 2B, could hardly be more dramatic. There is an extreme enhancement effect for the special high distortion. Indeed, in this condition, the endorsement rate for the special high distortion (.71) far exceeds that for the prototype itself (.58) [$t(148) = 2.77$, $p < .01$]. Likewise, the low distortions of the special high distortion also have high endorsement rates (.65), even exceeding those of the low distortions of the prototype (.57) [$t(148) = 2.11$, $p < .05$]. Comparing across conditions, there is little difference between the endorsement rate for the special high distortion in the modified condition (.71) and the prototype in the standard condition (.75) [$t(197) = 0.70$, $p > .10$]. In addition, there was no difference between the endorsement rates of the low-level distortions in the standard condition (.64) and the low-level distortions of the special high distortion in the modified condition (.65) [$t(197) = 0.09$, $p > .10$].

In addition to the extreme enhancement effect for the special high distortion, the typicality gradient for the regular patterns (prototype, low distortions, and high distortions) is far shallower in the modified condition than in the standard condition. A mixed model ANOVA with distortion level and condition as factors revealed a main effect of distortion, with participants tending to endorse items closer to the category prototype [$F(2,394) = 4.97$, $MS_e = .390$, $p < .01$]. More important, there was a significant interaction between condition and distortion level [$F(2,394) = 2.91$, $MS_e = .228$, $p = .05$], reflecting the fact that the slope of the typicality gradient was significantly shallower in the modified condition than in the standard condition.

The preceding results make dramatically clear that learning during transfer exerts a powerful influence on

Table 1
Observed Endorsement Probabilities for the Different Pattern Types in the First 10 Transfer Trials

	Knowlton–Squire Condition		Modified Transfer Condition	
Pattern Type	*M*	*n*	*M*	*n*
Prototype	.87	15	.62	16
Low	.59	119	.65	34
High	.64	132	.62	318
Random	.35	234	.33	675
Special high	–	–	.75	60
Special low	–	–	.69	387

Note—*M*, mean; *n*, number of observations upon which each mean is based. Note that different participants contribute different numbers of observations to these means, so detailed quantitative comparisons should be made with caution.

the pattern of performance in this paradigm. We also conducted an analysis of the first 10 trials of transfer in order to gauge the speed at which this learning during transfer occurred. (For similar analyses in a related category-learning paradigm, see Bozoki, Grossman, & E. E. Smith, 2006.) The data, presented in Table 1, are, naturally, somewhat noisy, due to the relatively small number of observations in some of the cells, but a clear pattern of learning during transfer emerges, even in these first few trials. In particular, the typicality gradient around the special high level distortion in the modified transfer condition is quite steep, suggesting that participants are very quick to augment their representations with information from the initial stages of the transfer test. In the following section, we consider the ability of a learning-during-transfer exemplar model to account for the full pattern of results.

Theoretical Analysis

A natural question is whether a formalized learning-during-transfer model can in fact account simultaneously for the varied effects displayed in Figure 2 and Table 1. As a first approximation, we used a relatively simple learning-during-transfer exemplar model. The general intuition behind this model is that the probability with which an item is classified as a member of the category is determined jointly by its similarity to the training patterns and its similarity to the previously seen test patterns, weighted by the relative strengths of the training and test patterns in memory.

Specifically, the evidence in favor of category membership of item i is given by

$$S = \left(y \cdot \sum s_{it}\right) + \left(\sum s_{if}\right), \tag{1}$$

where $\sum s_{it}$ is the summed similarity of test item i to the training items, y is a parameter reflecting the long term memory strength of the training exemplars, and $\sum s_{if}$ is the summed similarity of the test item to the preceding transfer items weighted by their individual memory strengths.

The individual short term memory strength for a previous transfer item j is given by

$$str(j) = v^{lag(i-j)-1}, \tag{2}$$

where v is a memory decay parameter that ranges between 0 and 1, and where $lag(i-j)$ is the number of trials intervening between the presentations of the current test item i and the previous test item j. Therefore, the test item presented on the immediately previous trial has a memory strength of 1, the test item presented two trials back has a memory strength of v, the test item presented three trials back has a strength of v^2, and so forth.

The probability with which the item is classified as a category member is then given by

$$P(\text{Cat}) = \frac{S}{S+k}, \tag{3}$$

where k is a response-criterion parameter.

We followed J. D. Smith's (2002) and J. D. Smith and Minda's (2001, 2002) approaches to computing the similarity between the individual pairs of dot patterns. First, the psychological distance (d_{ij}) between each pair of patterns was assumed to be functionally related to the physical distance between corresponding dots across the pairs (see J. D. Smith & Minda, 2001, for details). Second, the similarity between each pair of patterns was an exponential decay function of this distance (Shepard, 1987), such that

$$s_{ij} = \exp\left(-cd_{ij}\right), \tag{4}$$

where c is an overall sensitivity parameter. The model uses four free parameters: the training-exemplar weight parameter y (Equation 1), the memory decay parameter v for computing individual transfer item strengths (Equation 2), the criterion parameter k (Equation 3), and the overall sensitivity parameter c (Equation 4).

We fitted the model simultaneously to the data from all the transfer trials from both conditions by searching for the values of the free parameters that minimized the sum of squared deviations (*SSD*) between the predicted and observed category-endorsement probabilities. In conducting the fits, we derived predictions for each individual participant by using the precise sequence of transfer items that the individual participant experienced. The final predictions are averages computed over these individual-participant predictions.

The predicted endorsement probabilities are superimposed as dots on the observed probabilities in Figure 2. As can be seen in the figure, the model provides an excellent quantitative account of the complete set of data (*SSD* = .0028) and captures all of the key qualitative effects of interest. First, it predicts an extreme prototype enhancement effect in the standard condition. The reason is that the summed similarity of the prototype grows large due to the frequent presentations of the prototype and its low distortions at time of transfer. Likewise, the model predicts an extreme special high-distortion enhancement effect in the modified condition for analogous reasons. Finally, the typicality gradient for the standard patterns is relatively flat in the modified condition because the prototype and standard low distortions are not presented with high frequency in that condition. The best-fitting parameters are reported in Table 2.

Table 2
Best-Fitting Parameter Values for the Learning-During-Transfer Exemplar Model Fits to the Complete Set of Transfer Trial Data

Parameter	Value
c	1.438
k	0.272
y	0.029
v	0.843
SSD	0.0028

Note—c, sensitivity parameter; k, response criterion parameter; y, weight given to each training exemplar; v, memory decay parameter for the test exemplars; *SSD*, sum of squared deviations between the observed and predicted endorsement probabilities.

To test whether the model could account for the rapid learning during transfer (Table 1), we fitted a version of the model simultaneously to the data from the first 10 trials, as well as to the complete set of test trials. To account quantitatively for both the early test-trial data and the complete test-trial data, we needed to introduce assumptions about how observers adjust their criterion setting during the course of testing. Note that as testing continued, the overall summed similarity of any given test item to all previously presented items grew larger and larger (because more items entered into the sum). Presumably, observers gradually increased the magnitude of the criterion setting to compensate for this increase in summed similarity. We modeled the increase in the criterion by using a functional form analogous to the increase in summed similarities formalized in Equations 1 and 2. Specifically, the magnitude of the criterion on trial n of transfer was given by

$$k_n = k_0\left[1 + u + u^2 + \ldots + u^{n-1}\right] = k_0\left[\frac{1-u^n}{1-u}\right], \qquad (5)$$

where k_0 is the starting value of the criterion (on Test Trial 1), and u ($0 \le u \le 1$) governs the rate of growth of the criterion across trials. In general, Equation 5 describes a curvilinear increase in the magnitude of the criterion setting across trials; that is, one in which the criterion increases at a decreasing rate. (In the special case in which $u = 0$, the criterion is constant across trials of testing, whereas in the special case in which $u = 1$, the magnitude of the criterion is proportional to the number of test trials.) With this additional assumption of an increasing criterion, the exemplar model uses 5 free parameters: the overall sensitivity parameter c, the training-exemplar weight parameter y, the memory decay parameter v, the starting criterion parameter k_0, and the criterion growth parameter u.

Holding all parameters fixed across conditions and stages of testing, we fitted the model simultaneously to the data from the first 10 trials and the complete transfer data from both the standard and modified conditions. Due to the low number of observations of the prototype in the first 10 trials (see Table 1), we collapsed across the prototype and the low distortions when fitting the model. The predicted endorsement probabilities are shown next to the observed data in Table 3. The model does a good job of accounting for the data (*SSD* = .0100). Once again, the model pinpoints all of the data from the complete set of transfer trials in both the standard and modified conditions (see top panel of Table 3). Now, however, we also demonstrate that the model can capture the very rapid learning observed for the special high distortion and special low distortions during the first 10 trials of the modified condition (see the bottom panel of Table 3). The only shortcoming in the model is that it overpredicts the endorsement probabilities of the prototype and low distortions in the first 10 trials of the standard condition.[1] The best-fitting parameter values are shown in Table 4.

The preceding modeling analyses were intended to be illustrative, and it is important to make various caveats: In particular, various aspects of the Knowlton–Squire dot-pattern learning paradigm require the introduction of simplifying assumptions in order to conduct the model fits. A great deal more research is needed to develop more rigorous modeling accounts. One simplifying assumption involves the technique for computing similarities among the dot patterns. On the basis of previous research reported by Posner et al. (1967), J. D. Smith and Minda (2001) suggest that the physical dot-distance method closely approximates psychological distance. Although the approximation may be a good one, psychological similarities among the dot patterns are also almost certainly influenced by higher order configural properties and emergent features that cannot be captured solely by measures of the physical locations of the individual dots (e.g., Hock, Tromley, & Polmann, 1988; Ichikawa, 1985). Second, the present modeling account assumes that an observer sums the similarity of a test item to all previously presented test items, whether or not that observer has endorsed those items as being category members. A more complex possibility is that whether or not a transfer item becomes part of an observer's augmented category representation depends on the classification response that the observer provides to the item at time of test. Despite these qualifications, we

Table 3
Observed and Predicted Values for the Full Transfer Test and the First 10 Trials

	Knowlton–Squire		Modified	
Pattern Type	Pre	Obs	Pre	Obs
	All Trials			
Prototype	.72	.75	.59	.58
Low	.65	.64	.59	.57
High	.55	.57	.54	.56
Random	.31	.30	.30	.29
Special high	–	–	.73	.71
Special low	–	–	.65	.65
	First 10 Trials			
Prot/low	.69	.62	.64	.64
High	.60	.64	.59	.62
Random	.35	.35	.34	.33
Special high	–	–	.77	.75
Special low	–	–	.68	.69

Note—Modified, modified transfer condition; Pre, predicted values; Obs, observed values; Prot/low, weighted average for the prototype and low distortions.

Table 4
Best-Fitting Parameter Values From the Learning–During-Transfer Exemplar Model Fits to All Transfer Trials and the First-10-Trials Data

Parameter	Value
c	1.447
y	0.015
v	0.764
k_0	0.026
u	0.863
SSD	0.0100

Note—c, sensitivity parameter; y, weight given to each training exemplar; v, decay parameter for test item memory strengths; k_0, starting criterion parameter; u, criterion growth parameter; *SSD*, sum of squared deviations between the observed and predicted endorsement probabilities.

believe that the present analyses are valuable for demonstrating the plausibility of an exemplar-based, learning-during-transfer account of the results.

Finally, although the focus here is on the exemplar-based account, we do not, of course, consider that these results rule out alternative models. For example, a prototype theorist could posit that multiple prototypes are established at time of transfer. A dominant new prototype might be formed around the special high distortion, and a secondary prototype might be formed around the standard prototype defined by the training instances (Gureckis & Love, 2006). The crucial point, however, is that the category representation is being modified at time of test. Previous claims in the literature—that exemplar theory is disconfirmed by the pattern of transfer in the Knowlton–Squire paradigm—need to be reevaluated in light of the present empirical and formal-modeling results.

SUMMARY AND DISCUSSION

J. D. Smith (2002) and J. D. Smith and Minda (2001, 2002) have argued that exemplar models fail to predict the magnitude of the prototype-enhancement effect and the steepness of the typicality gradient observed in the Knowlton–Squire (1993) version of the prototype-learning paradigm. This work has had a profound impact on the field in terms of favoring prototype models over exemplar models (see Ashby & Maddox's, 2005, *Annual Review* chapter for a review). In particular, the dominant interpretation is that the steep typicality gradient provides evidence for the abstraction of a prototype from the training instances.

In the present research, however, we demonstrated in dramatic fashion that learning-during-transfer processes play a significant role in influencing performance in this paradigm, even after initial training has occurred. Whereas in the standard version of the task, observers are flooded with presentations of the prototype and its low distortions during transfer, in the modified condition tested here we instead selected an arbitrary high-level distortion to play an analogous role. Our observation of an extreme "high-level distortion enhancement effect," as well as the shallow typicality gradient for the standard patterns, makes clear that the category representation is updated in the transfer phase of the task. Moreover, a simple version of an exemplar-based learning-during-transfer model provides an excellent account of the complete set of results, both the extreme prototype enhancement effect in the standard paradigm and the high-distortion enhancement effect in the modified version. Thus, the dominant interpretation by the field—that the steep typicality gradient observed in the standard task disconfirms exemplar theory—must clearly be reevaluated.

Other Implications

Although this research is primarily aimed at addressing conclusions from the prototype-exemplar debate, it has other implications as well. The original purpose of Knowlton and Squire's (1993) seminal study was not to contrast the predictions from prototype and exemplar models; rather, their study involved a demonstration that amnesic participants, with poor explicit recognition memory, could nevertheless perform well on the dot-pattern category-learning task. Indeed, whereas the amnesics performed significantly worse than did matched normal controls on tests of recognition memory, they did not differ significantly from normal controls in their dot-pattern category learning. This demonstration of a dissociation between categorization and recognition performance was consistent with Knowlton and Squire's (1993) hypothesis that a separate implicit-learning system guides various forms of category acquisition. Similar demonstrations of such dissociations have been observed in closely related paradigms as well (e.g., Kolodny, 1994; Reber, Stark, & Squire, 1998; J. M. Reed, Squire, Patalano, E. E. Smith, & Jonides, 1999).

However, the interpretation of such dissociations has been a topic of intense debate. First, theorists have argued that many of the dissociations are in fact well predicted by single-system models that make allowance for parameter differences between groups (e.g., Love & Gureckis, 2007; Nosofsky & Zaki, 1998; Palmeri & Flanery, 2002; Zaki, 2004; Zaki & Nosofsky, 2001; for related ideas, see; Kinder & Shanks, 2001; Lamberts & Shapiro, 2002), thereby calling into question the need to posit separate explicit memory and implicit category-learning systems. Second, researchers have also argued, with supporting evidence, that observers can use forms of short-term working memory to acquire category knowledge at time of test (Bozoki et al., 2006; Palmeri & Flanery, 1999, 2002; Zaki & Nosofsky, 2001). The present study amplifies the latter point in dramatic fashion, and shows that the structure of the test phase can have a profound influence on performance, even after completion of initial training. As argued by E. E. Smith (in press), the best current evidence for implicit category learning in amnesics comes from studies in which the amnesics first receive incidental training on a single category, are then told about the existence of the category, and are finally tested with highly structured category members. Based on the present results, however, it seems clear that performance in such incidental-learning, single-category tasks involves an amalgam of memories of the training instances and of the particular category members experienced at time of test. Thus, this crucial role of learning during test needs to be carefully considered when comparing and interpreting the performance of normals and amnesics on such tasks.

In the Knowlton–Squire (1993) dot-pattern task, numerous factors likely conspire to weaken the relative impact of the training instances, including the weak category structure at time of training (high distortions with little resemblance to one another), the lack of a training criterion, the lack of trial-by-trial feedback, and participants' lack of awareness that they are participating in a category-learning situation. These factors are then combined with strong category structure at time of test (multiple presentations of a prototype and low-level distortions that are all highly similar to each other) and explicit knowledge of the goal of categorization. Under these conditions, the dramatic role of learning during test seems easy to understand. It remains an open question whether analogous forms of learning during test arise in other category learning paradigms. Although such forms are not likely to be as dramatic as those that occur in the present kinds of incidental, single-category learning tasks, researchers should be on the lookout for them by including appropriate control conditions that manipulate the composition of the training instances and of the transfer tests.

AUTHOR NOTE

This research was supported by NIMH Grant R01 MH48494 to Indiana University. Correspondence concerning this article should be addressed to S. R. Zaki, Psychology Department, Bronfman Science Center, Williams College, Williamstown, MA 10267, or to R. M. Nosofsky, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405 (e-mail: szaki@williams.edu or nosofsky@indiana.edu).

REFERENCES

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, **56**, 149-178.

Bozoki, A., Grossman, M., & Smith, E. E. (2006). Can patients with Alzheimer's disease learn a category implicitly? *Neuropsychologia*, **44**, 816-827.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 638-648.

Gureckis, T. M., & Love, B. C. (2006). Bridging levels: Using a cognitive model to connect brain and behavior in category learning. In R. Sun (Ed.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 315-320). Hillsdale, NJ: Erlbaum.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1264-1282.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.

Hock, H. S., Tromley, C., & Polmann, L. (1988). Perceptual units in the acquisition of visual categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 75-84.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 418-439.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 322-330.

Ichikawa, S. (1985). Quantitative and structural factors in the judgment of pattern complexity. *Perception & Psychophysics*, **38**, 101-109.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 924-941.

Kinder, A., & Shanks, D. R. (2001). Amnesia and the declarative/nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience*, **13**, 648-669.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, **262**, 1747-1749.

Kolodny, J. A. (1994). Memory processes in classification learning: An investigation of amnesic performance in categorization of dot patterns and artistic styles. *Psychological Science*, **5**, 164-169.

Lamberts, K., & Shapiro, L. (2002). Exemplar models and category-specific effects. In E. M. E. Forde & G. W. Humphreys (Eds.), *Category specificity in brain and mind* (pp. 291-314). Hove, U.K.: Psychology Press.

Love, B. C., & Gureckis, T. M. (2007). Models in search of a brain. *Cognitive, Affective, & Behavioral Neuroscience*, **7**, 90-108.

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

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.

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.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 924-940.

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.

Palmeri, T. J., & Flanery, M. A. (2002). Memory systems and perceptual categorization. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 41, pp. 141-189). San Diego: Academic Press.

Posner, M. I., Goldsmith, R., & Welton, K. E., Jr. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology*, **73**, 28-38.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353-363.

Ramsey, J. L., Langlois, J. H., & Marti, N. C. (2005). Infant categorization of faces: Ladies first. *Developmental Review*, **25**, 212-246.

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.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407.

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, **46**, 2977-2987.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, **121**, 278-304.

Smith, E. E. (in press). The case for implicit category learning. *Cognitive, Affective, & Behavioral Neuroscience*.

Smith, J. D. (2002). Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science*, **13**, 437-442.

Smith, J. D. (2005). Wanted: A new psychology of exemplars. *Canadian Journal of Experimental Psychology*, **59**, 47-53.

Smith, J. D., & Minda, J. P. (2001). Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 984-1002.

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 800-811.

Smith, J. D., Murray, M. J., Jr., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 659-680.

Storms, G. (2004). Exemplar models in the study of natural language concepts. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 1-39). New York: Academic Press.

Strange, W., Keeney, T., Kessel, F., & Jenkins, J. (1970). Abstraction over time of prototypes from distortions of random dot patterns: A replication. *Journal of Experimental Psychology*, **83**, 508-510.

Zaki, S. R. (2004). Is categorization performance really intact in amnesia? A meta-analysis. *Psychonomic Bulletin & Review*, **11**, 1048-1054.

Zaki, S. R., & Nosofsky, R. M. (2001). A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cognitive, Affective, & Behavioral Neuroscience*, **1**, 344-359.

Zaki, S. R., & Nosofsky, R. M. (2004). False prototype enhancement effects in dot pattern categorization. *Memory & Cognition*, **32**, 390-398.

NOTE

1. Even this limitation seems to involve a statistical anomaly associated with the very low endorsement rates of the low distortions during the first 10 trials of the standard condition (see Table 1). Note that a standard prototype model, which assumes that a prototype is abstracted from the training instances, would also predict high endorsement rates for the low distortions during the first 10 trials of transfer.

(Manuscript received October 13, 2006;
revision accepted for publication February 19, 2007.)