

False prototype enhancement effects in dot pattern categorization

SAFA R. ZAKI

Williams College, Williamstown, Massachusetts

and

ROBERT M. NOSOFSKY

Indiana University, Bloomington, Indiana

Results from the classic dot pattern distortion paradigm have sometimes yielded prototype enhancement effects that could not be accounted for by exemplar models of categorization. However, in these experiments the status of the prototype was confounded with certain stimulus-specific properties as well as with the frequency of presentation of the prototype during testing. In two mock-subliminal experiments, participants made categorization judgments to patterns that were generated as prototypes, low-level distortions, or high-level distortions. The participants rated the prototypes as being more likely to be members of a category, although no patterns were presented during training, and there was no objective category structure. In two other experiments, greater prototype enhancement effects were observed when the prototype and low-level distortions were presented with greater frequency during transfer. These results suggest that classic prototype enhancement effects may not be due to the abstraction of a prototype at time of original learning, but rather to other factors not formalized in extant models.

In categorization research, one of the most central and debated issues has been the nature of category representation. According to prototype theory, people abstract the central tendency of a category and use that prototype as a basis for future categorization decisions (Homa & Vosburgh, 1976; Posner & Keele, 1968; Reed, 1972; Smith, Murray, & Minda, 1997). In contrast, according to exemplar models, rather than forming a prototype, people represent categories by storing individual category exemplars in memory (Heit, 1994; Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986).

One of the primary paradigms in which these theories have been investigated is the dot pattern paradigm first introduced by Posner, Goldsmith, and Welton (1967). In this paradigm, a prototype is created by randomly placing nine points in the central 30×30 grid of a 50×50 grid. Patterns at various levels of distortion can then be generated by moving each of the nine dots. For a desired level of distortion, the distance and direction that each dot is moved is governed by a statistical decision rule such that dots tend to move further at higher levels of distortion. Researchers have been drawn to this paradigm because, as presumably occurs in the natural world, categories are formed that have a graded structure, a poten-

tially infinite number of members, and complex underlying dimensions (Homa, 1984).

This methodology has been used extensively to test theories of category representation. In the early stages of dot pattern category research, Posner and Keele (1968) observed that after studying various members of a category, participants endorsed the previously unseen prototype at levels equal to and sometimes higher than the old training items and other new category members. This high level of endorsement of the prototype relative to other category members is generally referred to as a *prototype enhancement effect* and was traditionally taken as evidence that the participants in these experiments had in fact abstracted a prototype. However, exemplar models also provide a natural explanation of such prototype enhancement effects. Because the prototype is highly similar to numerous old training examples, even models that assume an exemplar-based representational system predict that the prototype will be endorsed with higher probability than other category members (Busemeyer, Dewey, & Medin, 1984; Hintzman, 1986; Shin & Nosofsky, 1992).

However, in recent years prototype theorists have severely challenged exemplar models on grounds that they are unable to predict the magnitude of the prototype enhancement effects. These challenges stem from results observed in a particularly influential version of the dot pattern task introduced by Knowlton and Squire (1993) and used subsequently by a number of researchers (Nosofsky & Zaki, 1998; Palmeri & Flanery, 1999; Reber & Squire, 1999; Reber, Stark, & Squire, 1998; Smith & Minda, 2001). In the Knowlton and Squire paradigm,

This research was supported by National Institute of Mental Health Grant R01 MH48494. We thank Evan Heit, Andy Wills, and two anonymous reviewers for their comments on an earlier version of this article. Correspondence concerning this article should be addressed to S. R. Zaki, Department of Psychology, Bronfman Science Center, Williams College, Williamstown, MA 10267 (e-mail: szaki@williams.edu).

participants view 40 high-level distortions of a dot pattern prototype in the training phase of the experiment. In the test that follows, the participants are shown various new patterns and are asked to decide whether each of them belongs to the category experienced during training. The new test patterns include multiple instances of the prototype, some low-level distortions, some high-level distortions, and random patterns that do not belong to the category. The general pattern of results is that participants endorse the prototype with the highest probability, followed in turn by the low-level distortions, high-level distortions, and random patterns.

Although exemplar models correctly predict this ordering of endorsement (Nosofsky & Zaki, 1998), Smith (2002) and Smith and Minda (2001, 2002) have argued strongly that such models fail to predict the steepness of the observed typicality gradient. That is, they fail to predict the extent to which the prototypes are endorsed over the low-level distortions and that the low-level distortions are preferred over the high-level distortions. In a series of formal analyses, Smith (2002) and Smith and Minda (2001, 2002) fitted both prototype and exemplar models to data obtained in the Knowlton and Squire (1993) paradigm. In these analyses, Smith and Minda measured similarity between dot patterns in terms of physical distances between pairs of corresponding dots. The results of their analyses showed the prototype model to provide a good description of the data, whereas the exemplar model fell short. Specifically, the exemplar model failed to predict the magnitude of the observed prototype enhancement effects and predicted a typicality gradient for the prototypes, low-level distortions, and high-level distortions that was too flat.

Nosofsky, Zaki, and Palmeri (2002) raised questions about some of the physical dot-distance assumptions that Smith (2002) and Smith and Minda (2001, 2002) used for computing dot pattern similarity. Introducing alternative assumptions, Nosofsky et al. found that there was no difference in the ability of prototype and exemplar models to account for the categorization data. Critically, however, even in these revised analyses, Nosofsky et al. acknowledged that the exemplar model still tended to underpredict the steepness of the typicality gradient and the magnitude of the prototype enhancement effect. Moreover, Nosofsky and Zaki (1998) had previously acknowledged the same pattern of results in formal modeling analyses that made use of direct similarity ratings instead of measures based on physical dot distance.

Although the underprediction of the magnitude of the prototype enhancement effect is small, the effect is intriguing and could potentially indicate that some form of prototype abstraction is indeed taking place in the dot pattern category-learning paradigm. However, the central theme of our present research is to point up reasons why a certain degree of “false” prototype enhancement often takes place in this paradigm. Specifically, we demonstrate that the degree of prototype enhancement is artificially raised owing to factors that may have nothing to do

with the format of people’s category representations. The presence of such false prototype enhancement effects has profound implications, therefore, for the interpretation of results from this classic and highly influential paradigm.

The first factor we consider has to do with how these dot patterns are constructed. As was described earlier, in the method developed by Posner et al. (1967) one first constructs the prototype of a category by placing nine points in the central 30×30 grid of a 50×50 grid. To create the other category members, each of these points is perturbed according to a statistical decision rule. The reason for starting out with the prototype in the central 30×30 grid is so that no high-level distortion will then be created whose points lie outside the 50×50 grid. However, this method also ensures that the prototypes are the most compact of all patterns, followed, in turn, by the low-level and high-level distortions.¹ It is possible that participants may be influenced by some aspect of this stimulus-specific confound when making their categorization judgments. Similar stimulus-specific effects have been shown to operate, for example, in recognition memory tasks involving naturalistic faces (Zaki & Nosofsky, 2001).

A more critical factor that may have played a role in boosting the prototype enhancement effect involves the composition of the transfer test in the Knowlton and Squire (1993) paradigm. Recall that during training, participants experience only high-level distortions of the prototype. By contrast, at time of transfer, participants are “flooded” with numerous patterns located near the center of the category. Specifically, among the 44 category members that are presented, there are four repetitions of the prototype and 20 low-level distortions. When the models are fitted to the data, the simplifying assumption is made that because no feedback is given in the transfer phase, participants’ category representations are static. However, this assumption is probably false. For example, Palmeri and Flanery (1999) have demonstrated that participants can perform the Knowlton and Squire categorization task even when the study phase is omitted. Apparently, participants can glean the category structure simply by being exposed to the transfer items. It is certainly possible, therefore, that experience with the transfer test may steepen the typicality gradient even after participants are exposed to the training sequence. That is, given the composition of the transfer test, the effects of learning during transfer (LDT) may be amplified for the central items. In fact, Nosofsky et al. (2002) demonstrated that a version of the exemplar model that assumed that people continue to augment their category representation with the items presented during transfer provided excellent quantitative fits to the Knowlton and Squire data. A key question, therefore, is whether there is converging evidence that the typicality gradient may be artificially steepened simply because of the composition of the transfer test.

In this article, we examine the effects of these two factors on the rate of endorsement of the prototypes into a

category. In the first pair of experiments, we investigate whether the compactness of the prototype relative to other category members might influence the probability with which the prototype is endorsed. In the second pair of experiments, we investigate whether the composition of the transfer test influences the steepness of the typicality gradient.

EXPERIMENT 1A

Experiment 1A was designed to investigate the idea that, even without previous exposure to a set of dot patterns and even without the presence of any objective category structure, people would be more likely to endorse prototypes than high-level distortions as category members. To test this hypothesis, we measured the rate of endorsement of these two types of stimuli into hypothetical categories in a paradigm in which participants were told that they had seen various categories subliminally but in fact had seen no stimuli at all. In addition to the participants' seeing no stimuli in training, there was no categorical structure in the set of items shown in transfer. Therefore, any difference that might be observed between the endorsement rates of the prototypes and high-level distortions can be attributed only to the different way in which these items are constructed.

Method

Participants. One hundred seventy-nine Indiana University undergraduate students participated in the research in order to fulfill a requirement of their introductory psychology class.

Stimuli. There were no stimuli shown in the "subliminal training" phase. In the test phase, 44 dot pattern prototypes were generated for every participant according to the method outlined by Posner et al. (1967). Forty of these dot patterns were then distorted at a level of 7.7 bits/dot. These patterns are henceforth referred to as *high-level distortions*. The other 4 prototype patterns were not distorted and are henceforth referred to as *prototypes*. It is critical to note that these prototypes and high-level distortions are not related to each other. In a sense, they are from 44 different dot pattern categories. Finally, for each participant, the test also included the particular prototype stimulus used in Knowlton and Squire's (1993) study.

Procedure. The participants were tested individually in a computerized task. There were 30 trials in the "subliminal training" phase. On each trial, the participants saw a fixation point followed by a white rectangle. The white rectangle remained on the screen for approximately 50 msec. Therefore, what the participants essentially saw was a quick flash on the screen. The participants were asked to simply attend as best they could. They were told that we were testing their ability to subliminally perceive dot patterns.

In the phase that followed, the participants were told that all of the dot patterns in the first phase belonged to five categories of patterns in the same sense that, if they had seen pictures of dogs, cats, and birds, each picture would be a member of one of these categories. On each trial, one of the test patterns was shown on the computer screen. The pattern appeared as white dots on a black background. The participants were asked to judge how likely it was that each pattern was a member of one of the categories in the first phase on a scale from 1 (*not likely*) to 5 (*very likely*). The pattern remained on the screen until the participants made their responses. The participants received no feedback on their responses. For each participant, the test set was a unique set of patterns with the exception of

the Knowlton and Squire (1993) prototype. The patterns were shown in a random order.

Results

The mean categorization ratings for the Knowlton and Squire (1993) prototype, the randomly generated prototypes, and the randomly generated high-level distortions are shown in Figure 1. Overall, the mean categorization judgment for the randomly generated prototypes (3.13) was significantly higher than the mean categorization judgment for the randomly generated high-level distortion patterns [2.94; $t(176) = 3.537, p = .001$]. In addition, the mean categorization judgment of the Knowlton and Squire prototype (3.15) was comparable to the mean categorization judgment of the other prototypes and was significantly higher than the mean categorization judgment of the randomly generated high-level distortions [$t(176) = 1.997, p < .05$].

EXPERIMENT 1B

The results of Experiment 1A indicate that dot pattern prototypes may receive a small performance boost simply because of item-specific idiosyncrasies involved in Posner et al.'s (1967) stimulus construction method. The purpose of Experiment 1B was to replicate this effect in a situation that did not so severely contrast the prototypes and the high-level distortions. That is, we were concerned that, because the only items in the transfer test in Experiment 1A were prototypes and high-level distortions, we may have inadvertently highlighted the difference in the compactness of these items. In most transfer tests, including that of Knowlton and Squire (1993), a more graded set of patterns is included.

Therefore, in Experiment 1B we tested once again whether participants were more likely to endorse prototypical patterns over other types of items simply because of stimulus-specific characteristics. The general design was similar to that of Experiment 1A, except that the transfer test included items that were generated to be prototypes, low-level distortions, and high-level distortions of their own categories.

Method

Participants. Eighty Indiana University undergraduate students participated in the research in order to fulfill a requirement of their introductory psychology class.

Stimuli. No stimuli were shown in the mock subliminal training phase. For the transfer test, we generated a unique set of 44 random dot pattern prototypes for each participant. Ten of these patterns were then distorted at 4 bits/dot (low-level distortions), 30 were distorted at 7.7 bits/dot (high-level distortions), and 4 were left undistorted (prototypes).

Procedure. The procedure was identical to that of Experiment 1A.

Results

Figure 2 shows the mean categorization judgments for the different types of patterns in Experiment 1B. Although none of the items were related, there was a significant

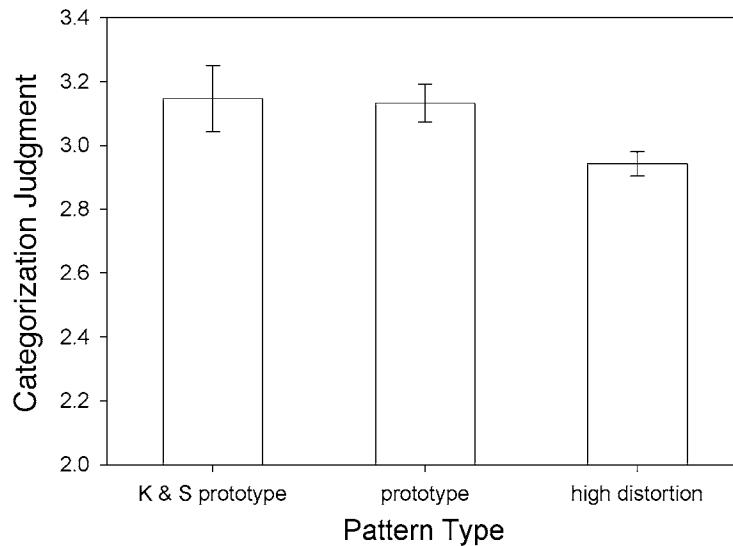


Figure 1. Observed mean categorization judgments for the Knowlton and Squire (K & S, 1993) prototype, randomly generated prototypes, and randomly generated high-level distortions in Experiment 1A. Error bars represent 1 standard error of the mean.

main effect of “distortion” of the items on categorization judgments [$F(2,79) = 8.486$, $MS_e = .165$, $p < .001$].

Discussion

The results from Experiments 1A and 1B suggest that participants show a typicality gradient in assigning categorization judgments even when there are no training items and each of the items in the transfer test belongs to its own category. These results suggest that artifacts of the stimulus-generation process may contribute to the magnitude of the prototype enhancement observed in the Knowlton and Squire (1993) paradigm.

EXPERIMENT 2A

The purpose of Experiment 2A was to investigate whether the composition of the transfer tests in previous studies could also have contributed to “false” prototype enhancement effects. Many categorization studies make the assumption that because no feedback is given during the transfer phase, participants do not modify their categories on the basis of the transfer items. This assumption is likely to be an oversimplification (Nosofsky, 1986). In fact, we know from previous studies that participants can acquire categorical information from the test phase of the Knowlton and Squire (1993) task when no training instances are provided (Palmeri and Flanery, 1999). Furthermore, Nosofsky et al. (2002) demonstrated that a version of the exemplar model that assumed people continue to learn during the transfer phase provided excellent quantitative fits to the Knowlton and Squire data. The question we ask here is whether, even after actual training has taken place, the typicality gradient is steepened

due to the repeated presentation of the prototype and low-level distortions at time of transfer.

To address this issue, we contrasted two conditions that differed only in terms of the composition of the transfer test. In both conditions, participants trained on the same 40 high-level distortions from the Knowlton and Squire (1993) paradigm. After training, in one condition the complete Knowlton and Squire transfer test was shown. In the other condition, only a subset of the test items containing very few items near the center of the category was shown. The hypothesis is that exposure to the complete set of test items results in a steeper typicality gradient because participants are augmenting their category representations with the patterns shown in the test phase.

Method

Participants. One hundred twenty-three Indiana University undergraduates participated in the full condition, described below. One hundred ninety-eight Indiana University undergraduates participated in the subset condition, described below.

Stimuli. The stimuli were the set of dot patterns used by Knowlton and Squire (1993). In the full condition, the transfer set consisted of all 84 items—namely, 4 presentations of the prototype, 20 low-level distortions, 20 new high-level distortions, and 40 random patterns. In the subset condition, the transfer test consisted of 1 presentation of the prototype, 2 low-level distortions, 20 new high-level distortions, and 20 random patterns. (We reduced the number of random patterns shown in the subset condition so that the base rate of category members vs. nonmembers would be roughly equivalent across the full and subset conditions.) For each participant in the subset condition, the specific items representing each stimulus type were randomly chosen from the full set of items.

Procedure. Due to the fact that each participant in the subset condition would provide only one binary response to the prototype and two binary responses to the low-level distortions, we chose to run a large number of participants. In the training phase, the par-

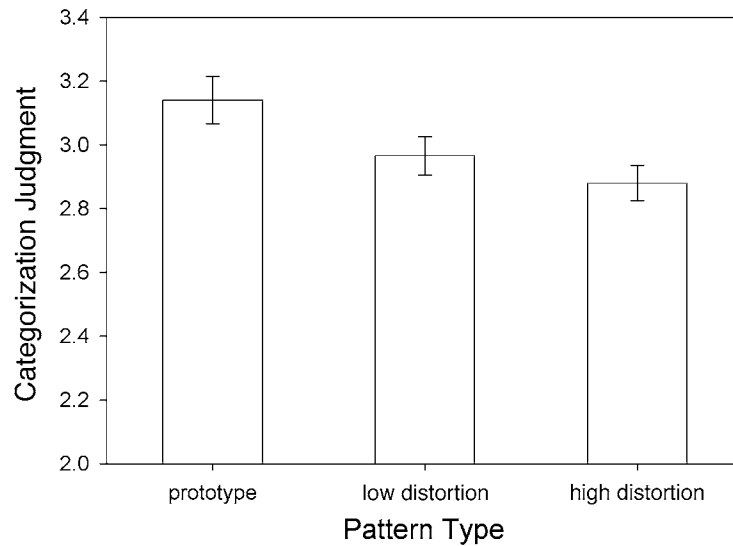


Figure 2. Observed mean categorization judgments for the prototypes, low-level distortions, and high-level distortions in Experiment 1B. Error bars represent 1 standard error of the mean.

Participants in both conditions were asked to observe the items as they flashed on the screen. Each pattern was shown for 5 sec. The patterns appeared in a different random order for each participant. In the transfer task, the participants were told that they had seen dot patterns that belonged to a category, just as if they had seen a series of pictures of the category DOG. On each trial, a pattern appeared on the screen and the participants' task was to decide whether the presented pattern was a member of the category. The pattern remained on the screen until the participant gave a response. Following Knowlton and Squire (1993), the participants received no feedback after their responses. All items appeared in a different random order for each participant.

Results

The typicality gradients for the subset and full conditions are shown in the left panel of Figure 3. Because we are interested primarily in the within-category typicality gradient, we limit our analysis to the performance on the prototype, low-level distortions, and high-level distortions. It is evident from inspection of the figure that, as hypothesized, this typicality gradient is steeper in the full condition than in the subset condition. We conducted a mixed-model ANOVA with distortion level and condition as factors. The results indicated a main effect of distortion, with participants tending to endorse items closer to the category prototype [$F(2,638) = 6.802$, $MS_e = .078$, $p < .001$]. Importantly, however, there was a condition \times distortion level interaction [$F(2,638) = 3.041$, $MS_e = .078$, $p < .05$], confirming the observation that the typicality gradient is steeper in the full condition than in the subset condition. However, some caution should be used in interpreting these results, because assumptions of normality and homogeneity of variance are likely to be violated in the subset condition, in which each participant contributes either one or two binary responses to two of the cells.

Therefore, to provide converging evidence, we computed for each individual participant the slope of the regression line that best fitted the endorsement probabilities of the prototype, low-level distortions, and high-level distortions. The mean slope in the full set condition ($M = -.067$) was significantly steeper than that in the subset condition [$M = -.011$; $t(319) = 2.38$, $p < .05$]. This result provides further evidence that the steepness of the typicality gradient is indeed influenced by the composition of the transfer test.

EXPERIMENT 2B

A potential concern with the results from Experiment 2A involves the reason for the flattened typicality gradient in the subset condition. Specifically, relative to the full set condition, most of the flattening arises because of increased endorsement probabilities for the high-level distortions, not decreased endorsement probabilities for the prototype. Because of possible shifts in response criteria across the conditions, our view is that the absolute endorsement probabilities of the individual pattern types are not as informative as the steepness of the overall gradient itself. Nevertheless, to follow up on this issue, we decided to repeat the basic experiment in a more sensitive paradigm. The design was identical to the previous one, except that instead of collecting binary-valued *yes/no* responses, we collected more continuous-valued category membership ratings.

Method

Participants. One hundred eighty-five Indiana University undergraduates participated in the subset condition, and 104 students participated in the full condition.

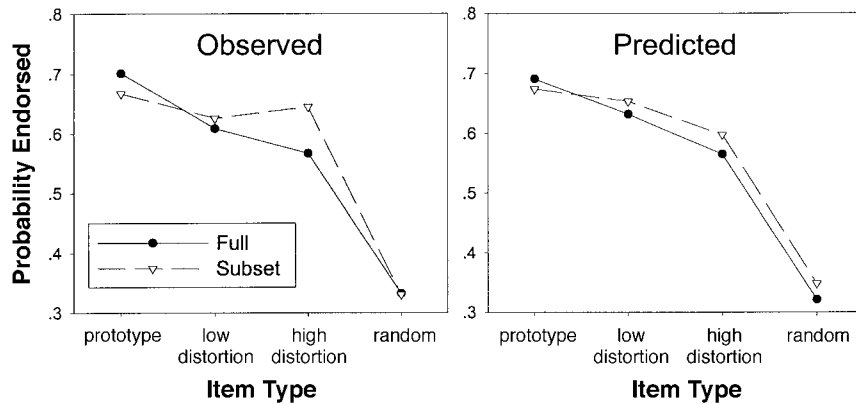


Figure 3. Observed and predicted endorsement levels of the various item types in Experiment 2A. Full = full condition; subset = subset condition.

Procedure. The procedure was identical to that in Experiment 2A, with one exception. In the transfer phase of Experiment 2B, instead of giving a *yes/no* response, the participants were asked to rate how confident they were that the pattern was a member of the category on a scale from 1 (*definitely no*) to 7 (*definitely yes*).

Results

The mean categorization judgments as a function of item type for the full and subset conditions are shown in Figure 4. Once again, we conducted an analysis of variance restricted to the prototype and the low-level and high-level distortions. As in Experiment 2A, there was a main effect of distortion [$F(2,574) = 20.046$, $MS_e = .524$, $p < .01$]. More importantly, there was also a significant condition \times distortion interaction, reflecting that the typicality gradient for the full condition was steeper than the typicality gradient for the subset condition [$F(2,574) = 7.159$, $MS_e = .524$, $p < .001$]. The regression slope analysis yielded the same pattern of results, with the mean slope in the full set condition ($M = -.301$) being significantly steeper than that in the subset condition [$M = -.066$; $t(287) = 3.54$, $p < .001$]. Finally, it is evident from inspection of Figure 4 that the flattened slope in the subset condition is due primarily to reduced endorsement levels for the prototype rather than to increased endorsement levels for the high-level distortions.

Discussion

The results of both Experiments 2A and 2B indicated that the participants who were exposed to the full set of items in the transfer test of the Knowlton and Squire (1993) paradigm produced steeper typicality gradients than did those who experienced only the subset of the items. These results are consistent with the idea that learning continues during the transfer phase and that learners augment their category representation with the new items experienced during transfer. The prototype enhancement effect is magnified because in the full condition of the Knowlton and Squire paradigm, participants

experience numerous patterns from the center of the category that had not been experienced during training.

THEORETICAL ANALYSIS

To corroborate our interpretation of the reason for the steepened typicality gradient, in this section we present a formal analysis involving an application of an exemplar model to the choice-probability data from Experiment 2A. We emphasize at the outset that the analysis is intended for illustrative purposes only. Rigorous tests of the model are not possible for several reasons. First, there is little consensus in the field regarding the appropriate manner of computing psychological similarity for these dot patterns. Second, we are at too preliminary a stage of research to specify the precise LDT process that operates. Our more limited goal is to provide an illustrative modeling application to bolster the plausibility of an exemplar-based LDT account.

According to the present LDT model, the evidence in favor of category membership is given by

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

where $\sum s_{it}$ is the summed similarity of test item i to the training items, $\sum s_{if}$ is the summed similarity of the test item to the transfer items (excluding itself), and y is a free parameter reflecting the weight given to the transfer items relative to the training items.² The probability that the item is endorsed as a category member is then given by

$$P(\text{Cat}) = \frac{S}{S+k}, \quad (2)$$

where k is a response criterion parameter.

The psychological distance between dot patterns is assumed to be based on the physical dot-distance measure used by Smith and Minda (2001). We refer the reader to Smith and Minda's (2001) article for a detailed presentation of the physical dot-distance measure. Finally, the

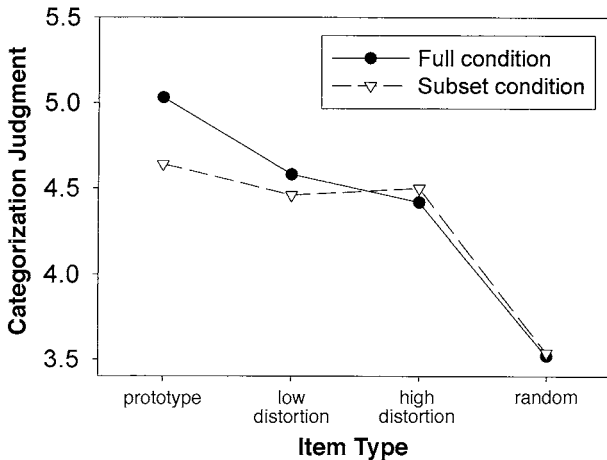


Figure 4. Mean categorization judgments of the various item types by participants in the full and subset conditions in Experiment 2B.

similarity between patterns i and j was assumed to be an exponential decay function of this distance (Shepard, 1987),

$$s_{ij} = \exp(-cd_{ij}), \quad (3)$$

where c is an overall sensitivity parameter.

We fitted the LDT model to the data by minimizing the sum of squared deviations between the predicted and observed endorsement probabilities for the four pattern types in the full and subset conditions. The free parameters were the overall sensitivity parameter c , the transfer-exemplar weight parameter γ , and the separate response-criteria parameters k_F and k_S for the full and subset conditions, respectively. The predicted endorsement probabilities are shown in the right panel of Figure 3. Overall, the model provides a good account of the pattern of results across conditions. Its main limitation is that it underestimates the endorsement rate for the high-level distortions in the subset condition. Nevertheless, the model captures the critical qualitative effect of a steeper typicality gradient in the full set condition in comparison with the subset condition.

GENERAL DISCUSSION

The dot-pattern prototype-distortion paradigm stands among the most venerable paradigms in the investigation of the nature of category representations. Historically, the finding that people classify untrained prototypes with high accuracy strongly motivated the idea that a prototype abstraction process operates during category learning. However, subsequent work indicated that exemplar memory models also account well for such prototype enhancement effects.

Nevertheless, the pendulum of research ideas has recently begun to sway back in the direction of prototype abstraction. Smith (2002) and Smith and Minda (2001,

2002) have argued strongly that exemplar models fail to predict the magnitude of observed prototype enhancement effects in Knowlton and Squire's (1993) influential version of the dot pattern paradigm. Although the question of whether prototype models account for the data better than exemplar models do is an issue of continued debate (Nosofsky et al., 2002; Smith, 2002), exemplar theorists have acknowledged that there is indeed a greater degree of observed prototype enhancement than is predicted by the standard models.

The key theme of the present research, however, has been to demonstrate that a "false" prototype enhancement may be involved. It is "false" in the sense that the reason for the extreme enhancement is not because a prototype abstraction process has taken place during initial category learning. Rather, performance on the prototype is boosted due to the operation of other factors that are not incorporated in the standard models.

The first factor is that there is a stimulus-specific confound associated with the dot distortion paradigm, such that the prototypes are the most compact patterns and items of high-level distortion are the least compact. In Experiment 1, we demonstrated that the mere method by which dot pattern prototypes and their distortions are generated can lead to enhanced endorsement rates for the prototypes. We suspect that beyond this particular example, other, similar stimulus-specific factors may underlie some of the prototype-enhancement effects reported in the literature. For example, in generating a set of stimuli for use in an experiment, it is unlikely that a researcher would start with a bizarre-looking dot pattern prototype, so some special care is often taken in the selection of the prototypes that are used. Such special care is unlikely to occur, however, for the numerous dot pattern distortions that are then generated from the prototype. Inherent stimulus-specific properties, such as pattern "goodness" (Garner, 1974), may yield performance boosts for the prototype that are not a result of a prototype abstraction process having taken place during category learning.³

The second factor is that the category representation developed at the end of the initial training phase is not a static representation but continues to evolve during transfer. Furthermore, in the Knowlton and Squire (1993) version of the paradigm, the added learning that takes place is amplified for the prototype and other items near the center of the category. The reason is that in this version of the paradigm, participants experience only high-level distortions during initial training but then are flooded with the low-level distortions and the prototype at time of transfer. If the participants continue to augment their category representations with these new exemplars experienced during transfer, then the natural consequence is that even exemplar models predict a magnified prototype enhancement effect and a steeper typicality gradient. Our present experimental work demonstrated clearly that the structure of the transfer phase does indeed exert a powerful influence on performance in accord with these ideas.

Furthermore, we reported an illustrative theoretical analysis involving an LDT exemplar model to bolster the plausibility of this account.

Note that our experimental results and model-based illustration do not rule out the idea that some form of prototype abstraction takes place. However, the results do challenge the idea that the prototype is formed at the end of initial training. They also indicate that the results from the Knowlton and Squire (1993) paradigm are not highly diagnostic for discriminating between prototype versus exemplar accounts of category representation.

Although the focus of the present work was on prototype enhancement effects in the dot distortion paradigm, we believe that the results have potentially far more wide-reaching implications. In numerous paradigms designed for study of the nature of category representations, there is an initial training phase followed by a transfer phase. The transfer phase often includes highly diagnostic stimuli designed to discriminate among alternative models. Because feedback on such items is generally withheld at time of transfer, the category representation that has evolved is assumed to remain stable. The present research, however, suggests strongly that category representations continue to evolve during transfer, even in the absence of explicit feedback. That is, an unsupervised form of category learning seems to be overlaid on the representation that evolved during supervised training. These results, therefore, corroborate findings from the unsupervised category literature that suggest that people make use of inherent category structure without the aid of feedback (Ahn & Medin, 1992; Clapper & Bower, 1994; Gureckis & Love, 2003; Homa & Cultice, 1984; Wills & McLaren, 1998). The novelty in these results is that they demonstrate that the unsupervised learning occurs even after the formal training stage in this paradigm.

The extent to which such unsupervised learning affects the category representation in other paradigms remains an open question. It seems reasonable that in the present Knowlton and Squire (1993) paradigm, unsupervised learning at time of transfer plays a major role. During the training phase of the paradigm, participants are unaware that they are even learning a category. It is only at the start of transfer that participants are made aware that they are in a category-learning situation. Under such conditions, it is reasonable to expect that the implicit feedback and unsupervised learning that occurs during transfer will dominate performance. Such unsupervised LDT may well operate more subtly in more standard paradigms as well, and researchers should be aware of its potential effects.

REFERENCES

- AHN, W.-K., & MEDIN, D. L. (1992). A two-stage model of category construction. *Cognitive Science*, *16*, 81-121.
- 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.
- CLAPPER, J. P., & BOWER, G. H. (1994). Category invention in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 443-460.
- GARNER, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- GURECKIS, T. M., & LOVE, B. C. (2003). Towards a unified account of supervised and unsupervised category learning. *Journal of Experimental & Theoretical Artificial Intelligence*, *15*, 1-24.
- 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.
- HOMA, D. (1984). On the nature of categories. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 49-94). Oxford: Academic Press.
- HOMA, D., & CULTICE, J. C. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 83-94.
- HOMA, D., & VOSBURGH, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning & Memory*, *2*, 322-330.
- KNOWLTON, B. J., & SQUIRE, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747-1749.
- MEDIN, D. L., & SCHAFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.
- NOSOFKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.
- NOSOFKY, 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.
- NOSOFKY, R. M., ZAKI, S. R., & PALMERI, T. J. (2002). *Exemplars, prototypes, and the categorization-recognition dissociation*. Unpublished manuscript.
- 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.
- 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.
- 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. L., & SQUIRE, L. R. (1998). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory*, *5*, 420-428.
- REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*, 382-407.
- SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.
- SHIN, H. J., & NOSOFKY, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, *121*, 278-304.
- SMITH, J. D. (2002). Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science*, *13*, 437-442.
- 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.
- WILLS, A. J., & MCLAREN, I. P. L. (1998). Perceptual learning and free classification. *Quarterly Journal of Experimental Psychology: Comparative & Physiological Psychology*, **51B**, 235-270.
- ZAKI, S. R., & NOSOFSKY, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1022-1041.

NOTES

1. To ascertain whether this stimulus-specific confound exists in the Knowlton and Squire (1993) stimulus set, we conducted the following analysis: We measured the dispersion of the dots in each pattern by calculating the absolute city-block distance of each dot to the center of the display. The average dispersions were 5.94, 6.37, 7.26, and 8.24 units for the prototypes, low-level distortions, high-level distortions, and random patterns, respectively.

2. This process is intended as an approximation, because it involves summing the similarity of individual test items to at least some transfer exemplars that have not yet been presented. Conducting a more rigor-

ous model-based analysis would involve the specification of a more detailed process by which the category representation is modified trial by trial, and extensive simulations of the different random orders in which transfer items are experienced.

3. We should acknowledge that such effects of "compactness" and "goodness" are likely to be most pronounced in the present type of single-category paradigm, in which participants judge whether test items are or are not category members. However, Smith's (2002) and Smith and Minda's (2001, 2002) challenges of exemplar models pertain only to the single-category paradigm, and the present results make clear that the stimulus generation method does indeed exert a significant effect in this situation. It is an open question whether similar stimulus-specific effects may also play some role in multiple-categorization paradigms, in which participants must discriminate to which of several categories a test item belongs. Although speculative, one possibility is that increased pattern goodness may reduce the probability with which participants use explicit or implicit "junk" responses. Such a process would also lead to increasing the magnitude of prototype enhancement effects.

(Manuscript received May 15, 2003;
revision accepted for publication September 26, 2003.)