Distinguishing Prototype-Based and Exemplar-Based Processes in Dot-Pattern Category Learning

Article *in* Journal of Experimental Psychology Learning Memory and Cognition · August 2002

DOI: 10.1037/0278-7393.28.4.800 · Source: PubMed

2 authors, including:

John Paul Minda
The University of Western Ontario

Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 4, 800–811

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.4.800

Distinguishing Prototype-Based and Exemplar-Based Processes in Dot-Pattern Category Learning

J. David Smith
State University of New York at Buffalo

John Paul Minda
University of Illinois at Urbana–Champaign

The authors contrast exemplar-based and prototype-based processes in dot-pattern categorization. In Experiments 1A and 1B, participants provided similarity ratings of dot-distortion pairs that were distortions of the same originating prototype. The results show that comparisons to training exemplars surrounding the prototype create flat typicality gradients within a category and small prototype-enhancement effects, whereas comparisons to a prototype center create steep typicality gradients within a category and large prototype-enhancement effects. Thus, prototype and exemplar theories make different predictions regarding common versions of the dot-distortion task. Experiment 2 tested these different predictions by having participants learn dot-pattern categories. The steep typicality gradients, the large prototype effects, and the superior fit of prototype models suggest that participants refer to-be-categorized items to a representation near the category's center (the prototype), and not to the training exemplars that surround the prototype.

Some descriptions of categorization suggest that humans average their exemplar experiences to form a category prototype, compare new items to it, and accept the items as category members if they are similar enough to the prototype. Other descriptions of categorization suggest that humans store the exemplars they experience as whole, independent memory traces, compare new items to these, and accept the items as category members if they are similar enough to the exemplars. This theoretical issue remains unresolved because both prototype and exemplar theories (and models) are powerful enough to explain many phenomena.

For example, consider the well-known dot-distortion task (Homa, Rhoads, & Chambliss, 1979; Homa, Sterling, & Trepel, 1981; Posner, Goldsmith, & Welton, 1967, Posner & Keele, 1970). Initially, the performance enhancement of the prototype relative to other category members seemed to show prototype abstraction by participants during category learning. But Shin and Nosofsky (1992) argued that an exemplar-based mechanism could produce prototype-enhancement effects. Similarly, early studies showed that the categorization of training exemplars decays more rapidly than the categorization of transfer items (especially prototypical transfer items—Homa, Cross, Cornell, Goldman, & Schwartz, 1973; Posner & Keele, 1970; Strange, Keeney, Kessel, & Jenkins, 1970). Initially, this result seemed to show that humans learn prototypes during category training and use them in transfer even while partially forgetting the training exemplars. However, subsequent formal analyses suggested that an exemplar-based system could explain this result, too (Hintzman, 1986; Hintzman & Ludlam, 1980).

More recently, research by Knowlton and Squire (Knowlton, Mangels, & Squire, 1996; Knowlton & Squire, 1993; Knowlton, Squire, & Gluck, 1994; Squire & Knowlton, 1995; Squire & Zola, 1996) showed that amnesics perform like controls when categorizing dot patterns that are statistical distortions of an underlying prototype. Knowlton and Squire concluded from this result that categorization performance relies on an implicit memory system—intact in amnesics—that represents category-level information in the form of prototypes. But Nosofsky and Zaki (1998) argued that this result could also be explained and modeled using a unitary system based on similarities to stored exemplars and lacking any prototypes or abstraction process (see Smith & Minda, 2001, for a theoretical analysis of Knowlton and Squire's results and of Nosofsky and Zaki's, 1998, reinterpretation of them).

Given this unresolved theoretical problem, it is important to consider cases in which prototype and exemplar theories make contrasting predictions so that the two representational hypotheses can be empirically differentiated. This is the article's empirical purpose. The case considered here is a version of the dot-distortion task that has received extensive empirical and theoretical attention (Knowlton & Squire, 1993; Nosofsky & Zaki, 1998; Palmeri & Flanery, 1999; Reber, Stark, & Squire, 1998a, 1998b; Smith & Minda, 2001). In this version of the task, participants are trained on 40 high-level distortions of an originating prototype, and are then asked to endorse (or not) previously unseen probe items as belonging in the category. These probe items are copies of the originating prototype, low-level distortions of it, high-level distortions of it, and random-unrelated items that are distant from the psychological space defined by the training exemplars and outside the category they represent.

This task raises simply the representational question about prototypes and exemplars. That is, given that we train participants on 40 exemplars that are spread out in psychological space around

J. David Smith, Department of Psychology and Center for Cognitive Science, State University of New York at Buffalo; John Paul Minda, Beckman Institute, University of Illinois at Urbana–Champaign.

Correspondence concerning this article should be addressed to J. David Smith, Department of Psychology, Park Hall, State University of New York at Buffalo, Buffalo, New York 14260. E-mail: psysmith@acsu.buffalo.edu

a central prototype, do they learn from this training and represent in memory the specific exemplars or the general prototype? The hope has been to discover which representation(s) control(s) categorization by probing participants' knowledge with the four transfer item types. This article shows that performance on those transfer item types may be dependent on the representation(s) learned and that one may be able to tell from performance on these probes what participants have learned and how they're processing.

The article proceeds as follows. Experiments 1A and 1B explore the organization of the psychological space of dot distortions, and the similarity relations among members of dot-distortion categories. Using a pairwise similarity-rating task, we find the shape of the similarity gradient that surrounds the central prototype and the shape of the similarity gradient that surrounds the training exemplars. It turns out that the psychology of dot patterns is such that the shapes of these gradients are very different. Simple geometrical intuitions explain why this difference obtains and why it generalizes over dot-pattern categories and probably over all psychologies of dot patterns. In turn, this difference causes prototype and exemplar theories to make contrasting predictions regarding the dot-pattern categorization task. Experiment 2 tests these different predictions by having participants learn dot-distortion categories and by modeling their performance profiles using equivalent prototype and exemplar models.

Experiment 1A

Experiment 1A explores the structure of the psychological similarity space that organizes families of random-dot polygons. This experiment was summarized briefly in Smith and Minda (2001) and is described fully here. Using a similarity-rating task, we measure the rated similarity between every pairwise combination of Level 7, Level 5, Level 3, Level 1, and Level 0 distortions of an underlying prototype (these levels will be defined shortly). Of critical interest is the shape of prototype-based and exemplar-based similarity gradients. That is, how does similarity change as the prototype is compared to distortions of Levels 0, 1, 3, 5, and 7? Prototype theory expects that categorization performance will reflect this similarity gradient. Likewise, how does similarity change as the training exemplars (which surround the prototype in stimulus space) are compared to distortions of the prototype at Levels 0, 1, 3, 5, and 7? Exemplar theory expects that categorization will reflect this similarity gradient. Thus, studying the character of psychological similarity in dot-distortion space may offer constructive insights about the predictions of prototype and exemplar theories.

A second purpose of Experiment 1A is to use the similarity-rating data to derive a measure of psychological similarity between dot patterns, a measure that is objective, theoretically neutral regarding prototype and exemplar theories, and that provides an equivalent set of inputs to prototype- and exemplar-based models.

Method

Participants. Thirty-seven undergraduates at the State University of New York at Buffalo participated to fulfill a course requirement.

Dot-pattern stimuli. The stimulus materials for the dot-distortion task are created with a well-established method that generates families of dot patterns from prototypes. In this method, nine points (the prototype) are randomly selected from within the central 30 × 30 area of a 50 × 50 grid. The distortions (the family members) are produced by applying a series of probabilities that determine whether each dot will keep the same position it had in the prototype, and if not, how far it will be displaced. Different series of probabilities let one produce dot patterns that are low-, medium-, or high-level distortions of the prototype and that form families of dot patterns that have strong, moderate, or weak family resemblance to the prototype. These different families of distortions can be thought of as lying on hyperspheres (N-dimensional shells) of different radii surrounding the central prototype, with the radius determined by the distortion level (Homa, Dunbar, & Nohre, 1991; Homa et al., 1981).

Specifically, distortions were built from prototypes by probabilistically moving each dot into one of five areas that covered the 20 × 20 grid of pixels that surrounded it. For Area 1, the dot kept its original position. For Area 2, the dot was moved to one of the 8 pixel positions immediately around its original position. For Area 3, the dot was moved to one of the 16 pixel positions in the second layer of pixels around it. For Area 4, the dot was moved into one of the 75 pixel positions in the third, fourth, and half of the fifth layer of pixels around it. For Area 5, the dot was moved into one of the remaining 300 pixel positions in the surrounding 20 × 20 pixel grid (i.e., to the 5th, 6th, 7th, 8th, 9th, or 10th layer of pixels around the dot's original position).

Different levels of distortions were arranged by adjusting the probabilities that dots would make small or large movements away from their original position. For Level 0 distortions (i.e., with 0 bits per dot of uncertainty attending the new position of the dot), the probabilities that dots would move to each of the five areas were 1.000, .000, .000, .000, and .000, respectively. Thus dots in Level 0 distortions never moved, and these patterns reproduced the prototype. For Level 1 distortions (i.e., with 1 bit per dot of uncertainty attending the new position of dots), the five probabilities were 0.880, .100, .015, .004, and .001. These distortions included mostly unmoved or slightly moved dots. The five probabilities were .590, .200, .160, .030, and .020 for Level 3 distortions, .200, .300, .400, .050, and .050 for Level 5 distortions, and .000, .240, .160, .300, and .300 for Level 7.7 (henceforth referred to as Level 7) distortions. These probabilities were those used in Posner et al.'s (1967, Table 1, p. 31) study for making distortions that mapped the range of the uncertainty variable (bits/dot) that was favored at the time.

With the dot positions chosen for a pattern, the constellation was magnified to be more visible. Each pixel position in the distortion algorithm was mapped to a 3 × 3 pixel square on the screen, and the dot was placed in the center of the appropriate 9-pixel cell on the screen. In this way the stimulus patterns were magnified threefold from being drawn in a virtual 50 × 50 coordinate space to being shown on an actual 150 × 150 pixel space on the screen. Finally, the DrawPoly procedure within Turbo Pascal 7.0 connected successive dots by lines and filled the resulting polygon shape in red. This followed the common practice of presenting the dot distortions as random polygon shapes (Homa et al., 1979, 1981).

Each trial consisted of two distortions presented side by side in the center of an 11.5-in. computer screen against a black background. Below each pair of shapes was centered the question: How different are these two shapes? Below this a 1 to 6 scale was shown, with the 1 labeled "No Difference" and the 6 labeled "Big Difference."

Twenty-five types of pairs were presented, representing every pairwise combination of distortion levels in the two orders in which they could occur (0-1, 1-0, 0-3, 3-0, etc.). The five same-level pair types (e.g., 1-1) were also presented in two "orders" (3-3, 3-3) to ensure that these kinds of trials were sampled just as often as were the other pair types. The second presentations of the same-level pairs were treated as trial types 26–30. These 30 trial types were presented in 16 successive random permutations that were presented without any apparent structure or break to participants. Each participant received 480 pairs in all, including 32 repetitions of each pairwise combination of distortion levels. Remember that the prototype for a trial was chosen at random on each of 480 trials for each of 37 participants. Thus this similarity-scaling experiment represents a comprehensive sample of dot-distortion space.

Participants were given these instructions. “In this experiment you will see two red polygon shapes on each trial. Your job is to look at each pair and decide how different they are. If there is no difference between them, give them a rating of ‘1.’ If there is a big difference between them, give them a rating of ‘6.’ Use intermediate ratings for intermediate differences. Use the number keys at the top of the keyboard to make your response. Make sure to use the whole range of the scale from 1 to 6.”

Results

Table 1 gives the average similarity rating for every pairwise combination of two distortion levels (together with two objective distance measures discussed below). Each entry in Table 1 summarizes 1,184 observations. Aspects of Table 1 are highlighted in the figures considered next.

Regarding exemplar-based similarity gradients, Figure 1A shows the average similarity rating provided by observers for three distortion levels (7, 5, and 3) compared to all lower distortion levels (e.g., Level 7 distortions compared to those at Levels 5, 3, 1, and 0). Here the exemplars in a shell around the prototype become the reference standard to which successive item types are compared. Subjective similarity barely changes for a given exemplar shell as the comparison item becomes less distorted and moves in toward the category center. Exemplar-based comparisons produce flat similarity gradients.

Regarding prototype-based comparisons, Figure 1B shows the average similarity rating provided by observers when a prototype is compared to the different distortion levels (i.e., 7, 5, 3, and 1). Here the central prototype (a Level 0 distortion) becomes the reference standard to which successive item types are compared. Subjective similarity changes sharply as the comparison item moves in toward the center of the category (i.e., as the comparison item becomes a lower- and lower-level distortion). Prototype-based comparisons produce steep similarity gradients.

Figure 1C highlights the contrast between exemplar- and prototype-based comparisons that arises if one uses Level 7 distortions in the training phase of a dot-pattern experiment. This is the most common empirical situation in the literature, so this contrast affects one’s sense of many experiments. A key theoretical question has been whether participants use the exemplars or the prototype as the comparative standard in making categorization decisions. The figure contrasts the expected similarity gradient under each hypothesis. Prototype-based similarity increases all the way to the center of the category as comparison items become less distorted. Exemplar-based similarity does not.

Though this effect is essentially qualitative, we also confirmed its significance through an analysis of variance (ANOVA) that compared exemplar-based gradients (Level 7 distortions compared to those at Levels 7, 5, 3, 1, and 0) to prototype-based gradients (Level 0 distortions compared to those at Levels 7, 5, 3, 1, and 0). Here the Level 7–Level 0 comparison—that belongs in both gra-

Table 1
Rated Similarity, Pythagorean Interdot Distance, and Logarithmic Interdot Distance for Pairs of Random-Dot Polygons at Different Levels of Distortion in Experiment 1A

Inner distortion level	Outer distortion level	Similarity rating	Pythagorean distance	Logarithmic Pythagorean distance
Exemplar-based comparisons				
0	0	1.31	0.00	0.00
0	1	1.60	0.19	0.16
1	1	1.81	0.36	0.29
0	3	2.58	1.01	0.66
1	3	2.62	1.11	0.72
3	3	3.11	1.69	0.95
0	5	3.23	2.04	1.09
1	5	3.25	2.16	1.13
3	5	3.66	2.56	1.24
5	5	3.85	3.14	1.40
0	7	4.55	4.90	1.76
1	7	4.58	4.98	1.77
3	7	4.59	5.19	1.80
5	7	4.71	5.54	1.86
7	7	4.97	7.24	2.09
Prototype-based comparisons				
0	0	1.31	0.00	0.00
0	1	1.60	0.19	0.16
0	3	2.58	1.01	0.66
0	5	3.23	2.04	1.09
0	7	4.55	4.90	1.76

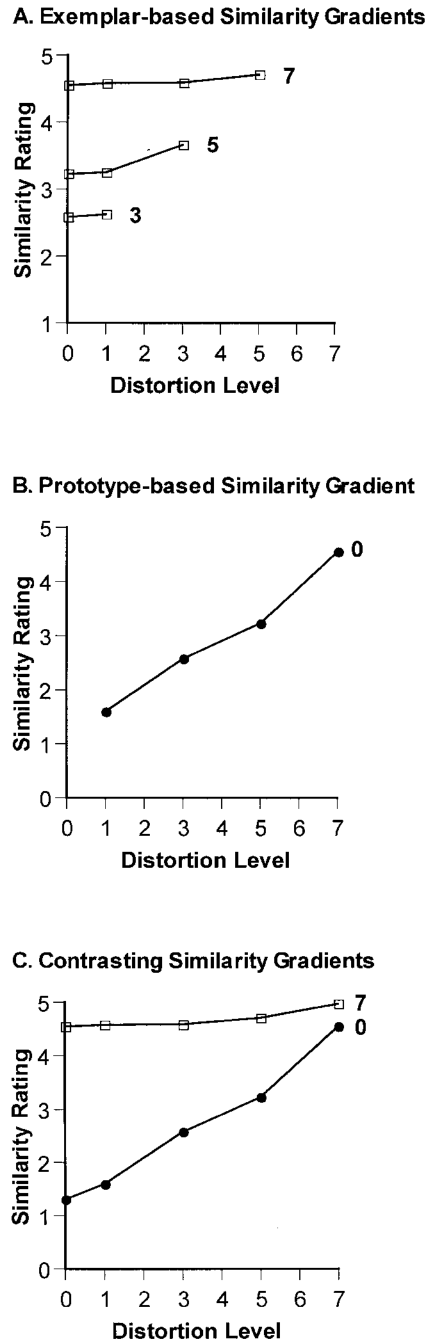

Figure 1. A: The mean similarity rating provided by observers in Experiment 1A when patterns at three levels of distortion (Levels 7, 5, and 3) were compared to all lower levels of distortion (e.g., Level 7 distortions compared to distortions at Levels 5, 3, 1, and 0). B: The mean similarity rating provided by observers in Experiment 1A when patterns at different levels of distortion (Levels 7, 5, 3, and 1) were compared to their prototype (Level 0 distortion). C: The prototype-based similarity gradient contrasted with the gradient based on comparisons to Level 7 training distortions.

dients—was included in both gradients. The interaction between the two gradients was overwhelmingly significant, $F(4, 144) = 245.29$, $p < .01$, $MSE = 0.0992$.

Objective measures of psychological similarity. Table 1 gives researchers an objective measure of the psychological similarity between dot polygons that is isomorphic to participants' ratings of similarity and that is theoretically neutral to the debate about prototypes and exemplars. Toward deriving this measure, Figure 2A shows the average rated similarity between patterns at different distortion levels plotted over the average Pythagorean distance moved by corresponding dots between patterns. To calculate this average distance measure, the Pythagorean distance between each pair of corresponding dots (with each dot specified by *x*-axis and *y*-axis coordinates) was found. Then these nine, two-dimensional distances were averaged. There appears to be a perfect logarithmic relationship between these two variables. Completing the derivation of the objective measure, Figure 2B shows the similarity ratings plotted over the measure ln (1 + average Pythagorean

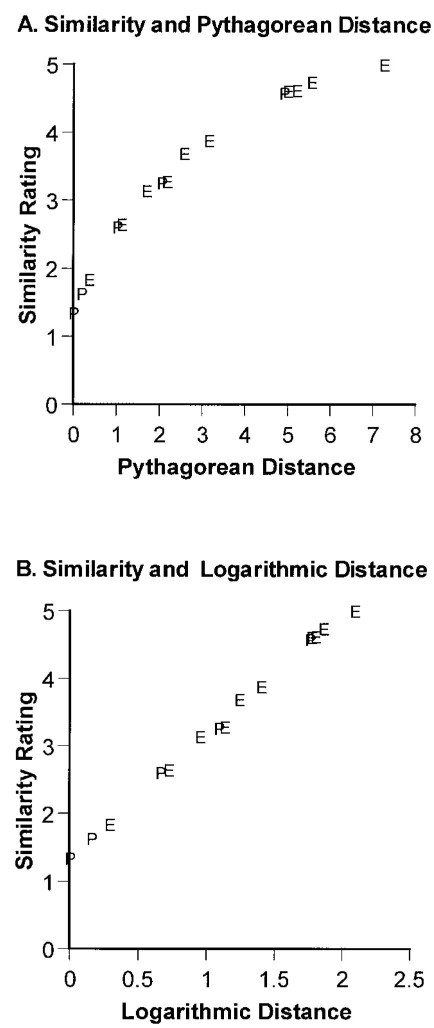

Figure 2. A: The relationship between the average Pythagorean distance moved by corresponding dots between pairs of patterns and the average similarity rating given them. Exemplar-based comparisons (E) that involve two distortions and prototype-based comparisons (P) that involve one distortion and the prototype are shown separately. B: The relationship between the measure ln (1 + average Pythagorean distance moved per dot) and average similarity rating, plotted as in Figure 2A.

distance moved per dot). Now there appears to be a perfect linear relationship with rated similarity. The correlation over the 15 pairs of observations is .998. Thus this objective, logarithmic distance measure is perfectly proportional to participants' subjective ratings of similarity (Posner et al., 1967).

Moreover, one can show that this measure is theoretically neutral regarding the prototype–exemplar debate. In both figures, the similarity ratings that involve exemplar–exemplar comparison pairs are labeled *E*; the similarity ratings that involve prototype–exemplar comparisons are labeled *P*. Clearly both types of comparisons obey the same psychophysics.

Experiment 1B

These results were replicated in Experiment 1B. Experiment 1B also explored the prototype-based and exemplar-based gradients relative to 30 specific originating prototypes. This allowed us to ask whether the present results are probably applicable to dot-pattern prototypes used in past or future research.

Method

Participants. Ninety undergraduates at the State University of New York at Buffalo participated to fulfill a course requirement. The larger sample size in this case allowed us to measure the similarity gradients relative to 30 individual prototypes.

Dot-pattern stimuli. The construction of prototypes and the rules for generating distortions were those described in Experiment 1A.

Procedure. The presentation of the trials, rating scale and instructions, and the presentation of the trials in random permutations of 30 pairwise combinations of distortion levels followed exactly the procedures described in Experiment 1A. The only difference was that the trials were also presented in random permutations of trials that used each of 30 originating prototypes. Thus each participant experienced 16 trials that involved each prototype.

Results

Table 2 gives the average similarity rating for every pairwise combination of two distortion levels (together with the two objective distance measures already discussed). Each entry in Table 2 summarizes 2,880 observations. Comparing Table 2 to Table 1, one sees that all results replicated exactly. The exemplar-based similarity gradients were flat again, as is clearly seen for the Level 7 distortions in Rows 11–15 of Table 2. The prototype-based similarity gradients were steep again, as is seen in Rows 16–20 of Table 2.

Although this difference is essentially qualitative, we also confirmed its significance through an ANOVA that compared exemplar-based gradients (Level 7 distortions compared to those at Levels 7, 5, 3, 1, and 0) to prototype-based gradients (Level 0 distortions compared to those at Levels 7, 5, 3, 1, and 0). The Level 7–Level 0 comparison—that belongs in both gradients—was included in both gradients. Again the interaction between the two gradients was overwhelmingly significant, $F(4, 356) = 505.69$, $p < .01$, $MSE = 0.112$.

Table 2
Rated Similarity, Pythagorean Interdot Distance, and Logarithmic Interdot Distance for Pairs of Random-Dot Polygons at Different Levels of Distortion in Experiment 1B

Inner distortion level	Outer distortion level	Similarity rating	Pythagorean distance	Logarithmic Pythagorean distance
Exemplar-based comparisons				
0	0	1.40	0.00	0.00
0	1	1.70	0.19	0.16
1	1	1.98	0.37	0.29
0	3	2.67	0.99	0.65
1	3	2.79	1.11	0.71
3	3	3.20	1.70	0.96
0	5	3.36	2.07	1.10
1	5	3.37	2.12	1.11
3	5	3.65	2.55	1.24
5	5	3.92	3.19	1.41
0	7	4.55	4.93	1.76
1	7	4.56	4.93	1.76
3	7	4.60	5.19	1.80
5	7	4.67	5.54	1.86
7	7	4.96	7.30	2.10
Prototype-based comparisons				
0	0	1.40	0.00	0.00
0	1	1.70	0.19	0.16
0	3	2.67	0.99	0.65
0	5	3.36	2.07	1.10
0	7	4.55	4.93	1.76

Finally, we confirmed that this difference applied to the gradients for all 30 originating prototypes in the experiment. Figure 3A shows the exemplar-based similarity gradient for each prototype; Figure 3B shows each prototype-based gradient. This result suggests that the profound difference between prototype-based and exemplar-based similarity gradients has been in effect for the prototypes used in past research, and will be in effect for the prototypes used in future research.

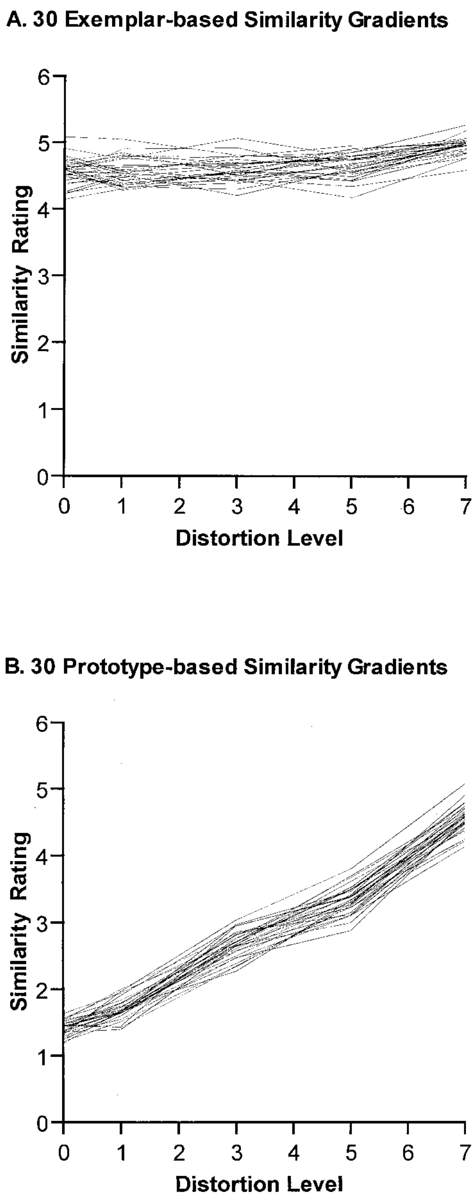

Figure 3. A: The mean similarity rating provided by observers in Experiment 1B when Level 7 distortions were compared to distortions at Levels 7, 5, 3, 1, and 0. This exemplar-based similarity gradient is shown separately for each of 30 originating prototypes. B: The mean similarity rating provided by observers in Experiment 1B when Level 0 distortions (prototypes) were compared to distortions at Levels 7, 5, 3, 1, and 0. This prototype-based similarity gradient is shown separately for each of 30 originating prototypes.

Discussion

The results of Experiments 1A and 1B have implications for prototype theory. Prototype theory assumes that psychological similarity to the category representation (the prototype) maps closely to category belongingness and to participants' level of category endorsement. Thus, prototype theory makes the following prediction to which it can appropriately be held. Participants (because the prototype is the comparative standard for to-be-categorized items) should show levels of category endorsement that increase strongly all the way to the center of the category. They should show steep typicality gradients and strong prototype-enhancement effects. Although we will show below that prototype models do produce this pattern, no model is needed to see the necessity of this prediction from the perspective of prototype theory.

These results also have implications for exemplar theory. Exemplar theory assumes that psychological similarity to the exemplar representations maps closely to category belongingness and to levels of category endorsement. Thus exemplar theory makes the following prediction to which it can appropriately be held. Participants (because the exemplars are the comparative standard for to-be-categorized items) should show levels of category endorsement that stay nearly flat all the way to the center of the category. They should show flat typicality gradients and negligible prototype-enhancement effects. Although we will show below that exemplar models do produce this pattern, no model is needed to see the necessity of this prediction from the perspective of exemplar theory.

One can understand geometrically why this distinction between prototype-based and exemplar-based comparisons must hold. To do so, visualize (Figure 4) what occurs as a hypothetical to-be-categorized item moves from out in the random-unrelated region of psychological space (Positions 8–4), then through the region of psychological space occupied by the high-level distortions (Position 3), the low-level distortions (Position 2), then all the way to center of the category (Position 1). At each step (from Position 8 to Position 1 in Figure 4), one can ask how the item's closeness or similarity to the exemplar shell changes, assuming that the exemplars might be the category representations, and how the item's belongingness changes to the category defined by the shell of exemplars. At each step one can also ask how the item's closeness or similarity to the prototype center changes, assuming that the prototype might also be the category representation, and how the item's belongingness changes to the category the prototype summarizes.

At first (Position 8), when the item is far from the category system, it will gain in closeness equally and strongly to both the exemplar shell and the central prototype, because at this distance the category system (the exemplar shell and central prototype) will nearly be a point source to the item. No matter which is the category representation, the item will gain strongly in similarity to the representation and in belongingness to the category represented.

Closer in to the category system (Position 5), the approaching item will still close on the prototype directly and gain strongly in similarity and belongingness to it. But now the item will close on the exemplar shell less directly because it will approach the flanks of the exemplar shell obliquely. If the exemplars are the category

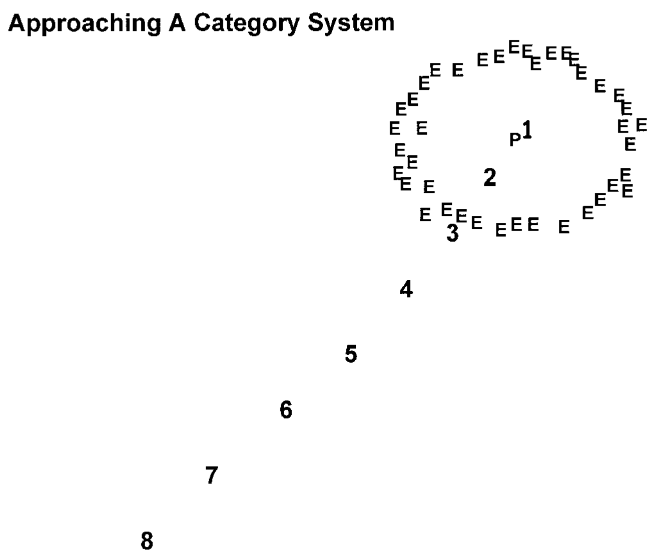

Figure 4. Schematic drawing of a series of hypothetical dot distortions (8 to 1) that are made to approach the psychological space of a category system containing a central prototype (P) and a shell of training exemplars (E).

representation, the item will gain more slowly in similarity to them and in belongingness to their category.

Finally, as the to-be-categorized item moves inside the exemplar shell (Position 2), it will still close on the prototype and gain in similarity and belongingness to it. But now it will not close on the exemplar shell. It will move toward some of the exemplars on the opposite side of the exemplar shell from where it entered. In compensation, though, it will move directly away from others on the side of the shell it entered. Many other exemplars will just slide past the item in a distance-neutral fashion. If the exemplars are the category representation, the item will not increase in similarity to them or in belongingness to the category they represent.

Experiment 1B showed that this difference between prototype-based and exemplar-based comparisons applies to many dot-pattern categories. More generally, this difference probably applies across all dot-pattern categories and even all psychologies of dot patterns, no matter the similarity metric of the space, the attentional strategy of the participant, or the shape of the exemplar cloud in psychological space (see also Smith & Minda, 2001). The 40 distortions will always be spread out somehow in psychological space. An approaching item will ultimately be unable to approach them all—at some point it must reach a steady state of moving toward some and away from others. In contrast, an approaching item will always be able to approach the prototype indefinitely closely, and become indefinitely similar to it, because the prototype is a single point in the space, not a cloud of points in the space. This difference between prototype-based and exemplar-based comparisons will hold even if (to take one extreme example) the 40 exemplars are stretched out in a line perpendicular to the path of the approaching to-be-categorized item or even if (to take the opposite extreme example) they are stretched out in line with its path.

The geometrical necessity of the situation converges with the present results to confirm that there is an important difference between prototype-based and exemplar-based comparisons in the dot-distortion task. It is felicitous—but probably necessary—that the geometry fits with participants' subjective experience and descriptions of similarity relations in dot-distortion space.

We point out that Posner et al. (1967) reported similar results decades ago. The present experiments build on their study in many ways. First, it extends their findings from dot patterns to the random-dot polygons that have been the other important class of stimulus materials in the dot-distortion literature. Second, it brings their result into the domain of the similarity and distance measures that are favored now, as opposed to the uncertainty and bits/dot measures that were favored then but made the pattern of their results difficult to discern. Third, the present study conducted a systematic examination of dot-distortion space by examining thousands of prototypes. Fourth, it went beyond the meaningful prototypes of Posner et al. (e.g., letters and shapes) to explore more broadly the random-dot prototypes that have been favored in experimental work since. Fifth, the present study emphasizes the sharp relevance of the results to the debate about prototypes and exemplars in the literature. The work of Posner et al. predated that debate. This fact alone is a strong endorsement of the theoretical neutrality of the objective similarity measure that emerged here and from their work.

It is also important to revisit Posner et al.'s (1967) research because the literature has not given due emphasis to the differences they found between prototype-based and exemplar-based similarity gradients, and to the theoretical implications these differences have. The different shapes of the exemplar-based and prototype-based distance/similarity gradients found here and in Posner et al.'s work suggest that the dot-distortion methodology allows situations to be created in which either prototype theory or exemplar theory could possibly be simply falsified. Experiment 2 evaluates this possibility.

Experiment 2

In Experiment 2, we train participants on a group of high-level distortions that lie in a shell around the prototype and then transfer them to prototypes (Position 1 in Figure 4), low-level distortions (Position 2), high-level distortions (Position 3), and to random items outside the region of psychological space occupied by the category exemplars (Position 8). If participants categorize by making comparisons to the members of the exemplar shell, they will show a flat gradient of category endorsements for high-level distortions, low-level distortions, and prototypes. But if participants categorize by making comparisons to the prototype (the category center), then they will show strongly increasing category endorsements over these three item types. Experiment 2 asks which gradient obtains, and asks whether prototype or exemplar models (made equivalent through the use of the neutral similarity inputs established in Experiment 1) fit the data better.

Method

Participants. Thirty undergraduates at the State University of New York at Buffalo participated to fulfill a course requirement.

Dot-pattern stimuli. The dot-distortion methodology was used to create families of random-dot polygons as described in the method for Experiment 1. Experiment 2 features prototypes (Level 0 distortions), low-level distortions (Level 5), high-level distortions (Level 7), and random items outside the category. These category-unrelated dot patterns were constructed by first randomly producing an unrelated prototype and then by producing a random Level 7 distortion of that prototype.

Training procedure. Participants received these instructions. "You will now see 40 red polygon shapes that belong to the same category of shapes. Study each shape for the 5 seconds that it appears so that you can learn to recognize members of the red category." Each of the 40 training trials then consisted of a Level 7 distortion centered in the top half of an 11.5-in. computer screen against a black background. The distortion was visible for 5 s.

Transfer procedure. Participants received these instructions. "Now you will see 84 polygon shapes. Half of them belong to the red category and half do not. Decide whether each shape is a member of the red category. Respond Y if YES; respond N if NO." Each shape was visible until the participant's response. The 84 transfer trials comprised 4 Level 0 distortions (prototypes), 20 Level 5 distortions, 20 previously unseen Level 7 distortions, and 40 random-unrelated items.

General procedure. Each participant served in five training-transfer cycles that involved different randomly chosen prototypes, exemplars, and transfer items. Each single transfer-training cycle required less than 10 min to complete. At each change of task, the exact instructions were repeated for training and transfer except that successively different colors were used to describe the category (red, yellow, green, blue, and purple, respectively). The stimuli in each task (both the 40 training items and the 84 transfer items) were then colored in that way. Each participant received training-transfer cycles based on 5 of the 30 originating prototypes whose similarity gradients were evaluated in Experiment 1B.

Models and Fitting Procedures

This section describes the equivalent prototype and exemplar models used here that take identical inputs, have identical parameters, and differ only in their underlying representational assumption. This section also describes the procedures used for fitting models to data.

The exemplar model. This model estimates the similarity a to-be-categorized item type (prototype, low-level distortion, high-level distortion, or random-unrelated) has to the high-level distortion training exemplars, and uses this similarity to estimate the strength of category endorsement that would result for that item type. In the model, the probability of endorsing an item type (i) into the training category (A) is

$$P(R_A|S_i) = \frac{\eta_{ih}}{\eta_{ih} + k}.$$

In this equation, the psychological similarity (η) between the to-be-categorized item type (i) and the high-level training distortions (h) is compared to a criterion quantity (k) that acts as a proportionalizing threshold for the model. That is, the stronger the similarity between a transfer item type and the training distortions, the more this threshold would be exceeded and the more strongly that item type would be endorsed into the category. The use of the threshold parameter k followed the model in Nosofsky and Zaki's (1998) study. The parameter k is one of the exemplar model's two free parameters.

The psychological similarity between each item type and the training distortions was calculated as follows. Following the psychophysical principle of dot-pattern distance established in Experiment 1, we began with the item type's average distance from Level 7 distortions such as those used in training. Given any transfer item type (i.e., prototype, low-level distortion, high-level distortion, and random-unrelated) and the training item type (high-level distortion), we simply found the average Pythagorean distance that corresponding dots were moved between patterns of the two types, and we set psychological distance equal to ln (1 + average Pythagorean distance). To estimate these distances powerfully, the measures used in the modeling of Experiment 2 were based on a sampling of 1 million of each of the relevant pair types. For exemplar-based comparisons, the average logarithmic distance between potential training distortions and prototype, low-level distortions, high-level distortions, and random-unrelated items potentially used in transfer were 1.761, 1.866, 2.098, and 2.894, respectively. As is standard in exemplar models, these distances were then scaled by a sensitivity parameter (c) that was the exemplar model's second free parameter. Then, as is also standard, the scaled distances were transformed into psychological similarities using an exponential-decay function. Thus the similarity between a transfer item type and the high-level distortion training exemplars was: $\eta_{ih} = e^{-cd_{ih}}$. This similarity—calculated identically for the four transfer item types—was then entered into the choice rule already described to predict the levels of category endorsement for the four transfer item types.

We used a hill-climbing algorithm to maximize the fit between predicted and observed profiles and thus to find the best-fitting predicted profile. This hill-climbing algorithm operated as follows. We chose a single parameter configuration (sensitivity and criterion value) and calculated the predicted categorization probabilities for the four item types according to that configuration. The degree of fit between the predicted and observed categorization probabilities was the sum of the squared deviations (SSD) between them. Then small adjustments were made to the provisional best-fitting parameter settings and the new settings were adopted if they produced a better fit (i.e., a smaller SSD between predicted and observed performance). During each iteration of the algorithm, a parameter and a directional change were chosen at random. These changes were small: gradations of 1/10,000 for sensitivity and 1/10,000 for the criterion value. These changes always respected the upper and lower bounds of the parameters (0.001 and 10.000). To ensure that local minima were not a serious problem, this fitting procedure was repeated by choosing four more configurations of the exemplar model and hill climbing from there.

We also briefly evaluated the gamma model, which was first introduced by Ashby and Maddox (1993), but which has come to be used as a more complex and mathematically powerful version of the exemplar model. This model is identical in every respect to the exemplar model just described, except that the gamma model's final choice-rule equation is

$$P(R_A|S_i) = \frac{\eta_{ih}^{\gamma}}{\eta_{ih}^{\gamma} + k^{\gamma}}.$$

That is, all of the quantities in the exemplar model's choice-rule equation can be raised to any power gamma that improves fit. The gamma parameter represents this model's third free parameter.

The prototype model. This model estimates the similarity a to-be-categorized item type (prototype, low-level distortion, high-level distortion, or random-unrelated) has to the category prototype, and uses this similarity to estimate the strength of category endorsement that would result for that item type. In the model, the probability of endorsing an item type (i) into the training category (A) is

$$P(R_A|S_i) = \frac{\eta_{ip}}{\eta_{ip} + k}.$$

In this equation (which is nearly identical to the exemplar model's corresponding equation), the psychological similarity (η) between the to-be-categorized item type (i) and the prototype (p) is compared to the proportionalizing threshold k. Given more prototype similarity, the threshold would be exceeded more and the level of category endorsement would grow stronger. The parameter k is one of the prototype model's two free parameters.

The psychological similarity between each item type and the prototype was calculated as follows. Following the psychophysical principle of dot-pattern distance established in Experiment 1, we began with the item type's average distance from the prototype. Given any transfer item type (i.e., prototype, low-level distortion, high-level distortion, and random-unrelated) and the prototype, we simply found the average Pythagorean distance that corresponding dots were moved between patterns of the two types, and we set psychological distance equal to ln (1 + average Pythagorean distance). On the basis of a sampling of 1 million of each of the relevant pair types, the average logarithmic distances between the prototype and prototypes, low-level distortions, high-level distortions, and random-unrelated items were 0.000, 1.093, 1.761, and 2.850, respectively. As in the exemplar model, these distances were then scaled by a sensitivity parameter (c) that was the prototype model's second free parameter. Then, as also in the exemplar model, the scaled distances were transformed into psychological similarities using an exponential-decay function. Thus the similarity between a transfer item type and the prototype was: $\eta_{ip} = e^{-cd_{ip}}$. This similarity—calculated identically for the four transfer item types—was then entered into the choice rule already described to predict the levels of category endorsement for the four transfer item types. In all respects, including the operation of the hill-climbing algorithm, the prototype model was identical to the exemplar model in structure, parameters, and in the character of the distance inputs it received. The many close similarities between these models make this pair of models balanced and appropriate for comparing prototype and exemplar theory.

Results

Levels of learning. Over all 150 training-transfer cycles, participants averaged 61.7% correct (SD = 13.5%). There were individual differences in overall performance (SD = 8.4%). Fourteen participants averaged below 60% correct over all 5 training-transfer cycles. There were also differences in performance within participants across the five training-transfer cycles. On average, participants had a standard deviation of 9.8% in the percentages correct they achieved on their 5 training-transfer cycles. These performance variations may reflect training sets that happened to instantiate a category more or less well, or transfer sets that probed category knowledge more or less well, or even fatigue. Participants averaged 64.5%, 65.9%, 59.2%, 59.8%, and 58.9% correct on training-transfer cycles 1–5, respectively. The variation in performance across and within participants recommended modeling each of the 150 performance profiles separately, and we report the results of this modeling next.

Results of modeling. Figure 5A shows the result when the best-fitting SSDs of the prototype and exemplar models were averaged within overall-performance bins—that is, within groups of performance profiles that showed consensually dismal performance (less than 40%), consensually poor performance (50%–55%), and so forth. This graph makes several points.

First, both models fit equally and very poorly the profiles that represent low levels of overall performance. This is because both

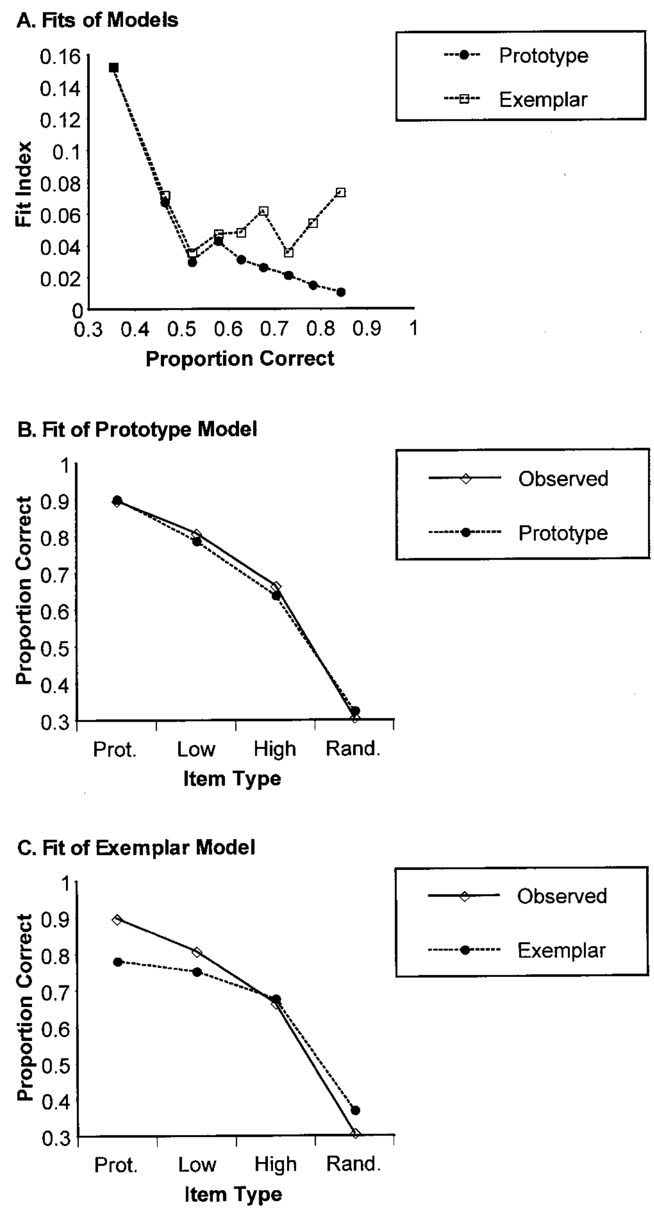

Figure 5. A: The average fit of the prototype and exemplar model across performance profiles that were binned by their overall proportion correct in the ranges of less than .4000, .4000–.4999, .5000–.5499, .5500–.5999, .6000–.6499, .6500–.6999, .7000–.7499, .7500–.7999, and greater than .7999. B: The fit of the prototype model to the observed categorization performances in Experiment 2. C: The fit of the exemplar model to the observed categorization performances in Experiment 2. Prot. = prototype; Rand. = random.

models are about the use of learned representations in categorization, whereas these profiles are about participants who didn't learn useful representations of either the prototype or exemplar kind. Second, once there is even a hint of category learning, as overall proportion correct rises above 50%, prototype models fit better (with a smaller SSD) and begin to be preferred as a description of performance over exemplar models. Third, for higher levels of learning, the preference for the prototype description grows much stronger. In fact, it seems that the higher the level of learning, the better the prototype model fits. In contrast, it seems that the higher the level of learning, the worse the exemplar model fits.

Over all 150 performance profiles, the prototype model fit significantly better than did the exemplar model (fit indices of 0.040 and 0.059, respectively), $t(149) = 7.482$, $p < .01$. The same was true of the 79 performance profiles that showed learning (i.e., overall performance greater than 60%). The fit indices for the prototype and exemplar models were 0.021 and 0.055, respectively, $t(78) = 8.247$, $p < .01$.

The character of the models' descriptions. The solid line in Figure 5B shows the composite observed performance profile taken over the 79 profiles that showed learning. One can see the steep (23%) typicality gradient over the three category-member item types (prototype, low-level distortion, high-level distortion) and the large (9%) prototype-enhancement effect. This curve mirrors the prototype-based similarity gradient shown in Figure 1C more than the exemplar-based similarity gradient shown there.

Figure 5B also shows the composite predicted performance profile obtained by averaging the best-fitting performance profiles when the prototype model was fit to the same 79 performances. The similar character of these two curves is clear. Figure 5C reprises the composite observed performance profile and also shows the composite predicted performance profile obtained by averaging the best-fitting performance profiles when the exemplar model was fit to these same 79 performances. The dissimilar character of these two curves is also clear. The exemplar model predicts only a 3% prototype-enhancement effect, one third of that observed. The exemplar model predicts only a 10% typicality gradient over the three category-member item types (prototypes, low-level distortions, high-level distortions), compared to the 23% gradient that is observed.

The gamma model. We also fit a more complex and powerful version of the exemplar model that contains the parameter gamma. Entering this discussion, we note that the use of this model here is problematic for several reasons. First, it grants the exemplar model asymmetric power and mathematical complexity over the prototype model and makes it difficult to compare fit indices. Second, it grants the exemplar model three free parameters as it tries to recover only four data points. This modeling situation risks being tautological. Third, the parameter gamma in many cases acts systematically to steepen typicality gradients and to enlarge prototype-enhancement effects. This means that adding gamma can be tantamount to adding a prototype process to the exemplar model, making the model theoretically ambiguous and potentially misleading psychologically (see Smith & Minda, 1998). Nonetheless, because exemplar theory has come to rely on the gamma parameter (Nosofsky & Johansen, 2000)—especially when modeling individual performance profiles as we did here—it is useful to consider the behavior of the model in the present situation.

Over all 150 performance profiles, the prototype model (with fewer free parameters) fit significantly better than did the gamma model (with more free parameters). The fit indices of the prototype and gamma models were 0.040 and 0.046, respectively, $t(149) = 3.645$, $p < .01$. The same was true of the 79 performance profiles that produced learning (fit indices of 0.021 and 0.030, respectively), $t(78) = 2.974$, $p < .01$.

Beyond its poorer fit, the character of the gamma model's predictions was still wrong for the data. The gamma model, like the standard exemplar model, predicted only a 3% prototype-enhancement effect, one third that observed. The gamma model predicted only a 14% typicality gradient over the three category-member item types (prototypes, low-level distortions, high-level distortions) compared to the 23% gradient observed. On average, it underpredicted prototype performance by 5.4%; it overpredicted performance on the high-level distortions by 3.7%. These errors are large given the ratio of free parameters to data points in these fits. But these errors are sensible—the gamma model showed some characteristics of exemplar-based processing that participants did not show.

General Discussion

Experiments 1A and 1B explored the shape of the similarity gradients that surround the training exemplars in a dot-distortion category task. They showed that exemplar-based comparisons create flat similarity gradients moving in toward the center of the category. This flatness is a geometrical and algebraic necessity (Smith & Minda, 1998, pp. 1428–1432). Experiments 1A and 1B showed that it is psychologically real. The exemplar-based similarity gradient covered only .4 of a rating scale point over the whole range of stimuli from Level 7 distortions to Level 0 distortions (see Tables 1 and 2). The prototype-based similarity gradient covered eight times that range.

This flatness of the exemplar-based gradient grounds exemplar theory's predictions about performance in the dot-distortion task. Exemplar theory links the level of exemplar similarity to the level of category belongingness and category endorsement. Accordingly, it predicts flat typicality gradients and small prototype-enhancement effects in a categorization task. The exemplar model behaves in this way (Figure 5C). Participants do not. They show large prototype-enhancement effects, three times what the exemplar model predicts. They show steep typicality gradients, twice what the exemplar model predicts. In both respects exemplar theory's predictions are qualitatively wrong for what participants do.

The present results are consistent with those of other dot-distortion studies. Knowlton and Squire's (1993) control participants showed 11% prototype-enhancement effects and 18% typicality gradients. Their amnesic participants showed a 13% prototype effect and a 16% typicality gradient. In two experiments, Reber, Stark, and Squire (1998a, 1998b) showed 17% prototype effects and 20% typicality gradients. Palmeri and Flanery (1999), in an interesting study that required participants to learn categories on-line during the transfer portion of the dot-distortion task, found a 10% prototype effect and a 20% typicality gradient. Each of these results is like the present result, and in each case the exemplar model would fail to recover the observed data pattern as well as a prototype model would. (See Smith & Minda, 2001, for further discussion and modeling results regarding these studies.)

We point out that the psychophysical principle regarding dot-pattern similarity that we established in Experiment 1 could be used to reconsider and reanalyze any other of the many dot-distortion studies in the literature. The simplicity of our approach is such that one does not even need to know the unique, individual coordinates of the nine-dot patterns used in the original research. The nature of Posner et al.'s (1967) distortion algorithm is such that the population estimates of similarities given above are extremely good estimators of the stimulus relationships in any study. Using these population estimates also has the virtues of being objective and neutral regarding the comparison between prototype and exemplar theory.

Pending further exploration, the results from the five dot-distortion studies just described and the results from the present experiment join the evidence from other category-learning tasks that raise concern about exemplar theory and exemplar models. For example, Smith, Murray, and Minda (1997) found that the exemplar model failed qualitatively to capture the performance profiles of half the participants in category-learning tasks. Smith and Minda (1998) found that the exemplar model failed qualitatively to fit the performance of whole samples early in category learning. Minda and Smith (2001) found that exemplar models always fit more poorly than comparable models across a wide range of categories varying in size, structure, and stimulus dimensionality. Minda and Smith (2002) found that exemplar models fit more poorly than the prototype model even for the 5–4 category structure that was most influential in motivating and sustaining exemplar theory (see also Smith & Minda, 2000).

Responding to these problematic findings, exemplar theorists have considered whether exemplar models can still succeed if they are made more complex and mathematically powerful. Most relevant here, they have proceeded by adding to the model the parameter gamma (e.g., Nosofsky & Johansen, 2000). For example, Nosofsky and Johansen (2000) argued that the exemplar model could explain Smith and Minda's (1998) problematic performance profiles if it was supplemented by gamma. Leaving aside the theoretical concerns about gamma that were already discussed, the present results are important because they show that the gamma model does not fit dot-distortion data as well as the prototype model does, even though the gamma model has more parameters and mathematical power, and even though the gamma model was thought to be particularly apt when fitting individual-participant data. The power, the parameters, and the individual-participant modeling are less important here, though, because the gamma model's failure is qualitative. Because its comparisons are inherently based in exemplars, its similarity and typicality gradients are too flat to capture what participants really do.

In short, the observed large prototype-enhancement effects and steep typicality gradients suggest strongly that participants are not referring to-be-categorized items to the shell of training exemplars as they make category decisions. Rather, the shape of the performance profiles suggests that they are referring to-be-categorized items to some representation that lies close to the center of the category. They are probably using prototypes.

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.

Hintzman, D. L. (1986). Schema abstraction in a multiple trace memory model. *Psychological Review*, *93*, 411–421.

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

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

Homa, D., Dunbar, S., & Nohre, L. (1991). Instance frequency, categorization and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 444–458.

Homa, D., Rhoads, D., & Chambliss, D. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 11–23.

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

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996, September 6). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402.

Knowlton, B. J., & Squire, L. R. (1993, December 10). 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 and Memory*, *1*, 106–120.

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, and Cognition*, *27*, 775–799.

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 275–292.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*, 375–402.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition memory 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.

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

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*, 304–308.

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998a). Contrasting cortical activity associated with category memory and recognition memory. *Learning and Memory*, *5*, 420–428.

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998b). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Science*, *95*, 747–750.

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, J. D., & Minda, J. P. (1998). Prototypes in the mist: the early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1430.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 3–27.

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, and Cognition*, *27*, 984–1002.

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

Squire, L. R., & Knowlton, B. J. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Sciences, USA*, *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, USA*, *93*, 13515–13522.

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

Received April 23, 2001
Revision received February 11, 2002
Accepted February 12, 2002 ■