Research Article

EXEMPLAR THEORY'S PREDICTED TYPICALITY GRADIENT CAN BE TESTED AND DISCONFIRMED

J. David Smith

University at Buffalo, State University of New York

Abstract—*One of exemplar theory's central predictions concerns the shape of typicality gradients. The typicality gradient it predicts is a consequence of its exemplar-based comparisons and appears no matter how the theory is evaluated. However, this predicted typicality gradient does not fit the empirical typicality gradients obtained in an influential version of the dot-distortion category task, and this is true even when the exemplar model is made more flexible and mathematically powerful. Thus, exemplar theory is disconfirmed in this domain of categorization. In contrast, prototype theories are consistent with the empirically obtained typicality gradients.*

The exemplar theory of categorization has a distinguished history (Estes, 1986; Medin & Schaffer, 1978; Nosofsky, 1987). It holds that people store category members as individuated memory representations and then judge whether novel items belong to the category by assessing their similarity to these representations (exemplars) in memory. Since 1978, exemplar theory has gained influence and broadened its reach by motivating models of, for example, recognition memory (Nosofsky, 1991), spoken-word recognition (Goldinger, 1998), and skill development (Palmeri, 1997). As the theory gains reach and influence, ensuring that it is an appropriate psychological description of human cognition becomes critical.

There is concern about exemplar theory's appropriateness. The exemplar model does not predict the performance of half the participants in category-learning tasks (Smith, Murray, & Minda, 1997), of whole samples early in category learning (Smith & Minda, 1998), or of participants with and without amnesia (Smith & Minda, 2001). Moreover, exemplar models fit more poorly than comparable models do when participants learn categories varying in size, structure, and stimulus dimensionality, including the categories that originally motivated exemplar theory (Minda & Smith, 2001, 2002). This article continues this reevaluation of exemplar theory (see also Nosofsky & Johansen, 2000; Nosofsky & Zaki, 1998; Smith & Minda, 2000).

My approach is to characterize the shape of the typicality gradient that exemplar theory predicts in a theoretically important version of the dot-distortion category task (Knowlton & Squire, 1993; Nosofsky & Zaki, 1998; Palmeri & Flanery, 1999; Reber, Stark, & Squire, 1998a, 1998b; Smith & Minda, 2001, 2002). The article shows that this predicted typicality gradient appears no matter how exemplar theory is evaluated and that the shape of this gradient is a natural consequence of the theory's exemplar-based processes. The article shows that exemplar theory's prediction is disconfirmed when it is tested against data from the dot-distortion task, even when the exemplar model is made more flexible and mathematically powerful. This demonstration has implications for judging the appropriateness of exemplar theory as a psychological description of categorization. In contrast, prototype theory is consistent with the empirically obtained typicality gradients.

Address correspondence to David Smith, Department of Psychology, State University of New York at Buffalo, Buffalo, NY 14260; e-mail: psysmith@acsu.buffalo.edu.

METHOD

Stimulus Materials

The stimuli used in dot-distortion category tasks are created with an established method that generates families of dot patterns from prototypes. This method originated with Posner, Goldsmith, and Welton (1967) and has been used extensively (e.g., Blair & Homa, 2001; Homa, Rhoads, & Chambliss, 1979; Homa, Sterling, & Trepel, 1981). In this method, nine points are randomly selected from within the central 30×30 area of a 50×50 grid. These nine dots are a prototype. Then low-level and high-level distortions of the prototype are produced by displacing the prototype's nine dots slightly or substantially, respectively. The prototype and its distortions define a dot-pattern category. In addition to the prototype and distortions of it, experiments usually include random nine-dot patterns that do not belong to the category and that participants should say do not belong. These random items are high-level distortions of unrelated prototypes. The details of these stimulus materials are given in Posner et al. (1967) and Smith and Minda (2001).

Category Task

The task considered here is an influential version of the dot-distortion task in which participants see 40 members of a dot-pattern category (each a high-level distortion of the prototype) and then endorse (or not) previously unseen probe items as belonging to that category. These probe items are copies of the originating prototype (P), low-level distortions of it (L), new high-level distortions of it (H), and random items (R) outside the category. The crucial data are the average levels of endorsement for each probe-item type.

Categorization Models

I fit categorization models to these endorsement profiles in the usual way. The models received as inputs the psychological distance between each probe-item type (P, L, H, and R) and the model's assumed category representation (prototype or exemplars). I took psychological distance to be ln(1 + the average Pythagorean distance that corresponding dots were moved between patterns of two types). (The addition of 1 ensured that 0 Pythagorean distance corresponded to 0 logarithmic distance.) This objective distance measure correlates beyond $r = .98$ with participants' ratings of psychological distance (Posner et al., 1967; Smith & Minda, 2001, 2002). These logarithmic distances were estimated by randomly sampling 1 million tokens of each pair type from dot-distortion space. For the exemplar model (with high-level distortions the category representation), the resulting average distances were 1.761 (P-H), 1.866 (L-H), 2.098 (H-H), and 2.894 (R-H). For the prototype model (with the prototype the category representation), the distances were 0.000 (P-P), 1.093 (L-P), 1.761 (H-P), and 2.850 (R-P). It was reasonable to use these average values in studying the operating characteristics of the models given the stability of the distortion-generating algorithm in Posner et al., given that participants see (and presumably store) a large, representative group of 40 training exemplars, and given that exemplar theorists have used average distances for this purpose (Nosofsky & Zaki, 1998).

These measures of psychological distance were inverted mathematically into measures of psychological similarity, so that more distance became less similarity or the reverse. I followed exemplar theory by taking similarity to be an exponentially decaying function of distance (d), with distance scaled by a free sensitivity parameter (c). Thus, the similarity (η) between a transfer item type (i) and the high-level distortions (h) or the category prototype (p) was, respectively,

$$\eta_{ih} = e^{-cd_{ih}} \quad \text{or} \quad \eta_{ip} = e^{-cd_{ip}}.$$

A choice rule received these similarities as inputs, converting them into endorsement levels ranging from 0 to 100% and yielding the level of category endorsement the model predicted for each probe-item type. These levels were found by dividing psychological similarity (η) by the sum of similarity and a free criterion parameter (k). The use of the proportionalizing quantity k as a free parameter followed recent applications of exemplar theory (Nosofsky & Zaki, 1998).

I evaluated the following three choice rules:

$$\text{(a)} \quad \frac{\eta_{ih}}{\eta_{ih} + k}$$

$$\text{(b)} \quad \frac{\eta_{ih}{}^{\gamma}}{\eta_{ih}{}^{\gamma} + k^{\gamma}}$$

$$\text{(c)} \quad \frac{\eta_{ip}}{\eta_{ip} + k}$$

For the exemplar model's choice rule (a), the relevant similarity was between a probe-item type (i) and the high-level distortions (h) seen during training. For the gamma model's choice rule (b), a mathematically powerful version of the exemplar model's choice rule, the same similarity applied, but the quantities in the choice rule could be raised to any power gamma that improved fit. The gamma parameter was this model's third free parameter (with sensitivity and criterion). For the prototype model's choice rule (c), the relevant similarity was between a probe-item type (i) and the prototype (p). The clear similarities between the prototype model and the exemplar model in their similarity calculations and choice rules make this pair of models balanced and appropriate for comparing the predictions of prototype and exemplar theory. (Ashby & Maddox, 1993, also discussed the importance of comparing models of categorization that are balanced and equivalently complex.)

Standard hill-climbing methods were used to find the parameter settings that allowed each model to recover best a categorization performance profile. These methods and the modeling procedures followed here are discussed more fully in Smith and Minda (1998, 2001).

I now sample in several ways the exemplar model's behavior and draw the shape of its predicted typicality gradient. This shape is not a result of the specific exemplar model evaluated here. This shape reflects the geometry of exemplar theory's exemplar-based comparison processes.

RESULTS

The Exemplar Model's Typicality Gradient

As a first approach toward drawing the shape of the exemplar model's typicality gradient, I sampled random configurations of the exemplar model (i.e., random values of its sensitivity and criterion parameters) and for each found the level of category endorsement predicted by the model for P, L, H, and R items. Performance profiles were saved and catalogued if they produced performance on the P, L, H, and R items in the range of 50% to 100%, 40% to 90%, 30% to 80%, and 20% to 70%, respectively. This let me consider only the performance profiles that were possibly human. Profiles were potentially included that contained P-L advantages (i.e., the endorsement advantage of P items over L items), L-H advantages, and H-R advantages of 60% down to −40%. I saved and catalogued 5,000 performance profiles.

Curve 1 in Figure 1 shows the averaged performance gradient of the exemplar model over the 5,000 profiles. This typicality gradient is one of exemplar theory's central predictions. A tiny P-L advantage yields to a larger L-H advantage and then a much larger H-R advantage. Conversely, the gradient flattens moving toward more typical item types. This shape is not an artifact of averaging many profiles with different characters. The average P-L advantage was 1.2% (SD = 0.9)—all the P-L advantages were small. Even this statistic understates the consistent shape of the model's performance gradient. The sizes of the three performance advantages (P-L, L-H, H-R) were highly correlated (minimum r = .95), so that the three values were yoked in a ratio of about 1 to 2 to 8 (precisely, 1.00 to 2.27, SD =

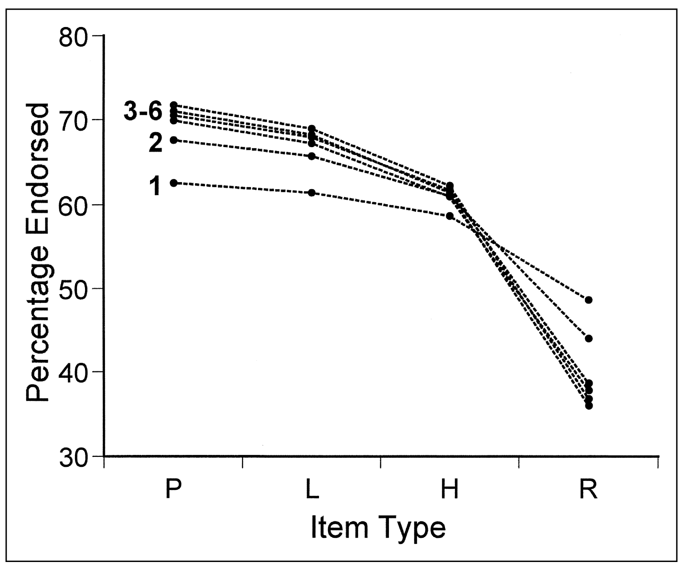

Fig. 1. The level of category endorsement predicted by the exemplar model for prototypes (P), low-level distortions (L), high-level distortions (H), and random items outside the dot-pattern category (R). The text describes the production of Curves 1 through 6.

Table 1. *Predicted mean category endorsements by categorization models for four item types, with accompanying derived measures*

Categorization model	Category endorsement				Advantage			Advantage ratio	
	P	L	H	R	P − L	L − H	H – R	(L − H)/(P – L)	(H − R)/(P − L)
1. Exemplar	62.6 (8.8)	61.4 (8.7)	58.6 (8.8)	48.6 (12.1)	1.2 (0.9)	2.8 (2.2)	9.9 (8.0)	2.27 (0.16)	8.03 (0.99)
2. Exemplar	67.6 (8.6)	65.7 (8.2)	61.0 (8.0)	44.0 (13.6)	2.0 (1.3)	4.7 (3.1)	17.0 (11.1)	2.35 (0.22)	8.60 (1.32)
3. Exemplar	70.5 (4.0)	67.9 (3.8)	61.7 (3.5)	38.6 (4.9)	2.6 (0.5)	6.2 (1.2)	23.1 (4.7)	2.38 (0.08)	8.83 (0.62)
4. Exemplar	69.9 (3.9)	67.2 (3.7)	60.9 (3.4)	37.8 (4.9)	2.6 (0.5)	6.3 (1.2)	23.1 (4.6)	2.37 (0.08)	8.72 (0.58)
5. Gamma	71.7 (4.5)	68.9 (4.2)	62.2 (3.7)	36.8 (5.6)	2.8 (0.5)	6.8 (1.5)	25.3 (5.8)	2.41 (0.10)	9.02 (0.78)
6. Gamma	71.0 (4.4)	68.2 (4.1)	61.4 (3.6)	36.0 (5.5)	2.8 (0.5)	6.8 (1.5)	25.3 (5.8)	2.41 (0.10)	9.02 (0.78)
7. Exemplar	70.9 (6.1)	68.5 (5.7)	62.8 (4.4)	40.5 (3.3)	2.4 (0.5)	5.8 (1.4)	22.3 (6.8)	2.39 (0.11)	9.11 (1.03)
8. Gamma	72.4 (7.3)	69.9 (6.8)	63.5 (5.2)	38.5 (4.9)	2.6 (0.5)	6.3 (1.8)	25.1 (9.4)	2.44 (0.18)	9.51 (1.42)
9. Prototype	78.8 (5.6)	65.9 (4.4)	56.2 (3.6)	39.6 (5.0)	12.8 (2.3)	9.7 (2.4)	16.6 (4.3)	0.75 (0.08)	1.28 (0.18)
10. Prototype	79.4 (8.6)	67.6 (7.2)	57.9 (4.1)	40.5 (4.3)	11.8 (1.6)	9.6 (3.3)	17.4 (7.4)	0.80 (0.18)	1.43 (0.43)

Note. Standard deviations are in parentheses. The text describes the formal procedures that were followed in constructing each row of the table. P = prototypes; L = low-level distortions; H = high-level distortions; R = random items outside the dot-pattern category.

0.16, to 8.03, *SD* = 0.99). This shows that exemplar theory can stretch or compress the overall performance gradient but cannot change that gradient's basic shape. The progressive flattening is always preserved. Row 1 in Table 1 summarizes these results.

One can visualize why this typicality gradient follows from the geometry of exemplar-based comparisons. Consider what occurs in Figure 2 as a hypothetical to-be-categorized item moves in from the random region of psychological space (e.g., Position 8), then through the region of psychological space occupied by the shell of training distortions (Position 3), and then on toward the prototype (Position 1). At each step, one can ask how the item's closeness or belongingness to the exemplar shell changes. At first (Position 8), when the item is far away, it will gain strongly in closeness to the exemplar shell because at this distance the shell will nearly be a point source to the item. Closer in to the category system (Position 5), 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 representation, the item will now gain more slowly in belongingness to their category. Finally, as the to-be-categorized item moves inside the exemplar shell (Position 2), it will hardly close on the exemplar shell at all. 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 exemplars 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 hardly increase in belongingness to the category they represent. The exemplar model simply translates these inevitable geometric facts into a performance gradient.

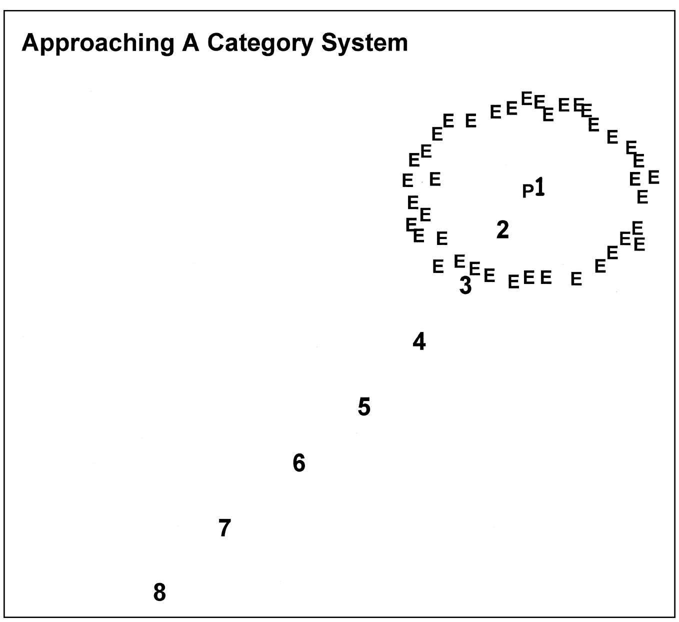

Fig. 2. Schematic drawing of a series of hypothetical dot distortions (8 to 1) that approach the psychological space of a category system containing a central prototype (P) and a shell of training exemplars (E).

Even so, a residual concern is that I inferred the model's operating characteristics based on its average behavior. An important purpose of models is to take on specific configurations that fit empirical data sets. For this purpose, the critical thing is how the model behaves as it tries to fit (i.e., predict or re-create) performance profiles. Accordingly, I generated performance profiles that included all those that are humanly possible. That is, I let P, L, H, and R performance vary systematically from 50 to 100%, 40 to 90%, 30 to 80%, and 20 to 70%, respectively, in 5% increments, and I let the exemplar model use its free parameters to fit each of these 14,641 performance profiles as best it could.

Curve 2 in Figure 1 shows the averaged performance gradient of the exemplar model as it fit these target data patterns. The exemplar model produced the expected typicality gradient as it fit data (Table 1, row 2), just as it did on average (Table 1, row 1).

A concern about the preceding analysis is that it sampled data patterns so inclusively that many were unlikely human data patterns. It might be better to ask how the model behaves when it fits plausibly human data patterns. Accordingly, I produced a corpus of hypothetical performance profiles based on the endorsement levels participants generally show for P, L, H, and R probe items. I built this corpus as follows. Using seven existing performance profiles (six summarized in Table 1, Smith & Minda, 2001, p. 994; one taken from Smith &

Minda, 2002), I found participants' average endorsement levels for the four probe-item types and the variability of these levels across studies. I did this only to establish a general baseline for humans' performance—no methodological or theoretical screen governed the choice of these data sets. The four average performance levels for P, L, H, and R items, respectively, were 77.7% (*SD* = 7.3), 65.8% (*SD* = 8.5), 58.4% (*SD* = 5.4), and 38.1% (*SD* = 5.6). From this baseline, I created 5,000 performance profiles that each represented a Gaussian disturbance of the four means by their standard deviations.

Curve 3 in Figure 1 shows the average result when the exemplar model fit these 5,000 performance profiles as best it could. The exemplar model still produced its characteristic typicality gradient (Table 1, row 3).

The previous analysis was friendly to encouraging the exemplar model's flexibility. The 5,000 data patterns tended to reproduce the steep—not flattening—typicality gradient that humans generally do produce (12% P-L advantage, etc.). This created a general pressure for steep—not flattening—predicted performance gradients as the model fit the data. To be sure that this pressure was strong and sufficient, I isolated from the 5,000 data patterns just described the 500 that had the largest prototype effects and that should exert the ultimate pressure on the model to produce steep performance gradients.

Curve 4 in Figure 1 shows the result when the exemplar model fit these 500 performance profiles. The predicted performance gradient (Table 1, row 4) was the same as in the preceding analyses. The small P-L advantage (2.6%, *SD* = 0.5) in this case is surprising and theoretically important given that the model ideally needed to produce P-L advantages 12 times as large (31.0%, *SD* = 4.4). Clearly, it is difficult for the exemplar model to produce the steep typicality gradient that is required.

The Gamma Model's Typicality Gradient

In analyzing the relationships among different models of categorization, Ashby and Maddox (1993; see also Maddox & Ashby, 1993) introduced the parameter gamma. Exemplar theory has come to rely heavily on gamma to improve the exemplar model's fit when the traditional model fails (Nosofsky & Johansen, 2000). Gamma can sometimes allow a model to predict steeper typicality gradients and larger prototype-enhancement effects than it would otherwise. Accordingly, I examined how this additional parameter would affect the model's behavior in this case. Including gamma was somewhat problematic because it granted the exemplar model three free parameters as it tried to recover only four data points—possibly making the model self-confirming. Moreover, adding gamma can amount to adding a prototype process to the exemplar model, making the gamma model theoretically ambiguous and misleading psychologically (see Smith & Minda, 1998). Nonetheless, it was important to consider the gamma model because exemplar theorists favor it strongly and because the results it produces are illuminating.

I let the gamma model fit as best it could the 5,000 performance profiles that were created (as described previously) through Gaussian disturbances of the average P, L, H, and R endorsement levels that humans produced in seven studies. Curve 5 in Figure 1 (also Table 1, row 5) shows that even the mathematically powerful gamma model produced the typicality gradient expected by exemplar theory.

As I did for the exemplar model, I followed up this analysis by isolating the 500 performance profiles that had the largest prototype effects. These profiles should exert ultimately strong pressure for the gamma model to predict a differently shaped gradient that fits the data patterns better. Curve 6 in Figure 1 (also Table 1, row 6) shows that the gamma model still traced the characteristic shape of exemplar processing.

Testing Exemplar Theory's Prediction

To test exemplar theory's prediction against real data, I let the exemplar and gamma models fit the five existing data sets whose methods accorded with the structure of the exemplar model used here. (Thus, I excluded the data sets from Knowlton & Squire, 1993, Experiment 2, and Palmeri & Flanery, 1999, though including them would make no difference to this analysis.) Curve O in Figure 3 shows the composite of the five observed profiles. Curves E and G show the average performance profiles predicted by the exemplar and gamma models when they fit as best they could the five data sets individually (also Table 1, rows 7 and 8). Exemplar theory predicts the wrong qualitative shape of typicality gradient for what participants show in a theoretically important and common version of the dot-distortion paradigm.

The Prototype Model's Typicality Gradient

Prototype theory also makes a prediction regarding performance in this paradigm. To show this, I fit a prototype model to the 5,000 performance profiles that the exemplar and gamma models fit in producing rows 3 and 5 in Table 1. The resulting typicality gradient (see Table 1, row 9)—as indexed by the levels of endorsement for P, L, H, and R items; or the P-L, L-H, and H-R advantages; or the ratios of these advantages—was very different from exemplar theory's gradient. There was no overlap between the distributions of ratios predicted by the prototype and exemplar models. To the contrary, there was a gap of about 6 *SD*s between the maximum value of the small ratios that the prototype model predicted and the minimum value of the large

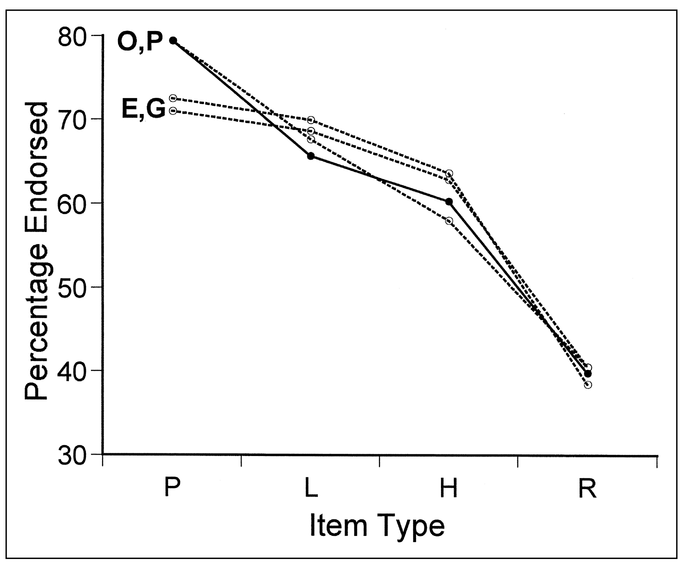

Fig. 3. The composite observed performance profile (O) from five dot-distortion studies whose method accorded with the structure of the models used in the present study. Also shown is the average of the five best-fitting predicted profiles when the exemplar model (E), the gamma model (G), and the prototype model (P) fit these five data sets individually.

ratios that the exemplar model predicted. This difference between the models may be the largest and most clearly qualitative difference between the predictions of prototype and exemplar theory that has ever been shown.

To test the prototype model's prediction against real data, I let it fit the same five data sets that the exemplar and gamma models fit to produce rows 7 and 8 in Table 1. The model's behavior is summarized in row 10. The characteristic shape of prototype processing appears again. Curve P in Figure 3 shows the average performance profile predicted by the prototype model as it fit the five performance profiles individually. Prototype theory predicts the correctly shaped typicality gradient.

SUMMARY AND IMPLICATIONS

This article establishes a central prediction of exemplar theory that would apply in other domains of categorization and in other areas where exemplar theory has been favored (e.g., memory, speech recognition, and skill-automaticity). In fact, exemplar theory's typicality gradient has a very wide-ranging intuition behind it. The theory's basic idea is that humans represent a cloud of individuated exemplars that are spread out in psychological space and that are the standard of comparison for new tokens impinging on the system. This implies that all impinging tokens that fall within the representational space of the exemplars will be like some exemplars and unlike others, and therefore about equally good members of the category the exemplars represent. If therapists store exemplars of a clinical syndrome like schizophrenia and compare new cases with these, there will be no strong typicality effects inside the space of the exemplars and no cases of schizophrenia that seem prototypical, because the cases that present will all be like some former cases and unlike others.[1] If art historians store individuated exemplars of Monet, then new Monets will seem to be about equally typical Monets, and there will be no steep typicality gradient leading up to a Monet stylistic prototype. Similarly, all members of exemplar-based phoneme categories or speaker-token categories would be about equally good. This is an interesting idea whether or not it is a true description of human cognition. The point here is that it is a basic prediction of exemplar theory that this article shows can be tested.

Moreover, this basic prediction is at the heart of exemplar theory's intuitive, conceptual, and geometric basis. This is important to say at a time when categorization models are becoming very complex and when the goal of fitting data best sometimes trumps the goal of giving data the best psychological description. Even if one could build an exemplar model that was so mathematically powerful that it could bend its typicality gradient in ways that violated this basic idea, it would not be usefully or sensibly or psychologically an exemplar model anymore because it would have forfeited the spirit, the intuition, the concept, and the geometry of exemplar theory.

Prototype theory makes a different prediction that is equally wide-ranging. Its basic idea is that humans collapse the cloud of exemplars into the representational point at their center, and that this prototype becomes the standard of comparison for new tokens impinging on the system. Tokens falling within the representational space of the exemplars could be indefinitely similar to the prototype and belong indefinitely strongly within the category it represents, because the category representation is a single, approachable point in the space, not a cloud of points in the space. This is the explanation of the strong prototype effects and the steep typicality gradients shown in rows 9 and 10 of Table 1 and Curve P in Figure 3. If therapists abstract the prototype of schizophrenia and compare new cases with this, there will be strong typicality effects and textbook cases. If art historians represent a Monet prototype, there will be marginal and beautiful examples of his style. This idea also raises interesting questions about human cognition in various domains, but the point here, too, is that this is a testable prediction of prototype theory.

When the predictions of prototype and exemplar theories were tested in the domain of dot-pattern categorization, prototype theory's prediction was clearly supported; exemplar theory's prediction clearly not. Exemplar theory's failure is important because this is one of the most venerable and influential categorization paradigms. Its failure is also important because exemplar theory failed even given the mathematical support of the parameter gamma. Gamma has ameliorated the exemplar model's failure in some cases. Here, given an appropriate test that evaluated a sufficiently broad range of typicalities, gamma failed to do so. Thus, in this case I was able to distinguish gamma-supported processing from prototype-based processing, and rule out that the former was occurring. The failure of the exemplar and gamma models here has potentially profound implications for judging the appropriateness of exemplar theory as a psychological description of categorization.

However, there is no guarantee that this failure will extend to all areas of human cognition. I do not know how psychotherapists represent their diagnostic categories or art historians their styles, and the shape of their typicality gradients could reveal either exemplar-based or prototype-based processing. An exciting possibility suggested by the present article is that one may be able to draw the shapes of the underlying typicality gradients in various domains, and infer from these the representational basis for decision making and classification within those domains. If so, then it will be a victory for exemplar theory and prototype theory—no matter the outcome. It will mean that the intuitions of both theories were faithfully translated into geometries, models, and testable predictions that allowed the theories to be differentially supported. Theory, formal approaches, and experimentation will have created the synergy that is cognitive science at its best.

Acknowledgments—The preparation of this article was supported by the National Institutes of Health (HD-38051).

1. There will be typicality effects outside the exemplar space as tokens move far away from the exemplars in memory (i.e., Positions 3–8 in Fig. 2). There may be some typicality effects inside the exemplar space of the category if one's exemplar experience has included a large proportion of highly typical exemplars. The degree of flattening of exemplar theory's typicality gradient can be affected by the distribution of training exemplars, but there will always be flattening of that gradient relative to the gradient predicted by prototype theory.

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.

Blair, M., & Homa, D. (2001). Expanding the search for a linear separability constraint on category learning. *Memory & Cognition*, *29*, 1153–1164.

Estes, W.K. (1986). Array models for category learning. *Cognitive Psychology*, *18*, 500–549.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

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., & Squire, L.R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.
Maddox, W.T., & Ashby, G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*, 49–70.
Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.
Minda, J.P., & Smith, J.D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, 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. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 87–108.
Nosofsky, R.M. (1991). Tests of an exemplar model for relating perceptual categorization and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 3–27.
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. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 324–354.
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*, 525–530.
Posner, M.I., Goldsmith, R., & Welton, K.E. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology*, *73*, 28–38.
Reber, P.J., Stark, C.E.L., & Squire, L. (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 Sciences, USA*, *95*, 747–750.
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., & Minda, J.P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 800–811.
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.

(Received 7/30/01; Revision accepted 11/6/01)