Categorization, Typicality, and Shape Similarity

Matthew A. Kurbat and Edward E. Smith
Department of Psychology, University of Michigan
330 Packard St.
Ann Arbor, MI 48104
kurbat@cog.psych.lsa.umich.edu

Douglas L. Medin
Department of Psychology
Northwestern University
2029 Sheridan Rd.
Evanston, IL 60208

Abstract

This work examines the contribution of shape features to subjects' judgments of typicality for visual categories. Shape was found to make a strong contribution to typicality, as evidenced by the strong correlation between results on pictures and those on silhouettes of the same pictures. Also, different measures of the contribution of shape template overlap, compactness, and number of parts - were shown to capture different aspects of that contribution. As one of the fundamental problems in category research is to determine the features used in categorization (e.g., Medin, 1989), the current work is important because it makes progress on this problem.

Introduction

The perceived world includes an indefinite number of discriminable objects. While we could treat each object as unique, often we do not - we categorize. Coding by category is fundamental to mental life because it greatly reduces demands on perception, reasoning, and memory (e.g., Medin, 1989; Smith, 1990).

One well-documented finding in categorization research is that not all members of categories are created equal: people perceive some members to be better or worse examples of a category than others (e.g., Mervis and Rosch, 1981; Smith and Medin, 1981). For example, people tend to think that a robin is a more typical bird than a chicken. Typicality appears to be a major determinant of the organization of memory, as evidenced by its ability to predict results in a variety of tasks using semantic and visual categories such as speed and order of learning category members, and speed of verifying them as instances of a category (e.g., Smith and Medin, 1981). Furthermore, variations in typicality cannot be explained solely (or largely) by familiarity (e.g., Malt and Smith, 1982).

A general interpretation of these findings is that the typicality of an instance is a measure of its similarity to its category, and categorization amounts to determining that an item is sufficiently similar to the target category (Rosch and Mervis, 1975; Smith, 1990). If typicality is really based on similarity, then we ought to be able to predict typicality as follows (Smith, 1990): (1) select a domain of instances; (2) estimate features of the instances by subjects' listings; (3) compute similarity of each exemplar to its category by using some explicit rule (such as Tversky's (1977) contrast model); (4) see if this similarity estimate predicts typicality.

While this method is clearly useful for semantic categories, its relevance to visual categories is less clear. Feature listings include few (if any) shape features; yet objects are often visually categorizable from their shape alone (Biederman, 1987; Rosch, Mervis, Gray, Johnson, and Boyes-Braem, 1976), or even from silhouettes alone (e.g., Rosch et al., 1976). So, while shape features are clearly important to visual categorization, categorization models based on listed features do not explain why. Thus it would be interesting to devise a model that accounts for the contributions of shape features to perceived similarity, typicality, and visual categorization.

It is not trivial to devise such a model, but one area of research that is useful for this purpose is object recognition. Researchers in object recognition have developed several classes of models, the most popular being template, feature, and structural description models (e.g., Pinker, 1984). Structural description models represent objects by their parts and relations between parts. While such structural descriptions are thought to be more psychologically veridical than feature or template representations (e.g., Pinker, 1984), it is also more difficult to devise a structural-description theory that is both broad in the scope of objects it can recognize, and specific about the means by which objects are represented and recognized (e.g., how objects are divided into parts, how parts are represented). The RBC/JIM model (Biederman, 1987; Hummel and Biederman, 1992) is probably the best structural description approach to date, but even it has severe limits on the generality of mechanisms for dividing objects into parts and representing parts. For parsing, RBC/JIM is limited for a number of reasons: its parsing rule is based on a misinterpretation of Hoffman and Richards (1984), it underdetermines parsing, it is overly sensitive to noise, it parses hierarchical objects incorrectly, etc.; for representing parts, RBC/JIM is also limited for numerous reasons: some geon features are absent from smooth objects or silhouettes, empirical evidence for some geon features might actually be evidence for lower level contour grouping mechanisms, algorithms for generating axes of symmetry are incompatible with RBC/JIM's representations of axes, etc. (Kurbat, 1994, discusses these and related points). Therefore we focus on template and feature models below.

While templates may be less psychologically veridical than structural descriptions, their recent resuscitation (e.g., Ullman, 1989) is likely due in part to their being simpler and easier to implement. Template models determine the similarity between two shapes in two steps (e.g., Ullman, 1989). The first step is an alignment step: shapes may differ in orientation, size, and location, so reorientation, resizing, and translation of the shapes may be necessary to align them. There are various methods for performing these transformations. For example, for reorientation (in the picture plane), one may first fit a line to the boundary of each object minimizing the sum of squared distances from the boundary, and then reorient all objects so each line has the same orientation (e.g., horizontal); for size, objects may be scaled to have the same area; for location, objects may be translated to have the same center of area (for further discussion, see Ullman, 1989). It is generally also necessary to compensate for changes in depth orientation of objects (e.g., a side view of a dog vs. a front view), but in the work discussed below we are only concerned with one possible depth orientation (i.e., side views of objects). The second step in template matching is the matching step. One simple method for this step is to count the number of pixels common to both superimposed objects ("intersection"), then count the number found in one or the other or both ("union"). Dividing intersection by union yields a shape similarity score from 0 to 100 - the higher the score, the greater the similarity in shape between the two shapes matches.

Feature models of object recognition are much like those in categorization, because they represent objects in terms of features and recognize them via featural similarity. Designers of such models must also confront the categorization issue of "what are the features?" (e.g., Murphy and Medin, 1985). Numerous candidate shape features exist (e.g., Ballard and Brown, 1982), but for present purposes we will concentrate on two that capture complementary aspects of the variability between shapes. The first is the notion of *compactness* (e.g., Zusne, 1970): the ratio of the square of the perimeter of an object to its area. Shapes with low compactness scores are compact (e.g., a circle), and those with high scores are not compact (e.g., a pencil). A second shape feature is the number of concavities in an exemplar's silhouette. Shape recognition research suggests we divide shapes into parts at concavities (Hoffman and Richards, 1984), so a count of the number of concavities is one measure of the number of parts produced by dividing at concavities. This measure may be implemented on computer by fitting the boundary of a shape with cubic spline curves, and then counting the number of places where the splines become concave (i.e., curvature of the spline becomes negative). We do not claim that these two features sufficiently characterize shapes; we do suggest that these features may capture some psychologically important differences between shapes.

As basic level categories are thought to be distinguished primarily via differences in shape[1] (e.g., Rosch et al, 1976; Biederman, 1987), such categories ought to be a good domain for testing shape-based models of similarity and categorization, hence the template and featural measures just discussed (while there is some question as to what is to "count" as a basic level category (e.g., Mandler, Bauer, and McDonough, 1991), we avoid this question here for sake of simplicity and use categories termed basic level by Rosch et al, 1976). For this test, 25 basic level categories were used (from Estin, Smith, and Medin, 1994), and two or more instances were selected from each category for a total of 66 exemplars (two instances were selected from 18 of 25 categories, and more than two were selected from the remaining seven). Each of the 66 exemplars was compared via template match. The template matching method used was similar to the one described in the introduction - to correct for differences in size, objects were scaled by computer to have the same area; to correct for differences in planar location, objects were translated by computer to have the same center of area (i.e., same average x and y coordinates for black pixels comprising each digitized silhouette). The only difference between the method used here and the one described in the introduction was that correction for planar orientation was not used because all objects were rotated in advance to have a horizontal orientation in the picture plane. This method was used to match each exemplar to every other exemplar to find the single best matching exemplar; each exemplaf was classified as belonging to the category of this best matching example. For example, if car exemplar 1 matched another car exemplar better than it matched all non-car exemplars, then it would be classified (correctly) as a car - otherwise, it would be classified incorrectly. Of the 66 total exemplars, 60 were classified correctly, for an accuracy rate of over 90%; of the 25 categories, every exemplar in 20 of them was classified correctly. (We are also working on a feature-based model - using the feature types discussed above, and others - to perform this same task).

Given that these shape measures make fairly accurate predictions about basic level categorization, it is interesting to examine how well they predict typicality - the issue with which we began. It seems unlikely that shape would predict typicality at all levels of abstraction - for example, relatively superordinate categories like 'mammal' and 'beverage' are

[1] There has been a controversy over whether shape is a necessary requirement for basic level categories. For example, Murphy (1991) presented results that he used to argue against the necessity of shape. However, the interpretation of the results presented by Murphy is less than clear - one problem is that categories used in Murphy's experiments had defining (non-shape) features at the basic level. This practice, which is common in research using artificial categories, undermines the idea that relevant structural properties of natural categories have been incorporated into Murphy's artificial materials (e.g., Lassaline, Wisniewski, and Medin, 1992). A more expanded version of the current paper will include further discussion of this issue (including, e.g., the comments of Tversky and Hemenway, 1991).

distinguished primarily by non-perceptual properties (Rosch et al., 1976). One interesting level at which to try these measures is again the basic level. This level is interesting because subordinate categories that are atypical within their basic level categories are more distinctive in their physical appearance than typical subordinates (Murphy and Brownell, 1985). In other words, the more distinctive something is in physical appearance, relative to other members of the category, the less typical it is. While Murphy and Brownell's results are limited by the fact that 'distinctiveness of appearance' is ambiguous, one obvious component of appearance is shape. So it seems natural to suggest that distinctiveness of appearance may be in part due to distinctiveness in shape, and that atypical exemplars may have more distinctive shapes than typical exemplars.

Experiment

This experiment had two purposes. The first was to compare subjects' typicality ratings of pictures with their typicality ratings of silhouettes of the same pictures. A high correlation between results with pictures and silhouettes would indicate a strong contribution of shape to judgments of typicality, because silhouettes include only shape information. The second purpose was to test different measures of shape similarity as predictors of typicality. As the shapes of members of different basic level categories may vary in different ways, attempts to capture these different sorts of variation may require different measures of shape.

Method

Subjects. The subjects were 138 undergraduates at the University of Michigan who participated as part of a course requirement.

Materials. The stimuli were black and white digitized pictures of exemplars of three basic level (according to Rosch et al, 1976) categories - birds, dogs, and fish. Pictures were side views of these animals chosen from wildlife and pet shop books. Within each of these three categories, subordinate categories were chosen to span the typicality range and to be relatively well-known to college students. For birds, 56 exemplars from 14 subordinate categories were used; for dogs, 52 exemplars from 17 subordinate categories were used; and for fish, 83 exemplars from 20 subordinate categories were used. Pictures were chosen to be from comparable viewpoints - all were side views. The pictures were digitized using an HP Scanjet II as black and white binary images (75 pixels per inch - each pixel either black or white). Pictures were normalized by computer to a constant size. Silhouettes were generated by computer for each picture. Silhouettes were created using an algorithm that first traced the boundary of each depicted object, and then made all pixels inside the boundary black.

Procedure and Design. Each subject generated data for either (1) birds and dogs, or (2) dogs and fish, or (3) fish and birds, with half the subjects in each of these three groups viewing pictures and half viewing silhouettes. An equal number of subjects was assigned to each of these six groups (for example, 23 subjects saw bird and dog pictures, 23 subjects saw dog and fish silhouettes, etc.). Each subject first gave typicality ratings for all instances of the first type of animal (presented in a random order), then for the second type of animal. Subjects were asked to rate how typical the pictured animal is of its basic level category (bird, dog, or fish) on a 1-7 scale, with 1 being very atypical and 7 being very typical.

Results and Discussion

Pictures versus silhouettes. Correlations between typicality ratings for pictures and silhouettes were .93, .89, and .96 for fish, dogs, and birds respectively ($p<.0001$ in all cases). So, shape information played a major role in ratings of typicality. This result is interesting in part because past work by Rosch et al (1976) showed that the differences in basic (not subordinate) level category were most correlated with differences in shape. The result is consistent with the suggestion of Murphy and Brownell (1985), that atypical subordinate categories are more distinctive in their physical appearance than typical subordinates, and our conjecture that distinctiveness of appearance may be in part due to distinctiveness in shape.

Several comments are in order on the generality of these results. On the one hand, the results should be qualified by the fact that the digitized images used in this experiment were black and white binary images (75 pixels per inch - each pixel either black or white), and so they lacked color and fine detail. Also, scaling the stimuli to a constant size led to a slight degradation of some pictures. Thus the presence of color, greater detail, and elimination of degradation due to scaling might have reduced the correlations. On the other hand, of the three basic level categories only birds have a great deal of color variation. Also, silhouettes exclude some shape information included in pictures, so silhouettes and pictures do not differ only in terms of non-shape information. Thus some of the picture-silhouette difference may be due to shape differences, and the contributions of shape to typicality may be even greater than indicated by our correlational results.

Shape measures versus typicality. Two versions of the template measure were used: exemplar and prototype (e.g., Smith and Medin, 1981). In the exemplar version, each pictured exemplar was compared via template match to all other instances of its basic level category; resulting scores were averaged to get the shape similarity score for that exemplar. In the prototype version, the subordinate category with the highest mean typicality rating was chosen as prototype for each of the three basic level categories (in all three cases, this subordinate category also had the highest or near highest score on the exemplar-based shape similarity measure just discussed). For example, 'labrador retriever' was the dog with the highest typicality rating, so it served as the dog prototype. Shape similarity scores were then determined for each nonprototype exemplar by comparing them via template match to each prototype exemplar. So, to again use dogs as an example, each nonprototype dog was compared against each labrador retriever exemplar using the template match. In both prototype and exemplar cases, average scores for each exemplar were then averaged within each subordinate category to get a mean score for that subordinate category. These shape similarity scores were then compared with typicality scores. For the exemplar version, shape similarity was not significantly correlated with typicality for pictures or silhouettes of fish, dogs, and birds (note that this is evidence against one particular exemplar model, not the class in general - see Smith and Medin, 1981; another sort of exemplar model worth investigating in this context is one that weights each exemplar by frequency of instantiation, following Barsalou, 1985). For the prototype version, template shape similarity correlated significantly with typicality for both bird pictures ($r=.70$, $p<.008$) and silhouettes ($r=.61$, $p<.03$) and for both fish pictures ($r=.65$, $p<.003$) and silhouettes ($r=.58$, $p<.009$), but not for dogs ($r=.35$ for pictures and .28 for silhouettes). Further improvements for birds were found by adding a parameter for normal orientation of birds. Some birds (i.e., penguins and owls) normally stand in a vertical orientation (and were so pictured), so the parameter value was set to 1 for these birds, but most birds stand in a more diagonal or horizontal orientation, so the parameter value was set to 0 for these birds. This parameter is sensible because most birds do not stand vertically, making those that do distinctive. Addition of this parameter increased correlations to $r=.87$ and $r=.85$ for bird pictures and silhouettes, respectively ($p<.002$ in both cases).

While the template measure yielded good results for birds and fish but not dogs, the opposite pattern occurred with the first of our feature-based shape measures, the compactness measure; so the two measures seem complementary. Specifically, correlations between compactness and typicality were not significant for birds and fish, but were significant for dog pictures (.57, $p<.02$) and silhouettes (.70, $p<.002$). The dog results make some intuitive sense: breeds with lower compactness scores like dachshunds and st. bernards were rated lower in typicality than breeds with more slender, elongated body parts (hence higher compactness scores), like retrievers.

A second feature measure used was number of parts in an exemplar's silhouette, as estimated by number of concavities. While this measure did not seem useful for birds and dogs, as their part structures remain essentially unchanged across subordinates, it did seem useful for fish, as the number of fins varies widely between subordinates. While number of parts per se did not correlate well with typicality for fish ($r=.08$ for pictures and $r=.03$ for silhouettes), a plot of parts versus typicality produced an asymmetrical inverted U-shaped function. Taking M to be the mean number of parts, the transformation $|\text{parts} - M|/\text{parts}$ is useful in this context. To justify the choice of numerator ($|\text{parts} - M|$), note that fish typically have 4-5 concavities on average, so those with fewer (e.g., eels) or more (e.g., catfish - whiskers create concavities) seem less typical. In other words, the numerator simply reflects degree of deviation of a given fish exemplar from the fish prototype, as measured in terms of number of parts. The choice of denominator was motivated by psychophysical studies of magnitude estimation. Specifically, perceptions of just noticeable differences in magnitude are commonly proportional to the magnitude, a relationship commonly known as *Weber's law* (e.g., Engen, 1972); so division by magnitude (number of parts) serves as a normalizing factor (one might argue that perhaps number of parts and mean number of parts should be averaged as the normalizing factor, but this linear transformation of the original normalizing factor would not change the correlation of the transformed predictor with typicality). In fact, the resulting transformed values correlated well with typicality of fish pictures ($r= -.74$, $p<.0003$) and silhouettes ($r= -.72$, $p<.003$). Finally, the transformed values correlated poorly with the template measure ($r= -.17$), so they seem to be capturing different aspects of shape variation.

General Discussion

In summary, shape made a strong contribution to the assessment of typicality for the pictures tested, as evidenced by the strong correlation between results on pictures and those on silhouettes. Also, we have several measures of the contribution of shape - template overlap, compactness, and number of parts - each of which appears to capture a different aspect of that contribution. Further, a more general test of the template shape measure using approximately two dozen basic level categories found that template-determined shape similarity was significantly correlated with typicality for most of the categories tested (Estin, Smith, and Medin, 1994). As one of the fundamental problems in category research is to determine the features used in categorization (e.g., Medin, 1989), the current work is important because it makes progress on this problem.

A number of directions for future research on this topic seem worth pursuing. First, it would be interesting to further develop the shape feature approach by trying additional feature measures, and ways to combine different measures. Second, despite the reasonably good predictions made by the template and feature models, they have clear limitations (e.g., Pinker, 1984). Thus one of our goals is to develop an improved structural description model that can account for the results presented here (further developments in shape feature measures might still be useful in that they could provide constraints on a structural description model). Third, we are interested in typicality because it can be used to estimate similarity which in turn can be used to predict category membership; but shape measures serve as direct measures of similarity, so shape similarity ought to predict categorization directly. Also, similarity as tapped by speeded categorization may differ from similarity as tapped by slower and perhaps more deliberative judgments of typicality (for a related point, see Barsalou, 1987). As we wish to account both for typicality and categorization, we are also examining shape model-based predictions of speed and accuracy of categorization. Our preliminary results suggest that the current shape measures predict speed and accuracy of categorization about as well as they predict typicality. Such a convergence of results is consistent with the claim that subjects did not determine typicality indirectly via subordinate categorization (e.g., by saying "that is a flamingo, and flamingoes are atypical"), because the categorization task does not allow this sort of slow, deliberative processing, and because the same measure predicts the results of both tasks. Fourth, we are also collecting ratings of familiarity and subordinate typicality to control for possible effects of these variables - we plan to assess contributions of these factors and different shape measures on typicality and subordinate categorization using multiple regression. Finally, as mentioned in the introduction, some current models for categorization of natural objects perform categorization via similarity as determined by listed features. As a way to examine the differences between verbal and visual categorization (and as another way to check if subjects are using background knowledge to perform typicality judgments), we plan to use both visual and verbal categorization paradigms, and use the feature-listing methods discussed in the introduction as well as our shape measures, to generate predictions of typicality. This would allow us to see the extent to which predictions based on feature-listing models correlate with those produced by the shape-based measures, and to examine relations between verbal and visual categorization.

References

Ballard, D., and Brown, C. (1982). *Computer Vision.* New York: Academic Press.

Barsalou, L. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **11**, 629-654.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, **94**, 115-147.

Engen, T. (1972). Psychophysics. In J.W. Kling and L.A. Riggs (Eds.), *Experimental Psychology*. New York: Holt, Rhinehart, and Winston.

Estin, P., Smith, E., and Medin, D. (1994). Typicality, shape similarity, and the basic-level advantage in categorization. Research in progress.

Hoffman, D., and Richards, W. (1984). Parts of recognition. *Cognition*, **18**, 65-96.

Hummel, J., and Biederman, I. (1993). Dynamic binding in a neural network for shape recognition. *Psychological Review*, **99**, 480-517.

Kurbat, M. (1994). Structural description theories of shape recognition. Manuscript submitted for publication.

Lassaline, M., Wisniewski, E., and Medin, D. (1992). Basic levels in artificial and natural categories: are all basic levels created equal? In B. Burns (Ed.), *Percepts, Concepts, and Categories*, Elvesier Science Publishers.

Malt, B.C., and Smith, E.E. (1982). The role of familiarity in determining typicality. *Memory & Cognition*, **10**, 69-75.

Malt, B.C., and Smith, E.E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, **23**, 250-269.

Mandler, J.M., Bauer, P.J., and McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, **23**, 263-298.

Medin, D.L. (1989). Concepts and conceptual structure. *American Psychologist*, **44**, 1469-1481.

Medin, D.L., and Schaeffer, M.M. (1978). A context theory of classification learning. *Psychological Review*, **85**, 207-238.

Mervis, C., and Rosch, E. (1981). Categorization of natural objects. In M.R. Rozenzweig and W.L. Porter (Eds.), *Annual Review of Psychology*, **32**. Palo Alto, CA: Annual Reviews.

Murphy, G.L. (1991). Parts in object concepts: experiments with artificial categories. *Memory and Cognition*, **19**, 423-438.

Murphy, G.L., and Brownell, H.H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **11**, 70-84.

Murphy, G.L., and Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289-316.

Pinker, S. (1984). Visual cognition: an introduction. *Cognition*, **18**, 1-64.

Rosch, E., and Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382-439.

Rosch, E., Simpson, C., and Miller, S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, **2**, 491-502.

Smith, E.E. (1990). Categorization. In D.Osherson, H. Lasnik, S. Kosslyn, E.E. Smith, and N. Block (Eds.), *Invitation to Cognitive Science*. Cambridge, MA: MIT Press.

Smith, E.E., and Medin, D.L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Tversky, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.

Tversky, B., and Hemenway, K. (1991). Parts and the basic level in natural categories and artificial stimuli: Comments on Murphy (1991). *Memory and Cognition*, **19**, 439-442.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, **32**, 193-254.

Zusne, L. (1970). *Visual Perception of Form*. New York: Academic Press.