Max-Planck-Institut
für biologische Kybernetik
Arbeitsgruppe Bülthoff

Spemannstraße 38 • 72076 Tübingen • Germany

Technical Report No. 40 September 1996

Features of the representation space for 3D objects

Shimon Edelman, Heinrich H. Bülthoff and Isabelle Bülthoff

Abstract

To explore the nature of the representation space of 3D objects, we studied human performance in forced-choice classification of objects composed of four geon-like parts, emanating from a common center. The two class prototypes were distinguished by qualitative contrasts (bulging vs. waist-like limbs). Subjects were trained to discriminate between the two prototypes (shown briefly, from a number of viewpoints, in stereo) in a 1-interval forced-choice task, until they reached a 90% correct-response performance level. In the first experiment, 11 subjects were tested on shapes obtained by varying the prototypical parameters both orthogonally (Ortho) and in parallel (Para) to the line connecting the prototypes in the parameter-space. For the eight subjects who performed above chance, the error rate increased with the Ortho parameter-space displacement between the stimulus and the corresponding prototype (the effect of the Para displacement was marginal). Clearly, the parameter-space location of the stimuli mattered more than the qualitative contrasts (which were always present). To find out whether both prototypes or just the nearest neighbor of the test shape influenced the decision, in the second experiment we tested 18 new subjects on a fixed set of shapes, while the test-stage distance between the two classes assumed one of three values (Far, Intermediate, and Near). For the 13 subjects who performed above chance, the error rate (on physically identical stimuli) in the Near condition was higher than in the other two conditions. The results of the two experiments contradict the prediction of theories that postulate exclusive reliance on qualitative contrasts, and support the notion of a metric representation space, with the subjects' performance determined by distances to more than one reference point or prototype (Edelman, 1995b).

Thanks to Mike Tarr for his comments on the ideas that led to these experiments. SE is currently at the Center for Biological and Computational Learning at MIT, on leave from the Weizmann Institute, where he is an incumbent of the Sir Charles Clore Career Development Chair. This work has been supported in part by the German-Israeli Foundation, through a grant to Heinrich H. Bülthoff, Nathan Intrator, and Shimon Edelman.

1 Introduction

To make sense of the world of shapes it encounters, the visual system must overcome two major computational difficulties. The first of these is the variability in the appearance of a 3D object (and, hence, in the stimulus it presents to the visual system), caused by the varying viewing conditions, such as illumination and pose with respect to the observer. Thus, the same 3D object may look completely differently when seen from different viewpoints; to realize that two views belong to the same object, the visual system must reveal their common origin, while ignoring the conditions that gave rise to their differences.

The second source of problems is the variability in the shape of individual objects belonging to the same category. Just as a series of views of the same object must be perceived as such, a collection of different shapes should be attributed to the same category, if they are sufficiently similar. There is, however, an important distinction between the two cases: whereas the changes in the object appearance precipitated by changing viewpoint can be fully characterized by a handful of parameters (as few as six, in the case of a rigid object), the variation in the shape of objects belonging to the same class is *a priori* unconstrained.

A convenient common approach to the description of the two kinds of computational problems mentioned above is to coach both in terms of class membership (Edelman et al., 1996). Recognizing an image as a view of some object then becomes the problem of deciding the membership of that image in the class of all views of that object, which we call its *view-space*. Analogously, the categorization of an image as produced by some member of a class of shapes amounts to pinpointing the location of the image in a *shape-space* spanned by all members of that class.

A considerable amount of attention has been given recently to the issues involved in the perception of different views as belonging to the same object, or, using the terminology we just introduced, in the processing of the view-spaces of individual objects. In contrast, much less work has been done on the processing of shape-spaces generated by object categories. In the present paper, we report two experiments intended to fill this gap. Our approach, the experimental results, and their interpretation are described in the next sections, following a brief survey of the previous work on the perception of view-spaces (*recognition*) and of shape-spaces (*categorization*) by human subjects.

1.1 View-space effects

Psychophysical studies conducted in the past few years led to the characterization of certain basic limitations of the visual system in generalizing shape-based recognition to novel conditions; see (Jolicoeur and Humphrey, 1996) for an extensive review and a discussion. Specifically, it was found that the recognition of novel views of objects tends to be slower and more prone to errors than the recognition of highly familiar views (Rock and DiVita, 1987; Tarr and Pinker, 1989; Edelman and Bülthoff, 1992; Humphrey and Khan, 1992; Cutzu and Edelman, 1994), and that it persists even when full 3D shape information is available to the subject through, e.g., binocular stereo cues (Edelman and Bülthoff, 1992).

The relevance of the above findings to the understanding of the processes of object recognition has been disputed on the basis of the difference between viewpoint-dependent performance exhibited by the subjects in these experiments, and the viewpoint-invariant performance found in other studies. In particular, Biederman and Gerhardstein (1993) reported essentially viewpoint-invariant performance on some of the objects used previously in (Edelman and Bülthoff, 1992), to which distinctive single parts have been added. A subsequent detailed investigation, in which the number of distinctive parts was manipulated in addition to object orientation, showed, however, that recognition always becomes poorer with increasing change in viewpoint, although this dependence is at its weakest for objects with one unique part (Tarr, Bülthoff, Zabinski, and Blanz, 1996).

1.2 Shape-space effects and the role of similarity

The assumption that the processes and the representations involved in identifying specific individuals are different from those used for categorization (Jolicoeur, 1990) has been recently put to an explicit test in a series of experiments, in which objective similarity between stimuli (and, consequently, the categorical level of their distinction) varied in a controlled fashion (Edelman, 1995a).

Subjects in those experiments were trained to discriminate between two classes of computer generated 3D objects, one resembling monkeys, and the other dogs. Both classes were defined by the same set of 56 parameters, which encoded sizes, shapes, and placement of the limbs, the ears, the snout, etc. Interpolation between parameter vectors of the class prototypes yielded shapes that changed smoothly between monkey and dog. Within-class variation was induced in each trial by randomly perturbing all the parameters. After the subjects reached 90% correct performance on a fixed canonical view of each object, discrimination performance was tested for novel views that differed by up to 60° from the training view. In all the experiments reported in (Edelman, 1995a), higher inter-stimulus similarity was associated with an increase in the mean error rate and, for misorientation of up to 45°, with an increase in the degree of viewpoint dependence. These results suggest that a geon-level (Biederman, 1987; Biederman and Gerhardstein, 1993) difference between stimuli is neither strictly necessary nor always sufficient for viewpoint-invariant performance.

The studies we mentioned so far concentrated on the quantification of the effects of viewpoint on recognition, and on the interaction between these effects and those of similarity among the objects that were to be recognized. While these studies explored the effects of the relative location of the stimuli both in the view-space and in the shape-space, the former exploration has been more thorough. For example, the experiments of (Bülthoff and Edelman, 1992) involved parametric control over viewpoint in two mutually orthogonal directions, whereas the study of (Edelman, 1995a) only manipulated the similarity between the two classes of stimuli, which is a scalar quantity. Thus, in the experiments reported below, we chose to concentrate on a parametric exploration of the effects of shape-space proximity (similarity) between the stimuli, the issues of viewpoint having been deemed of a secondary importance, in view of the previous findings in this field.

2 The ORTHO experiment

The first experiment involved two classes of objects, defined by prototypes P_1 and P_2 (see Figure 1). The objects were jointly parameterized by a number of variables that controlled their appearance; each object thus corresponded to a point in the parameter-space (Figure 2). The shape of the objects was manipulated by combining two orthogonal directions of displacement in the shape (parameter) space — in parallel and in perpendicular to the line connecting P_1 and P_2 (Figure 3).[1] Altogether, 15 exemplar objects for class 2 were formed by this procedure (Figure 4).

We chose this shape-space arrangement of stimuli because it offers an opportunity to test the predictions of a number of theories of object representation, and to evaluate these theories as models of recognition in human vision. Observe that the two class prototypes differed by a so-called qualitative contrast (Biederman, 1987): the sign of the bulge

Figure 1: Two prototypical hedgehog-like objects, similar to those we used in our experiments (both "hedgehogs" are shown at the same orientation). Each object is composed of a number of limbs protruding from a common center; the limbs are generalized cylinders, similar to Biederman's (1987) geons. The two prototypes are distinguished by qualitative (nonaccidental) contrasts (sign of bulge/waist of the limbs; see Figure 3). In addition, a number of quantitative parameters such as the degree of bulge/waist, the amount of taper, etc., control the exact shape of each instance object. Note that the qualitative contrasts emerge from the accumulation of quantitative changes, as illustrated in Figure 2.

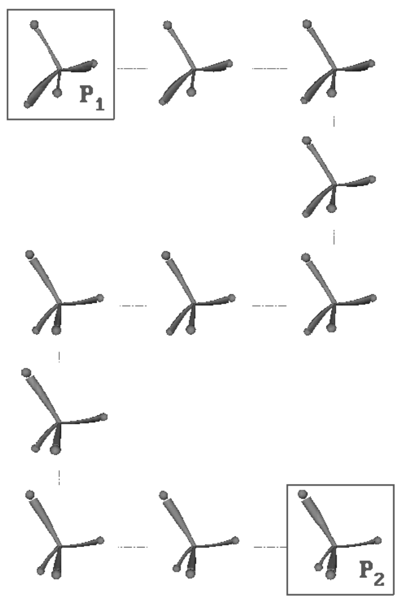


Figure 2: Each object can be represented by a single point in a parameter (shape) space. Changing the quantitative parameters ("morphing") corresponds to a movement of the shape-space representation of the object. This figure illustrates the morphing sequence that connects the two prototype objects. Although the changes between the successive images are minute, they accumulate to make up easily perceptible (and, eventually, "qualitative") differences between the endpoints of the sequence.

of the generalized-cylinder parts. Theories that postulate reliance on such contrasts (such as Biederman's Recognition By Components, or RBC) predict viewpoint-invariant near-perfect discrimi-

[1] Note the parallel between this experiment and the INTER/EXTRA/ORTHO experiments of (Bülthoff and Edelman, 1992).

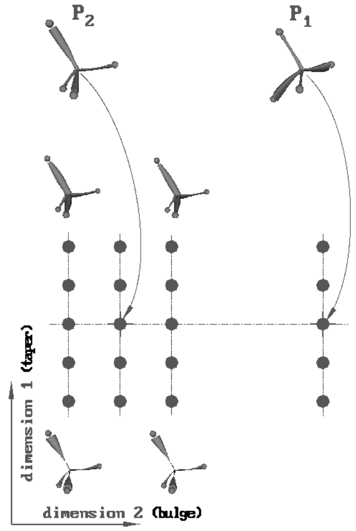


Figure 3: The parameter-space arrangement of stimuli. The parameter-space locations of the two prototypical objects are marked by P_1 and P_2. The two orthogonal directions of shape variation are bulge (increase/decrease) and taper (left/right). Specifically, the shift from P_1 to P_2 corresponds to a gradual change from a waist-like to a bulging profile of the hedgehog's limbs; the orthogonal direction corresponds to an equally gradual change of limb shape that tapers from the proximal towards the distal end to a shape that tapers in the opposite direction. See Figure 4 for an illustration of the entire array of stimuli corresponding to this parameter-space pattern.

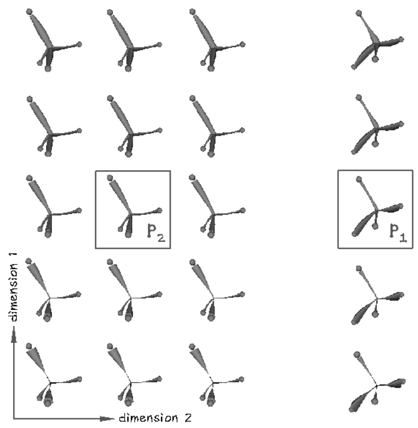


Figure 4: All the stimuli objects. The perception of variants of the learned prototype objects was probed with 15 exemplars made out of prototype P_2. All these exemplars, whose shape-space locations form a 3×5 grid centered on P_2, are illustrated here.

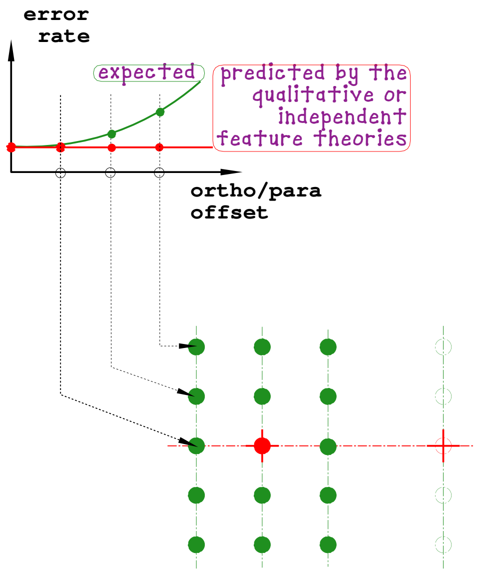


Figure 5: The parameter-space arrangement of the stimuli (see also Figure 1, right), and the expected performance in the ORTHO experiment. The subjects were trained to discriminate between the two prototypes, P_1 and P_2. They were then tested on the discrimination of stimuli produced by a shape-space variation orthogonal to the contrast between the two prototypes. See section 2 for a discussion of the predicted results and the actual findings.

nation performance for the two class prototypes, **and** for stimuli derived from the prototypes by a parameter-space displacement which is orthogonal to the line connecting P_1 and P_2. The same prediction can be derived from theories that postulate involvement of metric features, but deny the possibility of interaction between the different orthogonal dimensions of the feature-space (for an example of such a theory, see Ashby and Perrin, 1988). The reason for this prediction is that a variation which is orthogonal to the difference between P_1 and P_2 should not affect discrimination.

Not all theories postulating a metric feature-space as a substrate for recognition predict no effect for the proposed manipulation. For example, a theory according to which discrimination performance relies on the computation on distances to the class prototypes (Poggio and Edelman, 1990) and not to a decision surface (Ashby and Perrin, 1988) predicts that the performance will deteriorate the farther the stimuli are removed from the line connecting the two prototypes (see Figure 5).

2.1 Method

Eleven subjects were trained to discriminate between the two prototypes in a 1-interval forced-choice task. In each trial, an image of one of the two prototypes was briefly presented on the screen of a Silicon Graphics workstation, in binocular stereo (using LCD shutter glasses synchronized with the display). The subject was required to press the right or the left key on the computer mouse, depending on the class to which the stimulus belonged; an incorrect response triggered a beep (only during the training phase of the experiment). The presentation time was 300 *msec*. The object could be seen from any of four viewpoints, spaced evenly around the viewing sphere.

The subjects were trained for a minimum of 30 trials, until they reached a 90% correct-response performance level (computed on the trailing 30 trials of the session). They were then tested on shapes obtained by varying the prototypical parameters both orthogonally (ORTHO) and in parallel (PARA) to the line connecting the prototypes in the parameter-space, as described above.

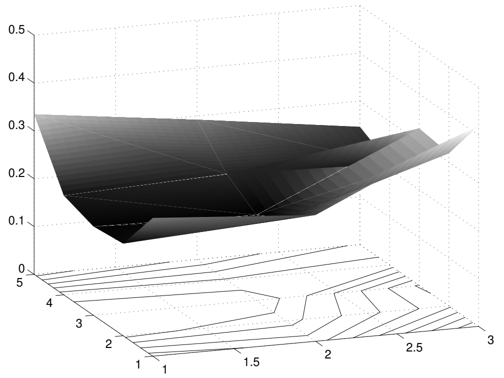
2.2 Results

Eight of the 11 subjects who participated in the experiment performed above chance in the test phase (the mean error rate of these was 23%). For these subjects, the error rate (computed over the four test views and the three repetitions per condition) increased with the ORTHO parameter-space displacement between the stimulus and the corresponding prototype. A General Linear Models analysis (using procedure GLM; SAS, 1989) showed this effect to be significant: $F(2,63) = 2.9$, $p < 0.08$. The effect of the PARA displacement was marginal: $F(2,63) = 1.7$, $p < 0.19$.[2]

A stronger effect was masked by the large individual differences (the error rates of the eight subjects ranged between 4% and 34%). When these were taken into account (by incorporating the subject variable into the analysis), the effect of ORTHO displacement became stronger: $F(2,28) = 5.0$, $p < 0.01$, and a significant effect of PARA displacement emerged: $F(2,28) = 3.3$, $p < 0.05$. Importantly, there was no interaction between these effects and that of subject. Figure 6 shows the mean performance, plotted against the ORTHO and the PARA displacement.

2.3 Discussion

The results of the first experiment clearly indicate that the parameter-space location of the stimuli mattered more than the qualitative contrasts (which were always present) between the two classes that had to be distinguished. Moreover, the stronger effect of the ORTHO relative to the PARA displacement suggests that both prototypes participated in determining the response to the test stimuli. This pattern exactly mimics the distinction



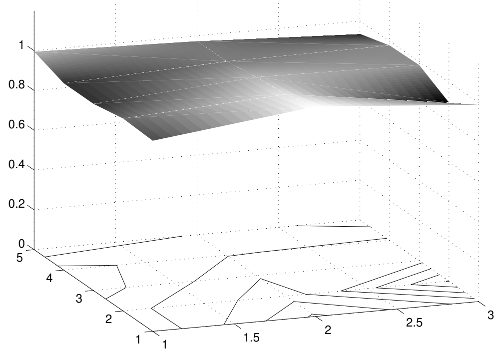


Figure 6: The mean performance of the eight subjects who responded above chance in the first experiment, plotted against the ORTHO and the PARA displacement (see section 2.2). The three PARA displacement values, denoted symbolically by the numerals 1,2,3, appear along the abscissa in the contour plots; the five ORTHO values correspond to the ordinate. The location of prototype P_2 corresponds to the point whose coordinates are $(2,3)$. Altogether, the 15 data points are arranged in a 3×5 grid around prototype P_2 (see Figure 3); the direction towards the other prototype in these plots is along the increasing abscissa values. *Top:* error rate; the adjacent lines in the contour plot are spaced at 2.5%. Note the general increase in the error rate for test stimuli that are closer to the other prototype. *Bottom:* response time; the line spacing is 25 *msec*.

between the ORTHO and the INTER/EXTRA effects in the experiments described in (Bülthoff and Edelman, 1992). There, however, the manipulation of the stimuli was carried out in the view-space (that is, the exemplars were rotated versions of the "prototype"), whereas in the present experiment the manipulation was in the shape-space (the exemplars differed from the prototype by their shape).

3 The NEAREST-NEIGHBOR experiment

The next experiment we describe was designed to gain further support for the idea that proximities to *both prototypes* contribute to the categorization process. If the visual system indeed relies on the computation of distances between the stimulus and

[2]In the computation of these effects, we collapsed the data over the two directions of ORTHO shift away from the prototype P_2, due to considerations of symmetry. Thus, the number of degrees of freedom in the F statistics was 2 in both cases.

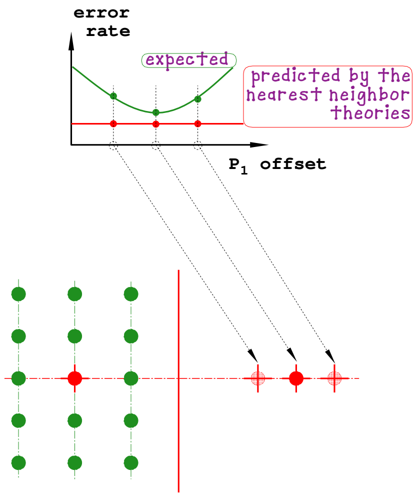


Figure 7: The parameter-space arrangement of the stimuli and the expected performance in the NEAREST-NEIGHBOR experiment. The two prototypes, P_1 and P_2, are as before. In this experiment, the location of P_1 relative to P_2 varied along the line connecting the two prototypes. Performance (discrimination between the two classes) was tested for the same physical stimuli, whose location in the parameter-space corresponds to the middle column in the 3×5 grid of points surrounding P_2. For a discussion of the expected performance, see section 3.

prototypical memory traces in some feature-space, it may use those distances in two different manners. The first possibility is that the identity of the prototype nearest to the stimulus is the sole determinant of the response (we term this the NEAREST-NEIGHBOR hypothesis). The second possibility is that a number of close neighbors of the stimulus jointly determine the nature of the response.

To distinguish between these two possibilities, we examined the responses of the subjects to a fixed set of stimuli, while manipulating the location of one of the two class prototypes (see Figure 7). Importantly, the manipulation left the test stimuli themselves always within the half-space dominated (by proximity) by the other, fixed, prototype. According to the NEAREST-NEIGHBOR hypothesis, the performance on the test stimuli should not change under the proposed manipulation. In contrast, theories that postulate the involvement of all sufficiently close prototypes (Poggio and Edelman, 1990) predict that performance will improve as the distance between the two prototypes increases (Edelman, 1995a).

3.1 Method

The stimulus set and the course of each trial were the same as in the first experiment. Of the 15 stimuli associated with prototype P_2, only the five belonging to the middle column were used (see Figure 7). These were crossed with three possible lo-

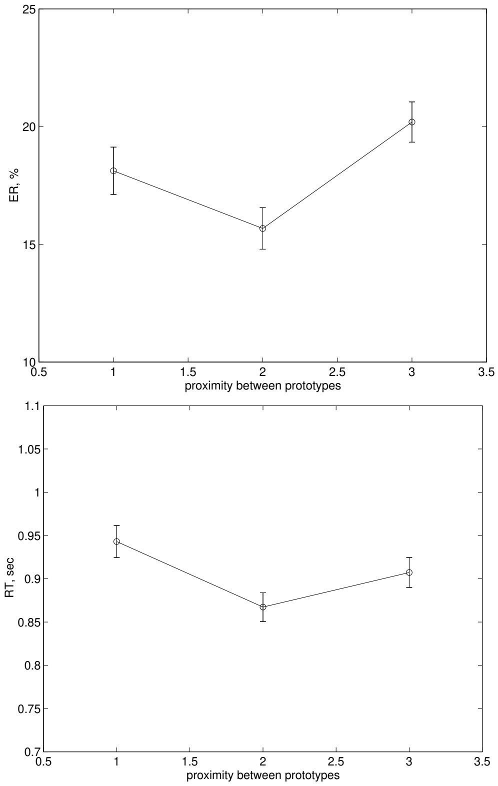


Figure 8: The mean performance of the 13 subjects who responded above chance in the second experiment, plotted against the PARA displacement of prototype P_1 (see section 3.2). *Top:* error rate. *Bottom:* response time. The error bars show the standard deviation of the corresponding means. The three values along the abscissa (prototype proximities 1, 2, 3) correspond, respectively, to the FAR, INTERMEDIATE, and NEAR conditions, described in the text.

cations of prototype P_1, which we termed FAR, INTERMEDIATE, and NEAR, yielding 15 test conditions (as in the first experiment).[3] Note, however, that the subject's performance was always assessed on the same five physical objects that belonged to class 2 (that is, the results reported below only pertain to the responses given to those objects).

3.2 Results

Thirteen of the 18 subjects who participated in this experiment performed above chance (mean error rate of these was 18.0%). When averaged over these subjects, the effect of moving prototype P_1 was marginal: $F(2, 180) = 1.6$, $p = 0.2$.[4]

[3]The location of P_1 in the test phase was manipulated by dividing the entire sequence of test trials into three blocks, each of which corresponded to FAR, INTERMEDIATE, or NEAR condition. The order of the blocks in the test session was randomized across subjects.

[4]Because this experiment concentrated on the effect of the displacement of prototype P_1 in the PARA direction, the data were collapsed over the ORTHO

As in the previous experiment, a stronger effect was masked by the large individual differences (the error rates of the 13 subjects ranged between 5% and 34%). When these were taken into account (by incorporating the subject variable into the analysis), the effect of moving P_1 became significant: $F(2,96) = 2.9$, $p < 0.06$; importantly, there was no interaction between this effect and that of subject. Figure 8 shows the mean performance, plotted against the PARA displacement of prototype P_1.

3.3 Discussion

The results of this experiment demonstrate the sensitivity of the visual system to the general setting of the categorization task with which it is confronted. If the classification decision were carried out by comparing a representation of the stimuli (which remained unchanged throughout the experiment) with that of the closest class prototype (which remained fixed as well), a constant performance would have ensued. We found, however, that the performance has been affected by the relocation of the second prototype relative to the first (closest) one, in clear violation of the prediction of the NEAREST NEIGHBOR hypothesis.

4 General discussion

The objective of the two experiments we described was to gather quantitative data regarding the process whereby the shape of an object is labeled as belonging to one of two classes. The foremost issue here is the nature of the representation-space wherein the decision takes place. To clarify this issue, we asked, specifically, whether the dimensions of the relevant space (i.e., the features involved in the representation) are independent. The results of experiment 1 suggest that they are not: our subjects performed worse on shapes that were progressively more different from the class prototype acquired during the training phase, even though this difference was orthogonal to the distinction between the two classes. This finding indicates that the visual system is not likely to rely solely on the single most distinctive contrast between the categories, even if this contrast is "qualitative" or "nonaccidental" (Biederman, 1987).

Even if the location of the stimulus in a shape-space spanned by the relevant features — and not merely its location along the line connecting the two class prototypes in that space — determines the subject's performance, the question of the nature of the shape-space (that is, the nature of the relevant features) still remains open. Rather than attempting to characterize the features explicitly (an undertaking that is notoriously resistant to a purely psychological approach), we chose to find out whether or not the features that the stimulus shares with the closest prototype alone determine the performance. The outcome of the second experiment reported above suggests that both prototypes contribute to the perceptual categorization decision.

An intriguing computational hypothesis consistent with our findings holds that the internal shape-space is spanned by a vector of proximities of the stimulus to a number of "reference" or prototypical objects, whose role here may be played by the class prototypes (Edelman, 1995b; Edelman et al., 1996). The implications of this hypothesis, according to which the features by which an object is judged are its similarities to other objects,[5] as well as a discussion of its compatibility with recent psychophysical and neurobiological findings on object representation, can be found in (Edelman, 1996).

In summary, the results of the two experiments we reported above contradict the prediction of theories of recognition that postulate exclusive reliance on qualitative contrasts (Biederman, 1987) or on proximity to a decision surface (Ashby and Perrin, 1988; Maddox and Ashby, 1993), and support the notion of a metric representation-space, with the subjects' performance determined by proximities to more than one reference point or prototype (Nosofsky, 1988; Nosofsky, 1991; Edelman, 1995b; McKinley and Nosofsky, 1996; Edelman et al., 1996).

displacement.

References

Ashby, F. G. and Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, 95(1):124–150.

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

Biederman, I. and Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for 3D viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, 19:1162–1182.

Bülthoff, H. H. and Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition. *Proceedings of the National Academy of Science*, 89:60–64.

Cutzu, F. and Edelman, S. (1994). Canonical views in object representation and recognition. *Vision Research*, 34:3037–3056.

[5]Note that more than one prototype is needed to pinpoint the location of the stimulus in the shape-space; thus, the computational scheme based on this notion is called the Chorus of Prototypes.

Edelman, S. (1995a). Class similarity and viewpoint invariance in the recognition of 3D objects. *Biological Cybernetics*, 72:207–220.

Edelman, S. (1995b). Representation, Similarity, and the Chorus of Prototypes. *Minds and Machines*, 5:45–68.

Edelman, S. (1996). Representation is representation of similarity. CS-TR 96-08, Weizmann Institute of Science. submitted to Behavior and Brain Sciences.

Edelman, S. and Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of 3D objects. *Vision Research*, 32:2385–2400.

Edelman, S., Cutzu, F., and Duvdevani-Bar, S. (1996). Similarity to reference shapes as a basis for shape representation. In Cottrell, G. W., editor, *Proceedings of 18th Annual Conf. of the Cognitive Science Society*, pages 260–265, San Diego, CA.

Humphrey, G. K. and Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Can. J. Psychol.*, 46:170–190.

Jolicoeur, P. (1990). Identification of disoriented objects: a dual-systems theory. *Mind and Language*, 5:387–410.

Jolicoeur, P. and Humphrey, G. K. (1996). Perception of rotated two-dimensional and three-dimensional objects and visual shapes. In Walsh, V. and Kulikowski, J., editors, *Perceptual constancies*, chapter 10. Cambridge University Press, Cambridge, UK. in press.

Maddox, W. T. and Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics*, 53:49–70.

McKinley, S. C. and Nosofsky, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 22:294–317.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14:700–708.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, 17:3–27.

Poggio, T. and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343:263–266.

Rock, I. and DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, 19:280–293.

Sas (1989). *SAS/STAT User's Guide, Version 6.* SAS Institute Inc., Cary, NC.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., and Blanz, V. (1996). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, pages –. in press.

Tarr, M. J. and Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21:233–282.