Similarity, connectionism, and the problem of representation in vision

Shimon Edelman and Sharon Duvdevani-Bar
Dept. of Applied Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel
{edelman,sharon}@wisdom.weizmann.ac.il

June 13, 1996

Abstract

A representational scheme under which the ranking between represented similarities is isomorphic to the ranking between the corresponding shape similarities can support perfectly correct shape classification, because it preserves the clustering of shapes according to the natural kinds prevailing in the external world. This note discusses the computational requirements of representation that preserves similarity ranks, and points out the relative straightforwardness of its connectionist implementation.

1 Introduction

1.1 Two problems of representation

It is possible to distinguish between two problems about mental representation (Cummins, 1989). The first of these, the problem of representations (plural), is basically empirical; two instances of this problem are the search for the representations employed by natural cognitive systems, and the design of the representational substrate for artificial cognitive modules. The second problem is called by Cummins the Problem of *Representation* (singular). Here, the central question is how, in principle, can an internal state of a system refer to anything at all in the external world.

Curiously, the Problem of Representation is rarely discussed in cognitive psychology. On the one hand, arguments put forward by Marr (1982) prompted cognitive scientists to contemplate the fact that not all representations are computationally equivalent; the classical example here is the difficulty of doing arithmetics with Roman numerals, compared to the ease of calculation using positional notation. On the other hand, it is generally taken for granted that a certain pattern of neuronal activity in someone's brain can represent the shape of a cat, or, amazingly, that physically distinct patterns of activity in the brains of two people can stand for the same shape (that of a cat).

1.2 First-order isomorphism: representation *by* similarity

Reviews in cognitive psychology tend to adopt extreme positions concerning representation. For example, Palmer (1978) asks "What is representation," then proceeds with an in-depth survey of the field, the upshot of which is that "we ... do not really understand our concepts of representation." In contrast, Suppes, Pavel and Falmagne (1994) state flatly "Representation of something is an image, model, or reproduction of that thing". This amounts to the Aristotelian idea of *representation by similarity*; according to it, a representation, say, of a cat, can only help you know about cats if it resembles one.

Among the problems raised by this notion of representation is the obligatory involvement of an observer: cartoon cats resemble real cats *to us*; the two kinds share no properties other than perceptual. Another problem is the inability of representations of this type to deal with abstraction (does having an image of a striped cat qualify as a representation of a calico cat?). Because of these and other considerations, in the philosophy of mind the idea of representation by similarity has been discredited so thoroughly as to inspire a rare consensus (Cummins, 1989).

1.3 Second-order isomorphism: representation *of* similarity

Barring resemblance, or "first-order structural isomorphism" (Shepard, 1968), between the object and the entity that stands for it internally, what relationship can qualify as representation of something by something else? Shepard (1968) suggests *"second-order" isomorphism*: "... the isomorphism should be sought — not in the first-order relation between *(a)* an individual object, and *(b)* its corresponding internal representation — but in the second-order relation between *(a)* the relations among alternative external objects, and *(b)* the relations among their corresponding internal representations. Thus, although the internal representation for a square need not itself be square, it should (whatever it is) at least have a closer functional relation to the internal representation for a rectangle than to that, say, for a green flash or the taste of a persimmon," (Shepard and Chipman, 1970, p.2).

More recently, the idea of representation by second-order isomorphism has been advanced, under various guises, in a number of fields in cognitive science. Typically, the researchers in these fields take for granted the implausibility of representation by similarity, that is, by first-order isomorphism. Consequently, the theories mention merely "isomorphism," it being implied that the isomorphism holds between structures (and is, therefore, "second-order," in Shepard's terms), and not between individual entities.

A particularly clear exposition of this point can be found in (Gallistel, 1990), where the example

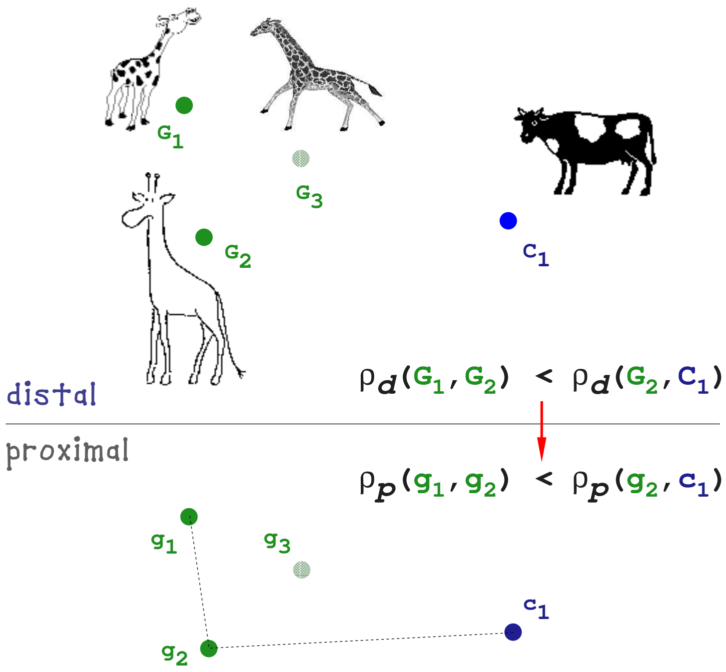

Figure 1: Clustering by natural kinds, and its representation that fulfills the requirement of second-order isomorphism (here, isomorphism between distance ranks in the two spaces; ρ_d and ρ_p are the distal and the proximal distance functions).

of numerical representation of physical quantities such as mass is discussed. On p.24, Gallistel points out that "The statement that the number 10 is isomorphic to some particular mass is meaningless. It is the system of numbers that is isomorphic to the system of masses." Gallistel then proceeds to defend the view that cognitive systems represent space and time by harboring structures that are isomorphic to corresponding structures in the world, subject to the above qualification regarding the importance of entire structures (that is, elements plus relations), as opposed to individual elements. A similar proposal, albeit in the context of representation of conceptual rather than perceptual spaces, can be found in (Holland et al., 1986). Here, the theory is couched explicitly in terms of a homomorphism between a collection of states and a set of possible transitions between states in the world, and in the internal representation. This version of representation by second-order isomorphism has been discussed very recently in (Cummins, 1996), where it is, unfortunately, called "The Picture Theory of Representation."[1] Importantly, Cummins echoes Gallistel in pointing out that constituents of a structure are meaningful only in the context of some structure or other.

Now, in the context of visual object representation, this observation translates into the notion that only certain relationships between the objects — not the shapes of the individual objects

[1] Unfortunately, because in vision research pictorial representations are traditionally associated with the Aristotelian notion of representation by similarity, or first-order isomorphism.

themselves — need to be represented. What could these relationships be? Shepard's example, stated in the preceding section, clearly suggests that the represented structure should be that of *similarity relationships* between objects. Accordingly, this approach can be termed "representation *of* similarity" (cf. Figure 1). Our goal in the rest of this note is to outline some of the mathematical issues raised by this approach to representation, and to examine its viability in the context of shape vision.

2 Towards veridical representation

2.1 Preliminaries

Let us now identify mathematical conditions under which isomorphism between distal and proximal relations gives rise to representation which is, in a sense, true to the original. Let $\mathcal{D}$ be the set of distal objects, and $\mathcal{P}$ — the set of internal tokens that participate in the process of representation. Let $f : D \subset \mathcal{D} \to \mathcal{X}$ and $g : P \subset \mathcal{P} \to \mathcal{Y}$ be functions defined, respectively, over sets of distal and proximal entities (no restriction needs to be placed at this stage on the number of arguments of f, g). According to the requirement of second-order isomorphism, $\mathcal{D}$ is isomorphic to $\mathcal{P}$ under a mapping M if $\forall D \subset \mathcal{D}, f(D) \sim g(M(D))$, where the relation $\sim$ is that of set isomorphism, defined over $\mathcal{X}, \mathcal{Y}$. To constrain the choice of M, we have to be more specific in defining the functions f, g. In the context of shape perception, it is natural to consider for this purpose similarity (actually, two similarity functions must be defined, one for the external objects and one for the internal tokens standing for those objects). Intuitively, we would like the representation to capture the similarity relationships within and between natural kinds (Quine, 1969). More specifically, similarity is seen to be relevant both to recognition, in which case resemblance between the viewed shape and some previously seen ones is to be assessed, and to classification, where similarities between a shape and a number of shape classes are compared (Nosofsky, 1988; Goldstone, 1994). For a pair of internal representations, similarity is not directly observable, but can still be defined operationally as the degree to which an optimal stimulus for one token activates the other one (Shepard and Chipman, 1970).

As noted in (Nicod, 1930; Quine, 1973), similarity should be construed as a triadic relation: knowing that A is more similar to B than to C is much more informative than merely knowing that A is similar to B and B to C. Furthermore, we may assume that similarity is defined qualitatively and not quantitatively, that is, the range of f, g is $\mathcal{X} = \mathcal{Y} = \{>, <\}$. This assumption limits neither classification (because knowledge of similarity for every triplet of points belonging to a metric space defines unequivocally the clustering of the points), nor recognition (because the location of each point can be recovered from triadic similarities using nonmetric multidimensional scaling; see section 3.1). We thus obtain $f : \mathcal{D} \times \mathcal{D} \times \mathcal{D} \to \{>, <\}$, and, analogously, $g : \mathcal{P} \times \mathcal{P} \times \mathcal{P} \to \{>, <\}$. The condition on M then becomes: $\forall D_1, D_2, D_3 \in \mathcal{D}, f(D_1, D_2, D_3) = g(M(D_1), M(D_2), M(D_3))$. In other words, the mapping M is required to preserve similarity ranks.

2.2 Constraints on the distal structure

The notion of similarity can be naturally quantified if the distal shape space and the internal representation space are both endowed with metric structure, in which case (dis)similarity in each space can be derived from the appropriate distance function. In the distal space, the existence of a metric for a given shape family (i.e., a subspace of the shape space) depends on the possibility to come up with at least one parameterization scheme common to all the shapes in the family (the issue of different parameterizations of the same set of shapes is discussed in section 2.3.3 below). A metric on shapes can then be defined simply as the Euclidean distance in the underlying parameter space R^n.

Obviously, only a small part of that space is occupied by objects that "look like something" — a combination of parameters picked at random is likely to define a nondescript collection of voxels. It is important to realize, therefore, that the goal behind the introduction of parameters is to formalize the idea of an "objective" similarity for shapes, and not to invent a language in which only some legitimate objects possess valid descriptions. With that qualification, the task of defining an appropriate parameterization becomes quite easy, and can even be automated to a certain extent.

The performance of one possible approach to a common parameterization of diverse shapes is illustrated in Figures 2, 3, and 4. According to this approach, the parameterization is computed in a straightforward manner, by assigning a dimension to each voxel in a fine subdivision of the bounding box of the object. As a result, it becomes possible to consider various objects as points in a common parameter space, and to morph (smoothly deform) one object into another, as shown in Figure 3 (this process corresponds to a movement in the parameter space along the line that connects the points corresponding to the two objects). Although the voxel-based parameter space is high-dimensional, a small number of dimensions may suffice to specify complex and diverse shapes, as illustrated in Figure 4, which examines the utility of principal component analysis on the space of shape parameters.

Note that the possibility of a common parameterization of a variety of object shapes provides a substrate for an objective similarity structure of the world, to be reflected internally by the representational process. Thus, it fulfills the first basic prerequisite for representation by second-order isomorphism. The second prerequisite, having to do with the manner whereby the distal similarities are mapped into the representational system, is the topic of the next section.

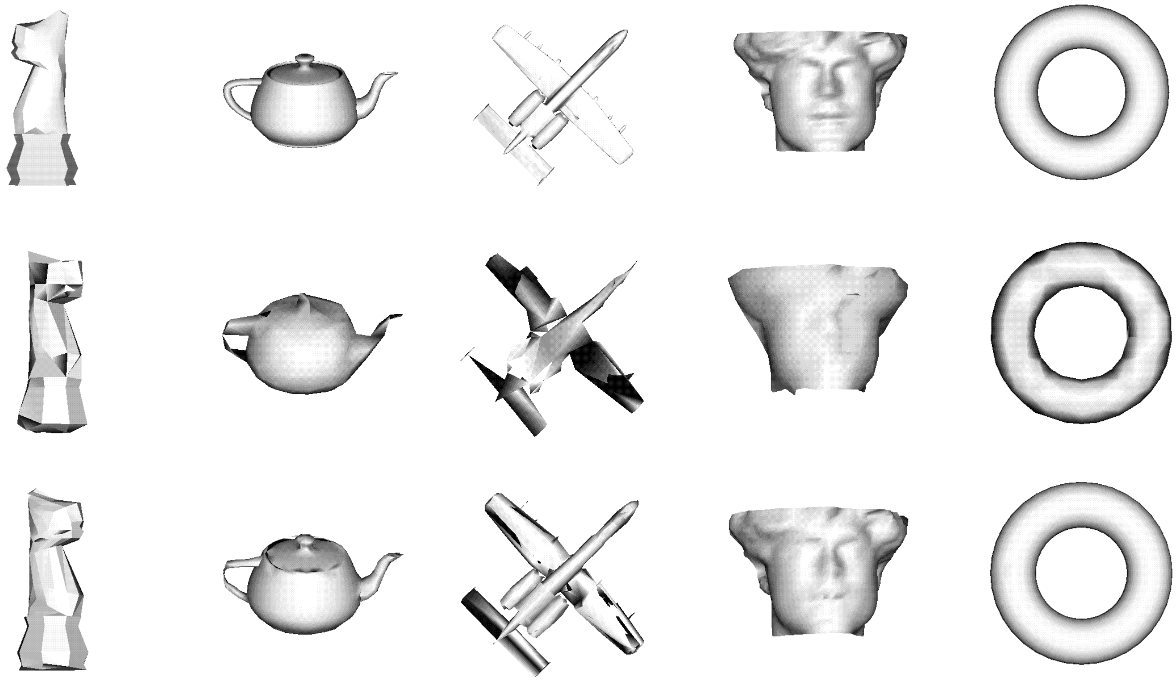

Figure 2: *Top:* images of several 3D objects. *Middle:* the reconstruction of the objects using linear resolution of 10 divisions per side of the bounding box (implying a 10^3-dimensional parameter space, as explained below). *Bottom:* reconstruction under linear resolution of 30 divisions per side of the bounding box (30^3 parameters). Given a linear resolution parameter u, the parameterization program subdivides the (bounding box of the) object into $u \times u \times u$ voxels. The vector of occupancy indices for the voxels thus constitutes a geometrical description of the object in terms of u^3 parameters. The resulting parameter space is universal (common to all the objects), as illustrated in the next figure, where the common parameterization is used to produce morph sequences between different objects.

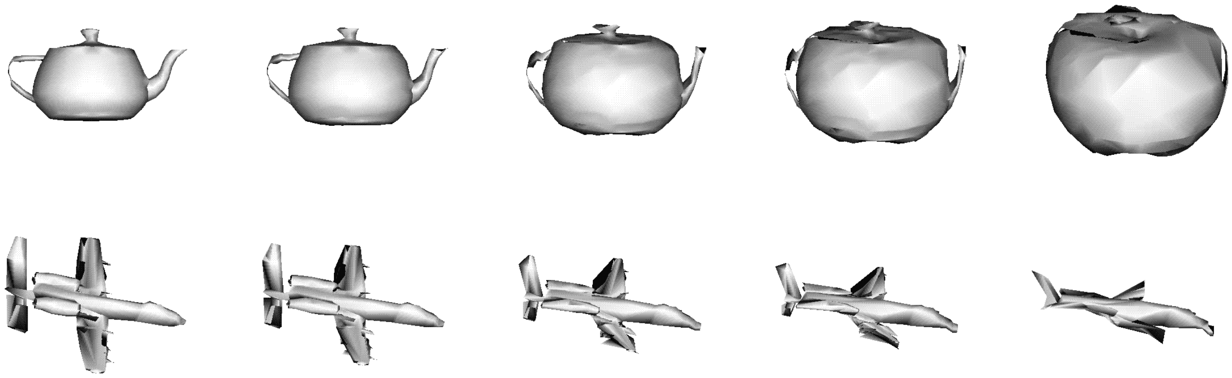

Figure 3: Successive stages in a morphing (smooth deformation) process linking distinct points in a common shape space. *Top:* morphing a teapot into an apple. *Bottom:* morphing an airplane into a shark. Both examples of morphing used a 30^3-dimensional parameterization of the objects.

2.3 Constraints on the distal-to-proximal mapping

2.3.1 Distance rank preservation

Strict rank preservation. The similarity ranking of three distinct parameter vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in R^n$ can be formalized via the notion of their *simple ratio*, defined as $\langle \mathbf{x}, \mathbf{y}, \mathbf{z} \rangle = |\mathbf{x}-\mathbf{y}|/|\mathbf{y}-\mathbf{z}|$. Obviously, a mapping $S : R^n \to R^n$ preserves distance ranks iff it preserves $\langle \mathbf{x}, \mathbf{y}, \mathbf{z} \rangle$ for any choice of points. A bijective mapping S with this property must be a similitude, that is, a mapping of the form $S(\mathbf{x}) = \lambda P(\mathbf{x})$, where $\lambda > 0, \lambda \in R$, and $P : R^n \to R^n$ is an orthogonal transformation (Reshetnyak, 1989). Thus, the requirement of rank preservation is quite restrictive in the class of mappings it allows.

Local rank preservation. In one or two dimensions, the rank preservation requirement is satisfied locally by *any* well-behaved mapping. Specifically, a mapping realized by an analytic function with a non-vanishing Jacobian in a given region is conformal there (Cohn, 1967). Such a function preserves similitude of small triangles; in particular, a scalene triangle formed by a triplet of points will be mapped into a triangle with the same ranking of side lengths (see Figure 1). In higher dimensions, conformality is much more restrictive. As proved by Liouville in 1850, already for $n = 3$ there are no globally conformal mappings from R^n to itself besides those which are composed of finitely many inversions with respect to spheres. Such mappings, called Möbius transformations, constitute a finite-dimensional Lie group which includes the group of motions in R^n and is only slightly broader than that group (Reshetnyak, 1989).

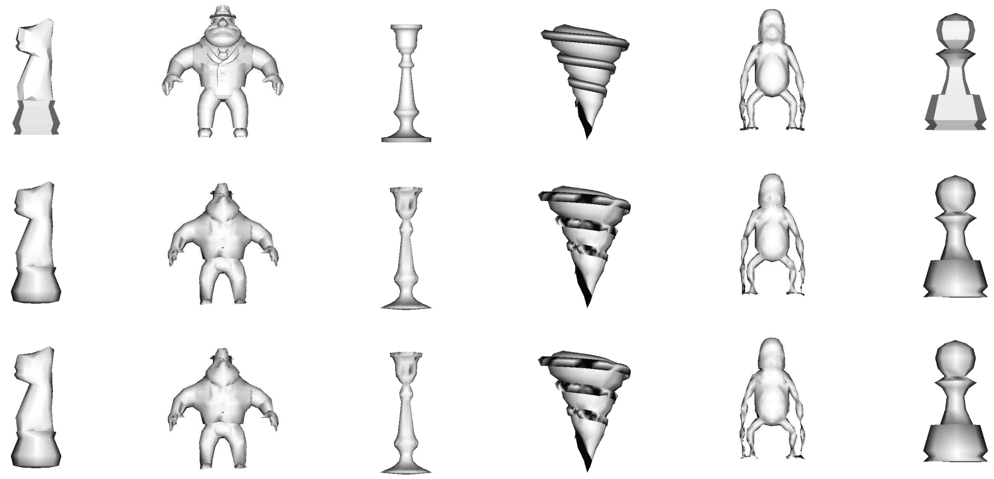

Figure 4: *Top:* images of several 3D objects. *Middle:* images of the same objects, parameterized with 25^3 parameters and re-rendered. *Bottom:* images rendered from representations of the objects in a common 5-dimensional parameter space, obtained from the high-dimensional voxel-based space using principal component analysis. The weights of the five leading principal components are given below:

knight	-0.19	4.60	6.82	5.37	6.94
al	-12.01	-16.19	1.00	-0.53	-0.32
canstick	-2.96	7.65	1.85	-12.87	-0.08
shell	19.36	-7.52	-2.37	-1.47	0.08
ape	-4.62	6.00	-13.46	3.79	0.30
pawn	0.43	5.45	6.15	5.72	-6.91

Local approximate rank preservation. A considerably broader class of mappings emerges if the requirement of conformality is replaced by that of quasiconformality. Intuitively, a regular topological mapping is quasiconformal if there exists a constant q, $1 \leq q \leq \infty$, such that almost any infinitesimally small sphere is transformed into an ellipsoid for which the ratio of the largest semiaxis to the smallest one does not exceed q (Reshetnyak, 1989). Under such a mapping, the ranks of distances between points are preserved approximately, on a small scale (Väisälä 1992, p.124).

2.3.2 The implications of locality

How relevant is local approximate preservation of distance ranks, offered by a quasiconformal distal-to-proximal mapping, to the representation of real-world shapes? An answer to this question is suggested by the realization that distance ranks need not be preserved globally, across the entire shape space. The limits to the extent of distance rank preservation expected of a perceptual system are suggested by the hierarchical structure of perceived categories, revealed in numerous studies in cognitive science. Within this hierarchy, there exists a certain level, called basic level, which is the most salient according to a variety of criteria (Rosch et al., 1976). Taking as an example the hierarchy quadruped, dog, terrier, the basic level is that of dog. Objects whose recognition implies less detailed distinctions than those required for basic-level categorization are said to belong to a superordinate level.

A reasonable assumption is that faithful representation of similarities is required *within* superordinate categories (e.g., within quadruped: giraffe to camel, leopard, horse), and, of course, within basic categories, but not *between* superordinate categories. In other words, the informativeness of the statement that a giraffe is more similar to a banana than to a beetle is questionable. At the same time, one expects that a giraffe be represented as more similar to other members of the quadruped category than either to a banana or to a beetle.[2]

2.3.3 The components of the distal-to-proximal mapping M

A mapping implemented by a perceptual system can be described generically as $M = \mathsf{dimens} \circ \mathsf{measure} \circ \mathsf{imag} \circ \mathsf{geom}$, where the first two components are dictated by the properties of the world, and the other two constitute part of the system:

Geometry. $\mathsf{geom}(\mathbf{p})$: the shape of the object, as a function of the distal parameter-space description $\mathbf{p}$.

Imaging. $\mathsf{imag}(\mathbf{p};\mathbf{z})$: the image on the receptor surface, as a function of the shape parameters $\mathbf{p}$ and the viewing conditions $\mathbf{z}$.

Measurements. $\mathsf{measure}(\mathbf{p};\mathbf{z})$: the set of internal measurements performed on the image.

Dimensionality reduction. $\mathsf{dimens}(\mathbf{p})$: the low-dimensional representation, independent of $\mathbf{z}$.

Let us now examine the constraints imposed on the mapping M by the requirement of second-order isomorphism between similarity patterns in its domain and its range. Note that some of the components of M are not under control of the perceptual system, while others are. The components determined by the system's environment are likely to have a detrimental effect on the proper

[2]Similarity across superordinate category borders may make sense, if the comparison concentrates on a few isolated dimensions, which differ from one case to another. For example, giraffes may be found similar to "tall" things (a palm tree, a lamp post), or to "spotted" things (a leopard, a camouflage net). Typically, similarity in those cases is markedly intransitive (a palm tree is not similar to a leopard). See (Tversky, 1977).

representation of similarity, e.g., by introducing extraneous sources of variability into the representations. Each of the components is, therefore, worth considering separately, to determine whether the controllable components can compensate for the effects of the uncontrollable ones.

Consider first the effect of the geometry mapping, geom. Although a perceptual system has no control over this mapping, no such control is necessary: parameterizations which are "unnatural" in the sense that they encode the geometry of objects in a non-smooth manner will be excluded from possible recovery by perceptual systems based on the proposed principle, but there will be always plenty of other, "natural" parameterizations that will be amenable to veridical representation.

Let $\mathcal{P}$ be the set of all parameterizations related to a given one p_0 by some conformal mapping T. The set $\mathcal{P}$ is an equivalence class (Väisälä, 1971); moreover, because the composition $T \circ M$ is conformal if M is, veridical representation of some $p \in \mathcal{P}$ is equivalent to the representation of any other $p \in \mathcal{P}$. Now, a conformal mapping M will give rise to a proper (i.e., second-order isomorphic) representation of object clustering under all parameterizations belonging to *some* class $\mathcal{P}_x$. The nature of that class will depend on the nature of the mapping (which can emphasize some distances among objects at the expense of others, with or without altering the distance ranks). A system that is a product of natural selection is expected to have evolved a mapping better suitable for the representation of those aspects of its habitat most important for its survival and behavior. Thus, while veridical representation is possible, it is also possible that two perceptual systems implementing different mappings will have incompatible (or even conflicting) pictures of the world.

The next component of M – the viewing mapping, imag – introduces a dependence on variables $\mathbf{z}$ which are extraneous to the shape parameters that are to be represented. These variables encode the pose of the object with respect to the observer, to the light sources, and to the other objects in the scene. Their influence must be counteracted by the perceptual system, through the combined action of measurement and dimensionality reduction, $\mathsf{dimens} \circ \mathsf{measure}$, to reduce the likelihood of two nearby parameter-space points (i.e., two similar shapes) being mapped into widely disparate points in the final representation space. Note that absolute invariance with respect to these variables is not necessary; it is only required that shape-space changes influence the measurements more strongly than view-space changes (more on this in section 3.2). Furthermore, not all the dimensions of $\mathbf{z}$ have to be treated by the same mechanism: image-plane translation can be compensated for by a covert shift of attention (Anderson and Van Essen, 1987) or an overt one (such as a saccadic eye movement), variation in apparent size — by global scaling using a hard-wired mechanism (Schwartz, 1985), and rotation in depth — by learning an appropriate normalizing mapping specific for each object class (Poggio and Edelman, 1990).

In addition to the requirement that $\mathsf{dimens} \circ \mathsf{measure}$ counteract the effects of the view-related variables, one may note that the preservation of distance ranks implies that any change in the distal parameter space must be reflected in the final low-dimensional representation (otherwise, if some of the original dimensions collapse under the representation, distances between points are likely to be distorted). To ensure that as many as possible of the original dimensions of variation among the distal objects are preserved, it is worthwhile to make as many varied measurements as possible. This makes the measurement space (defined by measure) high-dimensional, and necessitates subsequent dimensionality reduction (through the action of dimens; the existence of a low-dimensional parameterization of the distal shape space constitutes another computational justification for dimensionality reduction). In a flexible system, dimensionality reduction would have to involve learning to find informative dimensions, depending on the statistics of the input and (if available) of additional knowledge provided by the environment (see, e.g., Intrator, 1993).

To summarize, the message of this section is not mere reiteration of the importance of invariance or of dimensionality reduction. Rather, it is the possibility to achieve formally veridical representation, provided that those two problems are addressed with the proper computational goal in mind. According to the proposed theory of representation, that goal is preservation of parametrically defined similarity by the distal to proximal mapping. The next section presents empirical evidence in support of the proposed theory. The stress being on brevity and variety, only a short mention is made of selected results from psychophysics, computational modeling, and neurophysiology of object representation.

3 Empirical support

3.1 Psychophysics

In his discussion of cognitive representation, Palmer notes that "...in order to test the two models [first and second-order isomorphism], one must have access to the mapping function from the outside world" (Palmer, 1978, p.293). Such access can be realized by combining parametric shape generation with a psychological technique known as multidimensional scaling, or MDS (Kruskal and Wish, 1978; Shepard, 1980). Nonmetric MDS is an algorithm for embedding a set of points in a metric space, while preserving as closely as possible the ranking of pairwise distances between the points, specified in advance (in psychophysical applications, this ranking is derived from the subject's response data). A number of recent works have explored the veridicality of parametric shape representation using MDS (Edelman, 1995a; Cutzu and Edelman, 1996; Sugihara et al., 1996). In each of a series of experiments, which involved pairwise similarity judgment and delayed matching to sample, subjects were confronted with several classes of computer-rendered 3D animal-like shapes, arranged in a complex pattern in a common high-dimensional parameter space (cf. Figure 5, left). Response time and error rate data were combined into a measure of subjective shape similarity, and the resulting proximity matrix was submitted to nonmetric MDS. In the resulting solution, the relative geometrical arrangement of the points corresponding to the different objects invariably reflected the complex low-dimensional structure in parameter space that defined the relationships between the stimuli classes (Figure 5, middle).

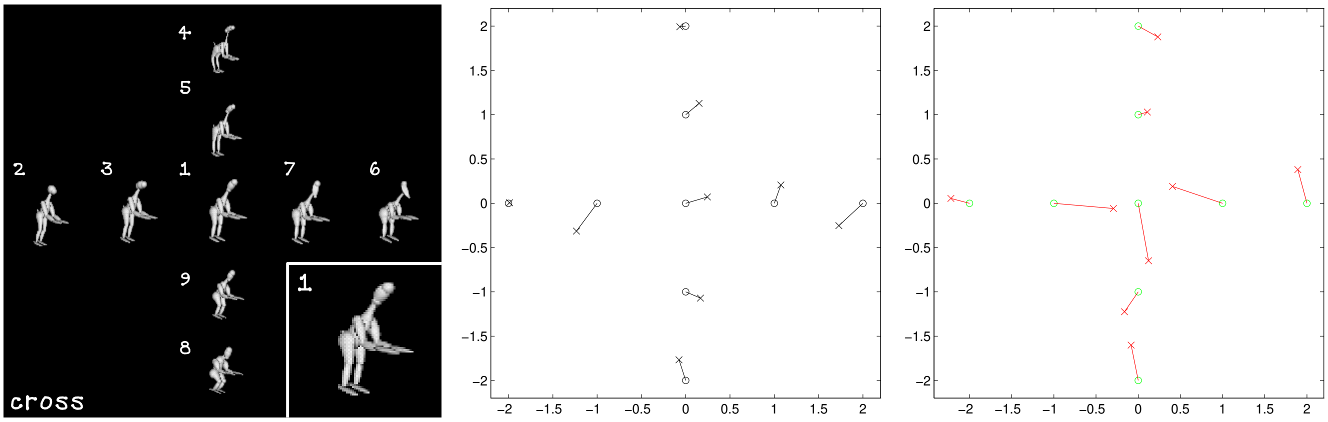

Figure 5:
Left: the parameter-space configuration used for generating the stimuli in one of the experiments described in (Cutzu and Edelman, 1996). *Middle:* the 2D MDS solution for the subject data. Symbols: $\circ$ – true configuration; $\times$ – configuration derived by MDS from the subject data, then Procrustes-transformed (that is, translated, rotated, and scaled; see Borg and Lingoes, 1987) to fit the true one. Lines connect corresponding points. The residual (Procrustes) distance between the MDS-derived configuration and the true one was 0.66 (expected random value, estimated by bootstrap from the data: 3.14 ± 0.15, mean and std dev; 100 permutations of the point order were used in the bootstrap computation). *Right:* the 2D MDS solution for the RBF ensemble model (see section 3.2); Procrustes distance: 1.11 (expected random value: 3.14 ± 0.17). Simulations showed that the recovery of the low-dimensional structure from *image-space* distances between the stimuli (as opposed to their representation in the RBF ensemble space) was impossible; see (Cutzu and Edelman, 1996) for details.

3.2 Computation

3.2.1 From Shepard's law of generalization to veridical representation of similarities

The results of the psychophysical experiments described above indicate that subjects internalize a veridical replica of the parameter-space configuration formed by the stimuli, which, in turn, can be recovered from the response data. Interestingly, the same mechanism that has been invoked in the past as a model of simple generalization with respect to a reference stimulus — a classifier whose response falls off exponentially with the distance between its preferred stimulus and the current input (Shepard, 1987; Shepard and Kannappan, 1993) — can also serve as a basis for modeling veridical representation of the shape space in humans, as explained below.

A monotonic fall-off of the likelihood of generalization with increasing distance between the test and the reference stimuli has been observed in a wide range of experiments, spanning visual and other modalities. Shepard (1987) surveyed empirical data on generalization gradients, and showed

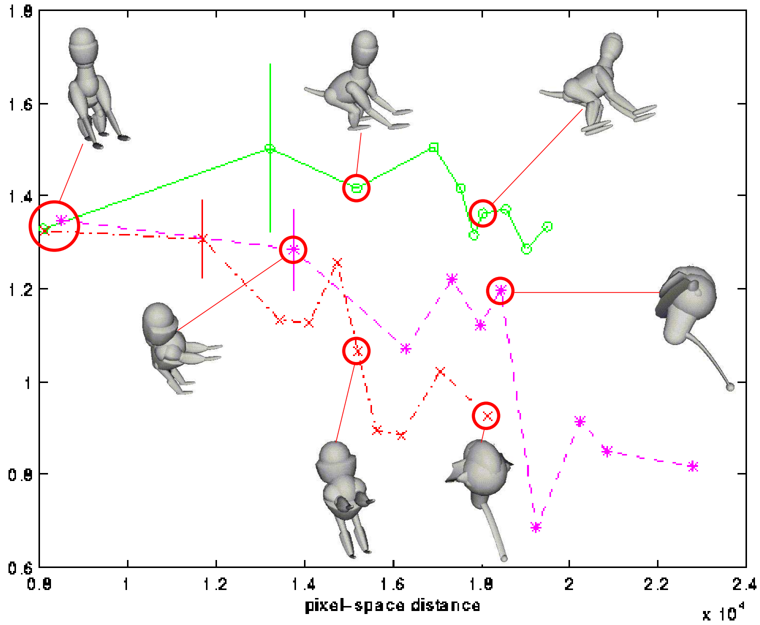

Figure 6: The response of a radial basis function module, trained on 10 random views of a parametrically defined object, to stimuli differing from a reference view of that object (marked by the big circle), in three ways: (1) by progressive view change, marked by $\circ$'s; (2) by progressive shape change, marked by $\times$'s; (3) by combined shape and view change, marked by $*$'s. The points along each curve have been sorted by pixel-space distance between the test and the reference stimuli (shown along the abscissa). Points are means over 10 repetitions with different random view-space and shape-space directions of change; a typical error bar ($\pm$ standard error of the mean) is shown for each curve. Note the insensitivity of the module's output to view-space changes, relative to shape-space changes.

that they can be described by a decaying exponential, $P(x;x_0) = e^{-d(x,x_0)/\sigma}$, where $P(x;x_0)$ is the probability of generalizing to stimulus x the response previously made to x_0, and $d(x,x_0)$ is the normalized distance in psychological (representation) space. Shepard (1987) also showed that this exponential generalization law can be derived from first principles, and is relatively insensitive to the form of the distance function entering the computation.

3.2.2 A multiple-classifier model

Now, consider a bank of k classifiers, each tuned to a particular point in the distal space (its optimal stimulus). Together, these classifiers implement a mapping M from the distal shape space to the proximal representation space, R^k. To fulfill the requirements for veridical representation discussed in section 2, the response of each classifier must be approximately constant over the range of different viewing conditions $\mathbf{z}$, and must fall off gradually and monotonically with parameter-space distance from its stimulus, as in Shepard's law of generalization. Such a system realizes a map M which is smooth and regular, and can, therefore, serve as a substrate for veridical representation of the original parameter space (any diffeomorphism restricted to a compact subset of its domain is quasiconformal; Zorich 1992, p.133).

A model of this type fully replicated the psychophysical results mentioned in the preceding section. Specifically, the MDS-derived configurations obtained from the performance data of the model showed significant resemblance to the true parameter-space configurations (see Figure 5, right). The model consisted of a number of classifiers (three or four, depending on the experiment), each of which incorporated a nearly viewpoint-invariant representation of a reference object, formed by training a radial basis function (RBF) network to interpolate a characteristic function for that object (Poggio and Edelman, 1990). Each RBF network was trained to output a constant value for a selection of views of the reference object; the views were encoded by the activities of an underlying receptive field layer. As a result, the classifier responded nearly equally to different views of the reference object it had been trained upon, and progressively less to objects that differed from the reference one in the shape space (see Figure 6). At the RBF ensemble level, the similarity between two stimuli was defined as the cosine of the angle between the vectors of outputs they evoked in the RBF modules.

3.2.3 Extension to a diverse collection of objects

The faithful representation of inter-object similarities acquires a somewhat different meaning if objects that do not belong to a homogeneous class are considered. In that case, it is not reasonable to expect that a system reflect any "proper" pattern of similarities inherent to the stimuli, simply because no such pattern exists for a collection of objects that is diverse enough (cf. section 2.3.2). Note, however, that the requirement of discounting the effects of the extraneous variables (associated with viewpoint, illumination, etc.) still stands. According to this requirement, a representational system must place views of the same object closer together in its internal shape space than views of different objects. Figure 7 illustrates the ability of the multiple-classifier model to do just that.

As a control, we first employed MDS to arrange four views of each of three objects (a human figure, a chess piece, and a candlestick) in 2D, according to their distances in the space spanned by a collection of receptive fields that covered the images. The left panel of Figure 7 shows that views in the resulting configuration were arranged by orientation of the object, and not by object identity. As in the model of the psychophysical findings described above, the correct clustering (by object identity) was obtained when the distances submitted to MDS had been computed in the space of activities of three RBF classifiers, each trained on other views of one of the objects (Figure 7, right).

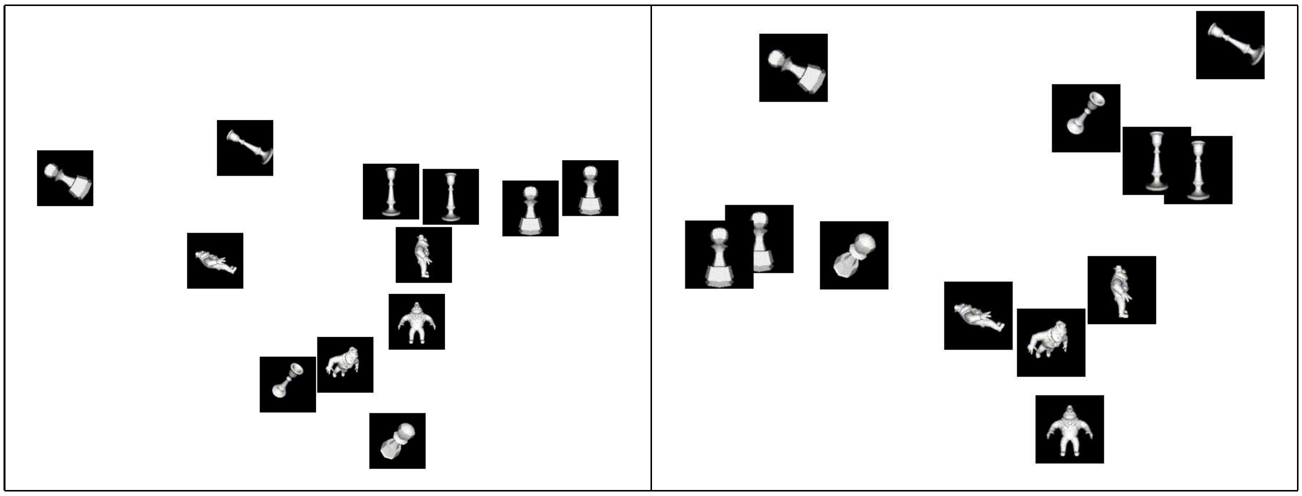

Figure 7: *Left:* The arrangement in 2D of different views of three objects, derived by MDS from distances computed in the space of activities of receptive fields covering the images (recall that the MDS procedure places views that are similar under the given metric nearby each other). 200 receptive fields, intended to serve as a crude model of the primate primary visual area, were used in this simulation. Note that the views are clustered by the orientation of the object. *Right:* The arrangement of the same views according to their distances in the space spanned by the outputs of three RBF classifiers, each trained on one of the objects (see sections 3.2 and 3.2.3). Note the clustering by object identity.

3.2.4 More on the multiple-classifier approach

In addition to its ability to replicate human performance in the perception of shape similarities, the multiple-classifier model described above possesses a number of other useful properties (Edelman, 1995b; Edelman et al., 1996). First, it is related to a powerful method of dimensionality reduction (Bourgain, 1985; Linial et al., 1994), in which points belonging to a multidimensional space are embedded into a space of much lower dimensionality, while preserving to a large extent the original interpoint distances. The importance of this relationship stems from the need for drastic dimensionality reduction that arises in biological visual systems, where the dimensionality of intermediate representation spaces (at the level of the optic nerve, for example) may be counted in the hundreds of thousands. Second, representing a stimulus by the responses of classifiers each of which is nonlinear (e.g., exponential, as in Shepard's law) in the measurement-space distance between the stimulus and some reference object introduces features polynomially related to the original measurement variables.[3] This may simplify subsequent classification, by nonlinearly enriching the representation space and making the possibility of linear separation between classes more likely (Cortes and Vapnik,

[3] Think of the Taylor expansion of $e^{-\|x-x_0\|}$ around x_0, where x is the stimulus and x_0 is a reference object.

1995). Third, responses of a number of classifiers can serve as a substrate for carrying out classification at superordinate, basic, or subordinate levels of categorization, depending on the manner in which these responses are processed. For example, summing the responses of classifiers tuned to members of a certain class gives a measure of the affinity of the stimulus to that class (Nosofsky, 1988), while considering the responses separately can allow fine individual-level distinction between stimuli. Fourth, if the saliency of individual classifiers in distinguishing between various stimuli is kept track of and is taken into consideration depending on the task at hand, then similarity between stimuli in the representation space can be made asymmetrical and non-transitive, in accordance with Tversky's general contrast model of similarity (Tversky, 1977).

It is interesting to compare the multiple-classifier approach to object representation to other recent proposals that avoid geometric reconstruction of individual objects. A typical example of the latter is representation by view interpolation. Because different views of a 3D object are known to lie on a smooth low-dimensional manifold embedded in the space of all views of all objects (Ullman and Basri, 1991), it is possible to employ standard function approximation techniques such as RBFs to determine whether an image belongs to a given object, the latter being represented merely by a set of its views. Indeed, this technique is incorporated into each classifier module in our multiple-classifier system. Note, however, that interpolation is useless if the object is novel. In fact, an interpolation module cannot be trained to produce the "right output" for novel objects, because such a thing is simply undefined. In contrast, the multiple-classifier system can still provide useful information in such a case (by signalling, e.g., the category affiliation of the stimulus rather than its identity, as suggested in the previous paragraph). In a sense, this is made possible by mimicking, on a higher level, the behavior of systems that carry out multiple feature measurements over the input image, and then classify it according to the histograms computed for the different features (Swain and Ballard, 1991; Mel, 1996); in our case, the measurements are similarities to entire objects. We conjecture that a combination of the two approaches with top-down processes capable of anchoring the representation activated by a stimulus in the retinal and the world coordinate frames would lead to the development of more comprehensive theories of vision.

3.3 Neurophysiology

The approach to representation based on a smooth distal to proximal mapping, and its implementation by the bank of classifiers, outlined in the preceding section, lead to explicit predictions regarding the processing of objects at the higher levels of the primate visual system. Specifically, one expects to find there units responding preferentially to certain objects, with the response falling off monotonically with dissimilarity between the stimulus and the preferred object, while staying nearly constant over different views of the preferred object (cf. Figure 6).

Although first reports of cells in the monkey inferotemporal cortex that respond preferentially to faces date back more than two decades (Gross et al., 1972), cells tuned to general objects have been found only recently. In particular, Tanaka and his group reported the desired selectivity for specific (mostly 2D) objects in recordings from the inferotemporal cortex (Fujita et al., 1992; Kobatake and Tanaka, 1994; Tanaka, 1992; Tanaka, 1993). More recently, Logothetis, Pauls and Poggio (1995) reported recordings from cells tuned to specific views of 3D objects on which the monkey had been trained. A small proportion of the object-tuned cells found by Logothetis et al. responded each to a limited subset of the objects, irrespective of view.

None of the above experiments involved parametric manipulation of the stimulus shape — a crucial component in testing the predictions of the theory of representation proposed here. In another study, where such manipulation was attempted, the stimuli were complex parametrically defined periodic 2D patterns (Sakai, Naya and Miyashita, 1994). In that study, the response of cells was found to decrease monotonically with parameter-space distance between the test stimulus and the preferred pattern to which the cells were tuned. With parametrically controlled 3D stimuli, it should be possible to look for cells that behave similarly to the RBF module whose behavior is illustrated in Figure 6. Such a cell would respond equally to different views of its preferred object, but its response would decrease with parameter-space distance from the point corresponding the shape of the preferred object. Furthermore, the responses of a number of such cells, each tuned to a different reference object, should carry sufficient information for classifying novel stimuli of the same general category. Finally, if the pattern of stimuli has a simple low-dimensional characterization in some underlying parameter space (as in Figure 5, left), it should be recoverable from the ensemble response of a number of cells, using multidimensional scaling.

4 Conclusions

A lawful correspondence between proximities of objects in an internal representation space and their distal similarities offers a principled solution to the Problem of Representation posed in the introduction, by showing that the internal space can mirror the visual world in a mathematically precise and behaviorally significant sense. This view of representation, derived from R. N. Shepard's (1968) idea of second-order isomorphism between the representation space and the world, possesses a number of advantages over the simplistic alternative ("pictures in the head"). It also constitutes a plausible model of representation in primate vision, as indicated by a series of recent psychophysical results.

We have seen that faithful representation by second-order isomorphism can be supported by a system of classifiers, whose graded-profile overlapping "receptive fields" are tuned to specific object classes in a common parametrically defined shape space. In sensory physiology, the notion of functional units selectively and reliably responsive to external stimuli (that is, units whose output covaries with the sensory input in a well-defined and predictable manner) has found a wide acceptance in the form of the feature detector doctrine (Barlow, 1979). A better understanding of the formal capabilities of a system of feature detectors in the context of shape vision, and of its performance as a psychological model, should help this doctrine evolve into a full-fledged theory of representation.

Acknowledgments

Thanks to A. Aertsen, G. Cottrell, F. Cutzu, N. Intrator, D. Mumford, T. Poggio, R. Shepard, and S. Ullman for useful discussions and suggestions, and to the two anonymous reviewers for constructive criticism. The experiments described in Figure 5 were carried out by F. Cutzu. K. Grill-Spector helped with the development of the shape morphing procedures. SE is an incumbent of the Sir Charles Clore Career Development Chair at the Weizmann Institute of Science.

References

Anderson, C. H. and Van Essen, D. C. (1987). Shifter circuits: a computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Science*, 84:6297–6301.

Barlow, H. B. (1979). The past, present and future of feature detectors. In Albrecht, D., editor, *Recognition of Pattern and Form*, volume 44 of *Lecture Notes in Biomathematics*, pages 4–32. Springer, Berlin.

Borg, I. and Lingoes, J. (1987). *Multidimensional Similarity Structure Analysis*. Springer, Berlin.

Bourgain, J. (1985). On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52:46–52.

Cohn, H. (1967). *Conformal mappings on Riemann surfaces*. McGraw-Hill, New York.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20:273–297.

Cummins, R. (1989). *Meaning and mental representation*. MIT Press, Cambridge, MA.

Cummins, R. (1996). *Representations, Targets, and Attitudes*. MIT Press, Cambridge, MA.

Cutzu, F. and Edelman, S. (1996). Representation of object similarity in human vision: psychophysics and a computational model. *Vision Research*. submitted.

Edelman, S. (1995a). Representation of similarity in 3D object discrimination. *Neural Computation*, 7:407–422.

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

Edelman, S., Cutzu, F., and Duvdevani-Bar, S. (1996). Similarity to reference shapes as a basis for shape representation. In Cottrell, G., editor, *Proceedings of COGSCI'96*, San Diego, CA. to appear.

Fujita, I., Tanaka, K., Ito, M., and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, 360:343–346.

Gallistel, C. R. (1990). *The organization of learning.* MIT Press, Cambridge, MA.

Goldstone, R. L. (1994). The role of similarity in categorization: providing a groundwork. *Cognition*, 52:125–157.

Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1972). Visual properties of cells in inferotemporal cortex of the macaque. *J. Neurophysiol.*, 35:96–111.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., and Thagard, P. R. (1986). *Induction: processes of inference, learning, and discovery.* MIT Press, Cambridge, MA.

Intrator, N. (1993). Combining Exploratory Projection Pursuit and Projection Pursuit Regression. *Neural Computation*, 5:443–455.

Kobatake, E. and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.*, 71:2269–2280.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling.* Sage Piblications, Beverly Hills, CA.

Linial, N., London, E., and Rabinovich, Y. (1994). The geometry of graphs and some of its algorithmic applications. *FOCS*, 35:577–591.

Logothetis, N. K., Pauls, J., and Poggio, T. (1995). Shape recognition in the inferior temporal cortex of monkeys. *Current Biology*, 5:552–563.

Marr, D. (1982). *Vision.* W. H. Freeman, San Francisco, CA.

Mel, B. (1996). SEEMORE: Combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition. Technical report, University of South California, Los Angeles, CA.

Nicod, J. (1930). *Foundations of Geometry and Induction.* Routledge & Kegan Paul.

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.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In Rosch, E. and Lloyd, B. B., editors, *Cognition and Categorization*, pages 259–303. Erlbaum, Hillsdale, NJ.

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

Quine, W. V. O. (1969). Natural kinds. In *Ontological relativity and other essays*, pages 114–138. Columbia University Press, New York, NY.

Quine, W. V. O. (1973). *The roots of reference.* Open Court, La Salle, IL.

Reshetnyak, Y. G. (1989). *Space mappings with bounded distortion*, volume 73 of *Translations of mathematical monographs.* Amer. Math. Soc., Providence, RI.

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.

Sakai, K., Naya, Y., and Miyashita, Y. (1994). Neuronal tuning and associative mechanisms in form representation. *Learning and Memory*, 1:83–105.

Schwartz, E. L. (1985). Local and global functional architecture in primate striate cortex: outline of a spatial mapping doctrine for perception. In Rose, D. and Dobson, V. G., editors, *Models of the visual cortex*, pages 146–157. Wiley, New York, NY.

Shepard, R. N. (1968). Cognitive psychology: A review of the book by U. Neisser. *Amer. J. Psychol.*, 81:285–289.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210:390–397.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323.

Shepard, R. N. and Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1:1–17.

Shepard, R. N. and Kannappan, S. (1993). Connectionist implementation of a theory of generalization. In Hanson, S. J., Cowan, J. D., and Giles, C. L., editors, *Advances in Neural Information Processing Systems 5*, pages 665–672. Morgan Kaufmann.

Sugihara, T., Edelman, S., and Tanaka, K. (1996). Representation of objective similarity among 3D shapes in the monkey. Invest. Ophthalm. Vis. Sci. Suppl. (Proc. ARVO). abstract.

Suppes, P., Pavel, M., and Falmagne, J. (1994). Representations and models in psychology. *Ann. Rev. Psychol.*, 45:517–544.

Swain, M. J. and Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision*, 7:11–32.

Tanaka, K. (1992). Inferotemporal cortex and higher visual functions. *Current Opinion in Neurobiology*, 2:502–505.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, 262:685–688.

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

Ullman, S. and Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:992–1005.

Väisälä, J. (1971). *Lectures on n-dimensional quasiconformal mappings.* Number 229 in Lecture Notes in Mathematics. Springer-Verlag, Berlin.

Väisälä, J. (1992). Domains and maps. In Vuorinen, M., editor, *Quasiconformal space mappings*, number 1508 in Lecture Notes in Mathematics, pages 119–131. Springer-Verlag, Berlin.

Zorich, V. A. (1992). The global homeomorphism theorem for space quasiconformal mappings. In Vuorinen, M., editor, *Quasiconformal space mappings*, number 1508 in Lecture Notes in Mathematics, pages 132–148. Springer-Verlag, Berlin.