Representation is Representation of Similarities

Shimon Edelman
Center for Biological and Computational Learning
Dept. of Brain and Cognitive Sciences
MIT E25-201
Cambridge MA 02142, USA
edelman@ai.mit.edu

February 1996
first revision, September 1996
second revision, February 1997

Abstract

Advanced perceptual systems are faced with the problem of securing a principled relationship between the world and its internal representation. I propose a unified approach to visual representation, based on Shepard's (1968) notion of second-order isomorphism. According to the proposed theory, a shape is represented internally by the responses of a few tuned modules, each of which is broadly selective for some reference shape, whose similarity to the stimulus it measures. The result is a philosophically appealing, computationally feasible, biologically credible, and formally veridical representation of a distal shape space. This approach supports representation of and discrimination among shapes radically different from the reference ones, while bypassing the need for the computationally problematic decomposition into parts; it also addresses the needs of shape categorization, and can be used to derive a range of models of perceived similarity.

Abstract

Advanced perceptual systems are faced with the problem of securing a principled (ideally, veridical) relationship between the world and its internal representation. I propose a unified approach to visual representation, addressing both the needs of superordinate and basic-level categorization and of identification of specific instances of familiar categories. According to the proposed theory, a shape is represented internally by the responses of a small number of tuned modules. Each of these modules is broadly selective to some reference shape, whose similarity to the stimulus it measures. This amounts to embedding the stimulus in a low-dimensional proximal shape space, spanned by the outputs of the active modules. This shape space supports representation of distal shape similarities which is veridical in the sense of Shepard's (1968) notion of second-order isomorphism (i.e., of correspondence between distal and proximal similarities among shapes, rather than between distal shapes and their proximal representations). Representation in terms of similarities to reference shapes supports processing (e.g., discrimination) of shapes radically different from the reference ones, while bypassing the need for the computationally problematic decomposition into parts, postulated by other theories. Furthermore, a general expression for similarity between two stimuli, based on comparisons to reference shapes, can be used to derive models of perceived similarity ranging from continuous, symmetric, and hierarchical, as in the multidimensional scaling models (Shepard, 1980), to discrete and non-hierarchical, as in the general contrast models (Tversky, 1977; Shepard and Arabie, 1979).

Keywords: representation, similarity, visual shape recognition, categorization.

1 Introduction and overview

1.1 Motivation

A common assumption underlying theories of vision is that a representation of the world — a geometrical replica (Marr, 1982), and possibly also affordances relevant to the required repertoire of actions (Gibson, 1966) — should be delivered to the decision-making stage of an intelligent system, natural or artificial. Achieving principled correspondence between the representation and the world is a challenging philosophical and computational problem. On the philosophical level, one would like to know how, in principle, is representation possible. In vision, for instance, one may ask, what can it be about the internal state of an observer seeing a cat on a mat that would make it refer to the shape of the cat?

A traditional answer to this question has been, for a long time, *similarity*. According to this view, which originated with Aristotle, an internal entity represents an external object by virtue of resemblance or isomorphism between the two: the representation of a tomato has something of the redness and of the roundness of the real thing.

Echoes of this idea, inherited by Berkeley and Hume from the Scholasts, can be found in present-day sources: "Representation of something is an image, model, or reproduction of that thing" (Suppes et al., 1994). Obviously, no one these days believes that a representation of a cat in an observer's brain is cat-shaped (or striped, or fluffy); rather, it is construed as a set of measurements which collectively encode the geometry and other visual qualities of a cat. Nevertheless, the philosophical foundation of the current theories of shape representation is still isomorphism: typically, it is assumed that structural (Biederman, 1987) or metric (Ullman, 1989) information stored in the brain reflects corresponding properties of shapes in the world, on a one to one basis.

Apart from philosophical troubles (Cummins, 1989), this approach also presents a formidable computational challenge, if the representation is to be veridical (i.e., if the geometry of each viewed shape is to be faithfully reconstructed from the proximal stimulus (Edelman, 1996)). Given the inherent imperfections of and the distortions introduced by the sensory channels (as manifested in the plethora of visual illusions), it is perhaps not too surprising that the human perception of shape is short of veridical in a variety of tasks, such as the estimation of local surface orientation (Koenderink et al., 1996), of local curvature (Phillips and Todd, 1996), or even of object size (Gregson and Britton, 1990). Now, it is certainly possible to learn fascinating lessons about the workings of the human visual system from the study of the cases in which it behaves non-veridically, nonlinearly, or downright peculiarly (Gregory, 1978; Gregson, 1988). Nevertheless, the central goal of the present paper — understanding how representation is at all possible — is probably better pursued by considering the cases in which the representations employed by the visual system do lead to veridical perception. As we shall see, lessons that can be drawn from these cases suggest a philosophically appealing and formally veridical approach to representation, that turns out also to be computationally feasible.

1.2 Representation by second-order isomorphism

In the processing of visual shape, some of the more striking instances of veridicality are found in experiments in which the subjects have to consider *similarities among shapes*, rather than the geometry of individual shapes (Shepard and Chipman, 1970; Shepard and Cermak, 1973; Edelman, 1995a; Cortese and Dyre, 1996; Cutzu and Edelman, 1996a). In these cases, the veridicality of the representation of the similarities among shapes is expressed in the consistency among subjects, and, when tested with parametrically controlled stimuli (Shepard and Cermak, 1973; Edelman, 1995a; Cortese and Dyre, 1996; Cutzu and Edelman, 1996a), in the agreement between the parameter-space patterns formed by the stimuli and their arrangement in a configuration obtained from the subject data by multidimensional scaling (more on this in section 7).[1] At the same time, human performance exhibits considerable departures from veridicality in perception (Koenderink et al., 1996; Phillips and Todd, 1996) and, especially, in recognition (Jolicoeur and Humphrey, 1996) of shapes (as opposed to perception and recognition of similarities among shapes).

How do people happen to be better judges of similarities among shapes than perceivers of shape? This state of the affairs should be expected if the visual system seeks a *second-order isomorphism* (Shepard, 1968) between similarities of shapes and similarities of the internal representations they induce, instead of a first-order isomorphism between the shapes and their representations: "...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). Essentially, this is a call for representation **of** similarity instead of representation **by** similarity (see Figure 1).[2]

[1]Agreement between patterns derived from visual and haptic perceptual data has been also reported (Garbin, 1990).

[2]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 (Palmer, 1978; Holland et al., 1986; Gallistel, 1990). Second-order isomorphism has been advocated recently in (Cummins, 1996), where it is called "The Picture Theory of Representation." This name

Place Figure 1 here

1.3 A computational theory of veridical representation

To provide a computational basis for the idea of representation of similarity, it is not enough merely to postulate, as J. J. Gibson did, that the relevant information is picked up or resonated to, without specifying the details of the pick-up process (Ullman, 1980; Marr, 1982). In the case of representation *by* similarity, the pick-up of external information amounts to a reconstruction of the visual world. Although it is quite easy to state, the reconstructionist goal is notoriously difficult to attain computationally, as signified by the limited success of Marr's research program in computer vision, and by the calls for alternative paradigms (Bajcsy, 1988; Aloimonos, 1990). Fortunately, as we shall see, reconstruction is not necessary, if representation *of* similarity is taken to be the goal of the visual system.

Computationally, the problem of representation can be addressed on several levels (cf. Marr, 1976). On the abstract level, the concern is to come up with an appropriate mathematical formulation, which would make the representation well-posed and tractable. The idea of second-order isomorphism does lead, in fact, to a well-defined computational notion of representation: according to this idea, to represent a collection of objects means to reflect in a consistent manner any change that an object may undergo.

By and large, this notion of representation is conceptually orthogonal to the reconstructionist approach: the tokens standing for objects need not resemble the objects themselves (see Figure 1). Although representation by second-order isomorphism does reduce to plain reconstruction if the represented quantities correspond to distances among densely spaced points situated on the surface of an object,[3] such a reduction is unwarranted; apart from placing a heavy computational burden on the perceptual system, it serves no useful purpose. As noted in (Shepard and Chipman, 1970), "it only attempts the absurdity of putting off until later the whole process of pattern recognition that must by definition precede the pivotal event in question" (i.e., the delivery of a representation capable of supporting perceptual judgment and categorization).

On the algorithmic level, representation by second-order isomorphism calls for assuring that the similarities between (necessarily proximal) perceived entities correspond in some orderly fashion to the distal similarities between objects. Now, a mechanism *tuned* to a particular shape provides a convenient way to estimate the similarity between the current stimulus and a reference one, if its response falls off monotonically with the extent of the (distal) deviation of the current stimulus from the preferred one. This monotonic relationship between proximal and distal similarities provides

is rather unfortunate, because in vision research pictorial representations are strongly associated with the Aristotelian notion of representation by similarity, or first-order isomorphism.

[3] As pointed out by S. Ullman in a personal communication.

the required algorithmic basis for veridical representation: as in nonmetric multidimensional scaling (Shepard, 1962; Kruskal, 1964), the rank order of the proximal similarities, being the same as the rank order of the distal similarities, allows the recovery of the distal configuration of the stimuli in some underlying parametric space (Edelman, 1995b).

On the implementational level, the challenge, then, is to identify a mechanism (biological or artificial), capable of responding selectively to certain shapes. A generic connectionist classifier, trained on the recognition of a particular class of objects, provides the required implementational substrate; a particular classification architecture (namely, the regularization networks of Poggio and Girosi, 1990) may be preferred on the grounds of biological plausibility.

An adequate computational solution, spanning all three levels, would exert a decisive influence on the philosophical outlook of the problem of representation. At the very least, familiar dogmas would have to be reassessed, and the relative merit of competing proposals reevaluated. The developments of the recent years in the computational, psychophysical, and neurobiological studies of visual representation suggest that the time for such a revision has come. In the remainder of this paper, I survey some of the relevant developments, and suggest a possible manner in which they can be related to some of the current views on the issue of representation in the philosophy of mind.

2 Representation of similarity: some preliminaries

I now proceed to describe in detail the computational-level approach to representation, outlined in the introduction. A standard answer to the central question at this level — what to represent? — is, not surprisingly, "shape." The surprise comes with the realization that an alternative answer is both plausible and preferable. The approach expounded below, which is closely related to Shepard's (1968) idea of representation by second-order isomorphism, offers such an alternative answer: *represent similarity between shapes, not the geometry of each shape in itself.*

2.1 Distal shape space

To be able to discuss second-order isomorphism, one must first define the two relevant similarity functions, one for the distal (represented) shapes, and the other for the proximal (representing) entities. I begin with the former.

Similarity between objects can be defined via an embedding of the objects into a metric space, where it is then determined by the distance between the points corresponding to each object. Rather than postulating a unique true distal similarity space for shapes (a notion that would appeal only to an extreme Platonist), I propose to consider an arbitrary space of the required kind, and to show later on that the exact choice of space is not critical.[4] What, then, should be required of such a distal *shape space*? Under second-order isomorphism, changes of shapes, and not the shapes themselves, are to be represented. According to this view, changing a shape corresponds simply to a movement of the point encoding the shape in an appropriate parameter space. To enable metamorphosis within a certain class of objects, all the members of that class must admit a *common parameterization.*

Although modern computer graphics offers a number of approaches to a common parameterization for a very wide spectrum of possible shape morphing (Pentland and Sclaroff, 1991; Galin and Akkouche, 1996) (see also appendix A), it is difficult to expect that a structure of similarities common to extremely disparate shapes will carry over into a cognitive system (the need to judge the similarity between objects from widely disparate categories arises rarely, if ever). Different object classes may, therefore, be encoded by different sets of parameters.

To some extent, the ease with which a common parameterization can be constructed for a set of objects probably depends on the degree of their membership in the same natural kind (Quine, 1969) of shapes (say, quadruped animals), or in the same artificial shape category (office tables). If any shape were equally likely (for a "medium-sized" count noun object), the burden of representing the visual world would be, I suspect, much heavier.

2.2 Proximal shape space

Defining similarity via proximity in an internal metric shape space is somewhat more problematic, as discussed by Gregson (1975, chapter 4). The main tool at the disposal of a psychologist who wishes to show that representations of a set of stimuli can be taken to form a spatial order is multidimensional scaling (see Shepard, 1980, for a review). Using this technique, it is possible to show that, in a wide variety of perceptual tasks, subjects behave as if they represent the stimuli by distributions of points in an internal similarity space of the kind that is needed here (Shepard, 1987; Nosofsky, 1992).

A degree of caution is called for when interpreting this state of affairs. First, the applicability of multidimensional scaling is ultimately determined by the relevance of the resulting solution: "Even though it is always the case that, if we are prepared to tolerate a high enough dimensionality and if we are prepared to tolerate degenerate, clustered, or lumpy configurations, we can get a spatial representation, ultimately, the criterion for accepting a representation is the sense that can be made of it, and the results that can be retrieved or predicted, by rules invariant over the space, from it" (Gregson 1975, p.134).

Second, one should not assume too lightly that the internal similarity space is metric in the full sense employed, e.g., in differential geometry. In that space, as pointed out by Clark (1993,

[4] The problem of alternative parameterizations is addressed in appendix C.

p.147), "Distances are monotonically related to similarities, but there is no presumption that sums or ratios of distances are interpretable. There may be no common unit to express distances along different axes." Fortunately, in visual shape processing these concerns seem to be largely mitigated; in section 7, we shall see that both the metric space assumption and the applicability of MDS are justified by the human performance data in a variety of shape perception tasks.

The metric-space definition of internal similarity seems to fall short of explaining such prominent phenomena in the perception of similarity as subjectivity, task dependence, and asymmetry (Tversky, 1977; Tversky and Gati, 1978; Nosofsky, 1991; Medin et al., 1993). These shortcomings are, however, only superficial. In particular, while the metric-space model makes it possible to speak about objective distal similarity (a prerequisite for a realist ontology of visual shapes), the perceptual system of the observer can warp the objective similarity space, according to his or her or its idiosynchrasies, and to the dictates of the task (Harnad, 1987; Goldstone, 1994). Furthermore, similarity need not remain restricted by the symmetry that it inherits from the underlying distance function; the metric-space model can be considered a starting point for a more realistic definition, e.g., of the kind proposed in (Krumhansl, 1978). Indeed, as I shall argue in section 5, a distance-based definition of similarity does not preclude one from modeling a considerable variety of similarity-related phenomena in human perception.

The possibility of a principled quantification both of the distal and of the proximal shape similarity addresses the first problem faced by the proposed theory of representation: what to represent. The next question — how to communicate similarity relationships induced by a given distal shape space structure across the gap separating the world from the observer — is addressed in the following section.

3 Representation of similarity: the problem

3.1 Levels of representation of similarity

Let us now consider the process of representation as a mapping from a distal to the proximal metric shape space. One may ask at this point, what properties must the mapping possess for the image of the original shape space to qualify as its faithful representation?

Distinctness. The minimal requirement appears to be that the mapping be one to one, so that distinct points in the original space are mapped to distinct points in the representation space.[5] To realize the implications of limiting the representational requirements to distinctness, note that a

[5]It is difficult to impose this requirement over all possible objects, unless the dimensions along which objects can vary are known in advance. Thus, any perceptual system is prone to the error of omission caused by the necessarily finite set of measurements that span its internal representational space.

major reason for maintaining internal representations is *generalization*: any system, at any point in time, will have encountered only a finite number of (labeled or rewarded) stimuli; for any other stimulus, the response will have to be generalized, based on memory traces of past experiences with related stimuli (Shepard, 1987). A representation whose fidelity is limited to distinctness provides no basis for generalization, because it does not contain information concerning relationships between stimuli, beyond the identity of each of them.

Nearest-neighbor preservation. A modicum of generalization capability is afforded by the requirement that the representation mapping preserve the nearest neighbor structure prevailing in the original space. In this case, two points that are nearest neighbors of each other before the mapping remain so after the mapping. This kind of representation preserves the structure of natural kinds, which, in turn, provides a basis for generalization (specifically, all objects more similar to some object O_1 than to O_2 will be represented as such, rather than merely as distinct both from O_1 and from O_2).

Full similarity spectrum preservation. If the identity of the k'th nearest neighbor of each point is preserved for some $k > 1$, the resulting representation will be in closer correspondence with the original space. At the limit, when the rank order of all interpoint distances for any finite set of points is fully preserved, the representation mapping becomes a similitude. The original shape-space configuration of the points can then be recovered from the distance rank information, up to rigid motion (Shepard, 1962; Kruskal, 1964; Shepard, 1980; Borg and Lingoes, 1987). A representation that possesses this degree of fidelity can support categorization at a number of levels, including the determination of the identity of the stimulus (see section 5).

The above hierarchy is clearly not the only possible way to define the fidelity of the representation mapping. If the representation is to be used mainly for classification, one may require that points that are separable under some parametric decision surface in the original space remain so following the mapping (this is in contrast with the distance-based requirements, which are nonparametric). For example, if points in the original shape space tend to form linearly separable clusters, one may require that the clusters remain linearly separable under the mapping. Moreover, one may also require that clusters that are not originally linearly separable become so under the mapping (Cortes and Vapnik, 1995). These considerations are beyond the main concern of the present section, which is to specify a *minimal* computational basis for the processes that operate on the representation space. Still, if the original-space configuration of stimuli allows an efficient remapping making explicit an underlying structure of linearly separable clusters, this possibility must remain open following the mapping into the representation space. Obviously, whereas the lowest-fidelity (distinction-preserving) representation does not necessarily preserve such properties, the highestfidelity (similarity-preserving) representation does.

3.2 Distal to proximal mapping M

In practice, the structure of the world is never perceived directly, but always through the more or less distorting channel of the distal to proximal mapping. If that channel lets some of the original dimensions of variation of stimuli collapse, the resulting representation runs the risk of not satisfying even the distinctness requirement, stated above. For example, in achromats, the perceptual dimensions of color are projected out of existence, giving rise to the emergence of a perceptual system separated from that of a normal person by a gap which cannot be bridged. A more complicated situation may arise when the transformation relating two representations is invertible but highly distorting. In that case, two systems may have widely different but not unbridgeable grasps on the world. A pair of stimuli which normally appear similar to one of the systems may seem dissimilar to the other.[6]

3.2.1 Constraints on the mapping M

Let us consider the constraints on the distal to proximal mapping M implied by the requirement that a representation should preserve similarity ranks everywhere in the shape space. A one to one mapping with this property must be a composition of scaling with rotation or reflection (Reshetnyak, 1989).[7] Thus, the requirement of global rank preservation is quite restrictive in the class of mappings it allows.

Locally, the rank preservation requirement is satisfied by *any* well-behaved (that is, smooth and invertible) mapping (Cohn, 1967). Such mappings are conformal, that is, they preserve angles, and, therefore, also the similitude of small triangles (see appendix B). In particular, a scalene triangle formed by a triplet of points in a distal shape space will be mapped into a triangle with the same ranking of side lengths in the proximal representation space (see Figure 1).

3.2.2 Component-wise analysis of M

How likely is a mapping M, implemented by a typical visual system, to meet the above requirements for distance rank preservation? Such a mapping can be described generically as a composition of four functions, $M = f_4 \circ f_3 \circ f_2 \circ f_1$, where the first two components, f_1 and f_2, are dictated by the properties of the world, and the other two constitute part of the system (see Figure 2):

[6]Two examples are the "other race" effect in face recognition (Brigham, 1986), and the distinction between the l and the r sounds, as it is perceived by a native speaker of Japanese, vs. a native speaker of English.

[7]One should keep in mind that scaling and other transformations mentioned in the present context pertain to configurations formed by objects in the shape space, and not to the objects themselves.

Geometry. The function $f_1(\mathbf{p})$ maps the distal parameter-space description $\mathbf{p}$ of the object into its geometry (e.g., the coordinates of the vertices of a fine mesh, suitable for rendering by a graphics system).

Imaging. The function $f_2(\mathbf{p}; \mathbf{z})$ maps the object's geometry into the image on the receptor surface of the visual system. Its dependence on the shape parameters $\mathbf{p}$ is determined by the prior action of f_1 and is written down explicitly for convenience; the dependence on the viewing conditions $\mathbf{z}$ is, however, peculiar to f_2.

Measurements. The function $f_3(\mathbf{p}; \mathbf{z})$ corresponds to the set of internal measurements performed on the image. In a typical model of biological vision, each measurement stage consists of a convolution with a number of filters, followed by the application of a nonlinearity.

Dimensionality reduction. The function $f_4(\mathbf{p})$ maps the measurement space into a low dimensional representation of the shape space, while removing the dependence on the viewing conditions $\mathbf{z}$. The low dimensionality of the ultimate internal shape space reflects the corresponding characteristic of the distal parameter space; it is also important for the reasons of computational tractability (Edelman and Intrator, 1997).

Place Figure 2 here

Note that the second component of M – the view mapping, f_2 – 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, $f_4 \circ f_3$, 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. 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 (Edelman and Duvdevani-Bar, 1997; more on this in section 4). 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; Lando and Edelman, 1995).

As pointed out above, the preservation of distance ranks implies that any change in the distal parameter space must be reflected in the final low-dimensional representation (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 the action of f_3) high-dimensional, and necessitates subsequent dimensionality reduction (through the action of f_4). 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 (for an introduction to this aspect of representation, see, e.g., Intrator, 1993).

4 Representation of similarity: a solution

4.1 Representation = measurement + dimensionality reduction

We have seen that veridical representation is theoretically possible, insofar as a low-dimensional subspace isomorphic (in Shepard's sense) to a distal shape space may be extracted from the high-dimensional space of measurements performed by the system. This situation is illustrated schematically in Figure 3. The input to an object recognition system – an $n \times n$ image – can be considered as a point in a n^2-dimensional image or *raster* space $\mathcal{R} = R^{n^2}$ (in biological vision, one may think of the space of patterns transmitted by the optic nerve to the brain). The task of a representational system is, given a pattern $\mathbf{X} \in \mathcal{R}$, to determine the location of $\mathbf{X}$ in a proximal shape space $\mathcal{S} \subset \mathcal{R}$.

Place Figure 3 here

The problem of locating $\mathbf{X}$ within $\mathcal{S}$ is analogous to the problem of determining the exact location of a point on a terrain, which arises in navigation and in the preparation of topographical maps. In topography, this problem can be solved by triangulation: the location of the point is computed from bearings taken to a number of landmarks whose coordinates are known. Likewise, the location of a point in the shape space can be found from its disposition with respect to a number of reference points, known to belong to the same space ("terrain"). This approach leads to a straightforward implementation of representation by second-order isomorphism, as described in the next section.

4.2 A Chorus of Prototypes

The main difference between triangulation in topography and in cognitive modeling is the quantity measured to provide the location of the test point. In topography it is easy to measure direction; in a biologically motivated model, distance (actually, a quantity monotonically related to distance). Consider a generic connectionist classifier, trained on instances of a certain shape class, which corresponds to a reference point or a prototype in the shape space. Note, first, that such a classifier can be made to learn from examples. A simple mechanism shown to be applicable, in particular, to visual object recognition is radial basis function (RBF) interpolation (Poggio and Edelman, 1990); other learning frameworks such as multilayer perceptrons trained by backpropagation are also applicable. An RBF module essentially interpolates the view space (see Figure 3) of the object on which is has been trained, starting from the exemplar views provided during training. As a result, the response of such a classifier is approximately constant over the range of the different viewing conditions.

If the classifier's response also falls off gradually and monotonically with parameter-space distance from the stimulus (the shape on which it has been trained; see Figure 4), it can be used to pinpoint the location of the test stimulus in the shape space, by a process related to triangulation and to nonmetric multidimensional scaling (Edelman, 1995b). Note that a number of classifiers, each tuned to a different reference point, must be activated (just as in triangulation a number of landmarks must be used for each measurement).

An ensemble or a *Chorus* (Edelman, 1995b) of k classifiers maps the distal shape space to a proximal representation space, R^k. If the response of each classifier degrades gracefully with the dissimilarity between the test stimulus and the preferred shape, the entire ensemble realizes a mapping M which is smooth and regular. Thus, the distal to proximal mapping is conformal[8] and can, therefore, serve as a substrate for veridical representation of the original parameter space, as argued in section 3.2.1.

The main reason to use a bank of classifiers rather than raw measurement-space distances to reference points for pinpointing the current stimulus is the possibility to train a classifier to ignore those directions in the measurement space that are irrelevant to the identity of the stimulus (e.g., directions corresponding to changes in the viewpoint parameters $\mathbf{z}$). Connectionist modelers have realized in the past that the response change caused by moving the stimulus away from a stored exemplar should depend on the direction of movement, if the space of admissible exemplars is a low-dimensional manifold immersed in the representation space. Specifically, moving along a tangent to that manifold should incur a smaller generalization cost than moving in a direction perpendicular to it. This insight has been incorporated into algorithms that train for invariance by differential reinforcement of stimuli removed in the tangent and the normal directions to the target manifold (Simard et al., 1992). In Chorus, invariance is not a goal, but rather a precondition that must be fulfilled for the resulting representation to be veridical. Furthermore, absolute invariance is not necessary: it suffices that the structure of categories, as defined by an appropriate metrics in the low-dimensional proximal representation space, not be distorted by the irrelevant components of distance, measured along the extraneous dimensions $\mathbf{z}$.

Training classifiers for particular stimuli, as it is done in Chorus, can be interpreted as downplay-

[8]Strictly speaking, it is quasiconformal (as is any diffeomorphism restricted to a compact subset of its domain; Zorich 1992, p.133), which means that it can be considered conformal on a small scale (see appendix B).

ing the irrelevant dimensions by switching from the measurement-space metrics to a representation-space metrics, induced by the class identities (Baxter, 1995). This property of the space spanned by the outputs of classifiers is important for devising better classification schemes. A typical example is vector quantization — a representational scheme in which the location of a point in a multidimensional space is coded by the identity of its nearest neighbor, chosen from a small set of points covering the space. In Baxter's (1995) canonical vector quantization, the distances to the covering points are computed according to the classifier metrics, and not the raw vector space metrics.

In comparison to the canonical vector quantization, in Chorus, the primary goal is representation, not classification. Accordingly, the computational question to be addressed is not whether the nearest-neighbor structure makes more sense when measured in the classifier space compared to the measurement space, but rather, to what extent the classifier-space distance structure of an arbitrary set of points reflects the corresponding structure in some low-dimensional distal parameterization. A preliminary empirical exploration indicates that classifier-space distances are indeed likely to behave in the desirable fashion (Edelman and Duvdevani-Bar, 1997b). The mathematical reason behind this property of Chorus may be its relationship 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. In Bourgain's embedding of a finite set of points, the locations of the points in the new space are encoded by their distances to randomly chosen subsets of the original set, which serve as reference entities. In Chorus, too, distances to reference points are measured: the response of a classifier trained on a reference pattern constitutes such a measurement, with the added advantage of tuning out the irrelevant dimensions. Thus, the employment of classifiers in Chorus makes Bourgain's principle of dimensionality reduction applicable in a situation where "noise" dimensions abound.

Place Figure 4 here

5 Uses of similarity

In the preceding section, we saw that the output of a Chorus of classifiers constitutes, under certain conditions, a veridical representation of a distal shape space to which the individual reference classes belong. I now turn to examine the extent to which this representation can be put to use in modeling the perception of similarity and its role in categorization. In this section, I will show that (1) the responses of a number of classifiers acting in parallel can serve as a substrate for carrying out classification at different levels of categorization, depending on the manner in which these responses are processed, and (2) 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).

5.1 Similarities at different levels of categorization

To understand the potential of the multiple-classifier representation to support shape categorization, it is necessary to consider the requirements of the relevant tasks at the different category levels.

5.1.1 Basic level

At the basic category level (Rosch et al., 1976), we are interested in the *identity* of the class $\mathbf{C}_j$ that is the closest neighbor of the stimulus $\mathbf{X}$ within the shape space $\mathcal{S}$. In some cases, the identities of several closest neighbors may be required (see Figure 5, middle). Note that at the basic level the identities of the neighbors should suffice for categorization, while at the subordinate level the knowledge of their disposition relative to the stimulus in the shape space may be required.

The major obstacle to be overcome at the basic level is the dependence of the appearance of the stimulus, $\mathbf{X}$, on factors such as illumination and viewpoint, in addition to the category membership index j. If $\mathbf{C}_j$ is taken to correspond to the image of a member of j in some canonical orientation, the viewing conditions can be seen to span a *view* space $\mathcal{V}_j$, which is transversal to the class space $\mathcal{C}$, and pierces it at $\mathbf{C} = \mathbf{C}_j$ (see Figure 3). A general-purpose function approximation module (Poggio and Edelman, 1990) trained to implement the "view normalization" mapping $T(j) : \mathcal{V}_j \rightarrow \mathbf{C}_j$ can perform basic-level categorization, because its response can be made largely independent of the viewing conditions.

5.1.2 Subordinate level

At the identity level, the task is to determine the exact location of the stimulus within the shape space, rather than its nearest neighbor(s) in the collection of known class prototypes. The central problem here lies in the fine resolution that must be attained despite the residual misalignment left over from the action of the normalizing transformation T. This problem can be approached by learning *hyperacuity* in the instance space. In hyperacuity-related visual tasks such as vernier discrimination (Westheimer, 1981), spatial resolution better than the spacing of the photoreceptors on the retina is attained by combined action of graded overlapping receptive fields (Snippe and Koenderink, 1992).

In shape-space localization, the response profile of each of the classifiers in Chorus defines a "receptive field" over the space $\mathcal{S}$. The vector of responses of a number of classifiers (Figure 5, right) contains the information necessary for pinpointing the location of the stimulus within $\mathcal{S}$, as argued in section 4. Moreover, because of graded nature of each response profile and the overlap between the different shape-space receptive fields, the localization is likely to be much more precise than what would have been possible if the responses of the classifiers were considered individually, in precise analogy to the spatial hyperacuity.

The required insensitivity of shape-space localization to viewpoint transformations stems from two sources. First, experience shows that hyperacuity can be attained despite considerable random misalignment of the stimulus as a whole, relative to its "home" or training pose, probably due to the shallow and overlapping profiles of the individual receptive fields (Poggio et al., 1992). Second, explicit training for invariance with respect to "irrelevant" transformations can complement the inherent tolerance of the receptive-field system. Importantly, once learned from examples, the normalizing transformation $T(j)$ can work even for stimuli not previously encountered by the system, provided that they belong to the same class as the examples used for training. The simplest approach here is to apply to a novel stimulus a transformation that is the average of the normalizing transformations learned for the class to which the stimulus belongs (Lando and Edelman, 1995).

5.1.3 Superordinate level

Consider now two tasks at a less specific level in a hierarchy of recognition tasks. The first of these is to decide whether the stimulus $\mathbf{X}$ is the image of *some* familiar object. For this purpose, it would suffice to represent the shape space $\mathcal{S}$ as a scalar field over the image space $S(\mathbf{X}) : \mathcal{R} \to R$, which would express for each $\mathbf{X}$ its degree of membership in $\mathcal{S}$. For example, one may set $S = \max_i\{p_i\}$ (the activity of the strongest-responding prototype module), or $S = \sum_i p_i$ (the total activity, as in Figure 5, right; cf. Nosofsky, 1988).

The second task is to characterize a superordinate-level category of the input image, and not merely decide whether it is likely to be the image of a familiar object. This can be done by determining the identities of the prototype modules that respond above some threshold. For example, if, say, the cat, the sheep and the cow modules are the only ones that respond, the stimulus is probably a four-legged animal.

Place Figure 5 here

5.2 Features of similarity

In Chorus, the response of each classifier p_i is, in a sense, a feature, whose value for a stimulus $\mathbf{A} \in \mathcal{R}$ is signified by the activation $p_i(\mathbf{A})$. Consider the similarity structure induced by this feature space over the universe of stimuli. With the qualifications stated in section 2, one can take the Euclidean distance between the feature vectors corresponding to two objects, $\mathbf{p}(\mathbf{A})$ and $\mathbf{p}(\mathbf{B})$, to be a default measure of the similarity between them: $s_E(\mathbf{A},\mathbf{B})^{-1} \sim \sum_{i=1}^{k} [p_i(\mathbf{A}) - p_i(\mathbf{B})]^2$. A uniform scaling in the responses of all prototype detectors $\mathbf{p} \to c\,\mathbf{p}$ (as in seeing through fog) should not, however, be interpreted as a change in the shape of the stimulus object. To make the similarity insensitive to such scaling, let us define similarity by the cosine of the angle between $\mathbf{p}(\mathbf{A})$ and $\mathbf{p}(\mathbf{B})$, in the space spanned by the prototype responses (cf. Ekman and Lindman, 1961):

$$s_a(\mathbf{A}, \mathbf{B}) \sim \sum_{i=1}^{k} p_i(\mathbf{A}) p_i(\mathbf{B}) \doteq \langle \mathbf{p}(\mathbf{A}), \mathbf{p}(\mathbf{B}) \rangle \tag{1}$$

This definition of similarity must, however, be further modified, at least for two reasons. First, s_a is independent of context, whereas perceived similarity depends on the "contrast set" against which it is to be judged. Second, s_a is symmetric, whereas human perception of similarity appears to be asymmetric in many cases (Tversky, 1977). To make s_a depend on the context, one can introduce a vector of weights, one per prototype, such that $w_i = w_i(\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots\})$. Thus, comparing $\mathbf{A}$ and $\mathbf{B}$ in two contexts, $\{\mathbf{A}, \mathbf{B} \mid \mathbf{C}, \mathbf{D}, \mathbf{E}\}$ and $\{\mathbf{A}, \mathbf{B} \mid \mathbf{F}, \mathbf{G}, \mathbf{H}\}$, may result in different values of similarity between $\mathbf{A}$ and $\mathbf{B}$. To model the asymmetry which frequently arises when subjects are required to estimate the similarity of some stimulus $\mathbf{A}$ to another stimulus $\mathbf{B}$, one may observe, following (Mumford, 1991a), that subjects in this case behave as if they take "$\mathbf{A}$ is similar to $\mathbf{B}$" to mean "$\mathbf{B}$ is some kind of prototype in a category which includes $\mathbf{A}$. Thus, the stimulus input $\mathbf{A}$ being analyzed is treated differently from the memory benchmark $\mathbf{B}$" (Mumford, 1991a; Medin et al., 1993). To give $\mathbf{B}$ the required distinction, each feature $p_i(\mathbf{B})$ can be weighted in proportion to its long-term saliency $\mathsf{sal}(p_i, \mathbf{B})$ in distinguishing between $\mathbf{B}$ and the other stimuli.[9] The resulting expression for similarity, which provides for the effects of context and for asymmetry, is

$$s(\mathbf{A}, \mathbf{B}) \sim \sum_{i=1}^{k} w_i p_i(\mathbf{A}) \left(\frac{p_i(\mathbf{B})}{\mathsf{sal}(p_i, \mathbf{B})} \right) \tag{2}$$

Note that this definition has the same form as the additive clustering (ADCLUS) similarity measure of (Shepard and Arabie, 1979), which, in turn, instantiates Tversky's (1977) discrete contrast model of feature-based similarity. At the same time, it is built on top of a continuous metric representational substrate – the shape space spanned by proximities to prototypes. The degree of compromise between these two approaches to similarity may depend on the demands of the task at hand, via the parameters of equation 2. At the one extreme, a Chorus-based system may behave as if it maps the stimuli pertaining to a task into a metric space, with the ensuing symmetric similarity and possible interaction among different dimensions; the other extreme may involve discrete all-or-none features, as in the examples surveyed by Tversky (1977).

[9]The computation of saliency can be carried out by a method such as Littlestone's (1988) *Winnow*.

6 Representation of similarity and other theories of what the brain may be doing

6.1 Making sense of novel objects

A central feature of the Chorus method is its ability to deal with novel objects (cf. Figure 7); once these are represented in terms of similarities to some of the reference objects, they can be remembered, recognized, or otherwise processed (Edelman and Duvdevani-Bar, 1997a). In theories of vision, this ability has been considered so far the prerogative of structural approaches to representation (Marr and Nishihara, 1978; Biederman, 1987). The structural approach employs a small number of generic primitives (such as the several dozen geons postulated by Biederman), along with spatial relationships defined over sets of primitives, to represent a potentially unlimited variety of shapes.

In principle, even completely novel shapes can be given a structural description, because the extraction of primitives from images and the determination of spatial relationships is supposed to proceed in a purely bottom-up, or image-driven fashion. In practice, however, both these steps proved so far impossible to automate, for reasons that may be non-accidental (Edelman and Weinshall, 1997). The few computer vision systems currently capable of unconstrained recognition from gray-scale images either ignore the challenge posed by the problems of categorization and of representation of novel objects (Murase and Nayar, 1995), or treat categorization as a byproduct of recognition (Mel, 1996).

In comparison to all these approaches, Chorus proposes to treat both familiar and novel objects equivalently, as points in a shape space spanned by similarities to a handful of reference objects. The viability of this method is attested to by the pilot implementation described in (Edelman and Duvdevani-Bar, 1997a), which achieved recognition performance on par with that of the state of the art computer vision systems, despite relying only on shape cues where other systems use shape and color or texture or both (Murase and Nayar, 1995; Mel, 1996; Schiele and Crowley, 1996). This performance was achieved with a low-dimensional representation (ten dimensions, compared to hundreds in other systems), whose extraction from raw images did not require, moreover, the problematic computation of a structural description. The employment of entire reference objects as high-level features suggests a link between Chorus and the studies of similarity and generalization in feature spaces, carried out by Shepard and others.

6.2 Similarity and memory-based generalization

Shepard's (1968, 1984) notion of second-order isomorphism is the closest to the present one among the previously proposed approaches to the understanding of representation. Interestingly, the computational approach to second-order isomorphism in Chorus is related to other work by Shepard — his law of generalization, which points out that the likelihood of obtaining the same response to two stimuli decreases exponentially with their separation in a psychological space, as defined, e.g., by multidimensional scaling (Shepard, 1987).

Within a connectionist framework, Shepard's law of generalization can be implemented in a straightforward manner, by constructing tuned units that exhibit radially symmetric exponential decay around the location of the preferred stimulus in a feature space (Hanson and Gluck, 1993; Shepard and Kannappan, 1993). However, it is computationally rather more interesting to note what happens when the radial "receptive field" of an exponential-decay unit is turned into an ellipsoidal one, by training the unit to ignore changes along some of the feature-space dimensions. In particular, if viewpoint-related changes in the appearance of a 3D shape to which the unit is tuned come to be ignored (e.g., through learning), the unit becomes a device capable of measuring the *shape-space* distance between the current stimulus and the optimal one. As we have seen in section 4, from here it is one step towards an implementation of the idea of representation by second-order isomorphism; all one has to do is have a number of tuned units acting in parallel.

A computational mechanism that is particularly suitable for the implementation of the tuned units is the regularization network (Poggio and Girosi, 1990). The simplicity of learning from examples in such networks, and the relatively straightforward manner in which they can be mapped onto the neurobiology of the brain prompted Poggio to revive the old notion of the function of the brain being largely that of a flexible memory, capable of learning from examples and of similarity-based classification (Poggio, 1990; cf. Hebb, 1949; Marr, 1970). It is important to realize, however, that by themselves neither these, nor many other learning-based approaches found in the literature can solve the problem of representation as posed in the introduction. The reason for this is that representation is not a problem of associating (whether by learning or otherwise) a proper output with a given input, simply because that which counts as "proper" differs from task to task (unless the world is represented by its replica, a choice that merely postpones the hard decisions by one stage). Thus, while different views of the same object are clearly to be associated with a constant response or mapped into a canonical view (Poggio and Edelman, 1990), there does not seem to be a useful universally valid specification of the proper response to a novel shape, e.g., one that is a parametric blend of two familiar shapes. Consequently, in a representational scheme learning must be augmented by generalization (a process whereby useful responses can be generated for novel stimuli). Thus, Chorus adopts the basic learning strategy by letting units become loosely tuned to certain familiar shape classes (invariantly over dimensions that are irrelevant to shape, such as viewpoint), **and** makes the existing tuned units collectively represent novel shapes, in a manner which allows their localization in an underlying low-dimensional shape space.

6.3 The new Pandemonium

The tuned modules of which Chorus is composed can be considered as "holistic" feature detectors, where the i'th feature of the stimulus is its similarity to the i'th reference object.[10] The concept of a feature detector originally developed under the influence of the discovery of "bug detectors" in the frog retina (Lettvin et al., 1959), and was linked to the notion of behavior-releasing mechanisms, borrowed from ethology (Barlow, 1979). Its generalization to higher perceptual functions such as shape recognition was subsequently attempted. A well-known proposal for an object recognition scheme based on feature detectors — the Pandemonium (Selfridge, 1959; Lindsay and Norman, 1977) — consisted of a three-level hierarchy: feature demons (responsible for the detection of lines, corners, etc.), cognitive demons (responsible for entire objects) and a master demon (responsible for the recognition decision). The limited influence of the Pandemonium model on computer vision (as opposed to psychological theories of shape processing) can be traced to two shortcomings.

The first problem with the Pandemonium is the choice of all-or-none primitive features, such as edges, corners, etc. This choice, which clearly violates Marr's (1976) principle of least commitment, is likely to lead to the loss of valuable information at an early processing stage; within the framework of section 2, it can be seen to render the distal to proximal mapping non-smooth, lessening the likelihood of veridical representation. This situation can be amended, if probabilistic features are used instead. According to this approach, sensory coding is "... the process of preparing a representation of the current sensory scene in a form that enables subsequent learning mechanisms to be versatile and reliable" (Barlow, 1990; Barlow, 1994). Specifically, a representation is useful for learning if it includes records of recurring and co-occurring events. In Barlow's Probabilistic Pandemonium, the response strength of a demon would be proportional to $-\log P$, where P is the probability of occurrence of the feature the demon detects (cf. Intrator and Cooper, 1992) .

The second problem with the Pandemonium lies at the level of decision-making (the "master demon"), where the stimulus is essentially described by the identity of the strongest-responding "cognitive demon." This winner-take-all decision (another violation of the principle of least commitment) does provide some information about the stimulus (namely, the identity of a reference stimulus to which the current one is the most similar), while discarding much more; the representation it provides qualifies only as nearest-neighbor-preserving, according to the terminology of section 3. Chorus improves on this by retaining the responses of a number of cognitive demons.

[10] The holistic nature of these features stems from the possibility of a reference shape being an entire object, rather than, say, a generic part.

6.4 Top-down effects and representation as explanation

A number of recent theories postulate an interplay between bottom-up and top-down influences in the processing of perceptual information (Carpenter et al., 1991; Carpenter et al., 1992; Mumford, 1991b; Mumford, 1992; Ullman, 1995; Hinton et al., 1995). Evidence from neurobiology (surveyed, e.g., in Ullman, 1995) strongly suggests that information can flow from the higher to the lower cortical areas and to the thalamus. The computational role of the top-down direction of flow of information may be clarified if one assumes that the goal of perceptual processing is to find a good (e.g., minimum description length) "explanation" for the stimulus (von Helmholtz, 1964; Dayan et al., 1995). Intuitively, it seems unquestionable that a human observer is capable of parsing even most complicated scenes into the constituent objects, in such a manner that every pixel eventually receives a label attributing it to this or that component. Such processing of scenes (as opposed to objects pre-segmented from their natural background) is a serious challenge for feedforward schemes such as Chorus.

The notion of "representation as explanation" does not contradict the idea that similarities between stimuli are to be represented, although in certain cases, such as scene processing, these two approaches offer largely orthogonal views on the problem of representation. On a conceptual level, the representation of a scene may well be a part of a cognitive schema (Rumelhart, 1980) in which it is embodied, and may, therefore, be encoded in terms of similarities to related schemata. Perceptually, however, scenes that fit the same schema (e.g., city street) are too diverse for the similarities to be informative, unless the computation of similarity involves explicit alignment of corresponding components (Markman and Gentner, 1993), or ignores shape details altogether. In the latter case, only gross violations of the schema structure, such as the appearance of a sofa levitating above a sidewalk (Biederman et al., 1982), are registered.

With some ingenuity, the theory behind Chorus may actually be interpreted in terms of the idea of representation as explanation. Specifically, the activity of the reference-object modules may be taken to model the probability distribution associated with the structure of the visual stimulus. In the case of single objects, this interpretation does not seem to be too problematic: a stimulus that is attributed both to the camel and the leopard modes in the probability (or explanation) space is simply taken to be a giraffe. In comparison, in the case of scenes (or, more generally, of objects that share common parts, which, in turn, come to be represented independently), an explanation of the stimulus requires an account of the spatial arrangement of the components, and not only of their identities. A natural approach to this problem is suggested by (Riesenhuber and Dayan, 1997), who propose to combine global configural and local template-like representations in a scheme that is driven by a top-down interpretation process (see also section 9.2).

In addition to dealing with compound objects and scenes, a Chorus-like scheme may benefit from top-down flow of information in deciding which stimuli are to be retained as reference objects, in gathering the statistical salience data for each reference object (section 5), and in control-related chores such as the computation of the target for the next fixation (cf. Koch and Ullman, 1985). By and large, however, Chorus embodies an attempt to find out how far a mostly bottom-up approach to representation can be taken. Perceiving the hidden causes of things is a feat worthy of Sherlock Holmes, and the human visual system seems to be capable of it, given enough time and a challenging task such as separating figure from ground in an underexposed photograph (Mumford, 1994, p.133). In less unique situations, including a variety of controlled experimental conditions, the performance of a perceptual Dr. Watson ("merely" making sense of the stimulus, as detailed in the next section, instead of accounting for each and every pixel, as expected from a Holmes) seems to be a goal both worthy of pursuit and more readily attainable.

7 Perception of similarity

According to the proposed theory of representation, to make sense of a stimulus means to locate it in a low-dimensional psychological space which (1) is inhabited by similar stimuli and (2) stands in a principled relationship to a low-dimensional physical space, such as a common parameterization of the stimulus set. The main tool in testing the predictions of this theory is multidimensional scaling (MDS), a computational procedure for embedding a set of points, one per stimulus, into a metric space in such a manner that the inter-point distances conform as closely as possible to perceived similarities (proximities) between the points, as measured in some psychophysical procedure (Kruskal and Wish, 1978; Shepard, 1980).

7.1 Background

Usually, MDS is employed in an exploratory mode, as follows. After the data are collected, the stimuli are embedded into a low-dimensional space, and the resulting configuration is inspected. The analysis is considered successful if the dimensions of the (psychological, or proximal) embedding space are correlated with some (physical, or distal) variables involved in the generation of the stimuli, and if the configuration of the stimuli in that space is meaningful. Among the examples of this procedure given in (Shepard, 1980), one finds the application of MDS to the processing of perceived similarities between Morse signals (the data were obtained by asking unskilled subjects to decide whether two consecutively sounded signals were same or different). The two dimensions of the embedding space in that example correspond to the number of components and the proportion of dots and dashes. Another example is the near-circular arrangement of colors in 2D, obtained by MDS from a table of judged similarities between color patches; this result supported Newton's suggestion to represent hues by points on a circle.

In the domain of shape perception, MDS has been applied in the analysis of perceived similarities among relatively simple 2D figures (rectangles, random irregular polygons), but the most spectacular results have been achieved in two studies that involved more complex shapes. In the first of these studies, subjects were requested to judge (from memory) the similarity of shapes of 15 of the US states (Shepard and Chipman, 1970). The 2D configurations obtained by MDS were surprisingly consistent across subjects, and also made sense geometrically (i.e., states of similar elongation and shape were grouped together). Shepard and Chipman point out that the findings of (1) very much the same configuration whether the states were pictorially displayed or only imagined, along with (2) the relationship, in both cases, between the recovered configuration and the actual cartographic shapes, support the idea of a second-order isomorphism between internal representations and their corresponding external objects.

In the second study, the stimuli (2D closed contours) were created parametrically, in such a manner that the set of shapes formed a toroidal configuration in the parameter space (Shepard and Cermak, 1973). The perceived similarities paralleled closely the parameter-space distances among the stimuli. Shepard and Cermak also report some interesting patterns of clustering that subjects imposed on the stimuli when prompted to consider possible categorical labels (such as "fish" or "jet plane") that could be applied to the (originally unmarked) 2D contours; these findings support the assertion, made in section 2.1, that a metric-space representation of similarity does not contradict the possibility of category-related effects, and, in fact, can provide the requisite substrate for the emergence of those effects.

7.2 Explorations of shape space

To obtain a more direct support for the second-order isomorphism idea, it is necessary to exert control over the original configuration built into the stimuli; the success of the recovery of that configuration from subject data can then be quantified and judged statistically. This corresponds to an application of MDS in a confirmatory, rather than exploratory, mode — an approach that can only be pursued with shapes that are generated with computer graphics and are controlled parametrically.

The veridicality of representation of parametrically defined 3D shapes in human subjects has been tested in two recent studies (Edelman, 1995a; Cutzu and Edelman, 1996a). In each of a series of experiments, which involved pairwise similarity judgment, delayed matching to sample, and long-term memory recall, subjects were confronted with several classes of computer-rendered 3D animal-like shapes, arranged in a complex pattern in a common parameter space. Response time and error rate data were combined into a measure of perceived pairwise shape similarities, and the object to object 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 stimulus classes (see Figure 6).[11]

Place Figure 6 here

The ability of the subjects to represent the low-dimensional pattern of similarities among stimuli did not extend to nonsense objects, as indicated by the results of control experiments involving "scrambled" shapes (Cutzu and Edelman, 1996a). The stimuli in these experiments were obtained by translating the parts of the animal-like shapes to a common center, resulting in star-like nonsense objects. For these objects, the similarity between true and MDS-recovered configurations was consistently lower than for animal-like shapes.

Computer simulations showed that the recovery of the low-dimensional structure from *image-space* distances between the stimuli was impossible, as expected. In comparison, the psychophysical results were fully replicated by a Chorus-like model, patterned after a higher stage of object processing, in which nearly viewpoint-invariant representations of familiar object classes (but, presumably, not of nonsense objects as in the control experiments; cf. Bülthoff and Edelman, 1992) are available; a rough analogy is to the inferotemporal visual area IT; see, e.g., (Young and Yamane, 1992; Tanaka, 1993; Logothetis et al., 1995). As pointed out in section 4, such a representation of a 3D object can be easily formed, if several views of the object are available, by training a mechanism such as a radial basis function network to interpolate a characteristic function for the object in the space of all views of all objects (Poggio and Edelman, 1990). A number of reference objects (in Figure 6, the corners of the parameter-space CROSS) were chosen, and a separate RBF network was trained to recognize each such object (i.e., to output a constant value for any of its views, encoded by the activities of the underlying receptive field layer; cf. Figure 4). At the RBF 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 trained on the reference objects (equation 1). The MDS-derived configurations obtained with this model showed significant resemblance to the true parameter-space configurations (see Figure 6, right).

7.3 Further predictions

The experiments mentioned above and the accompanying simulations indicate that the human visual system is capable of forming an internal representation of a set of stimuli which is second-order isomorphic to the original, furthermore, that a simple implementation of the Chorus scheme

[11] For further details, see (Cutzu and Edelman, 1996b). This finding has been recently replicated psychophysically in the monkey (Sugihara et al., 1996).

can exhibit a comparable capability for veridical representation. While the psychophysical findings support the idea of representation by second-order isomorphism, they are compatible with a number of possibilities of implementing the appropriate distal to proximal mapping, other than Chorus. In fact, given the claim that a veridical representation is obtained generically if the mapping is smooth (section 2), one should look in the data for traits that are peculiar to Chorus, and are not easily explained either by a reconstructionist interpretation (which seems unlikely, in view of the results of the control experiments), or by alternative mappings. Specifically, it should be possible:

- To predict, for each subject, the distortion in the MDS configuration for one parameter-space pattern, given the distortion of another pattern. A better prediction is expected from the Chorus model, compared to a generic warping scheme that does not rely on distances to reference points.
- To quantify the importance of parameter-space distances from the stimulus to preset reference points. A stronger effect of the change of these distances is expected, compared to a parameter-space movement that preserves the relative distances to the reference points; preliminary results compatible with this prediction have been reported in (Edelman et al., 1996).
- To test the nature of the reference shapes, using priming. Stronger priming is expected for familiar shapes (including the so-called "impossible" objects), relative to less familiar ones. In comparison, the generic reconstructionist hypothesis (Biederman, 1987), according to which representations are constructed "on the fly" by putting together universal primitives, seems to predict uniform priming for all objects, and less priming for the "impossible" ones.

8 Neurobiology of similarity

The approach to representation based on a smooth distal to proximal mapping, and its implementation by the bank of classifiers, lead to explicit predictions regarding the mechanisms of object processing 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 4).

Although reports of cells in the monkey inferotemporal cortex that respond preferentially to faces span decades by now (Gross et al., 1972; Perrett et al., 1989), 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 (IT) cortex of anesthetized monkeys (Tanaka et al., 1991; Fujita et al., 1992; Kobatake and Tanaka, 1994; Tanaka, 1992; Tanaka, 1993). The interpretation of such findings has been hampered traditionally by the unknown nature of the optimal stimuli for the discovered cells: if a cell responds as vigorously to a brush as to a face, it cannot be properly considered a face detector. Rather than attempting the impossible (i.e., ruling out all the stimuli that the cell does *not* like), Tanaka developed an ingenious method for narrowing down the range of features that are both present in a given stimulus and effective in eliciting a response from the cell. This method has yielded the first evidence of the parallel between the functional organization of the IT cortex, where cells responding to similar shapes are arranged in columns running perpendicular to the cortical surface, and the primary visual cortex, where the columnar structure reflects orientation selectivity and ocular dominance.

Although the columnar organization of the IT cortex has been interpreted in terms of an alphabet of "elementary" features, it seems to be equally compatible with the notion that entire objects are represented, as called for by the Chorus model (Tanaka, 1993). Under this interpretation, the several hundred columns that can be squeezed into the available cortical area correspond to so many classes of "reference" stimuli. If the tuning properties of the columns are such that any stimulus likely to be encountered activates a number (say, three or four) columns, the entire system should possess a considerable representational power. Moreover, this power would grow if the system were plastic enough to attune itself to novel object classes, as it may be the case in practice (Rolls et al., 1989; Kobatake et al., 1992).

More recent data support this interpretation of Tanaka's findings: working with awake monkeys, Logothetis, Pauls and Poggio (1995) reported recordings from cells tuned to specific views of 3D objects (other than faces) 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. Together with the previous reports of a hierarchical two-stage approach to (relative) invariance in the face cells (Perrett et al., 1989), these findings suggest that a cell that responds to a certain shape nearly irrespectively of its view (and that in a cortical implementation of Chorus would be termed a reference-object or "prototype" cell) may do so by integrating the responses of several cells each of which prefers another view of the same shape, as suggested in section 4 (Poggio and Edelman, 1990; Edelman and Weinshall, 1991).

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 4. The specific predictions are as follows:

- The cell will respond equally to different views of its preferred object, but its response will decrease with parameter-space distance from the point corresponding the shape of the preferred object (three such cells have been reported by Logothetis et al., 1995).
- The responses of a number of cells, each tuned to a different reference object, will carry sufficient information for classifying novel stimuli of the same general category as the reference objects.
- If the pattern of stimuli has a simple low-dimensional characterization in some underlying parameter space (as in Figure 6, left), it will be recoverable from the ensemble response of a number of cells, using multidimensional scaling.

9 Discussion

9.1 Similarity: the raw and the processed

In shape perception, traditionally, the foremost information-processing challenge has been to achieve object constancy, that is, to perceive the object's shape despite wide variations in its visual appearance, caused by changes in the illumination and in the object's position with respect to the observer. The proponents of constancy observe, with Heraclitus, who pointed out that one cannot step into the same river twice, that people literally never see the same object twice: objects are scaled up or down, translate, rotate, articulate, deform, are lit or shadowed, and are occluded by other objects or obscured by fog.

This observation is both true and misleading. Stressing the influence of the viewing conditions on the appearance of objects tacitly assumes that it is the exact shape of the object that a representational system should attempt to recover. However, as students of categorization know well, an intelligent agent is much better off representing an object on a number of hierarchical levels of abstraction (with the option of attending to high-resolution details, if the object happens to be present in front of the observer, and if the task demands it), than storing a high-resolution replica of the object, and facing the problem of separating the chaff (pixel-level information) from the wheat (classification information) every time a new instance of that same object class is encountered.

When considered with the goal of proper representation of similarity in mind, the problem of variability of object appearance assumes somewhat different proportions. At the computational level, instead of seeking absolute *invariance* with respect to the extraneous view-related parameters, a system can settle for mere *tolerance*, as determined by the interplay of within- and between-category similarities. At the implementational level, the availability of learning modules that can be trained to compensate for the variability in object appearance shifts the focus from the easier problems in vision (of which invariance seems to be an example) to the more challenging ones, such as making sense of objects not previously seen. The Chorus scheme, built around a theory of **representation of similarity**, and implemented by a bank of trainable modules tuned to reference objects, embodies both the computational and the implementational-level lessons stated above.

9.2 Some challenges

The holistic treatment of objects, adopted by the present theory, results in representations that can be easily learnable from examples, but must be further worked upon if required to support inferences concerning hierarchical structure. For example, one can perceive the numerals on the face of a bent clock in Dali's *Persistence of Memory* as shapes in themselves, as well as see them as parts of the whole. It may be possible to address this requirement, to some extent, by coupling mechanisms that are selective for scale and retinal location with those that are selective for shape (Edelman, 1994). A well-founded approach to such a coupling, built around a recently developed computational mechanism called the Helmholtz Machine (Dayan et al., 1995), has been implemented and tested (on stylized face images) by Riesenhuber and Dayan (1997).

According to the reasoning of Dayan et al., complex underconstrained perceptual tasks require an intimate cooperation between bottom-up, or data-driven processes, and top-down, or expectation-driven ones. Their arguments resemble those of other proponents of the Helmholtzian strategy, mentioned briefly in section 6.4, and are related to Grenander's notion of Pattern Theory (opposed to and complementing mere pattern recognition), as recently advocated by Mumford (1994). Returning to the example of Dali's painting, one may observe that people are not only aware of the clocks that appear in it, but also of their twisted and bent shapes. Indeed, making sense of this painting may require knowledge of the possibility of objects bending without losing their identity.[12] The extension of the Chorus framework to deal with this and similar cases will have to await future work; one possible direction that such a development could take would be based on the ideas of class-based processing (Moses et al., 1996; Lando and Edelman, 1995), and of example-directed metamorphosis (Beymer and Poggio, 1996).

[12] For a striking report of the malleability of object representations that emerge in a developing cognitive system, and, in particular, see the work of Landau, Smith, and Jones, 1988. In this study, it was found that children's assumptions regarding the admissible deformations that an object may undergo while retaining the same count-noun name depended on the object's appearance: deformations of furry convoluted objects (as compared to a single example view) were tolerated to a much larger extent than deformations of angular artifact-like things.

9.3 Philosophical implications

Some of the philosophical implications of the Chorus scheme have been mentioned briefly in (Edelman, 1995b); here, I discuss at a greater length the place of the proposed theory in the current philosophical debate on the nature of representation, stressing its relationship with the increasingly influential idea of the world as an external memory.

9.3.1 Locke's conformity and Shepard's second-order isomorphism

In describing the implementation of Chorus (section 4), I have suggested that the modules tuned to specific shapes can be considered as feature detectors, spanning a feature space in which each dimension codes similarity to a particular object class. The idea of a feature detector as a basic ingredient of a representational system can be traced back to John Locke, who was among the first to fully realize the infeasibility of Aristotelian representation by resemblance. Because the firing of a feature detector is an event which is internal to the representational system, this raises immediately the problem of grounding (cf. Harnad, 1990) the representation in reality:

> 1. Objection. "Knowledge placed in our ideas may be all unreal or chimerical." ...If our knowledge of our ideas terminate in them, and reach no further, where there is something further intended, our most serious thoughts will be of little more use than the reveries of a crazy brain ...
>
> 2. Answer: "Not so, where ideas agree with things."
>
> [Locke, 1690, *Book IV (Of Knowledge and Probability), Chapter IV.*]

The principle on which Locke based his answer to the grounding problem is that of "conformity," postulated to prevail between the representations and their objects. As is well known, Locke distinguished between simple and complex ideas, each kind with its own grounds for conformity. Consider first the former, somewhat less problematic, kind. The argument here was that "...the idea of whiteness, or bitterness, as it is in the mind, exactly answering that power which is in any body to produce it there, has all the real conformity it can or ought to have, with things without us. And this conformity between our simple ideas and the existence of things, is sufficient for real knowledge." (Locke 1690, Book IV, Chapter IV, 4). In terms of feature detectors, this is a statement of belief in the availability of reliable detectors for immediate perceptual qualities.

The finding of cells tuned to well-defined features such as patterns of motion (Movshon et al., 1985; Newsome and Paré, 1988), 2D shapes (Tanaka et al., 1991; Kobatake and Tanaka, 1994), or faces (Gross et al., 1972; Perrett et al., 1982) supports this part of Lockean doctrine, and, in fact, suggests that it may be extended from "simple" features to entire objects. The impact of this evidence seems to have been limited by a persistent concern that the feature detectors do not "really" detect the features they happen to be tuned to (Dretske, 1981; Fodor, 1987; Cummins, 1989).[13] Nevertheless, it has been suggested (Albright, 1991) that philosophical worries regarding possibility of Lockean conformity in the functioning of feature detectors found in the brain should be quelled to some extent by the successful manipulation of the organism's perception of a feature through the injection of current in the vicinity of the appropriate detector pool in the cortex (Salzman et al., 1990).

More importantly, in the light of the possibility of veridical representation of distal changes by proximal ones, as in Shepard's (1968) theory of second-order isomorphism, the philosophical lure of settling the question regarding what this or that individual feature detector "really" detects is significantly reduced. Moreover, the problematic distinction between simple and complex ideas suggested by Locke can be given up: in Chorus, the "feature detectors" can be tuned to arbitrarily complex objects, yet serve as primitives just as learnable[14] and as immediately perceivable as Locke's simple ideas. At the same time, if second-order isomorphism can be made to work, Locke's "conformity" acquires a new concrete meaning: the order and the connection of ideas is identical to the order and the connection of things.[15]

9.3.2 A new angle on compositionality

According to this view, a representational system need not possess a *combinatorial* mechanism for creating complex "ideas" out of simple ones. In vision, the hypothesis of the combinatorial structure of concepts assumes the guise of part-based theories of object representation (Biederman, 1987; Bienenstock and Geman, 1995). The debate between theories that involve dynamically bound generic parts and prototype-based theories parallels the classical dispute between Empiricist and Rationalist theories of concepts, in which the main argument against prototype-based theories is their alleged failure to support compositionality and productivity (Fodor, 1981, p.296). That argument, however, hinges on a logicist approach, which does not recognize any way of combining simple concepts into complex ones, short of logical/syntactical connectives.

On Fodor's interpretation of Empiricism, a system equipped with, say, three object-specific modules, tuned to the shapes of a tuna, a cow, and a car, has only three (indivisible) visual concepts: tuna, cow, and car. In fact, however, such a system turns out to be capable of representing a variety of other shapes, some of which are quite unlike the shapes for which dedicated modules are available (cf. Figure 7). Here and elsewhere in cognitive modeling, the logicist approach insists on indivisible primitives and logical connectives, effectively forcing the violation of the principle of least

[13] cf. the debate whether the simple cells in the mammalian primary visual cortex are really line detectors or local Fourier analyzers (Hubel and Wiesel, 1959; Maffei, 1978).

[14] By ostension, as in "this is a cat" (pointing to a cat); see Quine, 1969.

[15] "Ordo et connexio idearum idem est ac ordo et connexio rerum," Spinoza, Ethics II, 7.

commitment. As a result, logicists cannot but predict a representational capacity that falls far short of the empiricist predictions based on coarse coding, which, in this example, means falling short of the experimental observations.In contrast, if the stimulus is compared simultaneously to a number of graded prototypes, instead of being subjected to a Pandemonium-like all-or-none logical/syntactic analysis, the productivity problem vanishes, along with the premise for Fodor's argument.

Place Figure 7 here

9.3.3 The world as its own representation

In a passage intended to deflect criticism coming from the proponents of fuzzy-set interpretation of the notion of a prototype, Fodor (1981, p.297) admits that prototype theories may be able to handle the combinatorics of defining the *extension* of terms, but not their *sense*. Extension, however, may be all there is to a representation.

Indeed, the idea of second-order isomorphism places the burden of representation where it belongs — in the world. In Chorus, the ensemble of feature detectors responds (J. J. Gibson would say, resonates) to the environment (while extracting task-specific information), without reconstructing it internally. By merely mirroring proximally the similarity structure of a distal shape space, Chorus embodies the ideas of those philosophers who argued that "meaning ain't in the head" (Putnam, 1988) and that "cognitive systems are largely in the world" (Millikan, 1995), circumvents the severe difficulties encountered by the reconstructionist approaches in computer vision, and may explain the impressive performance of biological visual systems, which, in any case, appear to be too sloppy to do a good job of reconstructing the world geometrically (O'Regan, 1992). Thus, in an important sense, Chorus lets the world be its own representation.

9.3.4 Qualia

If the world is its own representation, how are we to explain phenomenological qualia (Goodman, 1977) such as the redness of a tomato or the shape of a pear, as perceived subjectively? The Aristotelian representation *by* similarity appealingly solves the qualia problem by equating these perceptual qualities with the physical qualities of the corresponding percepts (i.e., the internal representations). Thus, a shift towards the view of representation *of* similarity carries with it a price. The standard version of the problem of qualia seems to be actually exacerbated: on the face of it, it is more difficult to explain the apparent richness of the perceived world, if one denies that the shape of each of the constituent objects is in itself fully represented.

A partial solution to this problem is suggested by the realization that the apparent richness of the perceived world is, to a considerable extent, apparent (Dennett, 1991). The source of this illusion may lie in the immediate availability of the information in the world, which acts as an "external store" (O'Regan, 1992).[16] A growing number of psychophysical experiments supports this view (Pollatsek et al., 1984; O'Regan, 1992; Blackmore et al., 1995; Rensink et al., 1995; Grimes, 1995). In these experiments, subjects are typically found to be unaware of moderate, or, at times, major changes made in the visual stimulus during the "blanking" period associated with a saccade, or induced artificially, by presenting two stimuli frames in succession, with a short-duration gray-field mask interposed between them. For instance, changes such as the disappearance (or the appearance) of pieces of furniture in a room scene, or the sudden growth (by a significant fraction) of the tallest building in a city skyline scene, may go unnoticed. This suggests that, under normal viewing conditions (i.e., without scrutiny), much less information than previously assumed is taken away from each scene.[17]

While Dennett's insights do mollify the acuteness of the qualia problem to a certain degree, they do not appear to be able to do away with it. In particular, we are still left with the need to explain why and how a tomato looks round and red to the observer who represents directly only the differences between tomatoes and, say, pears and oranges (as opposed to the shape and the color of the tomato). Interestingly, an explanation here may be less elusive than commonly thought: an accomplished account of qualia in psychophysiological terms has been formulated recently around the notion of a quality space (analogous to the shape spaces discussed earlier in this paper), reconstructed from an observer's responses using multidimensional scaling (Clark, 1993). Adding to the thoughts of Carnap and Goodman a great deal of data from psychology and physiology, Clark shows that, in principle, it is not impossible to characterize a perceptual experience in objective terms, starting from relative similarity defined over tuples of objects — the very notion that constitutes the foundation of the second-order isomorphism theory (see appendix D).

9.4 Concluding remarks

I have presented a theory of shape representation, based on Shepard's notion of second-order isomorphism between the similarity structure of the internal representation space and that of the world of objects. The highlights of the proposed theory are as follows:

- *Formal veridicality.* Representations are grounded in physical reality. This is expressed by a correspondence between proximal and distal similarities, which, under certain conditions, allows for formal veridicality.

[16] Cf. Berkeley (1710): "Upon shutting my eyes all the furniture in the room is reduced to nothing, and barely upon opening them it is again created."

[17] See also (Biederman et al., 1974). In memory research, this point seems to be more widely accepted, in the form of the schema theories (Bartlett, 1932; Rumelhart, 1980). For some notes of caution, see (Cavanagh, 1995; Koriat and Goldsmith, 1995).

- *Unifying approach.* The representational substrate is a feature space, spanned by similarities to reference objects. The feature-space approach offers the possibility of a smooth integration between the processing of shape and of other visual dimensions. Furthermore, it provides a common representational substrate for cognitive tasks at different levels of categorization.

- *Learnability.* Representations can be learned from examples, using a well-understood computational mechanism (regularization networks).

- *Empirical support.* There is a natural mapping of representation of similarity onto well-defined neurophysiological mechanisms (ensembles of tuned units). This mapping is indirectly supported by psychophysical data, and by a functional-level simulation in an artificial neural network model.

- *Philosophical appeal.* The proposed theory takes a clear stand on philosophical issues that have been intensely debated for a long time. It also offers an opportunity to increase the productivity of the debate, by encouraging the consideration of relevant arguments from adjacent disciplines.

To conclude, let us return to the Riddle of Representation, as posed in the introduction: by virtue of what does the representational state of a human observer seeing a cat on a mat refer to that cat (Cummins, 1989)? A slightly different formulation of this riddle — what is common to two humans, a robot, and a Martian, who all see a cat on a mat? — may actually point towards a solution: it seems likely that the only thing that *can* be common to these four representational systems is the cat itself, sitting "out there" on the mat. One way to implement the idea of the world as its own representation is by constructing a system that has at its disposal tunable modules which can be trained to respond to cats or dogs or any other object. Such a system will be representing a cat when it sees one (by virtue of firing of the appropriate modules), and will be able also to dream of a cat or imagine one (if the modules are made to fire in the absence of an immediate sensory stimulation). Moreover, if a selection of modules (not more than a few hundred), each tuned to a different class of stimuli, is available, the system should be able also to represent (through the response of a small subset of the modules at a time) many more stimuli in addition to those actually stored in memory.

Acknowledgments

The title of this paper is a paraphrase on W. V. O. Quine: "To be is to be the value of a bound variable." I thank to A. Aertsen, G. Cottrell, F. Cutzu, P. Dayan, S. Duvdevani-Bar, N. Intrator, D. Lloyd, D. Mumford, A. O'Toole, T. Poggio, R. Shepard, and S. Ullman for useful discussions and suggestions. I am grateful to Sharon Duvdevani-Bar for Figures 4 and 7, and to Florin Cutzu

for Figure 6. Supplementary material for this article (including papers cited as "submitted" or "in press") can be found at http://eris.wisdom.weizmann.ac.il/~edelman/archive.html.

References

Albright, T. D. (1991). Motion perception and the mind-body problem. *Current Biology*, 1:391–393.

Aloimonos, J. Y. (1990). Purposive and qualitative vision. In *Proc. AAAI-90 Workshop on Qualitative Vision*, pages 1–5, San Mateo, CA. Morgan Kaufmann.

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.

Bajcsy, R. (1988). Active perception. *Proc. IEEE*, 76(8):996–1005. special issue on Computer Vision.

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.

Barlow, H. B. (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research*, 30:1561–1571.

Barlow, H. B. (1994). What is the computational goal of the neocortex? In Koch, C. and Davis, J. L., editors, *Large-scale neuronal theories of the brain*, chapter 1, pages 1–22. MIT Press, Cambridge, MA.

Bartlett, F. C. (1932). *Remembering: An Experimental and Social Study.* Cambridge University Press, Cambridge.

Baxter, J. (1995). The canonical metric for vector quantization. NeuroCOLT NC-TR-95-047, University of London.

Berkeley, G. (1710/1996). *A treatise concerning the principles of human knowledge.* Oxford University Press, Oxford.

Beymer, D. and Poggio, T. (1996). Image representations for visual learning. *Science*, 272:1905–1909.

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

Biederman, I., Mezzanotte, R. J., and Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14:143–177.

Biederman, I., Rabinowitz, J. C., Glass, A. L., and Stacy, E. W. (1974). On the information extracted from a glance at a scene. *Journal of Exp. Psychol*, 103:597–600.

Bienenstock, E. and Geman, S. (1995). Compositionality in neural systems. In Arbib, M. A., editor, *The handbook of brain theory and neural networks*, pages 223–226. MIT Press.

Blackmore, S. J., Brelstaff, G., Nelson, K., and Troscianko, T. (1995). Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception*, 24:1075–1081.

Bookstein, F. L. (1991). *Morphometric tools for landmark data: geometry and biology.* Cambridge Univ. Press, New York.

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.

Brigham, J. C. (1986). The influence of race on face recognition. In Ellis, H. D., Jeeves, M. A., and Newcombe, F., editors, *Aspects of face processing*, pages 170–177. Martinus Nijhoff, Dordrecht.

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.

Carne, T. K. (1990). The geometry of shape spaces. *Proc. Lond. Math. Soc.*, 61:407–432.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., and Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Trans. on Neural Networks*, 3:698–713.

Carpenter, G. A., Grossberg, S., and Rosen, D. B. (1991). Fuzzy ART: An adaptive resonance algorithm for rapid stable classification of analog patterns. In *Proc. Intl. Joint Conf. on Neural Networks*, pages 411–416.

Cavanagh, P. (1995). Vision is getting easier every day. *Perception*, 24:1227–1232. guest editorial.

Clark, A. (1993). *Sensory qualities.* Clarendon Press, Oxford.

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.

Cortese, J. M. and Dyre, B. P. (1996). Perceptual similarity of shapes generated from Fourier Descriptors. *Journal of Experimental Psychology: Human Perception and Performance*, 22:133–143.

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. (1996a). Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Science*, 93:12046–12050.

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

Dayan, P., Hinton, G. E., and Neal, R. M. (1995). The Helmholtz Machine. *Neural Computation*, 7:889–904.

Dennett, D. C. (1991). *Consciousness explained.* Little, Brown & Company, Boston, MA.

Dretske, F. (1981). *Knowledge and the flow of information.* MIT Press, Cambridge, MA.

Edelman, S. (1994). Biological constraints and the representation of structure in vision and language. *Psycoloquy*, 5(57).

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. (1996). Vision reanimated. In Aloimonos, Y., Carlsson, S., and Eklundh, J.-O., editors, *Proc. 7th Rosenön Workshop on Computer Vision.* L. Erlbaum, Hillsdale, NJ. forthcoming.

Edelman, S., Bülthoff, H. H., and Bülthoff, I. (1996). Features of the representation space for 3D objects. MPIK-TR 40, Max Planck Institute for Biological Cybernetics.

Edelman, S. and Duvdevani-Bar, S. (1997a). A model of visual recognition and categorization. *Phil. Trans. R. Soc. Lond. (B)*, 352:–. to appear.

Edelman, S. and Duvdevani-Bar, S. (1997b). Similarity, connectionism, and the problem of representation in vision. *Neural Computation*, 9:–. in press.

Edelman, S. and Intrator, N. (1997). Learning as extraction of low-dimensional representations. In Medin, D., Goldstone, R., and Schyns, P., editors, *Mechanisms of Perceptual Learning*. Academic Press. in press.

Edelman, S. and Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics*, 64:209–219.

Edelman, S. and Weinshall, D. (1997). Computational approaches to shape constancy. In Walsh, V. and Kulikowski, J., editors, *Perceptual constancies: why things look as they do*. Cambridge University Press, Cambridge, UK. in press.

Efron, B. and Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman and Hall, London.

Ekman, G. and Lindman, R. (1961). Multidimensional ratio scaling and multidimensional similarity. Reports from the Psychological Laboratories 103, University of Stockholm.

Fodor, J. A. (1981). *RePresentations*. MIT Press, Cambridge, MA.

Fodor, J. A. (1987). *Psychosemantics*. MIT Press, Cambridge, MA.

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

Galin, E. and Akkouche, S. (1996). Métamorphose d'objets tridimensionnels: quelques méthodes d'accélération. *Revue Techniques et Sciences Informatiques*, 15:329–350.

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

Garbin, C. P. (1990). Visual-touch perceptual equivalence for shape information in children and adults. *Perception and Psychophysics*, 48:271–279.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin, Boston, MA.

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

Goodman, N. (1977). *The structure of appearance*. Reidel, Dordrecht.

Gregory, R. L. (1978). Illusions and hallucinations. In Carterette, E. C. and Friedman, M. P., editors, *Handbook of Perception*, volume IX, pages 337–357. Academic Press, New York, NY.

Gregson, R. A. M. (1975). *Psychometrics of similarity*. Academic Press, New York.

Gregson, R. A. M. (1988). *Nonlinear psychophysical dynamics*. Erlbaum, Hillsdale, NJ.

Gregson, R. A. M. and Britton, L. A. (1990). The size-weight illusion in 2D nonlinear psychophysics. *Perception and Psychophysics*, 48:343–356.

Grimes, J. (1995). On the failure to detect changes in scenes across saccades. In Akins, K., editor, *Perception*, volume 5 of *Vancouver Studies in Cognitive Science*, chapter 4. Oxford University Press, New York.

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.

Hanson, S. J. and Gluck, M. A. (1993). Spherical units as dynamic consequential regions: implications for attention, competition and categorization. In Hanson, S. J., Cowan, J. D., and Giles, C. L., editors, *Advances in Neural Information Processing Systems 5*, pages 656–664. Morgan Kaufmann.

Harnad, S., editor (1987). *Categorical Perception: The Groundwork of Cognition.* Cambridge University Press, New York.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42:335–346.

Hebb, D. O. (1949). *The organization of behavior.* Wiley.

Hinton, G. E., Dayan, P., Frey, B. J., and Neal, R. (1995). The wake-sleep algorithm for unsupervised neural networks. *Science*, 268:1158–1161.

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.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *J. Physiol.*, 148:574–591.

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

Intrator, N. and Cooper, L. N. (1992). Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5:3–17.

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.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.*, 16:81–121.

Kendall, D. G. (1989). A survey of the statistical theory of shape. *Statistical Science*, 4:87–120.

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.

Kobatake, E., Tanaka, K., and Tamori, Y. (1992). Long-term learning changes the stimulus selectivity of cells in the inferotemporal cortex of adult monkeys. *Neuroscience Research*, S17:237.

Koch, C. and Ullman, S. (1985). Selecting one among the many: a simple network implementing shifts in selective visual attention. *Human Neurobiology*, 4:219–227.

Koenderink, J. J., van Doorn, A. J., and Kappers, A. M. L. (1996). Pictorial surface attitude and local depth comparisons. *Perception and Psychophysics*, 58:163–173.

Koriat, A. and Goldsmith, M. (1995). Memory metaphors and the laboratory/real-life controversy: correspondence versus storehouse views of memory. *Behavior and Brain Sciences*. in press.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density. *Psychological Review*, 85:445–463.

Krushkal', S. L. (1979). *Quasiconformal mappings and Riemann surfaces*. Wiley, New York.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27.

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

Landau, B., Smith, L. B., and Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3:299–321.

Lando, M. and Edelman, S. (1995). Receptive field spaces and class-based generalization from a single view in face recognition. *Network*, 6:551–576.

Le, H. and Kendall, D. G. (1993). The Riemannian structure of Euclidean shape spaces: a novel environment for statistics. *The Annals of Statistics*, 21:1221–1271.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proc. IRE*, 47:1940–1959.

Lindsay, P. H. and Norman, D. A. (1977). *Human information processing: an introduction to psychology*. Academic Press, New York.

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

Littlestone, N. (1988). Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2:285–318.

Locke, J. (1690/1994). *An essay concerning human understanding.* Modern Library, New York.

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

Maffei, L. (1978). Spatial frequency channels: neural mechanisms. In Held, R., Leibowitz, H. W., and Teuber, H.-L., editors, *Handbook of sensory physiology: Perception*, chapter 2, pages 39–68. Springer-Verlag, Berlin.

Markman, A. and Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25:431–467.

Marr, D. (1970). A theory for cerebral neocortex. *Proceedings of the Royal Society of London B*, 176:161–234.

Marr, D. (1976). Early processing of visual information. *Phil. Trans. R. Soc. Lond. B*, 275:483–524.

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

Marr, D. and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London B*, 200:269–294.

Medin, D. L., Goldstone, R. L., and Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100:254–278.

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.

Millikan, R. (1995). *White Queen Psychology and other essays for Alice.* MIT Press, Cambridge, MA.

Moses, Y., Ullman, S., and Edelman, S. (1996). Generalization to novel images in upright and inverted faces. *Perception*, 25:443–462.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., and Newsome, W. T. (1985). The analysis of moving visual patterns. In Chagas, C., Gattas, R., and Gross, C. G., editors, *Pattern Recognition Mechanisms.* Vatican Press, Rome.

Mumford, D. (1991a). Mathematical theories of shape: do they model perception? In *Geometric methods in computer vision*, volume 1570, pages 2–10, Bellingham, WA. SPIE.

Mumford, D. (1991b). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biological Cybernetics*, 65:135–145.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of the cortico-cortical loops. *Biological Cybernetics*, 66:241–251.

Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In Koch, C. and Davis, J. L., editors, *Large-scale neuronal theories of the brain*, chapter 7, pages 125–152. MIT Press, Cambridge, MA.

Murase, H. and Nayar, S. (1995). Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, 14:5–24.

Newsome, W. T. and Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.*, 8:2201–2211.

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). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23:94–140.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, 43:25–53.

O'Regan, J. K. (1992). Solving the real mysteries of visual perception: The world as an outside memory. *Canadian J. of Psychology*, 46:461–488.

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.

Pentland, A. and Sclaroff, S. (1991). Closed–form solutions for physically based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:715–729.

Perrett, D. I., Mistlin, A. J., and Chitty, A. J. (1989). Visual neurones responsive to faces. *Trends in Neurosciences*, 10:358–364.

Perrett, D. I., Rolls, E. T., and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.*, 47:329–342.

Phillips, F. and Todd, J. T. (1996). Perception of local three-dimensional shape. *J. Exp. Psychol.: HPP*, 22:230–944.

Poggio, T. (1990). A theory of how the brain might work. *Cold Spring Harbor Symposia on Quantitative Biology*, LV:899–910.

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

Poggio, T., Fahle, M., and Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, 256:1018–1021.

Poggio, T. and Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247:978–982.

Pollatsek, A., Rayner, K., and Collins, W. E. (1984). Integrating pictorial information across eye movements. *J. Exp. Psychol.: General*, 113:426–442.

Putnam, H. (1988). *Representation and reality.* MIT Press, Cambridge, MA.

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

Rensink, R., O'Regan, K., and Clark, J. J. (1995). Image flicker is as good as saccades in making large scene changes invisible. *Perception*, 24 (suppl.):26–27.

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

Riesenhuber, M. and Dayan, P. (1997). Neural models for the part-whole hierarchies. In Jordan, M., editor, *Advances in Neural Information Processing 9*, pages –. MIT Press. in press.

Rolls, E. T., Baylis, G. C., Hasselmo, M. E., and Nalwa, V. (1989). The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Exp. Brain Res.*, 76:153–164.

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.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In Spiro, R. J., Bruce, B., and Brewer, W. F., editors, *Theoretical Issues in Reading and Comprehension.* Erlbaum, Hillsdale, NJ.

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

Salzman, C. D., Britten, K. H., and Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346:174–177.

Schiele, B. and Crowley, J. L. (1996). Object recognition using multidimensional receptive field histograms. In Buxton, B. and Cipolla, R., editors, *Proc. ECCV'96*, volume 1 of *Lecture Notes in Computer Science*, pages 610–619, Berlin. Springer.

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.

Selfridge, O. G. (1959). Pandemonium: a paradigm for learning. In *The mechanisation of thought processes*. H.M.S.O., London.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with unknown distance function. part i. *Psychometrika*, 27(2):125–140.

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. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, 91:417–447.

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

Shepard, R. N. and Arabie, P. (1979). Additive clustering: representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86:87–123.

Shepard, R. N. and Cermak, G. W. (1973). Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology*, 4:351–377.

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.

Simard, P., Victorri, B., LeCun, Y., and Denker, J. (1992). Tangent prop – a formalism for specifying selected invariances in an adaptive network. In Moody, J., Lippman, R., and Hanson, S. J., editors, *Neural Information Processing Systems*, volume 4, pages 895–903. Morgan Kaufmann, San Mateo, CA.

Snippe, H. P. and Koenderink, J. J. (1992). Discrimination thresholds for channel-coded systems. *Biological Cybernetics*, 66:543–551.

Spinoza, B. (1677/1981). *The Ethics*. J. Simon Publisher, Malibu, CA.

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.

Sundararaman, D. (1980). *Moduli, deformations and classifications of compact complex manifolds.* Pitman.

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

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.

Tanaka, K., Saito, H., Fukada, Y., and Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.*, 66:170–189.

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

Tversky, A. and Gati, I. (1978). Studies of similarity. In Rosch, E. and Lloyd, B., editors, *Cognition and Categorization*, pages 79–98. Erlbaum.

Ullman, S. (1980). Against direct perception. *Behavioral and Brain Sciences*, 3:373–416.

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

Ullman, S. (1995). Sequence-seeking and counter-streams: a model for information flow in the cortex. *Cerebral Cortex*, 5:1–11.

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.

von Helmholtz, H. (1856/1964). Unconscious conclusions. In Dember, W. N., editor, *Visual perception: the nineteenth century*, pages 163–170. Wiley.

Westheimer, G. (1981). Visual hyperacuity. *Prog. Sensory Physiol.*, 1:1–37.

Young, M. P. and Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, 256:1327–1331.

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.

A Formalization of distal shape spaces

The idea that objects belonging to a given natural kind can be given a common parameterization has led to the emergence of the concept of a shape space in parallel, in a number of applied disciplines, ranging from biological morphometrics to computational molecular biology. In addition, concepts related to that of a shape space have been defined in different mathematical disciplines, such as statistics, complex analysis, and algebraic geometry.

Perhaps the most straightforward approach to the construction of a low-dimensional shape space is based on the notion of landmarks – fiducial points affixed to the object, whose location determines the object's shape (Bookstein, 1991). An orderly study of the geometry of shape spaces defined by locations of points has been initiated only recently, by D. G. Kendall (1984; Kendall, 1989), who pointed out that the notion of a shape must include a specification of the transformations which, by definition, leave the shape invariant. In Kendall's shape spaces, where objects are rigid configurations of points, a natural assumption is to define shape up to the action of the orthogonal group of transformations (that is, rigid motions plus reflection). From this, it follows that dissimilarity between two sets of points is to be measured by the Procrustes distance, defined by the sum of squares of residual distances between corresponding points, remaining after an optimal orthogonal mapping matching one set to the other has been applied (Borg and Lingoes, 1987).

An interesting consequence of allowing for a Procrustes transformation before computing shape-space distance is that it makes the topology of this space non-trivial. Consider the simple example of the space of all triangles in a plane, and a particular member of that space: the equilateral triangle. Start deforming this triangle by moving one of the vertices inwards, along the perpendicular to the opposite side; this deformation corresponds to a movement of the corresponding point in the shape space. At some stage, the chosen vertex will cross over the opposite side (at which moment the triangle will degenerate into a line), and will continue moving outwards. Finally, an equilateral triangle will be re-formed; this triangle is a rotated version of the original one, and therefore equivalent to it under the Procrustes metric. A continuous movement along a straight line in the triangle-vertex space corresponds, therefore, to a movement along a closed line in the shape space. It can be shown that this space is also not flat, and contains singularities (one of which is the triangle whose three vertices coincide); furthermore, the local Riemannian metric that takes these properties into account determines a global metric which is identical to the Procrustes distance (Carne, 1990; Le and Kendall, 1993).

In some cases, it may be desirable to define shape up to a group of transformations that is less restrictive than the orthogonal group, or, in other words, to allow deformation.[18] In that case, a suitable framework for the definition of a shape space is provided by the theory of Riemann surfaces (Krushkal', 1979). Specifically, any two surfaces (shapes) of a given genus related by a *conformal* mapping can be considered equivalent (belonging to the same class), with a *quasiconformal* mapping (see appendix B) taking one shape class into another. The resulting shape space (known as the Teichmüller space) possesses a Riemannian metric, defined by the deviation of the quasiconformal mapping from conformality (Krushkal', 1979). The Teichmüller space can be parameterized by a small set of real numbers, which provide a possible coordinate system for the resulting shape space (Sundararaman, 1980).

Place Figure 8 here

B Quasiconformal mappings

In two dimensions, a mapping realized by an analytic function with a non-vanishing Jacobian in a given region is conformal there (Cohn, 1967). In other words, any well-behaved function that maps a portion of the plane to itself is bound to preserve angles on a small scale (and, therefore, also ratios of side lengths of small triangles; see Figure 8). In higher dimensions, conformality is very restrictive. As proved by Liouville in 1850, already for $n = 3$ there are no everywhere conformal mappings from R^n to itself besides those which are composed of finitely many inversions with respect to spheres, or Möbius transformations. These constitute a finite-dimensional Lie group which includes the group of rigid motions in R^n and is only slightly broader than that group (Reshetnyak, 1989). This means that enforcing conformality in a mapping between high-dimensional spaces amounts to

[18]Consider, again, Dali's *Persistence of Memory*: our perception of the thing suspended from the tree branch as a deformed clock rather than an uninterpretable shape attests to the perceptual equivalence of some shapes related by deformations rather than transformations.

enforcing global isometry, or global preservation of distances (by analogy with the 3D Euclidean space, mappings that satisfy this constraint are called rigid motions).

A considerably broader class of mappings emerges if the requirement of conformality is replaced by that of quasiconformality. A regular topological mapping is quasiconformal if there exists a constant q, $1 \leq q < \infty$, 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). Intuitively, a conformal mapping is locally an isometry (i.e., a rigid motion; see Figure 8); a quasiconformal mapping is locally affine (i.e., a combination of motion with shearing deformation). Under such a mapping, the ranks of distances between points are preserved approximately, on a small scale (Väisälä 1992, p.124). The relevance of quasiconformality to the representation of real-world shapes stems from the realization that distance ranks need not be preserved globally, across the entire shape space, but rather within shape classes (just as the common parameterization that is the basis for the definition of distal similarity is required to hold within, but not extend across, the boundaries of natural kinds).

C Distal to proximal mapping and the possibility of different parameterizations

Consider the effect of the geometry mapping, f_1, defined in section 3.2.2. The properties of this mapping are to be defined with respect to a family of possible parameterizations of the distal shape space, rather than with respect to some illusory true and unique parameterization. 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, along with veridical representation, it is also possible that two perceptual systems implementing different mappings will have incompatible (or even conflicting) pictures of the world. Note that this effect cannot be distinguished from that of different parameterizations (discussed above).

D More on qualia

A simplified version of Clark's (1993) qualia account, coached in terms of the present approach, may be formulated, e.g., by considering the redness of a ripe tomato as a counterpoint to the greenness of an unripe one, and the shape of a pear — as a contrast to the shape of an apple. Obviously, a shape, a color, or some other quality considered in isolation can be represented in any desired manner whatsoever; it is the introduction of other objects that makes representation challenging. Now, a progressive reduction in the level of illumination would force the observer to switch gradually to scotopic vision, effectively losing both the ability to discriminate between the two tomatoes on the basis of their color, and all the color qualia. Likewise, gradual ripening of a green tomato would, by any sensible account of qualia, be perceived as an equally gradual turning of the quale of greenness into that of redness.

This suggests that it may be more productive to consider qualia such as "redness *vs.* greenness" and "pear-shape *vs.* apple-shape" as primitive, and redness or pear-shape – as derived (by a process computationally equivalent to multidimensional scaling). The "redness *vs.* greenness" quale may then be *identified* with the feature-space support for telling apart ripe and unripe tomatoes; although this reduction seems to hold only in the context of tomato discrimination, it is easily extended to apply to any other pair of stimuli, by projecting the difference between their feature-space representations onto the paradigmatic "red *vs.* green" direction. In shape perception, an analogous argument can be constructed using, e.g., the distinction between a pear and an apple; morphing a pear into an apple is the shape-space counterpart of the color shift induced by the ripening of the tomato in the color example. In summary, it seems sensible to accept the notion that qualia are qualia of similarities; this choice forestalls the embarrassing situation in which a quale can be anything at all, and points towards a potentially fruitful way in which the problematic issues associated with qualia can be addressed experimentally.

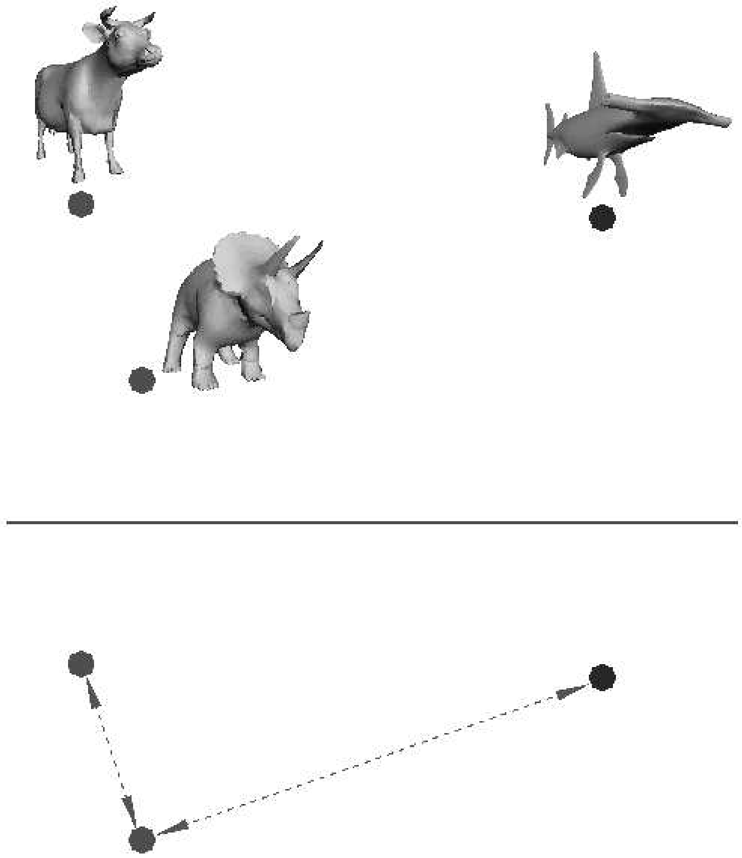

Figure 1: Clustering by natural kinds, and its representation that fulfills the requirement of second-order isomorphism, according to Shepard (1968). The disposition of the tokens corresponding to the three shapes in this illustration in the proximal representation space (bottom) reflects the disposition of the shapes in the distal shape space (top); the shapes of the tokens are immaterial to their representational capacities.

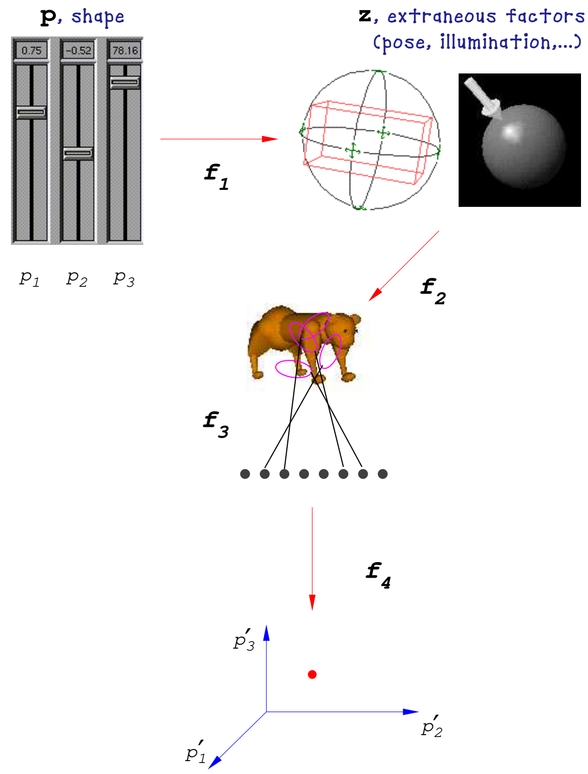

Figure 2: The components of the distal to proximal mapping, M (see section 3.2.2). In this schematic illustration, the shape of the object is determined by three parameters **p** (depicted by the three "slider" controls). The appearance of the object is governed by these parameters and by variables **z** that represent factors such as pose and illumination direction. To assure proper representation of the original parameter space, a typical perceptual system must carry out many measurements, then reduce the dimensionality of the resulting space, while getting rid of the extraneous variables **z**.

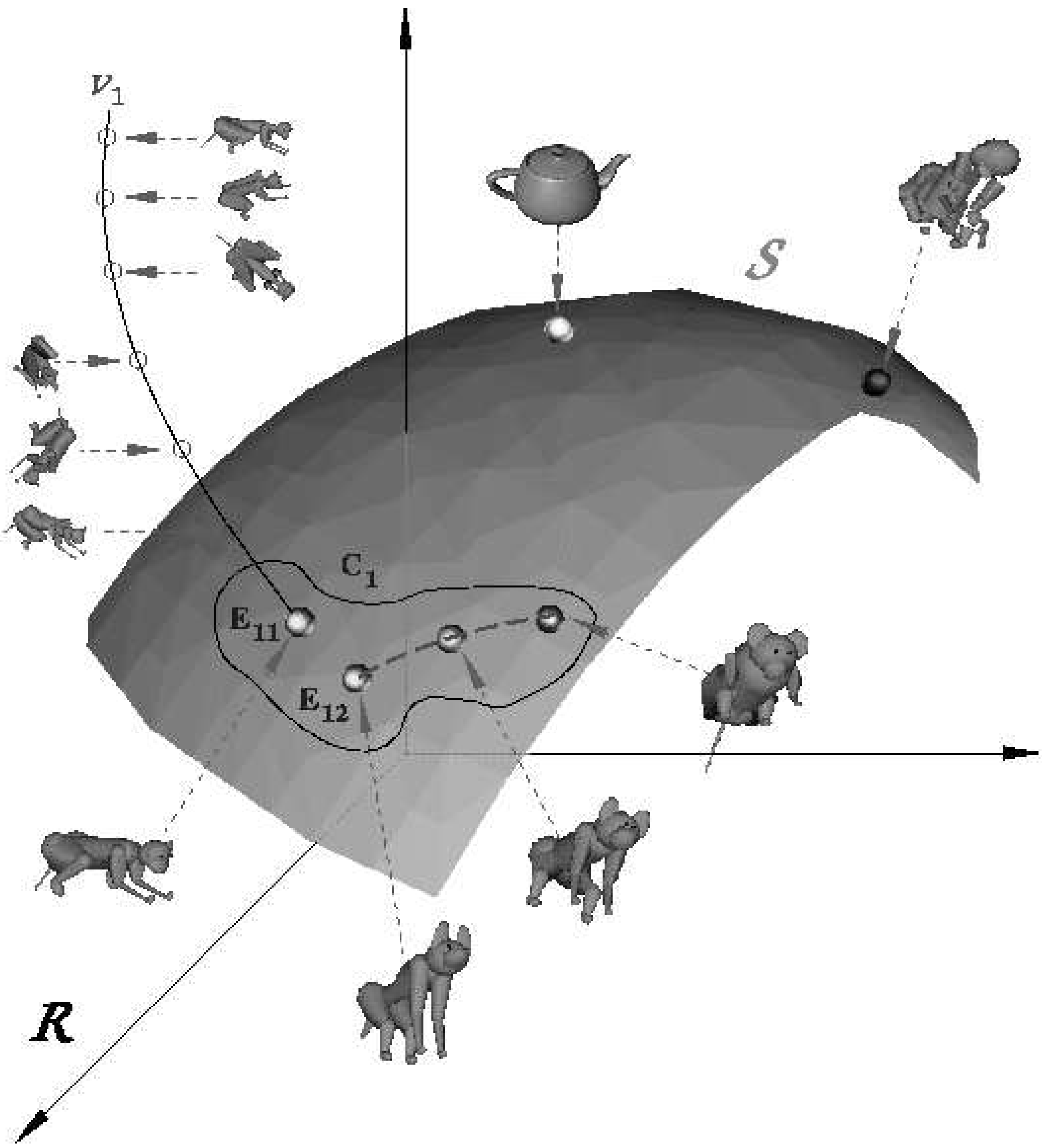

Figure 3: The image space, $\mathcal{R}$ (depicted here as 3-dimensional, to facilitate visualization), and some of its subspaces (see section 4.1). Two exemplars, $\mathbf{E}_{11}$ and $\mathbf{E}_{12}$, belong to the same class $\mathcal{C}_1$ (the class of 4-legged animal shapes). Some of the different views of $\mathbf{E}_{11}$ are shown (marked by open circles), along with its view space, $\mathcal{V}_1$. The surface patch represents a part of the shape space $\mathcal{S}$; the view spaces of the individual objects are transversal to it. A morphing sequence originating at $\mathbf{E}_{12}$ and leading to two other shapes is illustrated by the dashed curve contained in $\mathcal{S}$. Movement towards the upper right corner of $\mathcal{S}$ corresponds to a reduction in the resemblance between the resulting image and the images of coherently looking objects.

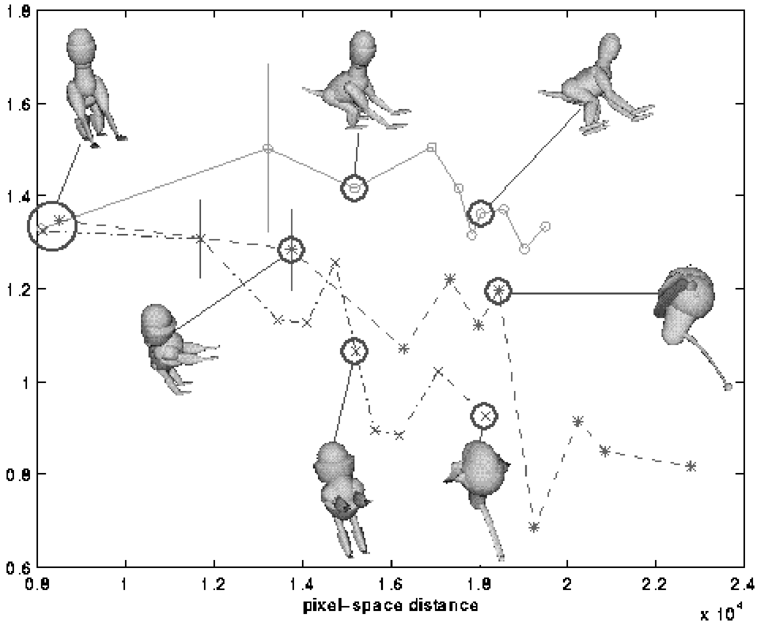

Figure 4: The response of a radial basis function module, trained on 10 random views of a parametrically defined object, to stimuli differing from a reference view of that object (marked by the big circle), in three ways: (1) by progressive view change, marked by $\circ$'s; (2) by progressive shape change, 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); the units along the ordinate are arbitrary (only the relative response in the three conditions matters). 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.

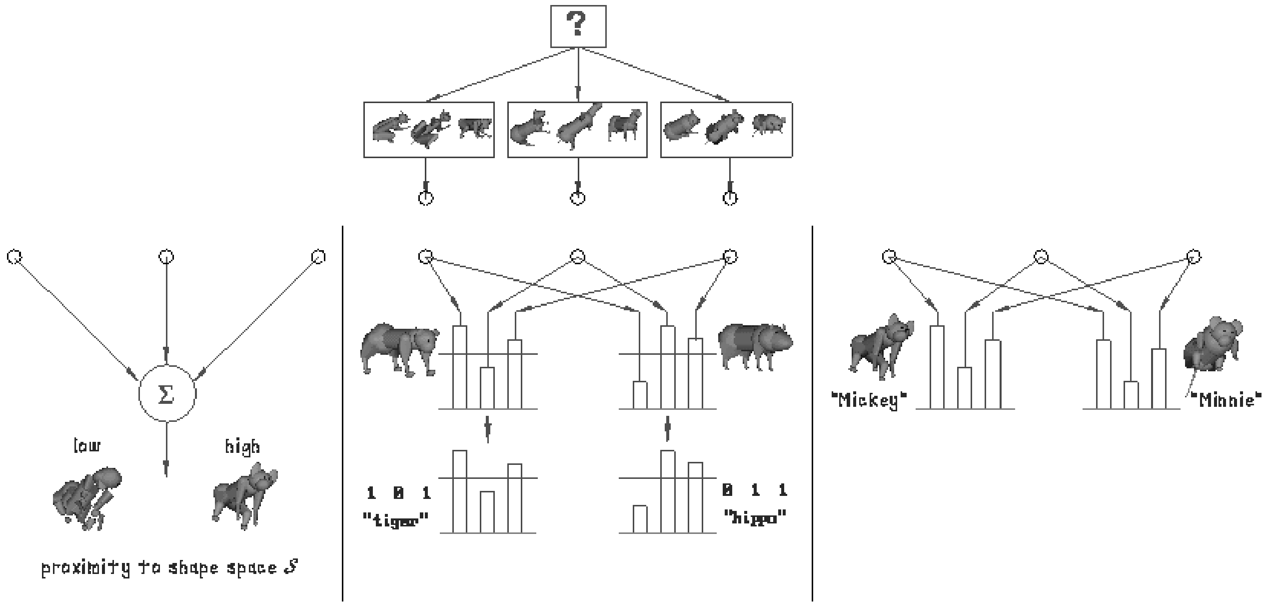

Figure 5: Using the Chorus representation (top) at different levels of categorization (see section 5.1). The three panes at the bottom show (left to right): a superordinate level; the basic level; a subordinate level.

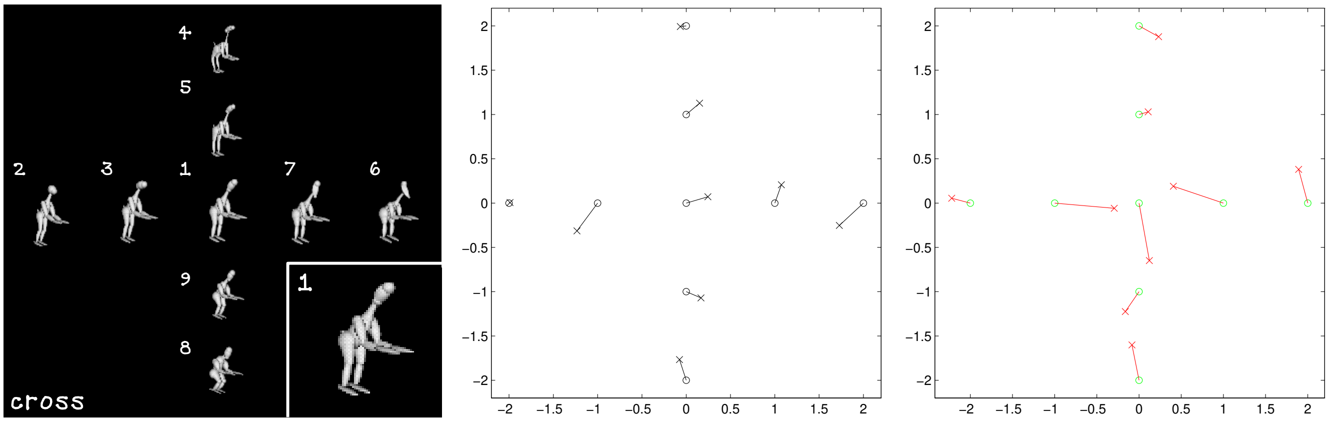

Figure 6: *Left:* the parameter-space configuration used for generating the stimuli in one of the experiments described in (Cutzu and Edelman, 1996a). *Middle:* the 2D MDS solution for all subjects. Symbols: $\circ$ – true configuration; $\times$ – configuration derived by MDS from the subject data, then Procrustes-transformed (Borg and Lingoes, 1987) to fit the true one. Lines connect corresponding points. The coefficient of congruence between the MDS-derived configuration and the true one was 0.99 (expected random value, estimated by bootstrap (Efron and Tibshirani, 1993) from the data: 0.86 ± 0.03, mean and std dev; 100 permutations of the point order were used in the bootstrap computation). The Procrustes distance between the MDS-derived configuration and the true one was 0.66 (expected random value: 3.14 ± 0.15). *Right:* the 2D MDS solution for the RBF model; coefficient of congruence: 0.98 (expected random value: 0.86 ± 0.03); Procrustes distance: 1.11 (expected random value: 3.14 ± 0.17).

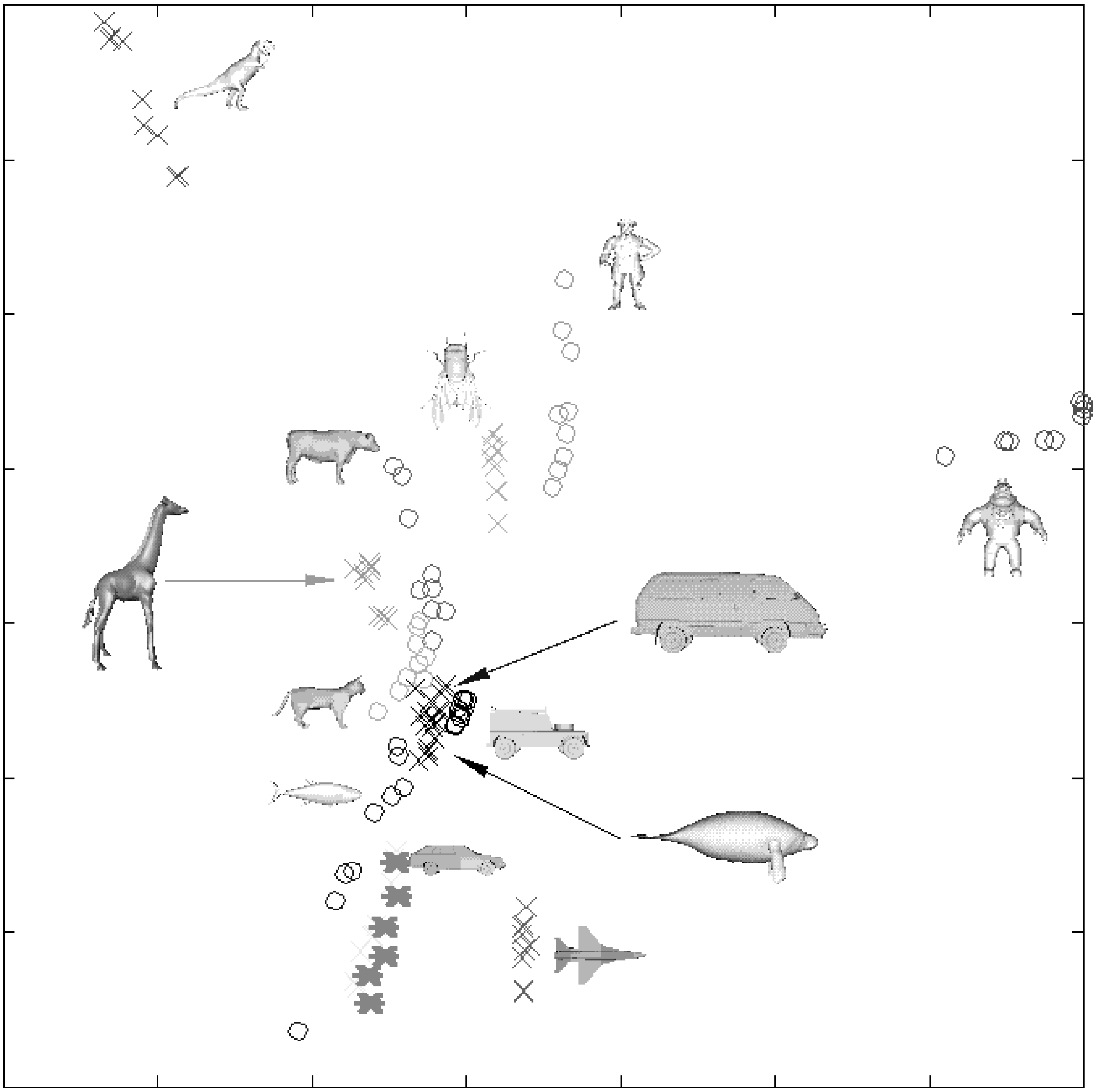

Figure 7: Representation of novel objects in terms of similarities to familiar ones. The plot is a two-dimensional rendition (produced by MDS) of a ten-dimensional space spanned by the outputs of ten prototypes modules, in a pilot implementation of the Chorus scheme. Each point in this plot corresponds to a view of an object; views belonging to the same object cluster together. The objects on which the modules have been trained are indicated by the small icons; the larger icons point to three novel test objects. Note that representations of similar objects (e.g., the quadrupeds) reside near each other; moreover, the novel quadruped (the giraffe) has been grouped with its likes. Because of the poor resolution of the front end of this system (implemented by a bank of 250 Gaussian receptive fields, each about 1/10 of the size of the stimulus images), objects that resemble each other at a coarse scale are sometimes confused (e.g., the manatee, or the sea cow, has been placed near the van; in the full ten-dimensional space, the manatee, which was a novel test object, was found to be similar to the tuna, the cow, and the automobile wagon, in that order). For a description of this system, see (Edelman and Duvdevani-Bar, 1997a).

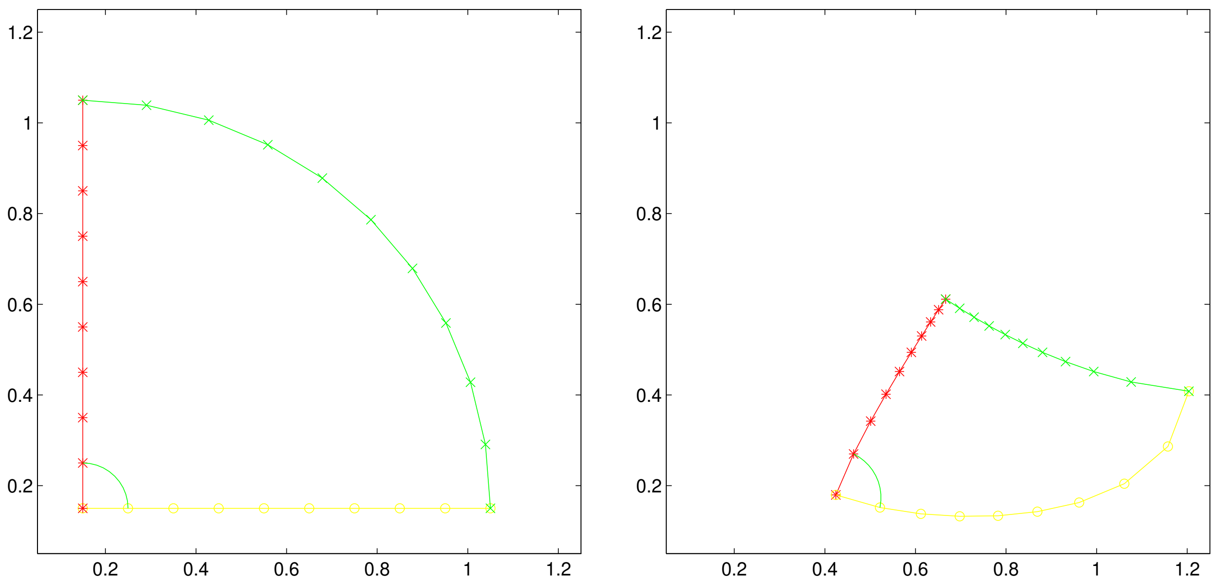

Figure 8: An illustration of the concept of conformal mapping, discussed in appendix B. *Left:* two similar "triangles" formed by two straight line segments and two circular arcs, all meeting at right angles. *Right:* the same two "triangles" under the action of the conformal mapping $\mathbf{z} = \sqrt{\operatorname{atanh}(\mathbf{x})}$ (the choice of the function here is arbitrary, and is intended merely to illustrate the concept of conformality). For the small triangle, both the isosceles shape and all the angles are preserved. At a larger scale, the triangle is distorted, although the angles remain right.