Representation of object similarity in human vision: psychophysics and a computational model

Florin Cutzu and Shimon Edelman*
Dept. of Applied Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel
[florin,edelman]@wisdom.weizmann.ac.il

Revised, December 1996

Abstract

We report results from perceptual judgment, delayed matching to sample, and long-term memory recall experiments, which indicate that the human visual system can support metrically veridical representations of similarities among 3D objects. In all the experiments, animal-like computer-rendered stimuli formed regular planar configurations in a common 70-dimensional parameter space. These configurations were fully recovered by multidimensional scaling from proximity tables derived from the subject data. We show that such faithful representation of similarity is possible if shapes are encoded by their similarities to a number of reference (prototypical) shapes, as in the computational model that accompanies the psychophysical data.

Keywords: Object recognition / Representation / Similarity / Multidimensional scaling

Running title: *Faithful Representation of Similarity*

*To whom correspondence should be addressed.

1 Introduction

The human visual system possesses an impressive ability to recognize and categorize complex three-dimensional objects. Recognition performance of human observers is the more striking in the light of the variability in object appearance which the visual system must overcome along the way from the retinal image to a characterization of the stimulus in terms of its geometrical structure, familiarity, etc. As in the famous aphorism of Heraclitus, who pointed out that one cannot step into the same river twice, we literally never see the same object twice. The variability in object appearance stems from several sources. First, images of objects taken under different viewing conditions (e.g., varying illumination or pose) generally look differently. Second, the appearance of different exemplars of the same category may vary, often by an amount exceeding the variation between categories of related objects.[1]

Clearly, the visual system must treat these factors differently: whereas illumination-related changes in object appearance are mostly ignored (unless the observer makes a special effort to determine the illumination under which a given image has been taken), view-related changes must be both marked (people are usually aware of the orientation of an observed object), and compensated for, if the object is to be recognized irrespective of viewpoint. In comparison, shape-related changes must be represented explicitly and, if deemed significant, acted upon; they may be dismissed only if deemed superficial or irrelevant to the task at hand.

Computational studies of visual recognition in the past tended to concentrate on the problem of reducing the effects of viewing conditions, leaving aside the complementary problem of dealing with shape changes. Indeed, competing theories of recognition surveyed in the next section are routinely compared mainly on the basis of the solution they offer to viewpoint-induced variability in object appearance (Biederman, 1987; Tarr and Bülthoff, 1995). Correspondingly, psychophysical studies of object recognition in human subjects normally explore the effects of parametric manipulation of stimulus orientation, but not of stimulus shape. The manipulation of shape, in comparison, is confined to the study of categorization and, in particular, of similarity — a central concept in theorizing about categorization.

A comprehensive theory of object representation and recognition is expected to account for the effects of both pose and shape changes on human performance. This goal — integrating the findings of numerous pose-manipulation experiments with the wealth of data obtained in categorization experiments — remains to date elusive. This may be blamed partly on the distinct sets of mechanisms postulated by models of recognition on the one hand, and models of categorization on the other. For instance, the central concept in most theories of recognition is that of object constancy

[1] As measured according to some straightforward image-based metric, such as the sum of squared differences of corresponding pixel values.

(Ellis et al., 1989; Jolicoeur and Humphrey, 1996), or invariance, whereas categorization necessarily addresses a complementary issue of generalization (Shepard, 1987) of response over distinct yet similar stimuli. Furthermore, because categorization experiments tend to employ stimuli that are extremely impoverished (and therefore easier to control parametrically), it is difficult to extend theories of categorization derived from these experiments to realistic three-dimensional objects.

The present work constitutes an attempt to confront the above difficulties and to facilitate the development of an integrated framework for the understanding of recognition and categorization. Our first goal is to quantify human performance in the perception of controlled shape change, just as the performance under view change has been quantified in a number of studies carried out in recent years, as reviewed in (Jolicoeur and Humphrey, 1996). Unlike the previous work on the perception of similarities among objects, which, with very few exceptions (Edelman, 1995b), involved two-dimensional shapes (Brown and Andrews, 1968; Shepard and Cermak, 1973; Garbin, 1990; Cortese and Dyre, 1996), here we use realistically rendered animal-like three-dimensional objects as stimuli. The second goal of the present work is to offer a theoretical framework for the understanding of human performance in the perception of shape change, and to use it to develop a model that would be able to replicate the pattern of human perception of shape similarity.

The rest of the introductory section contains an overview of those theories of recognition and categorization that have a direct bearing on the present study. We then outline our theoretical approach, and present a synopsis of the experimental results. Section 2 describes in detail the three experimental methods we employed in testing subjects, and our approach to the analysis of perceived similarities. The results obtained with these methods are recounted in the three subsequent sections (3, 4, and 5). Section 6 then describes the computational models based on the psychophysical findings. Finally, section 7 recapitulates the results and offers some conclusions; the exciting philosophical and computational implications of the present study are discussed in detail elsewhere (Edelman, 1995c; Edelman, 1996b).

1.1 Theories of recognition

Of the two main sources of variability of object appearance — changing pose and changing shape — the former has received by far the most attention in theorizing about the psychophysics of object recognition. Reviews of recognition theories (Ullman, 1989; Ullman, 1996; Edelman and Weinshall, 1996) distinguish between three different ways to compensate for the effect of varying pose: invariant features, alignment, and structural descriptions. The first of these attempts to reduce the problem of recognition to that of pattern classification (Duda and Hart, 1973), by representing objects as vectors of features that are, ideally, invariant with respect to pose and other variables irrelevant to object identity (Mundy and Zisserman, 1992). The second approach calls for aligning, or bringing into register, the stimulus and a model before estimating the degree of match between the two (Lowe, 1986; Ullman, 1989). The third approach places the burden of isolating object identity from the effects of irrelevant variables such as pose on the construction of "qualitative" representations (e.g., structural graphs composed of generic parts and relationships), which, by design, are invariant to pose (Marr and Nishihara, 1978; Biederman, 1987). A fourth approach, which emerged only recently (Poggio and Edelman, 1990; Ullman and Basri, 1991), treats recognition as a problem of interpolation of the space of all views of an object from a relatively small number of data points (i.e., labeled example views available in advance).

The relevance of these theories to the understanding of human vision stems from the possibility to use each of them to generate specific predictions concerning human performance, which can then be tested in psychophysical experiments. Many extensive studies of that kind have been conducted to date; reviews stressing different aspects of the experimental findings can be found in (Biederman and Gerhardstein, 1995; Tarr and Bülthoff, 1995). By and large, experimental data support the notion that human recognition performance is viewpoint-dependent, except when the target objects are easily distinguishable from each other, as in Biederman's original studies, which led him to champion the cause of viewpoint invariance (Edelman, 1995a; Tarr and Bülthoff, 1995; Jolicoeur and Humphrey, 1996).

Given its amenability to experimental manipulation through the control of inter-stimulus similarity (Edelman, 1995a; Tarr et al., 1996) and familiarity (Moses et al., 1996), the degree of viewpoint dependence appears to play a secondary role as an indicator of the nature of internal representations underlying the process of recognition. The primary characterization of these representations is to be gleaned through the study of the effects of variables that control viewpoint dependence — similarity and familiarity. As we shall see next, extensive data concerning these effects are available, albeit in a setting that excludes most issues normally addressed in the psychophysical studies of 3D object recognition; bridging this gap is one of the central goals of the present paper.

1.2 Theories of categorization

The identification of familiar objects under variable viewing conditions is only one of the challenges facing the human visual system; more often than not, it is required to make sense of a shape that is not identical to any of the shapes encountered before. In such situations, which call for categorization rather than identification, the system must act on the *similarity* between the stimulus object and representations of other objects stored in memory. Because of its central role in categorization, and its influence on recognition performance over changes in viewpoint, the concept of similarity is likely to figure in any comprehensive theory of representation.

The groundwork for such a theory is provided by Shepard's (1987) proposal of a universal law of perceptual generalization, based on an interpretation of similarity in terms of proximity between stimuli in an underlying representation space. Shepard observed that the likelihood of generalization between stimuli in a variety of tasks decays exponentially with distance between points corresponding to the stimuli in a postulated internal representation space. Beside providing a concrete interpretation of the otherwise vague notion of similarity, this proposal constitutes a crucial step towards consolidating the treatment of novel stimuli. Specifically, all objects, including ones never before encountered by the system, have the same status — their nature (in our case, shape) is inherent in their representation-space location relative to each other.

When considered within this framework, the theories of recognition mentioned in the previous section emerge as rather unsuitable for supporting categorization (Edelman, 1997). A typical recognition theory, such as recognition by alignment (Ullman, 1989), postulates something like a nearest-neighbor decision as the last step in the recognition process (following compensation for the effects of viewpoint): the identity of the stimulus is declared to be the same as that of the known object most similar to it. This amounts to discarding valuable information inherent in the similarity of the stimulus to other objects known to the system (Edelman, 1995c). Models designed to operate on a level of categorization less specific than that of identity, such as Recognition By Components (Biederman, 1987), similarly discard information by ignoring "metric" details of the stimulus in favor of a "qualitative" representation;[2] a recent attempt to circumvent this problem is described in (Stankiewicz and Hummel, 1996).

Classical models of categorization developed in fields other than visual psychophysics adopt a more constructive approach to making use of similarity information. For instance, both the similarity choice model of identification (Luce, 1963; Nosofsky, 1985), and the exemplar-based Generalized Context Model (GCM) of classification developed in (Nosofsky, 1992a; Nosofsky, 1992b) base categorization on similarities to multiple internal representations. In particular, according to GCM, classification is based on comparing summed similarities to exemplars of alternative categories, while recognition (i.e., judgment of familiarity) depends on the summed similarity of the stimulus to all stored exemplars from all categories. To model the role of selective attention, GCM assigns different weights to different dimensions of the representation space. As a result, this theory can treat categorization, identification and recognition memory tasks within a unified framework based on the notion of an internal representation space.

Ashby and his collaborators developed a somewhat different theory of categorization based on the concept of a representation space (Ashby, 1992; Perrin, 1992). According to their General Recognition Theory (GRT), objects are represented by multivariate normal distributions, rather than by single points. Categorization thus requires building decision boundaries to partition representation space into response regions corresponding to category labels; this process can be viewed

[2]Shepard (1987), in comparison, remarks that categorization — i.e., realization that a stimulus is a member of a category to which other stimuli may belong — is a result of deeming certain differences immaterial, not of discarding the information carried by these differences.

as an extension of the classical signal detection theory (Green and Swets, 1966) to the multidimensional case. Similarity between objects is defined as the degree of overlap between the corresponding representation-space regions (Ashby and Perrin, 1988).[3]

1.3 Internal representations and veridicality

The contribution of theories such as Shepard's law of generalization, Nosofsky's GCM, and Ashby's GRT to the understanding of visual object processing lies in their principled approach to the problem of representation. Clearly, making sense of an object should involve not only compensation for the effects of viewpoint (as suggested by theories mentioned in section 1.1), but also finding the proper place for it in a "space" of like objects. It is equally obvious that the internal shape representation space, postulated by the above theories, is only useful insofar as the patterns of proximities among object representations there reflect properly the patterns of similarities among the objects themselves, as measured along some well-defined physical dimensions. In other words, the representations must be *veridical* with respect to their target objects.

Finding out in what manner, if at all, the internal representations harbored by the human visual system are faithful to their objects is a crucial step in the understanding of high-level vision. Now, the representation of an object can be faithful to the original in more than one sense. In particular, a complete specification of the geometry of an object, whether qualitative or quantitative, clearly must be considered veridical. An example of an approach that adopts the former view is provided by the RBC model (Biederman, 1987). In comparison, quantitatively veridical representations are called for by object recognition systems in computer vision, which typically require exact and detailed information regarding the object's geometry; a typical example is the alignment model (Huttenlocher and Ullman, 1987; Ullman, 1989).

In practice, both qualitative and quantitative representations proved notoriously difficult to compute from images of objects, counter to the expectations of many workers in vision (Binford, 1982; Edelman and Weinshall, 1996). These computational difficulties, as well as accumulating psychophysical evidence against full internal reconstruction of the visual world in human vision (see (O'Regan, 1992) for a review), prompted some theorists to doubt the appropriateness of characterizing the internal representations as (geometrically) veridical (Dennett, 1991; O'Regan, 1992).

A complete reconstruction of the geometry of individual objects is, however, not the only way to veridical representation: geometric reconstruction is not necessary for tasks such as shape categorization or discrimination between objects. In those cases, faithful representation of contrasts or dissimilarities between objects is representation enough (Shepard, 1968; Ashby and Perrin, 1988). The *modus operandi* of visual systems that adopt this approach has been termed by Roger Shepard

[3] The idea of expressing similarity as degree of overlap between two regions is reminiscent of Tversky's feature model (Tversky, 1977), and thus it is not surprising that the predictions of GRT can violate the metric axioms.

second-order isomorphism: "…the isomorphism should be sought — not in the first-order relation between *(a)* an individual object, and *(b)* its corresponding internal representation — but in the second-order relation between *(a)* the relations among alternative external objects, and *(b)* the relations among their corresponding internal representations. Thus, although the internal representation for a square need not itself be square, it should (whatever it is) at least have a closer functional relation to the internal representation for a rectangle than to that, say, for a green flash or the taste of a persimmon" (Shepard and Chipman, 1970). In the rest of this paper, we provide experimental support for precisely this notion of representation.

2 Experimental approach

To substantiate the above idea of veridicality, we set out to assess the degree to which the relationships among representations of 3D objects in human vision reflect geometrical similarities among these objects. To this end, we generated sets of stimuli characterized by well-defined and tightly controlled similarity relationships. This was achieved by embedding the stimuli in a common objective *shape space* — a metric space spanned by a few dozen parameters, jointly controlling the appearance of each stimulus object. We then investigated the patterns formed by the representations of the stimuli in the internal psychological space (Shepard, 1987) of the subjects.

To recover the patterns of the relations among internal representations, we hypothesized that the subject data (similarity judgments, response times and confusion rates, depending on the experiment) collected psychophysically, are systematically related to the representation-space distances between the stimuli (cf. (Tomonaga and Matsuzawa, 1992)). We then recovered the relative arrangement of the stimuli in the representation space using multidimensional scaling (MDS), as explained in section 2.2. The pattern formed by the stimuli in the internal representation space, as computed by MDS, was then compared with the original shape-space configuration built into the stimulus set, to assess the degree to which the former is a faithful representation of the latter.

2.1 The stimuli

2.1.1 Parameterization

Animal-like shapes similar to those illustrated in Figure 1 were used as stimuli in the psychophysical experiments. All animals had the same body parts, each modeled by a generalized cylinder. Each generalized cylinder was encoded by several parameters, such as axis length, the radii at the ends of the cylinder, the degree of axis curvature in two perpendicular planes (the axis was either parabolic or exponential, depending on the body part), and the degree of asymmetry of the cross-section. The cross-section was not necessarily circularly symmetric, with two of its halves being in general of different radii (this feature was necessary in order to render correctly the bulge of the muscles). Usually, spheres were placed at the joints between the cylinders. In all, there were 70 shape parameters common to all animals, defining the shape of the individual parts, their orientation and relative positions within the animal. Within this shape description framework, an individual shape corresponds to a point in a vector space of 70 dimensions (one dimension per parameter); a set of shapes corresponds to a "cloud" of points in this shape space.

INSERT FIGURE 1 HERE

2.1.2 Configurations in shape space

We now describe the configurations formed by the experimental stimuli in the parameter space. Each configuration was first defined by several reference shapes, from which the other ones were generated by interpolation, or morphing, in the parameter space. Given two animal-like shapes (i.e., two points in the 70-dimensional parameter space), $\mathbf{A}$ and $\mathbf{B}$, any shape $\mathbf{M}$ on the line segment $\bar{\mathbf{AB}}$, corresponding to the convex linear combination $\mathbf{M} = \alpha\mathbf{A} + (1 - \alpha)\mathbf{B}$, with $0 \leq \alpha \leq 1$, is also animal-like, provided that $\mathbf{A}$ and $\mathbf{B}$ are sufficiently close to each other. The blending parameter α controls the relative contributions of the shapes $\mathbf{A}$ and $\mathbf{B}$ to the appearance of the hybrid $\mathbf{M}$.

INSERT FIGURE 2 HERE

To facilitate the subsequent analysis of the similarities between the stimuli (as perceived by the subjects), we arranged the stimuli in planar and regular configurations in the parameter space (see Figure 2). In that manner, the distinction between a random pattern of similarities, and one related to the configuration built into the stimulus set in advance (to be derived from the subject data, as described in section 2.2), would be more apparent to the eye. In each series of experiments, the test objects were arranged in one of four distinct planar configurations in parameter space: triangle, star, square, cross. In the TRIANGLE experiments, seven objects were positioned at the vertices, barycenter and at the midpoints of the edges of an equilateral triangle in the parameter space. In the STAR experiments, the parameter space configuration was obtained by placing four objects at the vertices and at the barycenter of the equilateral triangle, and the other three objects — at the midpoints of the segments joining the barycenter to the vertices of the triangle. In the SQUARE experiments, nine objects were positioned at the vertices, barycenter and at the midpoints of the edges of a square in the parameter space. In the CROSS experiments, the configuration was obtained by placing five objects at the vertices and at the barycenter of the square, and the other four at the midpoints of the segments joining the barycenter to the vertices of the square.

The orientation of the plane defined by each of the four stimulus configurations in the parameter space with respect to the 70 axes (that is, the 70 dimensions of variation of the shape of the stimulus) was chosen at random. As a result, morphing any of the shapes into its neighbor in the configuration corresponded to a simultaneous change in all the 70 parameters. In other words, the planar (2D) configurations were not degenerate, in the sense that they could only be fully described relative to the entire set of 70 parameters defining the shape space in which they were embedded.

2.1.3 Rendering

The set of images of a given object was created by generating the object from its parametric description, rotating the resulting 3D shape to the required orientation, and rendering it on the screen of a computer workstation (SGI Indigo2/Extreme). Shape generation and rendering were carried out by native software (SGI Inventor), which created an image by computing the appropriate intensity for each point on the object's surface, according to the Gouraud shading model.

We remark that the nonlinearities in the image creation process led to a complicated relationship between the shape-space representation of an object and its appearance on the screen (see section 2.1.2 for an example). In particular, the parameter values determined the geometry of the surfaces that comprised the shape via nonlinear (for example, trigonometric) functions. The object's rotation in 3D and its rendering, which took into account (simulated) illumination and surface material properties, introduced further complications into the mapping from the parameter values to the object's image-plane appearance. In view of the nonlinear nature of this mapping, three objects whose corresponding points are collinear in the parameter space need not give rise to views that are collinear in the image (pixel) space. For example, the animal shape in the middle pane of the Figure 1 is situated in the parameter space at the midpoint between the other two shapes; in the pixel space, however, the three objects are in general not collinear.[4] The importance of this observation will be clarified in section 6, where we discuss the computational difficulties that must be overcome by any system attempting to recover the parameter-space configuration of a set of shapes from their images.

2.2 Similarity data analysis

As stated in the introduction, the ultimate goal of the experiments we undertook was to assess the subjects' ability to recover the shape-space configuration built into the stimuli from the images of the stimuli. For that purpose, we needed a method of reconstructing the configuration of the stimuli in a subject's internal representation space. Assuming (1) that this configuration affects the similarities among the stimuli, as perceived by the subject, and (2) that the perceived similarities, in turn, affect the subject's responses (e.g., the confusion rates among the different stimuli), one can attempt to recover the internal configuration by multidimensional scaling (MDS); the recovered configuration can then be compared with the original shape-space pattern built into the stimuli.

[4] Given two arrays of pixels $\mathbf{P}$ and $\mathbf{Q}$ that constitute the right and the left panes in the figure, the image situated at the midpoint on the line $\bar{\mathbf{PQ}}$ in the pixel space would be the superposition of the two *images* taken with equal weights $\mathbf{R} = 0.5 \cdot \mathbf{P} + 0.5 \cdot \mathbf{Q}$, and not the view of the hybrid *shape* $\mathbf{M}$).

Multidimensional scaling is a generic label for a family of numerical algorithms that search for an optimal embedding of a set of points in a (usually low-dimensional) metric space, given only the interpoint distances. The embedding is sought subject to the requirement of minimal discrepancy (stress) between the distances in the resulting configuration and the distances specified in the data (see Shepard, 1980, for a review). In the nonmetric version of MDS, the ranks of the distances, rather than their values, are used in the computation (Kruskal, 1964a; Shepard, 1966). Details regarding the particular MDS algorithm we used and an assessment of its performance on simulated data can be found in appendix A.

Given a set of data (e.g., a table of experimentally measured distances or dissimilarities among perceptual objects), the MDS procedure always returns *some* configuration, corresponding to the minimum-stress solution it has found. It is the responsibility of the experimenter then to determine whether or not the solution is meaningful. It should be noted that mere low stress is not a sufficient indicator of the significance of the solution, because low-stress spurious solutions are relatively frequent when the number of points is small. This problem can be remedied by using MDS in a confirmatory mode (Borg and Lingoes, 1987); one of the possibilities in that case is to compare the MDS solution to a pattern to which the points are known to conform. This is precisely the approach we have taken in the analysis of our experimental data.

2.2.1 Estimation of perceived similarities

We now describe the methods we used for gathering the similarity data for processing by MDS. Many early studies relied on the estimation of subjective similarities between stimuli, through a process in which the observer had to provide a numerical rating of similarity when presented with a pair of stimuli. One drawback of this method is that many subjects do not feel comfortable when forced to rate similarity on a numerical scale. Another problem is the possibility of subjects modifying their internal similarity scale as the experiment progresses. We avoided these problems by employing two different methods for measuring subjective similarity: Compare Pairs of Pairs (CPP) and Delayed Match To Sample (DMTS).

In the CPP experiments, the subjects differentially rated pairwise similarity when confronted with two pairs of revolving objects, each presented in a separate window on a computer workstation (SGI Indigo2/Extreme) screen. In the long-term memory (LTM) variant of this experiment, the subjects were first trained to associate a nonsense word with each object, then carried out the CPP task from memory, prompted by the object labels rather than by the objects themselves. Thus, in these experiments similarity was measured using a simple two-alternative forced choice paradigm, which is more stable and natural than the numerical rating task. In the DMTS experiments, the subjects had to decide whether two consecutively displayed views belonged to the same object. Both response times and confusion rates were employed to construct a view proximity measure. Reaction times and error rates measured in such a speeded perceptual task are presumably more directly related to perceptual similarity than the subjective scores obtained in numerical rating, or even the subjective comparison data obtained in the CPP task. A detailed description of all the methods we employed will be given in sections 3.1 and 4.1.

2.2.2 Comparison with objective similarities

Following to the principle of confirmatory MDS, we quantified the similarity between the configurations derived by MDS from the subject data in each experiment, and the configurations built into the stimulus parameter space, using Procrustes analysis (Borg and Lingoes, 1987). For each data set, we first computed the optimal Procrustes transformation (combination of scaling, rotation, reflection, and translation) between the MDS-derived and the true configurations; these are similarity transformations that do not affect the shape of the configuration. Second, we computed the residual distance (what remains after the MDS-derived configuration has been Procrustes-transformed to fit the true one). Finally, we also quantified the similarity between two configurations by computing their coefficient of congruence — a correlation-like measure applied to the vectors of interpoint distances in the two configurations (Borg and Lingoes, 1987).[5]

To assess the significance of a similarity measure (Procrustes distance or coefficient of congruence) in a given experiment, we compared it to the expected mean similarity, obtained by a bootstrap procedure (Efron and Tibshirani, 1993) from repeatedly permuted data. Estimation of statistical significance of data analysis by bootstrap is the method of choice in situations in which no replications of the experimental data are available. Indeed, each of our experiments yielded a single measure of similarity S_{subj} between the parameter-space and the subject-derived configurations in each category (Procrustes distance, coefficient of congruence). To estimate the significance of the deviation of this measure from the chance level, we computed the dispersion of the chance-level similarity by Monte Carlo simulations, in which the MDS procedure was run repeatedly on proximity tables obtained by permuting the rows[6] of the (single available) subject data table. This yielded an estimate of the mean chance-level similarity, S_{chance}, and of its standard deviation, std_{chance}. The null hypothesis of $S_{subj} = S_{chance}$ could then be rejected on the basis of the comparison between $|S_{subj} - S_{chance}|$ and std_{chance}. In other words, if the measure of similarity is significantly different from the expected mean chance value obtained with the permuted original data, the similarity between the true and the MDS-derived configuration is unlikely to be due to chance.

[5] We note that when simultaneously analyzing the data of all subjects in an experiment, the shape space configuration was Procrustes-compared with the common MDS configuration derived by individually weighted MDS (INDSCAL; see Carroll and Chang, 1970) from the concatenated similarity tables of all subjects.

[6] Using permuted real data, rather than random numbers drawn in some independent manner, made sure that the random data fed to the Monte Carlo simulations came from the same distribution as the real data.

2.2.3 Grounding objective similarity in geometry

The informativeness of the comparison between the MDS-derived configuration and the one built into the stimulus shape space depends on the parameterization of the latter being objective; at the very least, a change of parameterization should not affect the outcome of the comparison too strongly. We considered a parameterization to be generic in this sense, if it captured the geometrical similarities between the various members of the parameterized family of shapes. The geometrical similarity between two shapes was defined, in turn, as the Euclidean distance[7] between their high-resolution mesh representations. Under such representation, a solid shape is encoded as the ordered list of the (x, y, z) coordinates of the vertices of a high-resolution triangular mesh closely approximating its surface (see Figure 3).[8]

INSERT FIGURE 3 HERE

Now, under an objective parameterization distances among shapes measured in the parameter space should correlate with the corresponding distances measured in the space of 3D coordinates of the surface mesh points. We found this to be true (see Figure 4), confirming the hypothesis that our parameterization of the shape space can be considered objective.

INSERT FIGURE 4 HERE

3 Exploring the representation of similarity relationships among several objects

We now proceed to describe the experimental results. In the first series of experiments, we addressed the question whether a planar (two-dimensional) arrangement of several shapes in the (70-dimensional) parameter space can be recovered from the table of subjective similarity rank data, collected using the CPP paradigm.

3.1 Experimental method: CPP

Each CPP experiment involved a set of N objects ($N = 7$ for the star and triangle configurations and $N = 9$ for the cross and square configurations), for which we measured all pairwise similarities

[7] This decision was motivated by the characterization of the different dimensions of our shape space as inherently integral, rather than separable; see (Shepard, 1987) for a discussion of the issues relevant to the choice of the metrics.

[8] This definition of geometrical similarity assumes that the shape space can be considered a vector space. This, in turn, requires that correspondence between the elements of two vectors (i.e., vertices in the two meshes) be defined. Because the animal-like shapes we used are structurally identical and differ only metrically, the vertices of the triangular meshes of different shapes are in correspondence by construction. When two such shapes are brought into register by a sequence of similarity transformations that eliminate differences in orientation and position, the residual Euclidean distance in the space of mesh vertex coordinates directly measures shape similarity.

S_{ij}, $1 \leq i \leq N$, $1 \leq j \leq N$. The measured similarities formed $N \times N$ proximity tables, one for each configuration and subject. In each trial, the subject was shown simultaneously two pairs of shapes – say, pair A, consisting of shapes i and j, and pair B, consisting of shapes r and s. Each shape, subtending about 2° of visual angle, was displayed in a separate window on a computer screen, rotating continuously, at a rate of one complete revolution in six seconds, around its vertical axis. The subject had to select, by pressing one of two keys on the computer keyboard, the pair consisting of the most similar shapes, that is, to decide whether the similarity between shapes i and j is larger or smaller than the similarity between shapes r and s. There was no time limit: the subjects were allowed to study the revolving objects at will; most responses were made within 10 seconds of the beginning of the trial. After the subject made the choice by pressing the proper key on the computer keyboard, the next set of four shapes was shown, until all pairs of pairs were tested for the respective configuration. The choice of the subject was recorded together with the corresponding object labels.

This method presents a technical problem for large test sets: the number of trials required to fill an $N \times N$ proximity table grows as N^4, getting rapidly out of hand as N increases (the number of pairs of objects is $P = N \cdot (N-1)/2$, and the number of pairs of pairs which is the number of trials, is $D = P \cdot (P-1)/2$). Fortunately, only about 30% of the total number of comparisons (randomly chosen) usually suffice for obtaining an acceptably precise MDS solution (although we did test all pairs of pairs in our experiments).

Four series of CPP experiments were conducted, corresponding to the four shape space configurations described in section 2.1.2. The CPP TRIANGLE and CPP STAR experiments (7 objects) involved 210 comparisons. The CPP SQUARE and the CPP CROSS experiments (9 objects) involved 630 comparisons. The stimulus configurations with the appropriate labels are displayed in Figure 2.

3.2 Analysis of results

Following each experiment, the proximity tables were constructed from the rank data as described in appendix B.1, and were subjected to a verification of transitivity of choice. The latter step was necessary because transitivity in differential similarity judgments indicates that subjects tend to use the same features for all object pairs. If different features or criteria are used when comparing different pairs of objects, or if subjects are responding randomly, then some of the rankings would be intransitive, compromising the validity of the geometrical model of similarity space, and precluding the use of MDS, which assumes that the representation space in which the data points are to be embedded is metric (Beals et al., 1968).

We tested transitivity for triplets of object pairs. Given that the results were of the form "objects (i, j) are more/less similar to each other than objects (p, q)", it is straightforward to verify transitivity for triplets of object pairs. Specifically, if objects (i, j) were deemed more similar than (p, q), and (p, q) more similar than (v, w), then objects (i, j) should appear more similar than objects (v, w). The percentage of intransitive triplets of pairs was used to measure the degree of transitivity in each subject's data. An analysis of the results shows that the intransitivity was typically low (less than 4% in all experiments).

As explained in section 2.2, Procrustes analysis and bootstrap statistics were then applied to measure the quality of the reconstruction of the shape-space configurations. The results are listed in the figure captions.

3.2.1 The CPP TRIANGLE experiments

Eight subjects were tested with seven objects arranged in a planar triangle-like configuration in the parameter space. The configuration displayed in Figure 6 was obtained by running MDS on the pooled data of all subjects.[9] The recovery of the parameter space configuration is almost perfect, despite the various deformations that were present in the individual configurations.

INSERT FIGURE 6 HERE

3.2.2 The CPP STAR experiments

Six subjects were tested with seven objects arranged in a planar star-like configuration in the parameter space. The plot presented in Figure 7 was obtained by running MDS on the proximity data of all the subjects. The recovered configuration is a somewhat deformed version of the original configuration: two of the extremal points are displaced to the left; the qualitative star structure is, however, correct, as attested by Procrustes analysis.

INSERT FIGURE 7 HERE

3.2.3 The CPP SQUARE experiments

Six subjects were tested with nine objects arranged in a rectangular configuration in the parameter space. The pooled-subjects MDS plot is shown in Figure 8. The recovery of the parameter space configuration is nearly perfect.

INSERT FIGURE 8 HERE

[9]The degree of agreement among subjects in a given experiment is expressed by the dispersion of the individual dimension coefficients, produced by individually weighted MDS (see appendix A). Table 1 lists the individual dimension coefficients for all the experiments.

3.2.4 The CPP Cross experiments

Six subjects were tested with nine objects arranged in a cross-like configuration in parameter space. The MDS plot is displayed in Figure 9. The recovery of the parameter space configuration is qualitatively correct, even though a certain deformation is visible.

INSERT FIGURE 9 HERE

3.3 Recovery of the shape-space configuration for nonsense objects

It is possible that the remarkably accurate reconstruction of the true low-dimensional structure of the stimulus space was facilitated by our use of relatively natural-looking stimuli. To assess the contribution of the subjects' familiarity with the stimuli, we conducted experiments in which the test objects were scrambled versions of the animal shapes used in the experiments described in the preceding section.

3.3.1 Making "scrambled" objects

Each animal-like shape was taken apart by retaining one half of each symmetrical left-right pair of parts; each part was then translated so that one extremity touched a common central point, from which all the components radiated. One important consequence of this procedure is that the resulting shapes, unlike the original animals, had no symmetries, and were perceptually more complex than the original animals. Two examples of the resulting scrambled shapes are shown in Figure 10.

INSERT FIGURE 10 HERE

3.3.2 CPP experiments involving scrambled objects

Two subjects participated in a CPP Scrambled Star experiment, and two other subjects — in a CPP Scrambled Triangle experiment. Only two subjects per experiment were used here, the goal being merely to demonstrate that the recovery of shape space configurations is *possible* for complex, unfamiliar stimuli.[10] The subjects did not report any particular difficulty in making the differential similarity decisions, although they did take longer to make a decision in each trial. Although the subjects successfully recovered the original configurations, the quality of the recovery was somewhat less good, compared to the results obtained with the regular animal-like shapes (see Figure 11).

INSERT FIGURE 11 HERE

[10] A more thorough quantitative comparison of the processing of scrambled and intact shapes was reported previously in (Edelman, 1995b).

3.4 The CPP experiments: discussion

Apart from the remarkable recovery of the parameter-space patterns from the subject data, an interesting finding in the CPP experiments was the transitivity of the subjective similarity ratings. It should be noted that transitivity is a property of the metric space (geometrical) models of similarity, and does not necessarily hold, e.g., in the framework of Tversky's set-theoretic model (Tversky, 1977). That model was, in fact, devised to account for the circumstances under which similarity is neither symmetric nor transitive. The prevalence of transitivity in our data indicates that set-theoretic models relying on discrete features need not be invoked in the context of the high-level perceptual tasks (similarity judgment and object recognition) performed by our subjects.[11]

Unlike in the object discrimination task reported in (Edelman, 1995b), the subjects here recovered the shape-space configurations not only for the animal-like shapes but also, to some extent, for scrambled objects. This suggests that the CPP task may have involved a reasoning-like problem-solving strategy, and not the fast perceptual decision-making that is the hallmark of natural object recognition. Indeed, debriefing revealed that the subjects resorted to an analytic strategy, by considering at length (at least for one complete revolution) the four objects displayed on the screen, which allowed them to discover and compare the corresponding parts of the shapes (entire or scrambled).

4 Exploring the representation of relationships among views of several objects

The CPP experiments described above support the hypothesis of veridical representation of similarity, by demonstrating that it is possible to recover the true low-dimensional shape-space configuration of complex stimuli from proximity tables obtained from subjects who made forced-choice similarity judgments. It may be claimed, however, that the task facing the subject in the CPP experiments was not entirely natural from a perceptual standpoint. In particular, the unlimited time at the subject's disposal in those experiments allowed (if not encouraged) the subject to employ a deliberative "cognitive" strategy. Moreover, the stimuli were rotated on the screen, exposing the subjects to much more information than what is available in a normal encounter with an object. Our next series of experiments was designed to constitute a "perceptual" counterpart to the CPP paradigm, by asking whether the similarity relationships among complex 3D objects are reflected in the response times and error rates obtained in a fast discrimination task performed on *static views* of these objects.

In the delayed match to sample (DMTS) experiments described below, the subjects were briefly

[11] A possible approach to the derivation of asymmetrical models of similarity from a metric-space substrate is outlined in (Edelman et al., 1996).

exposed to static views of the objects, and never to the rotating objects themselves. The subjects received no training prior to the experiment and no feedback during the experiment. The reason for not training the subjects with rotating stimuli was our desire to determine whether the perceived similarity between views of the same object is *a priori* higher than the similarity between views of different objects (showing the subjects rotating stimuli would have reduced the significance of a positive result).

The purpose of these experiments was twofold:

1. Test whether the subjects are able to group correctly together different views of the same object, and to tell apart correctly views of different objects, in the absence of prior exposure to the objects;

2. Test whether the relative arrangement of the groups of views in the representation space (as derived from the response times and confusion rates) reflects the shape-space arrangement of the objects.

Note that both tasks are computationally nontrivial, because, as discussed in section 2.1.2, the relationship between the perspective projections of two objects does not simply reflect the relationship between the two corresponding sets of parameters that control the 3D geometry of the objects.

4.1 Experimental method: DMTS

A variant of the delayed match to sample procedure was used (see Figure 12). Pairs of static views of the same object or of different objects were consecutively and briefly flashed on the screen of a computer workstation (SGI Indigo2/Extreme). The two object images were separated by a mask. The subject's task was to decide (by pressing one of two keys on the computer keyboard) whether or not the two consecutive views showed the same object under different orientations. The response time was recorded. Upon key press, a mask was briefly displayed and the next pair of images of the experimental sequence was shown. The subjects received no training, that is, they were never shown the objects rotating on the screen. No feedback was given. Short presentation times were employed (300 $msec$). As explained below, three or four distinct viewpoints (corresponding to the vertices of a regular tetrahedron inscribed in the viewing sphere) were used for each object, depending on the type of experiment. Throughout the experiment, the stimuli were presented in binocular stereo, using liquid-crystal shutter glasses synchronized with the display.

Four series of experiments were carried out, corresponding to the TRIANGLE, STAR, SQUARE, CROSS configurations described in section 2.1.2 (see Figure 2). In the experiments which involved seven objects (TRIANGLE, STAR), four viewpoints per object (snapshots taken from the vertices of a regular tetrahedron inscribed in the viewing sphere) were used, and all possible pairs of views were tested. The experimental sequence consisted of $378 = 28 \cdot 27/2$ distinct pairs of views. Because only $42 = 7 \cdot 4 \cdot 3/2 = 11\%$ out of these 378 pairs belonged to the same object, the pairs of views belonging to the same objects were repeated five times to render the proportion of "same" and "different" trials more balanced. This brought the proportion of "same" pairs to 39%, out of a total of 546 view pairs. In the experiments which involved nine objects (SQUARE, CROSS), only three views per object were used, to keep the length of the experimental sequence within reasonable limits. The considerations stated above led to repeating the pairs of views belonging to the same objects five times, bringing the proportion of "same" pairs to 40% out of a total of 459 view pairs.

4.2 Analysis of results

The object and view proximity tables (constructed as described in appendix B.2) were analyzed using both multidimensional scaling and discriminant analysis. First, as in the previous experiments, nonmetric MDS was applied both to the views and to the object tables, to determine the psychological-space configuration of the stimuli. Second, to appreciate the usefulness of the MDS view configuration, the output of the MDS analysis of the view data was submitted to nonparametric discriminant analysis, using the SAS procedure DISCRIM (Sas, 1989). Because no assumption (such as normality) regarding the distribution of the data could justifiably be made, the nearest neighbor option (a nonparametric method for density estimation) was used in the DISCRIM procedure. Based on the view configuration produced by MDS, in which each point represents a view and is labeled with the name of the corresponding object, the DISCRIM procedure develops a discriminant criterion for classifying the observations (views) into groups (objects). To estimate the misclassification error, the derived discriminant is applied to the very data from which it is derived, using a procedure called cross-validation. The results for each experiment are shown in Table 2.

As in the CPP experiments, we employed Procrustes analysis and bootstrap statistics to assess the fidelity of the reconstruction of the shape space configurations. The results are included in the figure captions. Table 1 lists the dimension weight coefficients for individual subjects for all experiments.

4.2.1 The DMTS TRIANGLE experiments

Eight subjects were tested on shapes arranged in a triangular configuration in the parameter space. The configurations displayed in Figure 13 were obtained by running MDS on the pooled data of all subjects. The recovery of the parameter space configuration is almost perfect, even though the individual configurations (not shown) are slightly deformed in various ways. Nearest-neighbor discriminant analysis confirmed that the views were well clustered by object identity (global misclassification rate was 0%). The actual pooled-subjects mean error rate was 19.0%. The mean response time was 880 *msec* for correct same trials, 1080 *msec* for incorrect same trials, 840 *msec* for correct different trials and 1177 *msec* for incorrect different trials.

INSERT FIGURE 13 HERE

4.2.2 The DMTS STAR experiments

Seven subjects were tested on shapes arranged in a star-like configuration in the parameter space. The configurations displayed in Figure 14 were obtained by running MDS on the pooled data of all subjects. Views tended to cluster by object identity: the discriminant analysis indicated an error rate of 29%; the actual pooled-subjects mean error rate was 25%. The mean response time was 895 *msec* for correct same trials, 1100 *msec* for incorrect same trials, 845 *msec* for correct different trials and 1036 *msec* for incorrect different trials.

INSERT FIGURE 14 HERE

4.2.3 The DMTS SQUARE experiments

Nine subjects were tested with nine objects positioned at the vertices of a square in the parameter space. The configurations displayed in Figure 15 were obtained by running MDS on the pooled data of all subjects. The recovery of the parameter space configuration is almost perfect. Pooled-subjects mean error rate was 25%, and the global misclassification rate obtained by discriminant analysis, 41%. The mean response time was 836 *msec* for correct same trials, 1203 *msec* for incorrect same trials, 924 *msec* for correct different trials and 1139 *msec* for incorrect different trials.

INSERT FIGURE 15 HERE

4.2.4 The DMTS CROSS experiments

Seven subjects were tested with nine objects arranged in a cross-like configuration in parameter space. The configurations displayed in Figure 16 were obtained by running MDS on the pooled data of all subjects. The mean error rate was 23%; the global error rate computed by discriminant analysis, was 15%. The mean response time was 822 *msec* for correct same trials, 1002 *msec* for incorrect same trials, 888 *msec* for correct different trials and 1251 *msec* for incorrect different trials.

INSERT FIGURE 16 HERE

4.3 Recovery of the shape-space configuration for nonsense objects

As in the CPP experiment series, we next addressed the question whether or not the recovery of the shape-space configuration is possible for unfamiliar or nonsense objects. To that end, we conducted two DMTS experiments with scrambled animal shapes: one experiment involved a star-like parameter-space configuration, and the other — a triangular configuration.

4.3.1 Scrambled objects: configuration STAR

Five subjects participated in these experiments. The average correct response rate was 70%, which means that the subjects separated correctly the different stimulus views into objects. However, to reach a success rate comparable to the performance level in the DMTS experiments with regular objects, the exposure time had to be increased from 300 (at which subject performance was little better than random) to about 750 *msec* (value determined from pilot data). The mean response time was 1090 *msec* for correct same trials, 1267 *msec* for incorrect same trials, 1140 *msec* for correct different trials and 1322 *msec* for incorrect different trials. The configuration derived by MDS from the pooled data of all subjects is depicted in Figure 17; it is strongly distorted.

Because identical experiments involving a similar number of subjects and intact shapes did result in good recovery of the objective configuration, the failure to recover that configuration for scrambled objects is unlikely to reflect individual subject differences. Indeed, the higher degree of difficulty of the present experiments compared with the other DMTS tasks is consistent with the longer decision times involved the CPP experiments with scrambled objects.

INSERT FIGURE 17 HERE

4.3.2 Scrambled objects: configuration TRIANGLE

Five subjects participated in this experiment. Again, to obtain a success rate in line with that of the experiments involving intact objects (72%), the exposure time had to be increased to 750 *msec*. The mean response time was 998 *msec* for correct same trials, 1159 *msec* for incorrect same trials, 1043 *msec* for correct different trials and 1234 *msec* for incorrect different trials. The configuration derived by MDS from the pooled data of all subjects is depicted in Figure 18, right; note that the original structure of the parameter space is completely lost.

INSERT FIGURE 18 HERE

4.4 The DMTS experiments: discussion

The main goal of these experiments was to establish that the findings regarding the representation of similarity were not peculiar to the CPP task described in the preceding section. The results of the DMTS experiments demonstrate that behavioral correlates of similarity — response times and error rates obtained in a split-second decision task — fall into a pattern predicted by a metric-space model of representation of similarity. Even though the subjects received neither prior training nor feedback, and were exposed only to a randomized sequence of briefly flashed static views belonging to a variety of objects, views of the same object were clustered together and were well separated from views of other objects in the MDS configuration derived from the behavioral data. Moreover, the relative geometrical arrangement of the objects (view clusters) in that configuration was close to the true parameter-space pattern built into the stimuli.

In the scrambled-shape DMTS experiments, a level of performance (proportion of correct responses) comparable to that obtained with animal-like shapes could only be reached with substantially longer stimulus presentation times. Furthermore, the deformation of the objects configurations produced by MDS from scrambled-shape data was much more prominent. This finding is compatible with the similarly poor recovery of the true shape-space configuration for wire-like objects in the DMTS experiments reported in (Edelman, 1995b).

5 Representation of similarities in long-term visual memory

Unlike the CPP and the DMTS experiments described above, in which subjects could rely on immediate percepts or short-term memory representations of the stimuli, the LTM experiments required that the subjects judge similarity between long-term memory traces of the stimuli.

5.1 Experimental method: LTM

The long-term memory (LTM) experimental paradigm was derived from that of the CPP experiments. The subjects were first trained to associate a nonsense word with each object, then carried out the pairs of pairs comparison task from memory, prompted by the object labels rather than by the objects themselves.

Two sets of seven objects were arranged according to the star and the triangle parameter-space configurations, and were assigned names. The names were three-letter nonsensical words (e.g., "BON", "POM", "TOC" or "ROX") and were permuted over the object sets for different subjects.

Training phase. Six subjects participated in the LTM STAR experiment and another six in the LTM TRIANGLE experiment. The subjects were taught each of the seven labeled shapes in a separate session, and were required to discriminate between that shape and six similar non-target shapes, from various viewpoints. First, the target object (with its label displayed on top) was shown rotating on the screen, the camera describing a spiral going from one pole of the viewing sphere to the other. The subject then was confronted with a series of static images, belonging either to the target object or to similar non-targets, and had to decide "target" or "non-target" by pressing one of two keys on the computer keyboard. At this stage, the target images were presented without the object label. If the subject gave an incorrect response, correcting feedback was provided by displaying the appropriate target object image together with the object label. A training session was completed when the subject achieved 90% recognition rate of the target shape.

The training sessions were repeated three or four times for each target shape, over a period of about ten days. The subjects were never exposed to more than one target in one session, and they were ignorant as to the ultimate purpose of the experiment.

Testing phase. After two to three days of rest, the subjects were tested with questions of the type: "is the BON more/less similar to POM than TOC to ROX?", for all pairs of pairs of target shapes. The questions were printed on the screen, the subjects having to type their responses. The available response options were: "more similar," "less similar," and "don't know". Testing all pairs of pairs took on the average about one hour per subject. Even though the task was demanding and required intense concentration, none of the subjects reported special difficulties in performing it.

5.2 Analysis of results

Following each experiment, the individual data were subjected to a verification of transitivity of choice, using the procedure described in section 3.2. The percentage of intransitive judgments was quite low (8%), indicating that the geometrical model of similarity was appropriate, and validating the use of MDS.

The similarity data (computed from the subjects' decisions as described in section B.1) were then entered into proximity tables and submitted to MDS. The results of the LTM experiments are shown in Figures 19 and 20. The configurations recovered by MDS were similar to the parameter-space configurations built into the stimuli, though less so than in the corresponding CPP experiments.

INSERT FIGURE 19 HERE

INSERT FIGURE 20 HERE

5.3 The LTM experiments: discussion

Perceptual tasks, such as CPP or DMTS, whose solution requires that the subject consult merely the immediate percept or at most a transient memory trace of the stimulus, do not address the issue of object representation in all its generality, because they do not tap into the stable long-term visual representations formed by repeated encounters with objects. The LTM experiments were designed to address this problem, by forcing the subjects to form and to use long-term memory representations of objects.

The results of the LTM experiments suggest that long-term memory traces are indeed structured according to similarity (and are, at least locally, metric, as indicated by the transitivity of the similarity judgments). These results may be compared to the veridical representation of similarities among shapes of US states in subjects tested by Shepard and Chipman (1970); to the best of our knowledge, no other study considered the veridicality of representation of equally complex 2D or 3D stimuli in long-term memory.

6 Computational models

It is important to realize that the major computational accomplishment in the experiments we have described so far is that of the human visual system and not of the MDS procedure used to analyze the data. Note, first, that none of the complex shape-space configurations we have tested was ever revealed to the subjects in its entirety. The two dimensions of variation built into the stimuli were well hidden, first in the 70 dimensions of the parametric shape space, then in the hundreds of thousands of the pixel-wise dimensions of the images rendered on the screen. Furthermore, as we have already pointed out, the relationship between the parametric representation of a stimulus and its physical appearance (i.e., the values of the pixels in an image of the stimulus) was highly nonlinear.

The computational feat of the recovery of the two relevant dimensions and of their faithful representation is much more difficult than finding the proverbial needle in a haystack. This is a manifestation of a general property of high-dimensional spaces that has been termed "the curse of dimensionality" (Bellman, 1961) (for an illustration of problems associated with finding structure in multidimensional spaces, see (Huber, 1985)). This feat was performed by the subjects' visual system; the role of MDS was merely to help visualize the relevant information present in the subjects' response patterns. To support this claim, we conducted a simulation study, in which two computer models of similarity perception processed the same stimuli seen by the human subjects in the DMTS experiments. The responses of each model were entered into a proximity table, which was then processed by MDS, exactly as in the analysis of the data from the DMTS experiments with human subjects.

6.1 A simple image-based similarity model

In the first model, designed to illustrate the behavior of a raw image-based measure of similarity, the images of the stimuli were convolved with an rectangular array of overlapping receptive fields (RFs) with Gaussian kernels. Thus, each image was encoded by the vector of activities of the RFs. The view-wise proximity table was constructed for each parameter-space configuration by computing the Euclidean distances between the RF activity vectors encoding the tested images. In the MDS-derived view configurations, images of different objects were grouped together by object orientation, not by object identity (see Figure 21). In other words, within the category of our animal-like shapes, image-based similarity among corresponding views of different objects is greater than the similarity among the different views of the same object. Thus, the simple image-based similarity measure we examined (which may be considered roughly analogous to an initial stage of processing in the primate visual system, such as the primary visual area V1), could not reproduce the results observed with human subjects.

INSERT FIGURE 21 HERE

6.2 An RBF network similarity model

The second model we evaluated corresponded to a higher stage of object processing, in which nearly viewpoint-invariant representations of individual objects are available (a rough analogy is to the inferotemporal visual area IT; see, e.g., (Young and Yamane, 1992; Logothetis et al., 1995)). Such a representation of a 3D object can be relatively easily formed, if several views of the object are available (Ullman and Basri, 1991). This can be done, e.g., by training a radial basis function (RBF) 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 recent studies indicate that responses of several such object-specific modules (a "Chorus of Prototypes" (Edelman, 1995c)), each coarsely tuned to a different reference shape, may be able to support veridical representation of a range of shapes similar to the reference ones (Edelman and Duvdevani-Bar, 1997; Edelman, 1996a; Edelman, 1995b).[12] Accordingly, we chose R reference objects (e.g., for the parameter-space STAR configuration, the endpoints of the limbs were chosen, making $R = 3$), and trained an RBF network to recognize each such object.

Each RBF network was trained to output 1.0 for any of the views (encoded by the activities of the underlying receptive field layer) of its preferred object, and to return a smaller value for views of different objects. For a given object set, in a given parameter space configuration, R RBF networks were trained to recognize the R reference objects of the configuration. An input view of any of the test objects evokes different levels of activity in the R RBF networks, and thus is encoded by a vector of R components.

At the RBF level, the similarity between any two views was defined as the Euclidean distance between the R-dimensional vectors of the evoked network outputs. Thus, the view proximity table was constructed by measuring distances among the R-dimensional vectors of RBF network outputs. Unlike in the case of the simple image-based similarity measure realized by the first model, the MDS-derived configurations obtained with this model showed significant resemblance to the true

[12]The decision to implement a module tuned to a particular object by an RBF network was dictated mainly by practical considerations (e.g., RBFs are much easier to train, compared to other network architectures). In (Edelman and Duvdevani-Bar, 1997), it is shown that any mechanism that fulfills certain functional requirements can play the role of the object-specific module in Chorus. It should be noted that the resulting scheme addresses only issues having to do with categorization, and skirts problems such as feature selection and segmentation.

parameter-space configurations (see Figure 22).

INSERT FIGURE 22 HERE

7 General discussion

The detailed recovery from subject data of complex similarity patterns imposed on the stimuli supports the notion of veridical representation of similarity, discussed in the introduction. Although our findings are not inconsistent with a two-stage scheme in which geometric reconstruction of individual stimuli precedes the computation of their mutual similarities, the computational model that accompanies these findings offers a more parsimonious account of the psychophysical data. Specifically, representing objects by their similarities to a number of reference shapes (as in the RBF model described in section 6.2) allowed us to replicate the recovery of parameter-space patterns observed in human subjects, while removing the need for a prior reconstruction of the geometry of the objects. In the rest of this section we survey the implications of the results of the various experiments reported above for theories of object representation, and offer an intuitive explanation of the operation of the model we used in replicating these results. An extensive discussion of the mathematical underpinnings of the model is beyond the scope of the present paper, and can be found in (Edelman and Duvdevani-Bar, 1997).

7.1 Related psychophysical approaches

Our approach to the study of representation is based on the concept of a psychological space (Shepard, 1987; Clark, 1993), on the geometrical model of similarity, and on the multidimensional scaling (MDS) techniques (Torgerson, 1958; Shepard, 1974; Shepard, 1980). In this framework, represented objects are encoded as vectors of features in a low-dimensional psychological space[13] endowed with a metric. Assuming that perceptual similarities decrease monotonically with psychological space distances, multidimensional scaling algorithms derive the psychological space configuration of the stimulus points from the table of the observed similarities. In the more general, nonmetric version of MDS, the ranks of the proximities, rather than their values, are used in the computation (Shepard, 1962a; Shepard, 1962b; Kruskal, 1964a; Kruskal, 1964b; Shepard, 1966). Technically, MDS finds a monotonic function that transforms the similarities into numbers interpretable as distances in a space of low dimensionality by minimizing the discrepancy (stress) between the distances in the target space and the distances specified in the data. Two psychological space metrics were found to fit most similarity data: Euclidean – used for integral stimuli (such as lightness-saturation), and

[13]There are good reasons to keep the representation space low-dimensional (Edelman and Intrator, 1997), as well as evidence that it is, as a rule, indeed such (Shepard, 1980).

city-block – used when the perceptual dimensions of the stimulus are compelling and easily separable (Attneave, 1950). A more general metric is the Minkowski metric: $d_{ij} = [\sum_k |x_{ik} - x_{jk}|^r]^{1/r}$. This formula reduces to the Euclidean metric for $r = 2$, the city-block metric for $r = 1$, and for $1 < r < 2$ yields intermediate metrics, more appropriate in some experimental situations other than the two limit cases (Torgerson, 1958).

Physical distances (i.e., objectively measurable differences) between stimuli are not directly involved in the MDS model: distances are taken in the psychological space, not in a physical features space. This is not to say, however, that objectively measurable similarities are irrelevant. To infer the features used in the similarity judgments, the MDS configuration is usually interpreted in terms of various physical measurements that characterize the stimuli. This methodology was applied, for example, in studies of face representation (Rhodes, 1988; Young and Yamane, 1992). The process of interpretation of the MDS configuration can be facilitated by using stimuli whose objective similarities to each other are relatively clear (as in the case of the US states, used by Shepard and Chipman, 1970). Even better, stimuli for which objective similarities are precisely defined should be used; this approach was taken by Shepard and Cermak, 1973, as well as here. In the present study, we went on to correlate the MDS configurations with patterns of similarities built into the stimuli, which amounted to the use of MDS in confirmatory mode (Borg and Lingoes, 1987), and allowed us to impart statistical significance to the notion of veridicality of representations employed by our subjects.

7.2 Implications for theories of recognition and categorization

Our experiments were designed to assess the (hitherto unknown) degree of veridicality of internal representation of similarity among 3D objects, and not to test particular theories of recognition and categorization. Nevertheless, the outcome of these experiments (and, in particular, of the computational simulations) can be interpreted as evidence in favor of some such theories, and against others, as argued below.

7.2.1 Distinctive features

It is unlikely that our subjects simply stored a list of distinctive features for each object. Such diagnostic-feature representations can only subserve object recognition, i.e., correct labeling of previously seen inputs, and are, therefore, of limited use. To assign correctly a novel input (e.g., a view of an object that appears, for the first time, in a trial in the DMTS experiment) to its class, one has to compute similarities between the stimulus and the representations of other objects stored in memory. If only the distinctive features were stored, successful *recognition* would indeed be possible, but the proper *classification* of novel stimuli, and the observed recovery of the objective contrasts between the stored stimuli, would be impossible.[14]

7.2.2 Structural descriptions

As we already pointed out, our experimental results do not rule out the possibility that the relevant similarities were computed following the formation of full-fledged representations of individual objects. In the preceding section, we saw that lists of distinctive features are an unlikely candidate for such an intermediate representation. We observe now that geon structural descriptions (GSD's), postulated in (Biederman and Gerhardstein, 1993), are equally unlikely here. All our objects shared the same structural description (that is, the "anatomy" of all the animals was the same). Thus, the distance between any two of our objects as measured in the space of such descriptions is nominally the same, and is equal to zero. Consequently, veridical representation of (parametric) similarity among objects cannot be explained in terms of distances measured in the GSD space.

It may be noted that the accumulation of metric changes that accompanied the transformation of one object into another sometimes resulted in a qualitative difference between the objects (in Figure 1, this is apparent as the difference between the head shapes). This means that the task that our subjects faced fulfills an important precondition set by Biederman and Gerhardstein for the applicability of the GSD framework. The precondition is, roughly, that the stimuli to be discriminated differ on the basic, and not merely the subordinate, level of categorization. Could these qualitative differences (the only ones available under a GSD representation, where the metric information is suppressed) account for the observed patterns of veridicality? This is possible, but not likely, for two reasons. First, the qualitative differences are only accidentally related to the underlying parametric changes, making the former a poor indicator of the latter.[15] Second, as attested by independent studies, subjects are capable of veridical perception of similarities even in situations where all the shapes are qualitatively absolutely identical (e.g., are all rectangles that differ in size and aspect ratio; see Shepard, 1980, for references).

7.2.3 Similarities to reference objects

Theories of categorization, mentioned briefly in section 1.2, fare much better as models of similarity perception in our experiments. In particular, one may observe that the Chorus model we used in

[14] Note that if the task called merely for discrimination between objects, and if distinctive features such as unique coloration of different objects were available, then, as pointed out by a reviewer, one should not expect the visual system to store the objects in a form sufficiently detailed to support veridical representation of shape similarities. We shall return to this point in section 7.4.

[15] Note, also, that the GSD framework does not provide a natural metric for the structure space (although many such metrics can, of course, be defined *ad hoc*; e.g., one may decide that the change of one geon corresponds to a distance of one unit, etc.). Without such a metric, one can say that two distinct GSD's are different, but not by how much.

replicating the performance of our human subjects resembles Nosofsky's Generalized Context Model (GCM), in that it represents objects by similarities to other, reference objects (which play the role of GCM's exemplars). A central feature shared by Chorus and GCM is the distinction between persistent and ephemeral representations (Edelman, 1995c): to represent a wide variety of stimuli, it is enough to store a relatively small number of (persistent) object representations explicitly, with the rest of the stimuli being treated as ephemeral, and characterized by their shape-space proximities to the persistent objects.

Whereas GCM concentrates on the decision stage of the process of object categorization, which is postulated to occur in an internal shape representation space, Chorus provides a principled link between this space and the external (distal) objects, treated as points in a parameter space. As shown in (Edelman and Duvdevani-Bar, 1997), a Chorus-like mechanism is bound to result in a veridical representation of the distal parameter space, subject to two main conditions. First, the response of the individual modules comprising Chorus (in our case, the RBF networks) must decrease smoothly and monotonically with parameter-space distance between the current input and the module's optimal stimulus (i.e., the object on which the RBF module has been trained). Second, the module's response must be much less sensitive to changes in viewpoint than to changes in object shape (in our case, the nearly constant response of the RBF module to different views is enforced by the training algorithm). As we saw in section 6.2, a model that fulfills these conditions indeed exhibits the ability to form a veridical representation of a distal shape space.

7.3 Conditions for veridicality

The importance of our psychophysical findings for the understanding of object representation would have been greatly reduced if veridicality were an automatic consequence of some simple operation such as the measurement of distances between stimuli in the "raw" image space. The failure of the image-based distance model (section 6.1) to replicate human data shows, reassuringly, that images of objects must be further processed before distance computation can result in veridicality. Interestingly, one need not go *much* further: the Chorus (multi-RBF) model turned to be capable of veridical representation (note that whereas the image-based model fulfills only the first of the two conditions for veridicality mentioned above, Chorus fulfills both; cf. Edelman and Duvdevani, 1997).

An immediate implication of the Chorus model is that veridicality should break down if the stimuli on which the system is tested are sufficiently unlike any of the reference (persistent) objects, for which dedicated modules are available. This observation provided the motivation for the experiments with scrambled shapes. If the subjects represent the scrambled stimuli in terms of similarities to familiar objects, the recovery of parameter-space similarity patterns should be impossible, unless there is a chance to study the new objects and to form representations thereof.

Indeed, we found that in the DMTS experiments with scrambled shapes the recovery was poor, and in the CPP experiments, where the subjects could scrutinize the scrambled objects for a relatively long time, the recovery was somewhat better (although not as good as for the intact shapes, which looked more like real, familiar animals).[16]

7.4 Limits of veridicality

Although veridicality is certainly a desirable feature of a representational system, in some situations it may be deemed superfluous, and may be given up altogether. As an extreme example, consider a system that has to represent a single object. Clearly, in this case it does not matter what form the representation takes; any entity or symbol whatsoever will do. If the visual world in which the system is embedded contains several objects, each painted a different color, then, too, veridicality in the shape domain is not needed (if the goal of the system is merely discrimination among the objects).

A more realistic situation is one in which there are classes of shapes, with the between-class differences relatively pronounced, compared to the within-class differences (think of several basic-level categories of real-world objects, e.g., cats, hats, and airplanes). In this situation, veridicality is elusive, because the proper ranking of distal similarities is difficult to define: is a cat really more similar to a hat or to an airplane? Finally, when the system is required to represent categories whose members resemble each other to a considerable degree (e.g., cats, pigs, squirrels), veridicality is a must: a system which represents cats as more similar to pigs than to squirrels surely needs a tune-up. Thus, like many other characteristics of visual representation and processing (Schyns, 1996), the need for veridicality can be seen to depend on the task at hand.

7.5 Conclusions

We reported a series of experiments involving animal-like shapes, in which we were able to recover the objective parameter-space configuration imposed on the stimuli from the subject data, using multidimensional scaling. Thus, the portion of the subjects' internal representation space devoted to these shapes appears to be a faithful replica of the distal parameter space.[17] While comparable recovery of objective similarity relationships has been demonstrated in the past in the classical MDS

[16] As pointed out by a reviewer, the scrambled objects differed from the intact ones in a number of ways (e.g., symmetry and occlusion), any of which could have contributed to the difference in the subjects' performance. Our explanation of that difference in terms of a specific computational model seems, however, to be more parsimonious than an invocation of separate mechanisms specially sensitive to symmetry, occlusion, etc. The resolution of this issue will have to await a future work.

[17] A similar faithful recovery of parameter-space configurations has been recently demonstrated in the monkey, using the DMTS paradigm (Sugihara et al., 1996).

experiments involving simple perceptual qualities such as color (Shepard, 1962b), as well as some 2D contours (Shepard and Cermak, 1973; Cortese and Dyre, 1996), and even 3D shapes (Edelman, 1995b), our results show for the first time that such recovery is possible with complex 3D objects, arranged in a variety of shape-space configurations. This finding is especially significant vis a vis the complicated relationship between the parameter values that defined the location of an object in the multidimensional shape space, and the image of the object as it appeared on the screen.

We note that the full power of the MDS as a tool for mapping the internal representation space of subjects can be realized only if the experimental approach constrains the interpretation of the potential outcome to a sufficient degree; cf. (Shepard and Cermak, 1973). An application of MDS always produces a solution, which, furthermore, will have a low stress (i.e., low residual discrepancy with the data) if the number of points is small. It is important to realize that our conclusions are based not on low stress *per se*, but rather on the recovery of a specific pattern built into the stimuli. This effectively allowed us to invoke MDS in a confirmatory mode (Borg and Lingoes, 1987), and to demonstrate the statistical significance of the solution.

The remarkably faithful reconstruction of the parameter-space arrangements of the stimuli from the subject data would have been unlikely if the subjects stored distinctive features or structural descriptions of the input shapes. On a more positive note, our psychophysical findings and computational simulations, as well as related work by others (Tanaka, 1992; Logothetis et al., 1995), suggest that the biological substrate for object representation may be not unlike a "Chorus of Prototypes" — a collection of recognition devices, each coarsely tuned to a reference shape (Edelman, 1995c; Edelman, 1996b).

Acknowledgments

We thank S. Ullman for constructive comments in the course of the research described in this paper. We have also benefited from discussions with S. Duvdevani-Bar, K. Grill-Spector, D. Mumford, T. Poggio, and R. N. Shepard. Comments by two anonymous reviewers were of great help in improving the presentation of our findings. This research was supported in part by a grant from the Basic Research Foundation, administered by the Israel Academy of Arts and Sciences. SE is an incumbent of the Sir Charles Clore career development chair at the Weizmann Institute of Science.

A Multidimensional scaling

A.1 Quality of the MDS solution

The output of the MDS procedure includes, in addition to the recovered configuration, two numbers representing the quality of the solution. The first of these is the *Badness-of-Fit* index, corresponding to the MDS stress (Kruskal and Wish, 1978). This index ranges between zero and one, and measures how far (in a least-squares sense) do the interpoint distances in the metric psychological space of the given dimensionality deviate from the measured proximity table. Increasing the dimensionality of the solution causes a decrease in the Badness-of-Fit, by providing additional degrees of freedom for fitting the data. The second parameter pertinent to the quality of the solution is the *Distance Correlation*, also ranging between zero and one. This is a measure of goodness of fit, defined as the correlation between the recovered and the given distances.

We used the nonmetric version of the MDS procedure that is a part of the SAS statistical package (Sas, 1989). A graphical example of the power of nonmetric MDS is shown in Figure 5: the original 7-point configuration is presented on the left and the configuration recovered from the ranks of interpoint distances on the right. The recovery improves with increasing number of points; seven points was the minimal size of the configurations used in our experiments. To test the reconstruction accuracy statistically, we ran the the SAS MDS procedure on 100 random planar random configurations of seven points (that is, on 100 dissimilarity tables containing the ranks of the interpoint distances). The Pearson correlation coefficient characterizing the relation between the length values of the line segments joining corresponding points in the true and the recovered configurations, was 0.87, at a significance value of 0.0001, demonstrating that the MDS solution accounts for much of the variance of the original data.

INSERT FIGURE 5 HERE

A.2 Pooling the subject data

We analyzed the similarity data both on a subject by subject basis and for all the subjects simultaneously. In this latter case, we pooled the individual data sets, by concatenating the subject matrices and using the condition=matrix option in the SAS MDS procedure; this corresponds to an individually weighted MDS analysis (Carroll and Chang, 1970), which returns, in addition to the configuration common to all the subjects, a set of individual weights called dimension coefficients, one per axis, which represent the scaling transformations specific for each subject. Large differences in the values of the weights indicates lack of concordance among subjects.

B Construction of the proximity tables

B.1 The CPP experiments

The entry (i, j) in the $N \times N$ proximity table encodes the perceived similarity between objects i and j. Because the proximity table was to be submitted to nonmetric MDS, only the rank order of the inter-object similarities was needed. Assume that the object pair i, j involved in $P - 1$

comparisons with the rest of the pairs was considered to be the more similar pair $v_{i,j}$ times. Then the object pairs can be ranked by similarity in the order of the values of $v_{i,j}$: the higher the value of $v_{i,j}$, the higher the similarity between the objects i, j. Therefore the (i, j) entry in the proximity table is simply $v_{i,j}$. One problem with this simple procedure is that it usually yields many ties (equal values for v). To break the tie between two pairs i, j and r, s, one can use the outcome of the direct comparison between these two pairs, giving the higher rank to the pair which the subject deemed more similar. Another possibility is to sum the v values of all the pairs which were judged to be less similar than the i, j, do the same for the r, s pair and then give the higher rank to the pair with the larger value for this sum.[18]

B.2 The DMTS experiments

The data from the DMTS experiments consisted of the response times and the "same object" or "different object" decisions for all pairs of views that were tested. A view pair, tested in a DMTS trial, consisted of views of either different or of identical objects. As explained below, both the response time and the error rate associated with a view pair were used to measure pairwise view similarity. The basic idea behind this joint measure was that similarity depends both on the confusion rate between the two stimuli and on the speed of the decision: the more often and the faster two stimuli are confused, the more similar they are. Thus, both incorrect and correct decisions were used. Object similarity was obtained by collapsing across the different views of a given object.

Two square proximity tables were constructed for each data set: a view proximity table and an object proximity table.

View proximity table. The entry (i, j) in the view proximity table was intended to measure the perceived similarity[19] between the views i and j, views which may or may not belong to the same object (the index i runs between 1 and $7 \cdot 4 = 28$ in the 7-object experiments and between 1 and $9 \cdot 3 = 27$ in the 9-object experiments). In determining the proximity between the views i and j, the main factor is the proportion of times the subject responded "same object" when confronted with the i, j pair. At equal proportions, the response times can be used to make a finer distinction: the shorter the response time in the "same object" decision, the higher the similarity; the shorter the response time in the "different object" decision, the lower the similarity. Thus, the maximum of

[18] An analogous problem is ranking P teams which participate in a "complete" tournament. The outcome of the game could be either victory (which brings 1 point) or defeat (0 points). Each team plays all other teams, and the number of points (victories) can be used to rank the teams. The outcome of the direct games can be used to break the ties. The other tie-breaking method considers the sum of points of all the teams beaten by each of the tied teams – that is the tie is broken in favor of the team which has beaten the more valuable adversaries.

[19] Actually, the term similarity here is not entirely appropriate, because the subject was asked not to estimate similarity between views, but to decide whether or not a pair of views could belong to the same object.

similarity corresponds to maximum confusion rate and minimum "same object" response time, and the minimum of similarity corresponds to minimum confusion rate and minimum "different object" response time. Based on these arguments, the view proximity matrix was constructed as follows:

1. if the "same object" decision was predominant, then similarity was defined as the sum: a large weight $\times$ confusion rate $+$ the inverse of the average "same object" response time.

2. if the "different objects" decision was predominant, then similarity was defined as the sum: a smaller weight $\times$ confusion rate $+$ the "different objects" response time.

The two weights were chosen so as to ensure that the proximity in the cases when the "same object" decision predominated was always greater than the proximity in the cases when the "different objects" decision predominated, irrespective of the response times.

Object proximity table. The entry (i, j) in the object proximity table was intended to measure the perceived similarity between the objects i and j (the index i runs between 1 and 7 in the 7-object experiments and between 1 and 9 in the 9-object experiments). In determining the proximity between the objects i and j, the different pairs of views of these objects were considered as instances of pairings of the objects themselves – that is, the all responses to the pairing of views of object i with views of object j were pooled together in a overall measure of similarity between the objects themselves. If, for instance, four views per object were used, there were $4 \cdot 4 = 16$ instances of pairing objects i and j. Out of these 16 decisions, some were "same objects" decisions (these defined the confusion rate), the rest being "different objects" decisions. Considerations similar to those made above for the view proximities lead to the following definition of object proximity:

1. if the "same object" decision was predominant, then similarity was defined as the sum: a large weight $\times$ confusion rate $+$ the inverse of the average "same object" response time.

2. if the "different objects" decision was predominant, then similarity was defined as the sum: a smaller weight $\times$ confusion rate $+$ the "different objects" response time.

The two weights were chosen so as to ensure that the proximity in the cases when the "same object" decision predominated was always greater than the proximity in the cases when the "different objects" decision predominated, irrespective of the response times.

References

Ashby, F. G. (1992). Multidimensional models of categorization. In Ashby, F. G., editor, *Multidimensional models of perception and cognition*, pages 449–484. Lawrence Erlbaum, Hillsdale, NJ.

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

Attneave, F. (1950). Dimensions of similarity. *American Journal of Psychology*, 63:546–554.

Beals, R., Krantz, D. H., and Tversky, A. (1968). The foundations of multidimensional scaling. *Psychological Review*, 75:127–142.

Bellman, R. E. (1961). *Adaptive Control Processes.* Princeton University Press, Princeton, NJ.

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

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

Biederman, I. and Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6):1506–1514.

Binford, T. O. (1982). Survey of model-based image analysis systems. *International Journal of Robotics Research*, 1:18–64.

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

Brown, D. R. and Andrews, M. H. (1968). Visual form discrimination: multidimensional analysis. *Perception and Psychophysics*, 3:401–406.

Carroll, J. D. and Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of the Eckart-Young decomposition. *Psychometrika*, 35:283–319.

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

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.

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

Duda, R. O. and Hart, P. E. (1973). *Pattern classification and scene analysis.* Wiley, New York.

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

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

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

Edelman, S. (1996a). Receptive fields for vision: from hyperacuity to object recognition. In Watt, R., editor, *Vision*. MIT Press, Cambridge, MA. in press.

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

Edelman, S. (1997). *Representation and recognition in vision.* forthcoming.

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

Edelman, S. and Duvdevani-Bar, S. (1997). 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. (1996). 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.

Ellis, R., Allport, D. A., Humphreys, G. W., and Collis, J. (1989). Varieties of object constancy. *Q. Journal Exp. Psychol.*, 41A:775–796.

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

Green, D. M. and Swets, J. A. (1966). *Signal detection theory and psychophysics.* Wiley, New York.

Huber, P. J. (1985). Projection pursuit (with discussion). *The Annals of Statistics*, 13:435–475.

Huttenlocher, D. P. and Ullman, S. (1987). Object recognition using alignment. In *Proceedings of the 1st International Conference on Computer Vision*, pages 102–111, London, England. IEEE, Washington, DC.

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.

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

Kruskal, J. B. (1964b). Non-metric multidimensional scaling: a numerical method. *Psychometrika*, 29:115–129.

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

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

Lowe, D. G. (1986). *Perceptual organization and visual recognition.* Kluwer Academic Publishers, Boston, MA.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, E. G., editor, *Handbook of mathematical psychology*, pages 103–190. Wiley, New York.

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.

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

Mundy, J. L. and Zisserman, A., editors (1992). *Geometric invariance in computer vision.* MIT Press, Cambridge, MA.

Nosofsky, R. M. (1985). Overall similarity and the identification of separable-dimension stimuli: a choice model analysis. *Perception and Psychophysics*, 38:415–432.

Nosofsky, R. M. (1992a). Exemplar-based approach to relating categorization, idenitification, and recognition. In Ashby, F. G., editor, *Multidimensional models of perception and cognition*, pages 363–394. Lawrence Erlbaum, Hillsdale, NJ.

Nosofsky, R. M. (1992b). Similarity scaling and cognitive process models. *arp*, 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.

Perrin, N. A. (1992). Uniting identification, similarity and preference: general recognition theory. In Ashby, F. G., editor, *Multidimensional models of perception and cognition*, pages 123–146. Lawrence Erlbaum, Hillsdale, NJ.

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

Rhodes, G. (1988). Looking at faces: first-order and second-order features as determinants of facial appearance. *Perception*, 17:43–63.

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

Schyns, P. G. (1996). Diagnostic recognition: task constraints, object information, and their interactions. submitted.

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

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with unknown distance function. part ii. *Psychometrika*, 27(2):219–246.

Shepard, R. N. (1966). Metric structures in ordinal data. *J. Math. Psychology*, 3:287–315.

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

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 39:373–421.

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

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

Shepard, R. N. and 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.

Stankiewicz, B. and Hummel, J. (1996). MetriCat: a representation for basic and subordinate-level classification. In Cottrell, G. W., editor, *Proceedings of 18th Annual Conf. of the Cognitive Science Society*, pages 254–259, San Diego, 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.

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

Tarr, M. J. and Bülthoff, H. H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple-views? *Journal of Experimental Psychology: Human Perception and Performance*, 21(6):1494–1505.

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

Tomonaga, M. and Matsuzawa, T. (1992). Perception of complex geometric figures in chimpanzees (*pan troglodytes*) and humans (*homo sapiens*): analyses of visual similarity on the basis of choice reaction time. *J. Comparative Psychol.*, 106:43–52.

Torgerson, W. (1958). *Theory and Methods of Scaling.* John Wiley & Sons, Inc., New York.

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

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

Ullman, S. (1996). *High level vision.* MIT Press, Cambridge, MA.

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

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

Table 1

The individual-subject dimension weights for each experiment. "MATRIX" indexes the similarity matrix of each individual subject. The weights for each psychological space dimension vary somewhat across subjects, as can be seen under "Dimension Coefficients".

See Figures 9,14,19,20, showing examples of pooled-subject MDS solutions.

CPP Triangle:

	Dimension Coefficients	
MATRIX	1	2
1	1.24	0.68
2	1.10	0.89
3	1.12	0.87
4	0.97	1.03
5	1.18	0.78
6	1.12	0.86
7	1.13	0.85
8	0.99	1.01

CPP Star:

	Dimension Coefficients	
MATRIX	1	2
1	1.36	0.39
2	0.97	1.03
3	1.13	0.86
4	0.92	1.07
5	1.29	0.58
6	1.12	0.87

CPP Square:

	Dimension Coefficients	
MATRIX	1	2
1	0.85	1.13
2	0.97	1.03
3	1.00	1.00
4	1.01	0.99

5	1.14	0.84
6	0.67	1.24

CPP Cross:

	Dimension Coefficients	
MATRIX	1	2
1	1.12	0.86
2	1.07	0.92
3	1.22	0.72
4	0.98	1.02
5	0.90	1.09
6	0.90	1.09

DMTS Triangle (objects):

	Dimension Coefficients	
MATRIX	1	2
1	1.13	0.85
2	1.24	0.69
3	1.03	0.96
4	1.10	0.89
5	1.05	0.94
6	1.21	0.74
7	1.08	0.91
8	0.87	1.11

DMTS Triangle (views):

	Dimension Coefficients	
MATRIX	1	2
1	1.05	0.95
2	1.18	0.78
3	1.02	0.98
4	0.06	0.94
5	0.98	1.02
6	1.08	0.91
7	1.10	0.89
8	0.96	1.04

DMTS Star (objects):

	Dimension Coefficients	
MATRIX	1	2
1	1.03	0.97
2	1.40	0.21

DMTS Star (views):

	Dimension Coefficients	
MATRIX	1	2
1	1.01	0.99
2	1.15	0.82

3	1.16	0.82	3	1.02	0.98
4	1.02	0.98	4	0.99	1.01
5	1.11	0.88	5	1.01	0.99
6	1.02	0.98	6	1.07	0.92
7	1.04	0.96	7	1.00	1.00

DMTS Square (objects):

	Dimension Coefficients	
MATRIX	1	2
1	1.01	0.99
2	0.87	1.11
3	0.90	1.09
4	1.05	0.95
5	1.01	0.99
6	1.00	1.00
7	1.00	1.00
8	1.11	0.87
9	0.97	1.02

DMTS Square (views):

	Dimension Coefficients	
MATRIX	1	2
1	1.05	0.96
2	0.84	0.87
3	1.12	0.98
4	1.09	1.02
5	0.80	0.80
6	1.19	1.14
7	1.02	1.01
8	0.95	0.94
9	1.10	1.02

DMTS Cross (objects):

	Dimension Coefficients	
MATRIX	1	2
1	0.97	1.03
2	1.10	0.88
3	1.18	0.78
4	1.00	1.00
5	1.16	0.81
6	0.98	1.02
7	0.99	1.01

DMTS Cross (views):

	Dimension Coefficients	
MATRIX	1	2
1	1.02	0.98
2	1.09	0.90
3	1.00	1.00
4	0.97	1.03
5	1.03	0.97
6	1.02	0.98
7	0.91	1.08

LTM Star:

Dimension Coefficients

LTM Triangle:

Dimension Coefficients

MATRIX	1	2
1	0.98	1.02
2	1.19	0.76
3	0.93	1.07
4	1.15	0.82
5	1.11	0.88
6	1.21	0.73

MATRIX	1	2
1	1.28	0.61
2	1.09	0.90
3	1.21	0.73
4	0.90	1.09
5	1.14	0.83
6	1.23	0.69

Table 2

Discriminant analysis of the views configuration for the pooled subject data in the DMTS experiments. The "Rates" in the tables are the error count estimates for classification of views into seven objects classes (for the star and triangle configurations), or nine object classes (for the cross and square configurations). "Total" refers to the global misclassification error; "Priors" are the frequencies of appearance of the various stimuli.

DMTS TRIANGLE

Object	1	2	3	4	5	6	7
Rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Priors	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429

Total = 0.0

DMTS STAR

Object	1	2	3	4	5	6	7
Rate	0.2500	0.2500	0.5000	0.2500	0.2500	0.2500	0.2500
Priors	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429

Total = 0.2857

DMTS SQUARE

Object	1	2	3	4	5	6	7
Rate	0.3333	0.0000	0.6667	0.0000	0.3333	0.6667	1.0000
Priors	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111

Object	8	9
Rate	0.6667	0.0000
Priors	0.1111	0.1111

Total = 0.4074

DMTS CROSS

Object	1	2	3	4	5	6	7
Rate	0.0000	0.0000	0.3333	0.6667	0.3333	0.0000	0.0000

Priors	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111	0.1111

Object	8	9
Rate	0.0000	0.0000
Priors	0.1111	0.1111

Total = 0.1481

DMTS STAR, scrambled

	1	2	3	4	5	6	7
Rate	1.0000	1.0000	0.3333	0.6667	0.3333	0.0000	0.6667
Priors	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429

Total = 0.5714

DMTS TRIANGLE, scrambled

	1	2	3	4	5	6	7
Rate	0.3333	0.6667	1.0000	1.0000	1.0000	0.0000	0.6667
Priors	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429

Global error rate 0.6667

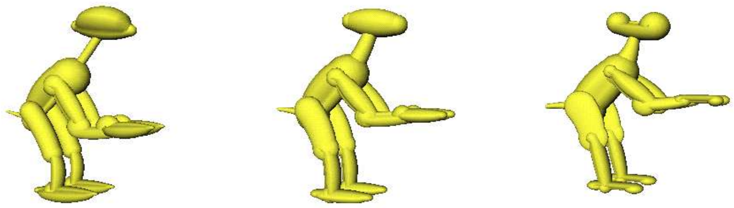


Figure 1: Three of the parameterized animal-like shapes used in the psychophysical experiments.

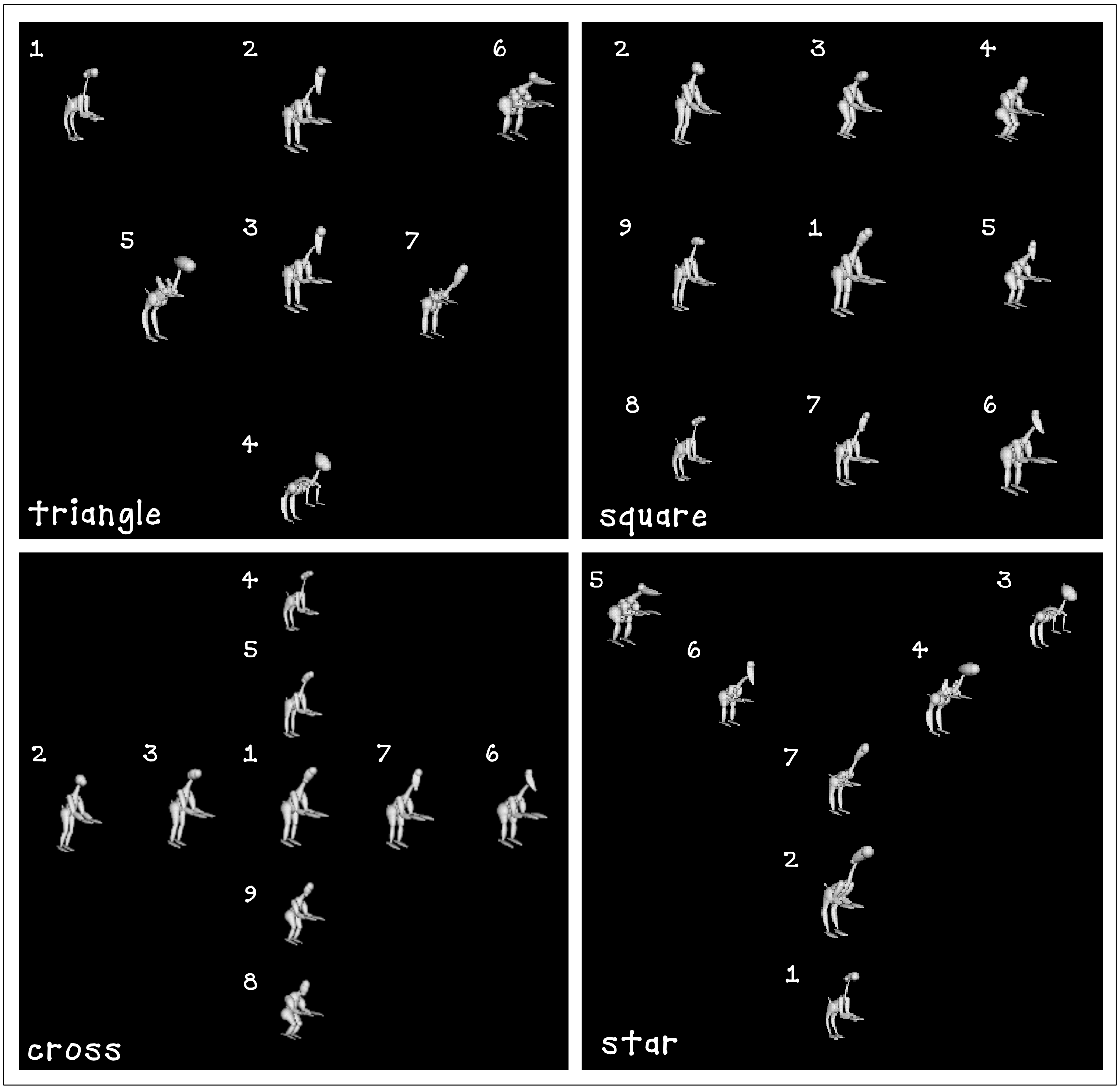


Figure 2: The four parameter-space configurations illustrating the similarity patterns built into the experimental stimuli.

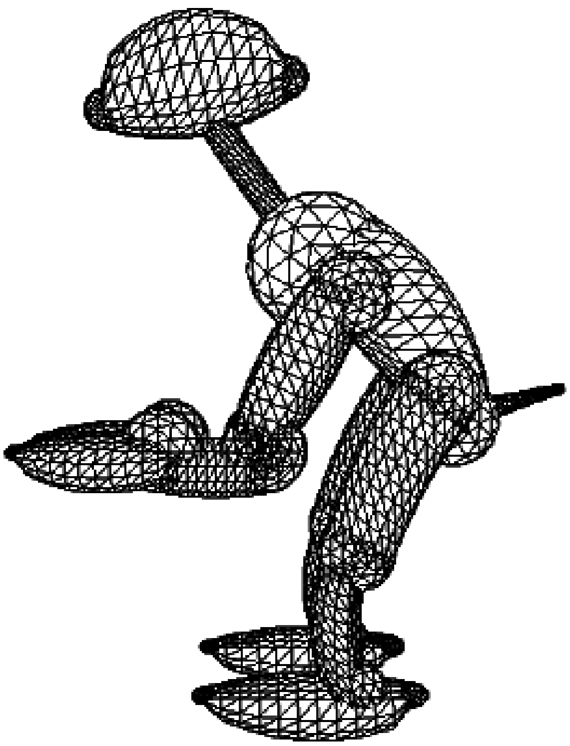


Figure 3: The surface of the object is approximated by a fine triangular mesh. The 3D coordinates of the vertices, taken in a predefined order, constitute a high-resolution objective geometrical encoding of the shape.

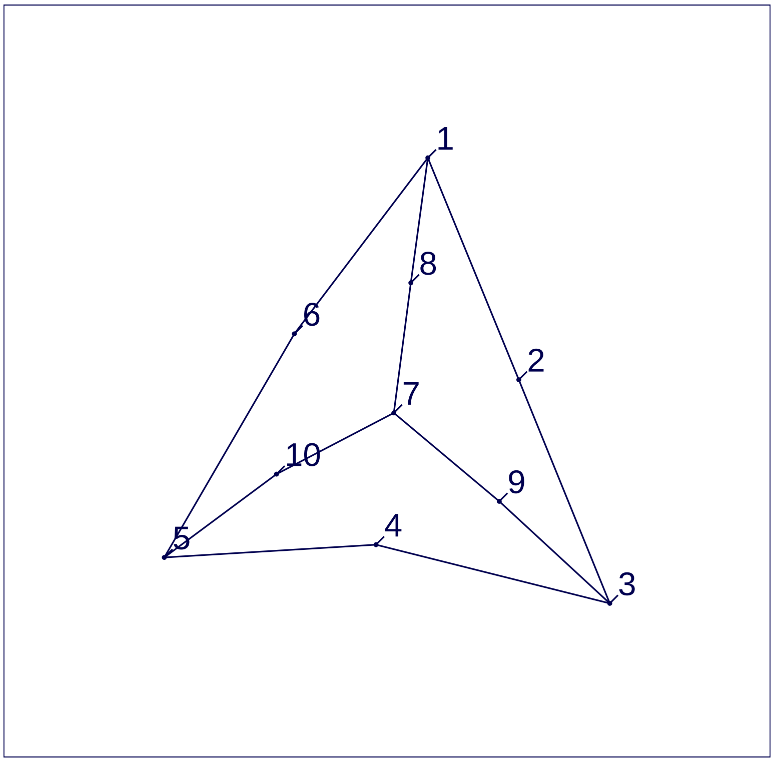
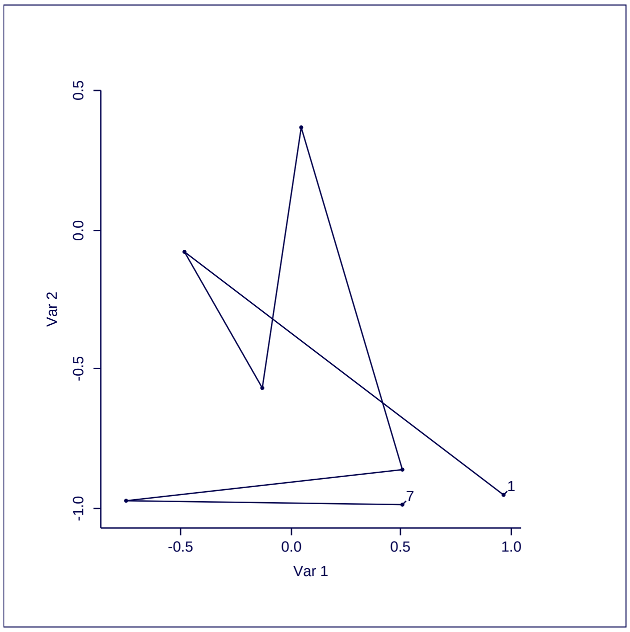


Figure 4: TRIANGLE and STAR parameter-space configurations recovered by MDS from distances computed in the space of the coordinates of the object-surface mesh (see Figure 3).
The animal shapes (corresponding to the star and triangle configurations in this example) were represented as lists (arrays) of 3D coordinates x_i, y_i, z_i $(i = 1 \ldots N)$ of the vertices of their surface meshes. The total number of points N of all meshes is the same and, by construction, the ordering of the points insures the correct feature correspondence between different animal shapes. Under this direct encoding (as opposed to the indirect encoding provided by the parameter vectors), the geometrical distance between any two animal shapes was defined as the Euclidean distance between their arrays of 3D point coordinates. Multidimensional scaling was then applied to the table of pairwise distances thus computed; the resulting configuration is shown in this figure. The triangle and the star configurations are clearly visible. The perfect recovery of both parameter-space configurations allows us to consider our parameterization of the shapes to be objective, in the sense that it reflects the precise geometrical differences between the shapes.



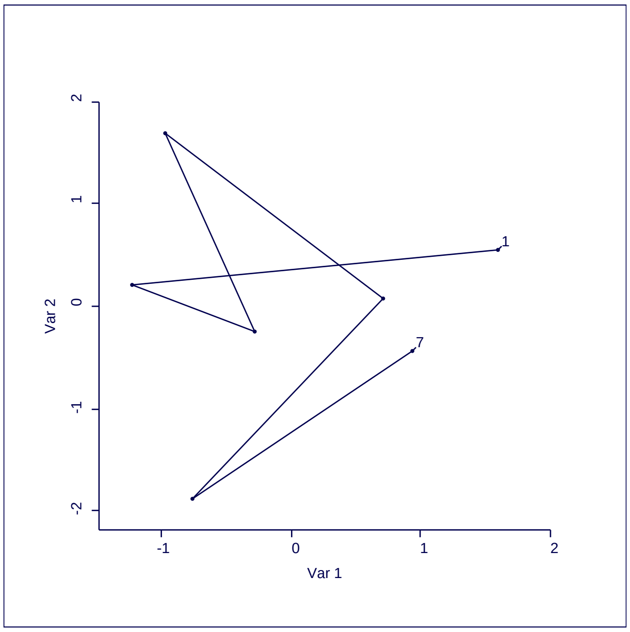


Figure 5: *Left:* a random configuration of seven points in the plane. *Right:* its reconstruction by nonmetric multidimensional scaling, from ranks of inter-point distances.

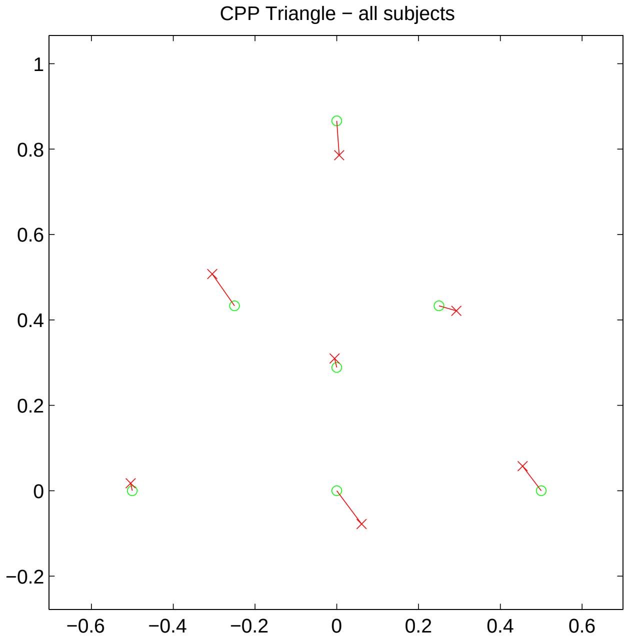


Figure 6: The CPP experiment, configuration TRIANGLE. Symbols: $\circ$ – true configuration; $\times$ – 2D configuration derived by MDS from the pooled subject data (eight subjects), then Procrustes-transformed to fit the true one. Lines connect corresponding points. Stress is 0.20.
The coefficient of congruence between the MDS-derived configuration and the true one was 0.99 (expected random value, estimated by bootstrap from the data: 0.88 ± 0.03, mean and std dev; 50 permutations of the point order were used in the bootstrap computation). Procrustes distance: 0.17 (bootstrap-estimated expected random value: 0.77 ± 0.07).

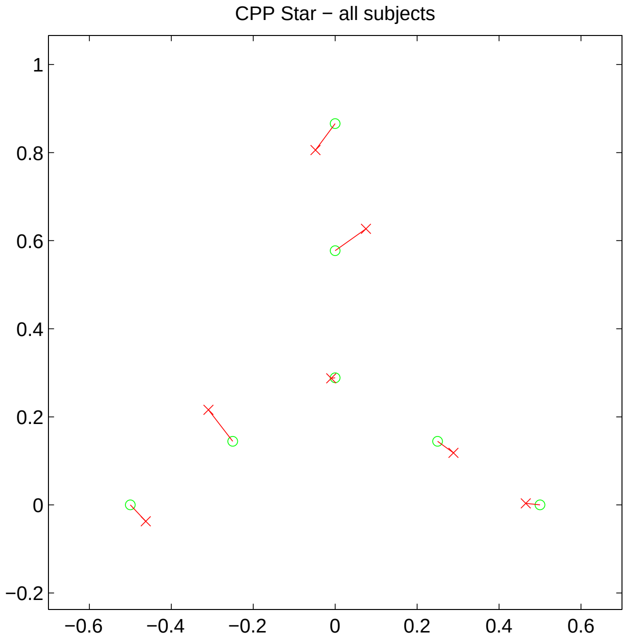


Figure 7: The CPP experiment, configuration STAR. Symbols: $\circ$ – true configuration; $\times$ – 2D configuration derived by MDS from the pooled subject data (six subjects), then Procrustes-transformed to fit the true one. Lines connect corresponding points. Stress is 0.11.
The coefficient of congruence between the MDS-derived configuration and the true one was 0.99 (expected random value, estimated by bootstrap from the data: 0.86 ± 0.03, mean and std dev; 50 permutations of the point order were used in the bootstrap computation). Procrustes distance: 0.12 (bootstrap-estimated expected random value: 0.77 ± 0.09).

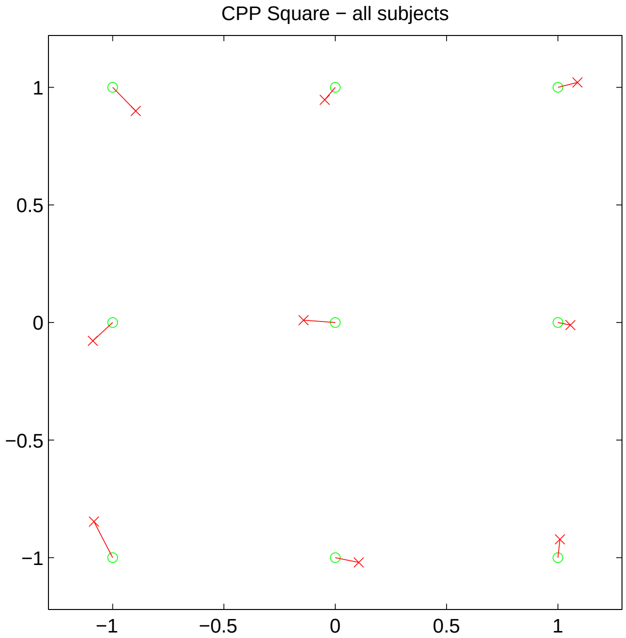


Figure 8: The CPP experiment, configuration SQUARE. Symbols: ∘ – true configuration; × – 2D configuration derived by MDS from the pooled subject data (six subjects), then Procrustes-transformed to fit the true one. Lines connect corresponding points. Stress is 0.14.
The coefficient of congruence between the MDS-derived configuration and the true one was 0.99 (expected random value, estimated by bootstrap from the data: 0.89 ± 0.02, mean and std dev; 50 permutations of the point order were used in the bootstrap computation). Procrustes distance: 0.28 (bootstrap-estimated expected random value: 2.42 ± 0.15).

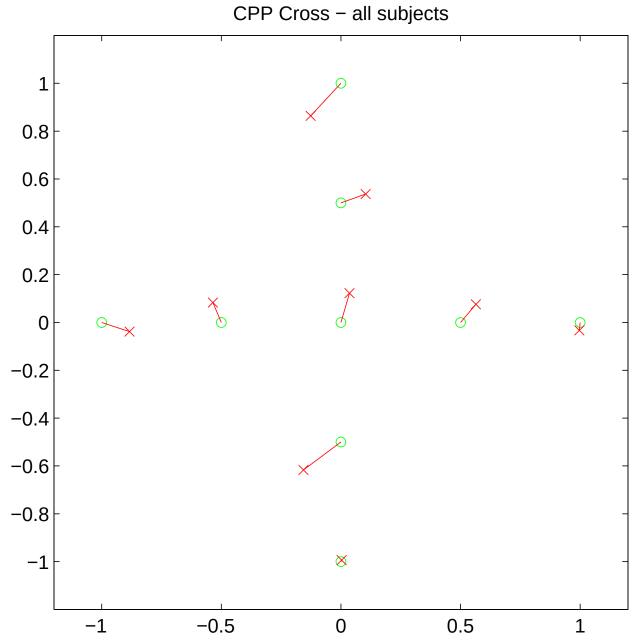


Figure 9: The CPP experiment, configuration CROSS. Symbols: $\circ$ – true configuration; $\times$ – 2D configuration derived by MDS from the pooled subject data (six subjects), then Procrustes-transformed to fit the true one. Lines connect corresponding points. Stress is 0.14.
The coefficient of congruence between the MDS-derived configuration and the true one was 0.99 (expected random value, estimated by bootstrap from the data: 0.86 ± 0.03, mean and std dev; 50 permutations of the point order were used in the bootstrap computation). Procrustes distance: 0.33 (bootstrap-estimated expected random value: 1.59 ± 0.07).

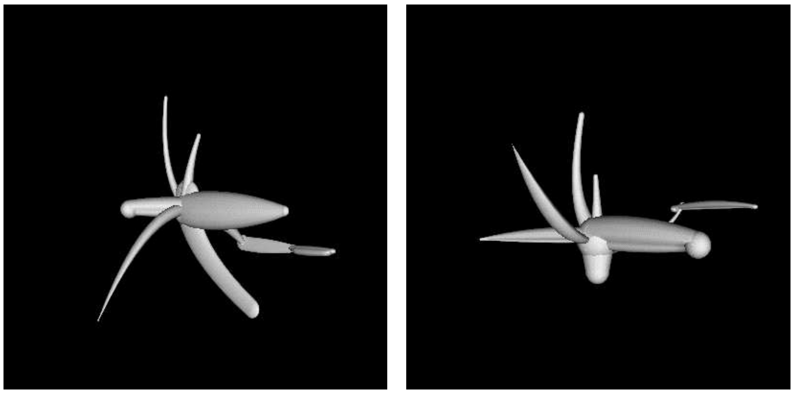


Figure 10: Two scrambled shapes derived from animal-like objects (see section 3.3).

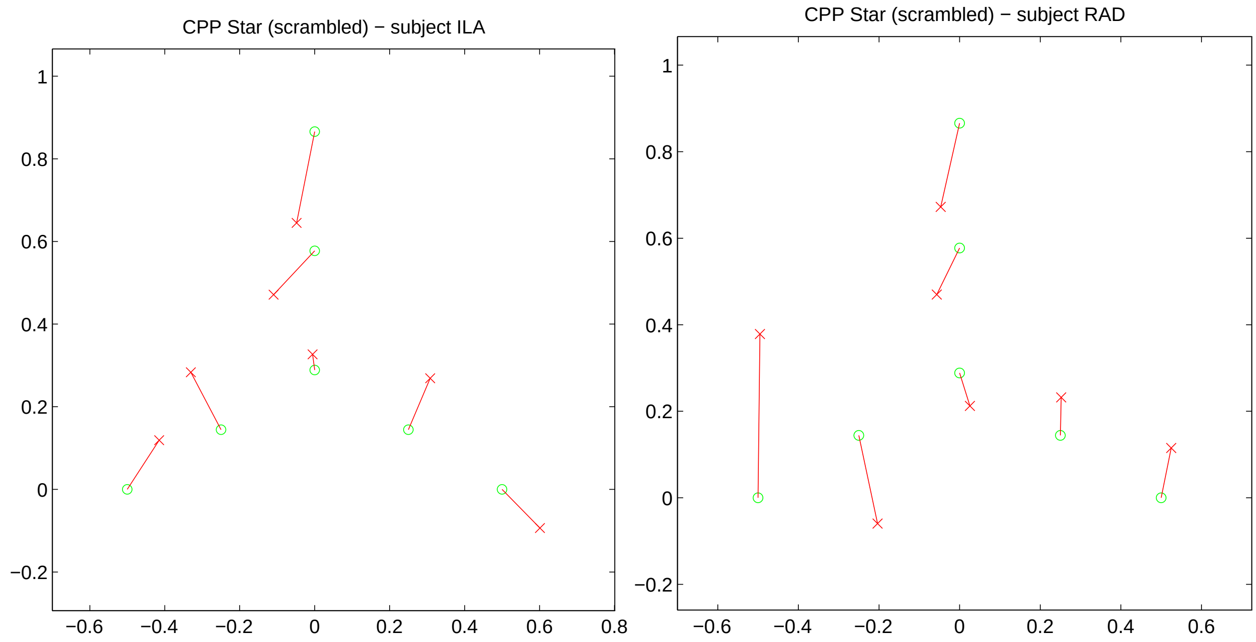


Figure 11: The CPP experiment, configuration STAR, scrambled. Symbols: $\circ$ – true configuration; $\times$ – 2D configuration derived by MDS from the subject data, then Procrustes-transformed to fit the true one. Lines connect corresponding points.

Left: the MDS configuration for subject *ILA*. Stress is 0. The coefficient of congruence between the MDS-derived configuration and the true one was 0.96 (expected random value, estimated by bootstrap from the data: 0.86 ± 0.04, mean and std dev; 50 permutations of the point order were used in the bootstrap computation). Procrustes distance: 0.35 (bootstrap-estimated expected random value: 0.77 ± 0.07).

Right: the MDS configuration for subject *RAD*. Stress is 0.1. The coefficient of congruence between the MDS-derived configuration and the true one was 0.97 (expected random value 0.85 ± 0.05, mean and std dev; 50 permutations of the point order were used in the bootstrap computation). Procrustes distance: 0.51 (bootstrap-estimated expected random value: 0.77 ± 0.07). Note that scrambled shapes yield smaller difference between real and bootstrap-estimated control values of the Procrustes distance, relative to the difference obtained with intact objects.

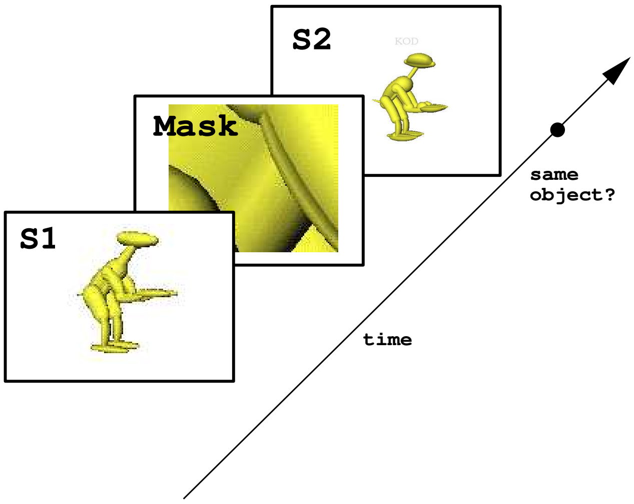


Figure 12: The time course of a trial in the DMTS experiment. The subject had to decide whether or not the images S_1 and S_2 belong to the same object. Image S_1 was shown for 300 msec, the mask for 750 msec, and image S_2 for 150 msec. Subjects were instructed to respond as quickly and as accurately as possible.

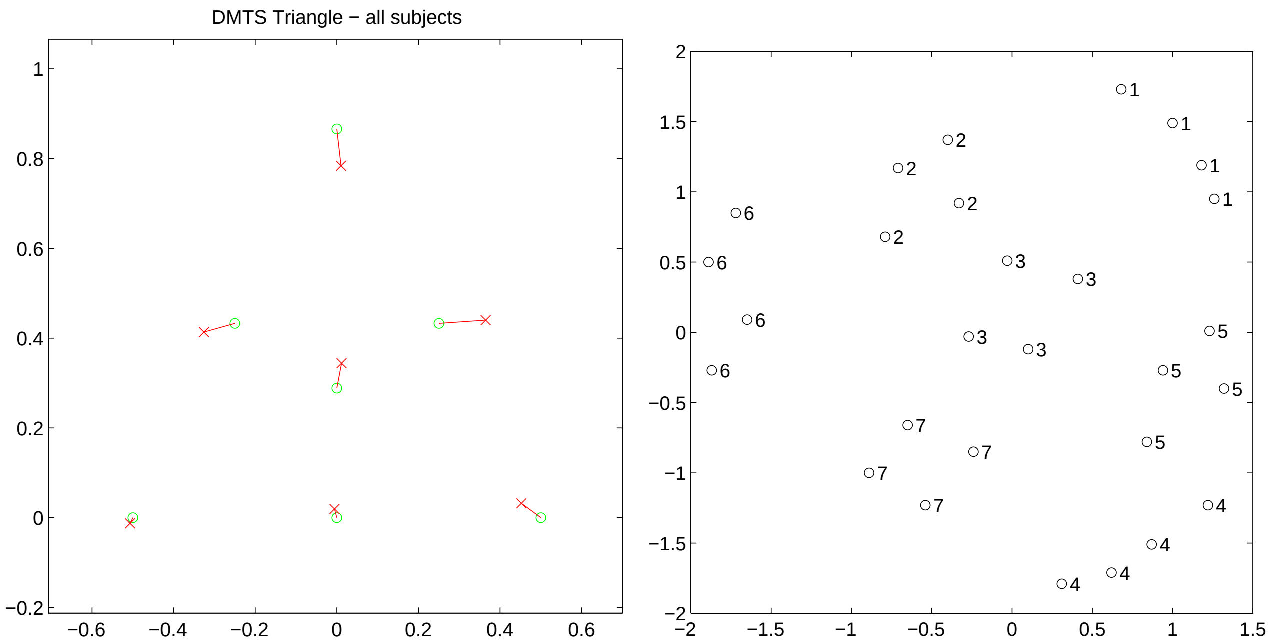


Figure 13: The DMTS experiment, configuration TRIANGLE. All eight subjects. *Left:* the 2D MDS **object** solution (stress 0.16). Symbols: $\circ$ – true configuration; $\times$ – 2D configuration derived by MDS from the pooled subject data, then Procrustes-transformed to fit the true one. Lines connect corresponding points.
Coefficient of congruence: 0.99 (expected random value: 0.87 ± 0.04); Procrustes distance: 0.15 (expected random value: 0.74 ± 0.08). *Right:* the 2D MDS **view** solution (stress 0.29). All four views of object k are labeled by the index k, where $k = 1 \ldots 7$. The object labels appear in Figure 2, under TRIANGLE.

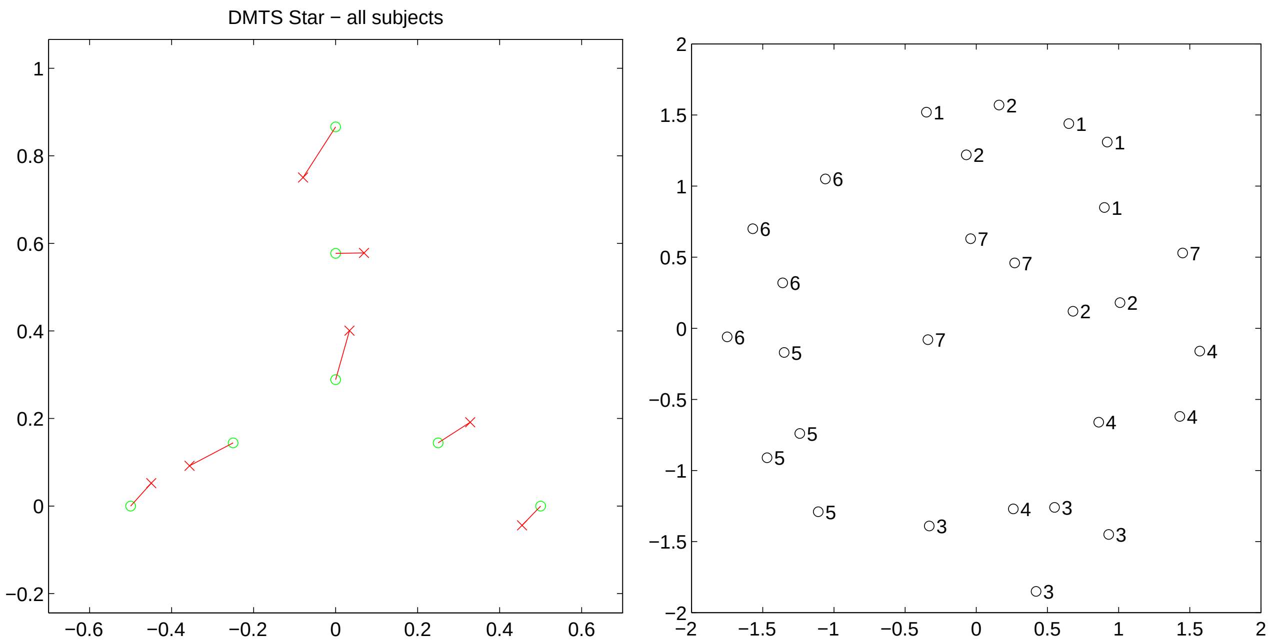


Figure 14: The DMTS experiment, configuration STAR. All seven subjects. *Left:* the 2D MDS **object** solution (stress 0.14). Symbols: $\circ$ – true configuration; $\times$ – 2D configuration derived by MDS from the pooled subject data, then Procrustes-transformed to fit the true one. Lines connect corresponding points.
Coefficient of congruence: 0.98 (expected random value: 0.86 ± 0.04); Procrustes distance: 0.25 (expected random value: 0.78 ± 0.07). *Right:* the 2D MDS **view** solution (stress 0.33). All four views of object k are labeled by the index k, where $k = 1 \dots 7$. The object labels appear in Figure 2, under STAR.

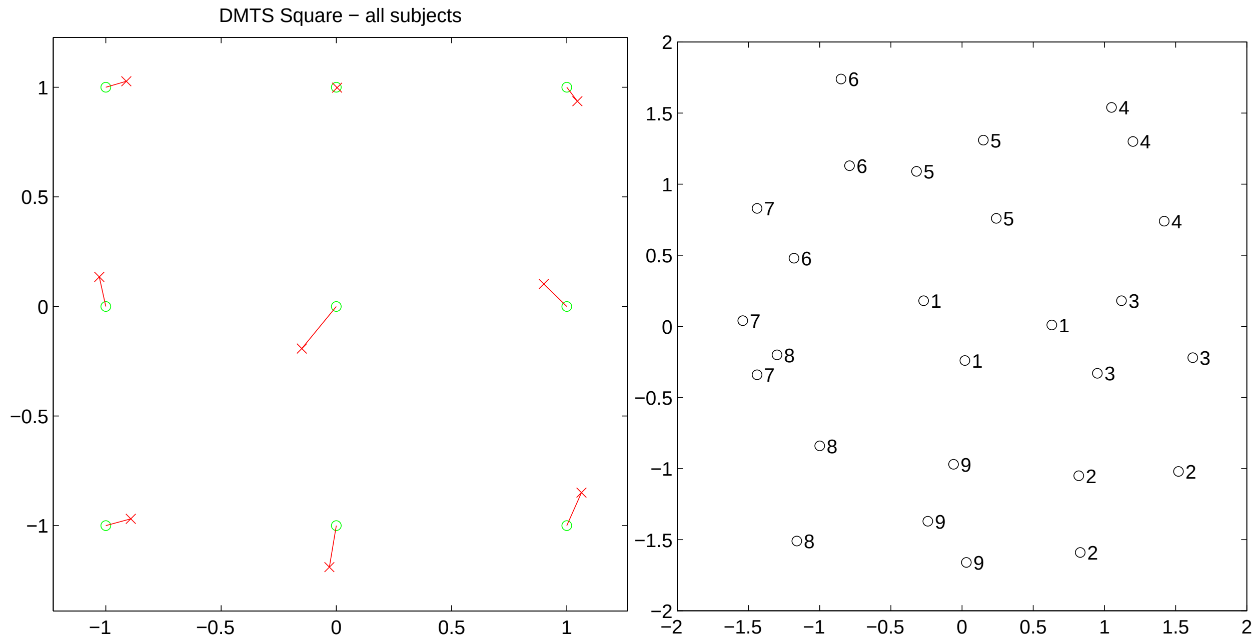


Figure 15: The DMTS experiment, configuration SQUARE. All nine subjects. *Left:* the 2D MDS **object** solution (stress 0.19). Symbols: $\circ$ – true configuration; $\times$ – 2D configuration derived by MDS from the pooled subject data, then Procrustes-transformed to fit the true one. Lines connect corresponding points.
Coefficient of congruence: 0.99 (expected random value: 0.89 ± 0.02); Procrustes distance: 0.38 (expected random value: 2.43 ± 0.17). *Right:* the 2D MDS **view** solution (stress 0.29). All three views of object k are labeled by the index k, where $k = 1 \ldots 9$. The object labels appear in Figure 2, under SQUARE.

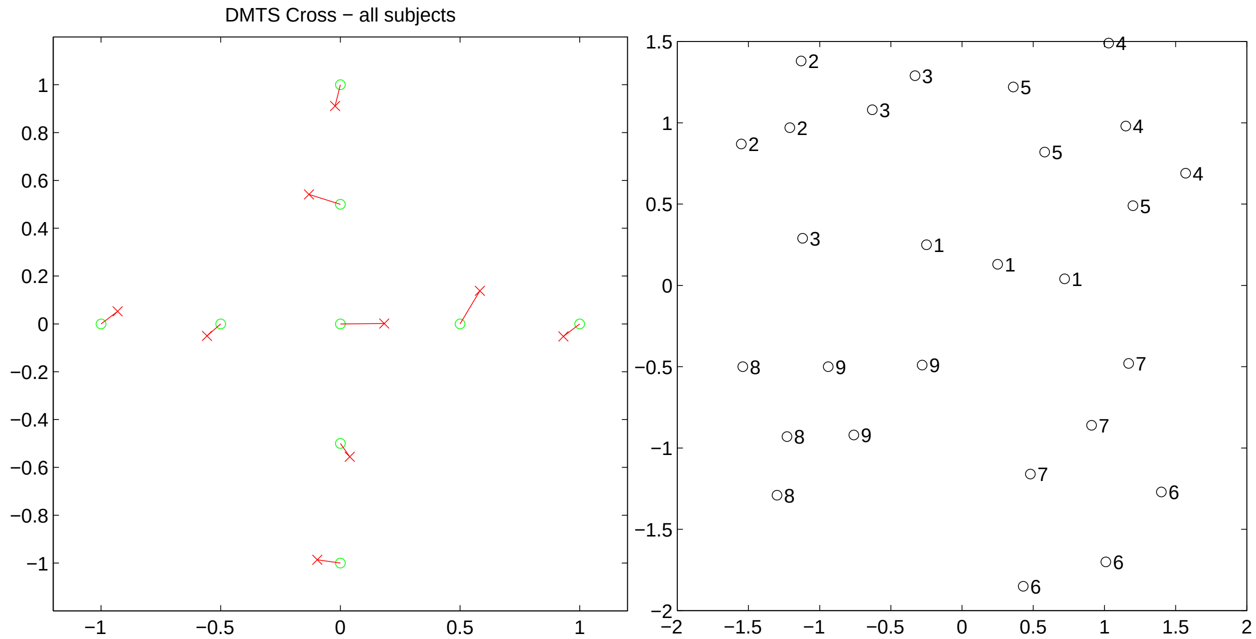


Figure 16: The DMTS experiment, configuration CROSS. All seven subjects. *Left:* the 2D MDS **object** solution (stress 0.18). Symbols: $\circ$ – true configuration; $\times$ – 2D configuration derived by MDS from the pooled subject data, then Procrustes-transformed to fit the true one. Lines connect corresponding points.
Coefficient of congruence: 0.99 (expected random value: 0.87 ± 0.04); Procrustes distance: 0.29 (expected random value: 1.56 ± 0.13). *Right:* the 2D MDS **view** solution (stress 0.29). All three views of object k are labeled by the index k, where $k = 1 \ldots 9$. The object labels appear in Figure 2, under CROSS.

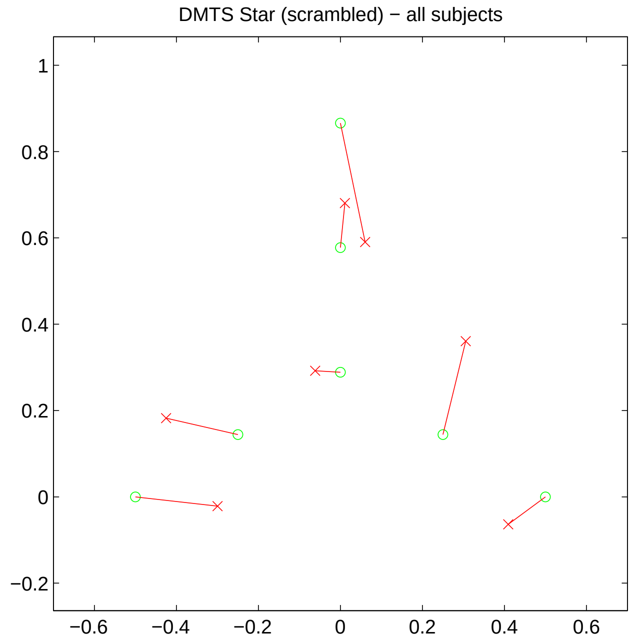


Figure 17: The DMTS experiment, configuration STAR, scrambled. All five subjects: the 2D MDS **object** solution (stress 0.13). Symbols: $\circ$ – true configuration; $\times$ – 2D configuration derived by MDS from the pooled subject data, then Procrustes-transformed to fit the true one. Lines connect corresponding points. Coefficient of congruence: 0.96 (expected random value: 0.86 ± 0.04); Procrustes distance: 0.38 (expected random value: 0.75 ± 0.13). Note that the configuration is distorted, and that scrambled shapes yield smaller difference between real and bootstrap-estimated control values of the Procrustes distance, relative to the difference obtained with intact objects.

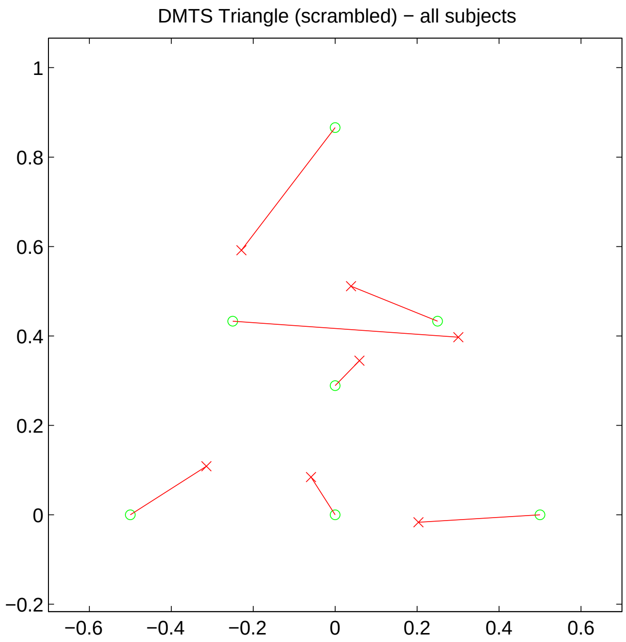


Figure 18: The DMTS experiment, configuration TRIANGLE, scrambled. All five subjects: the 2D MDS **object** solution (stress 0.21). Symbols: ∘ – true configuration; × – 2D configuration derived by MDS from the pooled subject data, then Procrustes-transformed to fit the true one. Lines connect corresponding points. Coefficient of congruence: 0.96 (expected random value: 0.89 ± 0.03); Procrustes distance: 0.73 (expected random value: 0.76 ± 0.13). Note that the configuration is distorted, and that scrambled shapes yield an insignificant difference between real and bootstrap-estimated control values of the Procrustes distance, relative to the difference obtained with intact objects.

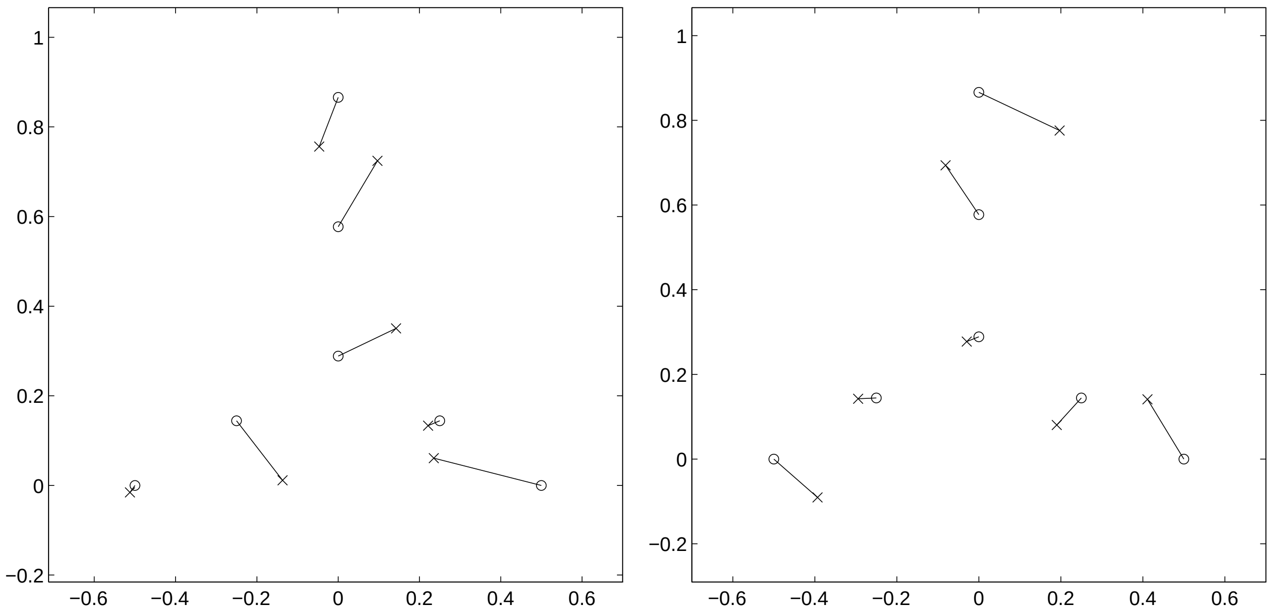


Figure 19: The LTM experiment, configuration STAR. *Left:* Subject *LIR* (stress 0). Coefficient of congruence: 0.96 (expected random value: 0.85 ± 0.04); Procrustes distance: 0.34 (expected random value: 0.78 ± 0.07). *Right:* All subjects (stress 0.12). Coefficient of congruence: 0.99 (expected random value: 0.86 ± 0.04); Procrustes distance: 0.34 (expected random value: 0.78 ± 0.07).

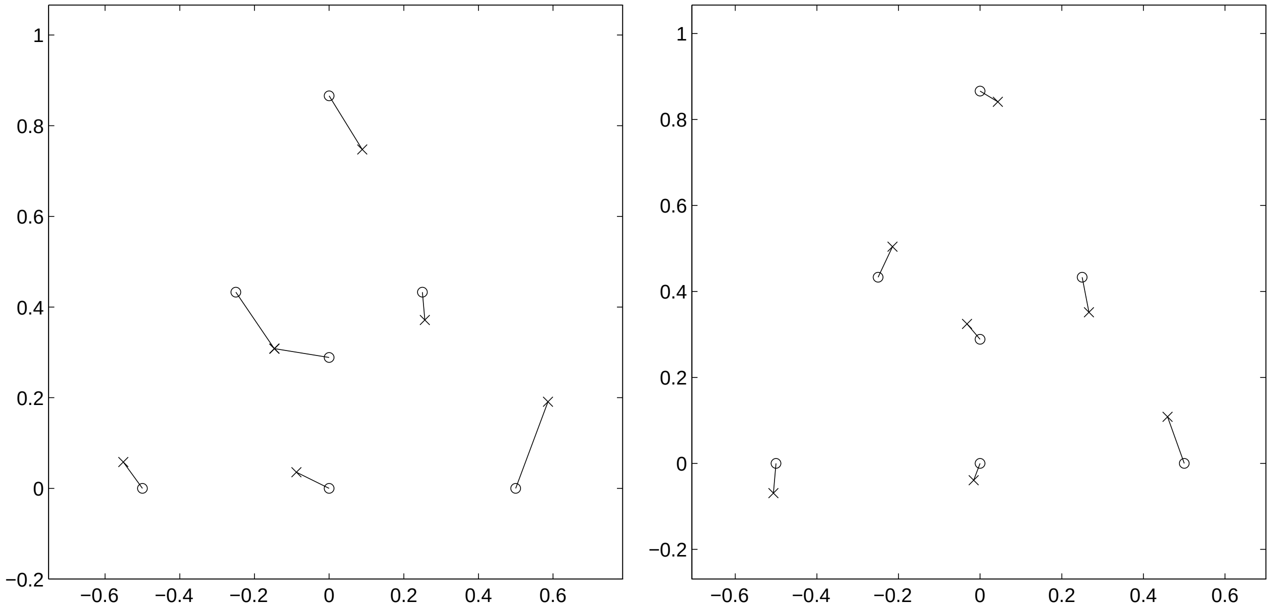


Figure 20: The LTM experiment, configuration TRIANGLE. *Left:* Subject *MOR* (stress 0). Coefficient of congruence: 0.99 (expected random value: 0.87 ± 0.03); Procrustes distance: 0.29 (expected random value: 0.78 ± 0.05). *Right:* All subjects (stress 0.12). Coefficient of congruence: 0.99 (expected random value: 0.87 ± 0.04); Procrustes distance: 0.18 (expected random value: 0.78 ± 0.05).

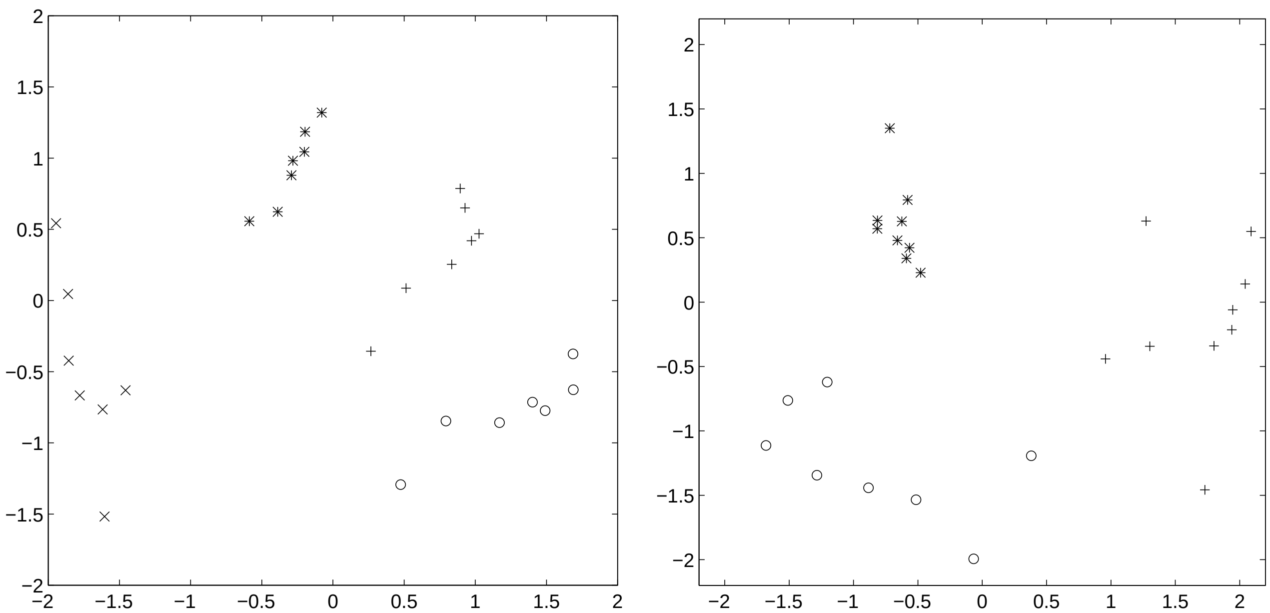


Figure 21: Simulated DMTS experiment; view solution derived from similarities among receptive field activities (the first model) evoked by the stimuli. *Left:* The STAR configuration. Note the four clusters, one for each orientation (marked by the different symbols). Each cluster contains seven views, one for each object in the same orientation. The object solution was essentially random: coefficient of congruence 0.89 (expected random value: 0.86 ± 0.04); Procrustes distance 0.64 (expected random value: 0.78 ± 0.07). *Right:* The CROSS configuration. There are three clusters, corresponding to the three different object orientations tested in this experiment. Here too the object solution was essentially random: coefficient of congruence 0.86 (expected random value: 0.86 ± 0.03); Procrustes distance 3.19 (expected random value: 3.09 ± 0.23).

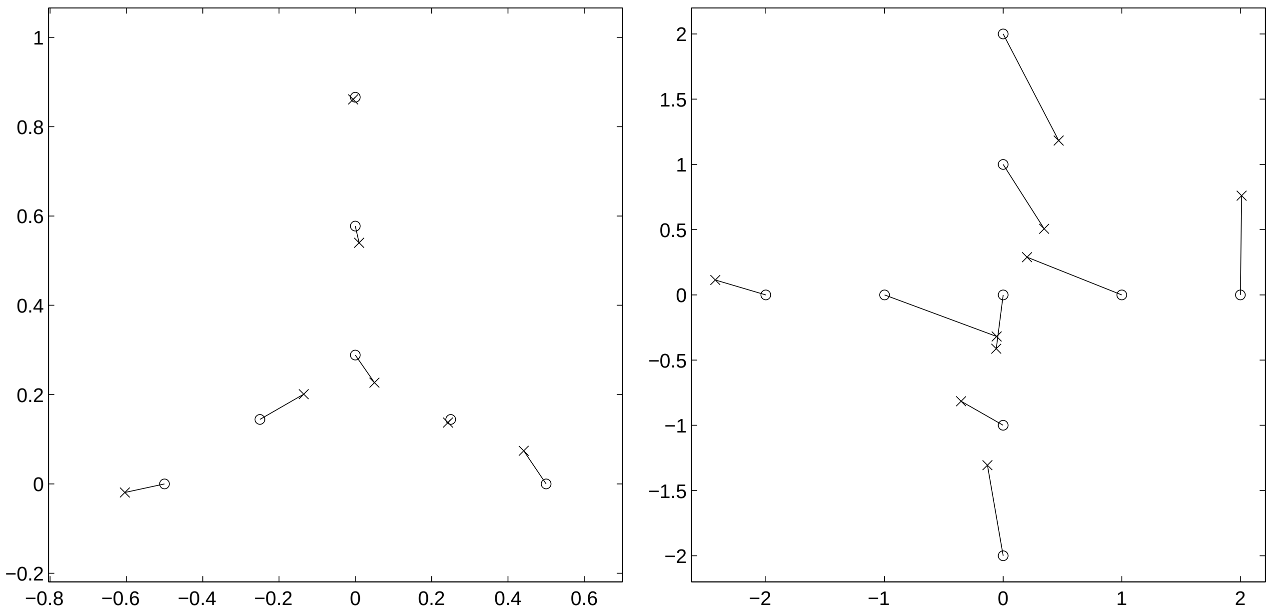


Figure 22: Simulated DMTS experiment; **object** solution derived from similarities among RBF network activities (the second model) evoked by the stimuli. Unlike Figure 21, here the clustering in the **view** solution (not shown) was by object identity, not by orientation. *Left:* The STAR configuration; three RBF networks were employed, one for each extreme points spanning the star configuration. The solution was highly significant: coefficient of congruence 0.99 (expected random value: 0.86 ± 0.04); Procrustes distance 0.17 (expected random value: 0.76 ± 0.09). *Right:* The CROSS configuration; the model contained four RBF networks, one for each extremal vertex. The solution was significant, although less so than for the STAR: coefficient of congruence 0.94 (expected random value: 0.86 ± 0.03); Procrustes distance 1.88 (expected random value: 3.13 ± 0.15).