Explorations of shape space

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

January 27, 1995

Abstract

Using a small number of prototypical reference objects to span the internal shape representation space has been suggested as a general approach to the problem of object representation in vision (Edelman, 1995c). We have investigated the ability of human subjects to form the low-dimensional metric shape representation space predicted by this approach. In each of a series of experiments, which involved pairwise similarity judgment, and delayed match to sample, subjects were confronted with several classes of computer-rendered 3D animal-like shapes, arranged in a complex pattern in a common high-dimensional parameter space. We combined response time and error rate data into a measure of view similarity, and submitted the resulting proximity matrix to nonmetric multidimensional scaling (MDS). In the two-dimensional MDS solution, views of the same shape were invariably clustered together, and, in each experiment, the relative geometrical arrangement of the view clusters of the different objects reflected the true low-dimensional structure in parameter space (star, triangle, square, line) that defined the relationships between the stimuli classes. These findings are now used used to guide the development of a detailed computational theory of shape vision based on similarity to prototypes.

1 Introduction

1.1 Background

The elucidation of the nature of representations used by the visual system is a long-standing central goal of visual psychophysics. The greatest challenge in recognizing 3D objects is the variability of their appearance, caused by factors such as changing illumination, viewpoint, and, in the case of articulated or flexible shapes, by the changes in the very geometry of the shape itself. This variability goes largely unnoticed in everyday life, because of the great flexibility of the brain in dealing with the naturally occurring visual stimuli.

Psychophysical studies conducted in the past few years led to the characterization of certain basic limitations of the visual system in generalizing shape-based recognition to novel conditions (Rock and DiVita, 1987; Tarr and Pinker, 1989; Edelman and Bülthoff, 1992; Humphrey and Khan, 1992; Edelman, 1995a; Cutzu and Edelman, 1994); see (Jolicoeur and Humphrey, 1994) for an extensive review and a discussion. Specifically, it was found that the recognition of novel views of objects tends to be slower and more prone to errors than the recognition of highly familiar views, that the effect of viewpoint increases when the objects to be distinguished are more similar to each other, and that it persists even when full 3D shape information is available to the subject through, e.g., binocular stereo cues.

Attempts to model the human recognition performance led to a certain progress in understanding the mathematics and the neurobiological mechanisms behind these psychophysical data. In particular, it is becoming increasingly clear that human performance in recognizing objects under changes of viewpoint can be accounted for to a significant degree by a theory based on the notion of interpolation among view-specific representations (Poggio and Edelman, 1990; Edelman and Weinshall, 1991; Bülthoff and Edelman, 1992; Edelman, 1995a). Detailed computational modeling showed that this approach to representation can be implemented by putting together simple building blocks such as the nonlinear graded overlapping receptive fields (RFs) found in biological visual systems (Edelman, 1995a; Edelman, 1995b).

1.2 View prototypes and the representation of 3D objects for recognition

The computational basis of representation by multiple views is twofold. On the one hand, the space of all views of a 3D object can be characterized mathematically as a smooth low-dimensional manifold in the space of all views of all possible objects (Ullman and Basri, 1991). Thus, a "black box" capable of recognizing an object can be built by learning to interpolate the view-space manifold from examples (Poggio and Edelman, 1990; Poggio, 1990). On the other hand, interpolation from examples can be considered as taking a linear combination of responses of "units" each encoding a particular example, that is, a prototypical input/output relationship (Poggio and Girosi, 1990). Furthermore, the response profiles of such units need not be more complicated than the familiar Gaussian receptive fields, commonly found in biological information processing systems. Thus, the prototypical-view approach is appealing both mathematically and biologically.

1.3 Shape prototypes: from recognition to classification

A more challenging problem arises if one wishes to represent unfamiliar objects, for which particular view prototypes are not likely to exist in the system. Because the view-space theorems do not apply in this case, no principled predictions regarding the structure of novel object view manifold can be made. Nevertheless, representations of a sufficient variety of objects, or prototypes, taken together, can serve as a basis for a provably veridical distributed (ensemble) representation of a completely unfamiliar shape, as noted in (Edelman, 1993). This approach to representation relies on the concept of *shape space* (superseding that of view space), and is centered around the observation that veridical proximal representation of shape contrasts allows the reconstruction of true distal shape contrasts, which, in turn, can serve as a basis for solving problems such as shape classification (see Figure 1).

The computational principle behind this idea is the same one that underlies nonmetric multidimensional scaling (Shepard, 1980): for a set of points (in the present case, points in the shape space), the ranks of interpoint distances constrain the point configuration sufficiently to allow its retrieval from the rank data. If the measured (proximal) ranks approximate well the true (distal) ranks of dissimilarities between object shapes, the location of each object in the true shape space can be effectively estimated. Some indications that this approach to representation is computationally viable are already available (Weiss and Edelman, 1995; Duvdevani-Bar and Edelman, 1995).

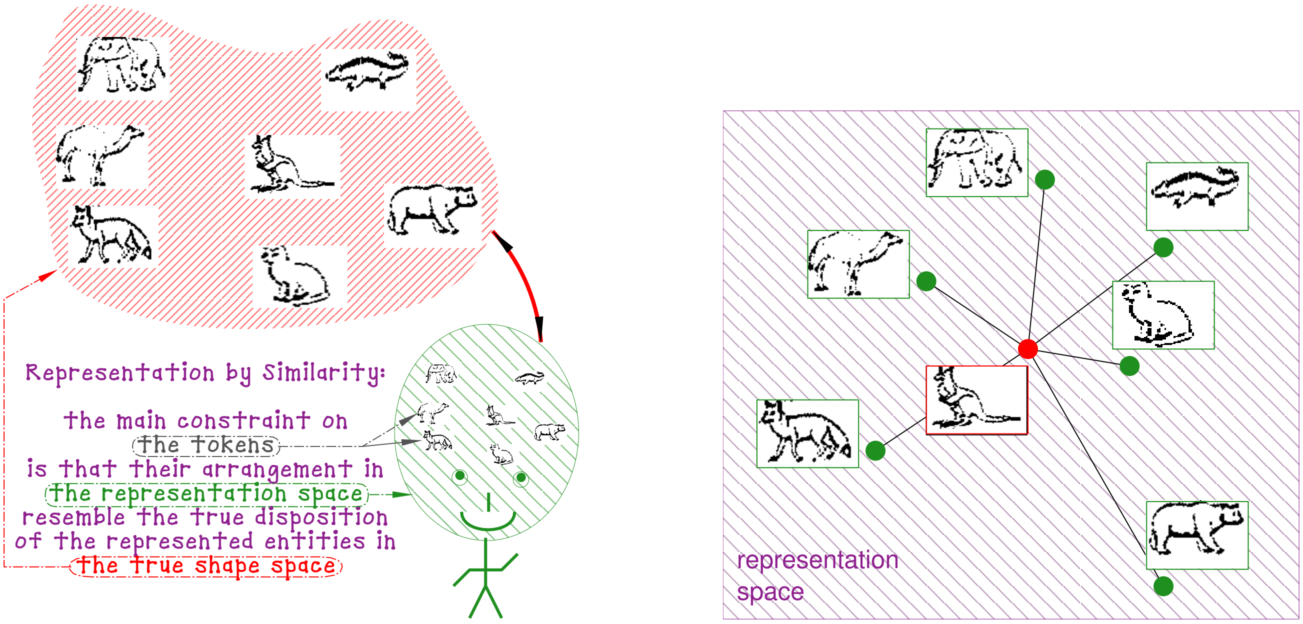


Figure 1: *Left:* the proposed scheme calls for the representation of similarities among objects, instead of the detailed geometry of each object. The advantages of this approach over that of representation by geometric reconstruction are discussed in (Edelman, 1994; Edelman, 1993). *Right:* this approach is equivalent to representing a given shape by taking its bearings in the shape space with respect to a number of reference or prototypical shapes.

2 General approach

The hypothesis of representation by similarity implies that the similarity relationships among objects presented to a visual system must be reflected in the distances among tokens standing for those objects in the system's internal representation space (which is assumed to be a metric[1] space). A natural tool for testing this hypothesis is multidimensional scaling, which has been employed in the past in the exploration of the representation space for simple stimuli such as colors (see (Shepard, 1980) for a review), as well as a class of closed 2D curves obtained by Fourier synthesis (Shepard and Cermak, 1973). This approach has been extended recently to 3D shapes in a study that employed two-alternative forced-choice discrimination between two classes of animal-like objects, jointly parameterized by several tens of parameters (Edelman, 1995b). The results of this study — the near-perfect recovery from the subject data of the true parametric (shape-space) similarity pattern inherent in the stimulus configuration — support the hypothesis of representation by similarity.

The shape-space pattern used in (Edelman, 1995b) consisted, basically, of two clusters of points, corresponding to the two classes of objects in each of the discrimination experiments reported there. In the present work, we explore the geometry of the psychological space structures induced by perceived similarity relations between a considerably more diverse classes of 3D animal-like shapes, arranged in much more complex patterns in the shape space. As before, the objective similarity relations between the test objects — that is, the Euclidean distances in the shape space — were under experimental control and could be varied at will. The psychophysical part of the present work explores the relationship

[1] "Metric" here is taken to mean merely "equipped with a norm" (an expression which is somewhat awkward to use in a psychological setting), and not necessarily "obeying the metric axioms."

between the internal representation of the object set recovered from subject data by multidimensional scaling (MDS), and the true configuration in the shape space, controlled by the experimenter. A computational model of the psychophysical results is reported elsewhere (Duvdevani-Bar et al., 1995).

2.1 Method

The experiments described here were based on the following basic design:

1. Measure perceived similarities for all pairs of input items (either static views or rotating objects), and construct a proximity table.

2. Use MDS to recover the psychological space structure of the visual representation set from the proximity table.

In the second step, MDS is run both on the individual data of each subject, and on the pooled data obtained by concatenating the proximity tables of all subjects (Carroll and Chang, 1970). We note that our approach to measuring perceived similarity is superior to the subjective rating methods of the classical MDS experiments. We used two methods: pairs of pairs comparison, described in section 3.2, and a variant of the delayed match to sample paradigm, described in section 4.2.

2.2 The stimuli

Our stimuli were complex 3D animal-like shapes, having, albeit in a rather schematic form, all the major body parts of real mammals: head, neck, body, jointed limbs, and a tail. The stimuli were created by a program based on the SGI Inventor 3D graphics toolkit, and were rendered for presentation to subjects on a Silicon Graphics Indigo^2/Extreme workstation. The rendering procedure used Gouraud shading, and a uniform-color matte surface material. In the experiments based on the Delayed Match to Sample (DMTS) paradigm the objects were viewed stereoscopically. Two examples of typical stimuli are shown in Figure 2.

All animals had the same body parts, modeled by generalized cylinders. The shape of an animal was controlled by 70 parameters, which defined the shape of the individual parts, their orientation and relative positions within the animal. Each generalized cylinder was encoded by several parameters: axis length, the radii at the ends of the cylinder, the degree of axis curvature in two perpendicular planes (the axis of the cylinder being either parabolic or exponential, depending on the body part), 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.

2.3 The shape-space configurations

An object encoded by 70 values of its shape and pose parameters can be represented as a point in a 70-dimensional Euclidean shape space. A set of animals is then represented by a cloud of points in the shape space. Euclidean distance between points corresponding to two shapes in this space is an objective measure of (dis)similarity between these shapes (metrics other than Euclidean can, of course, be used, but we saw no reason to resort to alternative metrics).

Given this conception of shape space, one can used elementary methods of analytic geometry to generate sets of objects endowed with well-defined, complex similarity structure, induced by their geometrical arrangement in shape space. In the experiments reported here, the sets of objects were

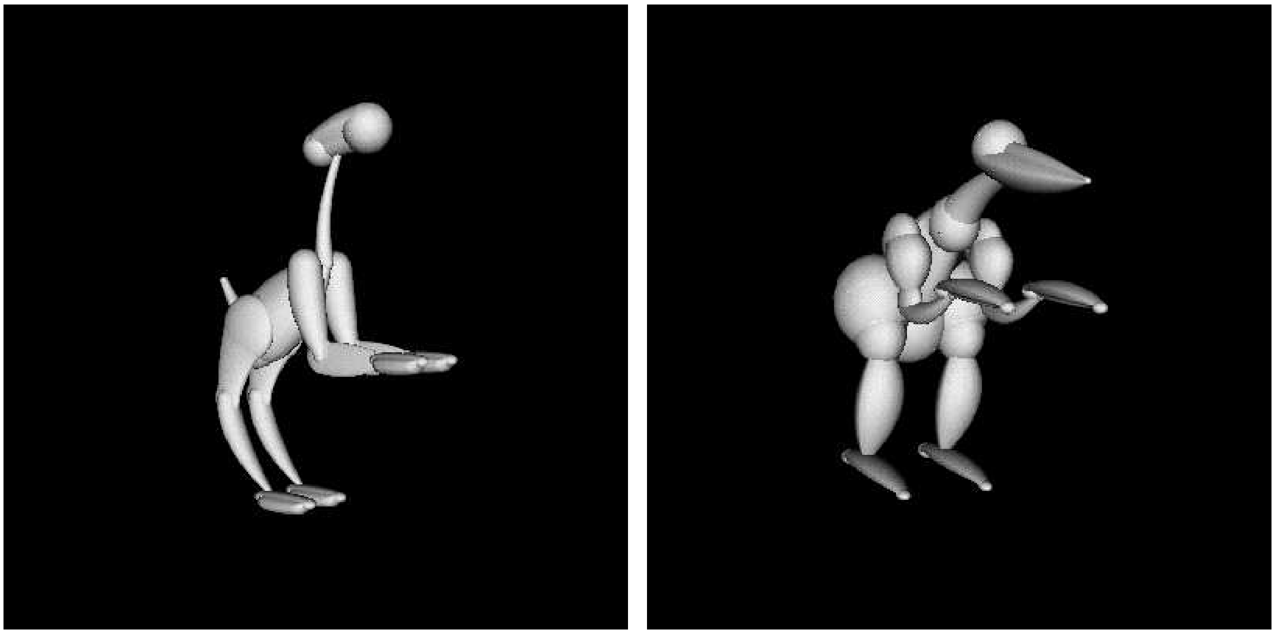


Figure 2: Two animal shapes

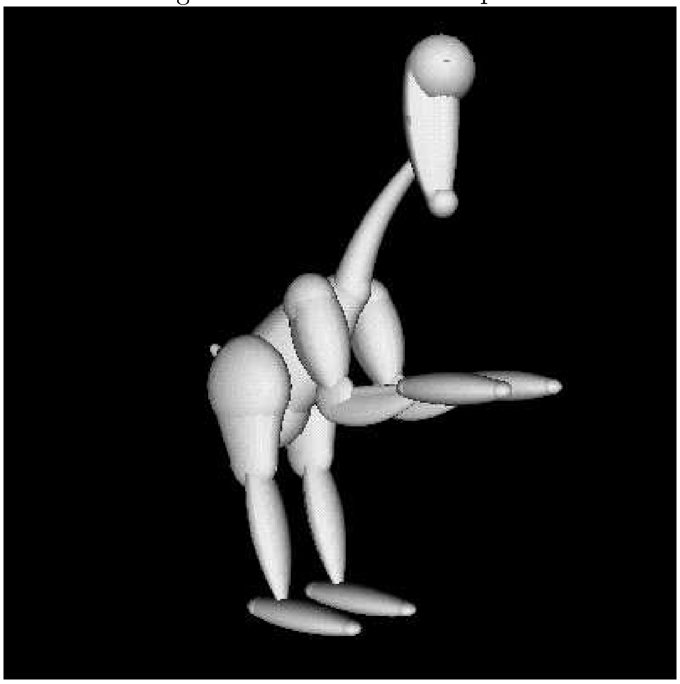


Figure 3: A hybrid animal

clouds of points arranged in regular configurations on a line or in a 2D plane in the 70-dimensional shape space. The two basic formulas used to generate the configurations are:

- Given two animals $\mathbf{A}$ and $\mathbf{B}$, any animal $\mathbf{M}$ on the line between $\mathbf{A}$ and $\mathbf{B}$ can be represented as $\mathbf{M} = \alpha\mathbf{A} + (1-\alpha)\mathbf{B}$, with $0 \le \alpha \le 1$. The parameter α controls the relative contributions of the "parents" $\mathbf{A}$ and $\mathbf{B}$ in the appearance of the hybrid $\mathbf{M}$. For example, the animal in Figure 3 is at the midpoint (in shape space) between the two animals shown in Figure 2.

- Given N animals, $\mathbf{A}_1, \ldots, \mathbf{A}_N$, the centroid animal $\mathbf{C}$ can be generated as $\mathbf{C} = (1/N)\sum \mathbf{A}_i$.

In each series of experiments, the test objects were arranged in a distinct configuration in parameter space: line, triangle, star, square, cross (see Figure 4; the labels in the figures correspond to the labels of the objects in the experimental plots, and are used consistently throughout the paper). In the LINE experiments, there were seven objects spaced at equal intervals and arranged on a line in the parameter space. In the TRIANGLE experiments, seven objects were positioned at the vertices, barycenter and at the midpoints of the edges of an equilateral triangle lying in a plane 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 lying in a plane in 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.

It is important to stress the complexity of the relation between the shape-space representation (i.e., the vector of parameter values) and the screen appearance (the array of pixel values) of the corresponding animal shape from some vantage point. The parameter values determine the geometry of the closed surface that models the animal via nonlinear (for example, trigonometric) functions, by specifying the 3D position and orientation of all the surface elements of the animal's body. The animal is then rotated in 3D and projected onto the screen by the rendering software, which creates an image by computing the appropriate shading for the surface. Thus, three shapes that are collinear in the shape (parameter) space certainly need not give rise to views that are also collinear in the image (pixel) space. For example, given two images (arrays of pixels) $\mathbf{P}$ and $\mathbf{Q}$, that constitute the profile views of the two animals of Figure 2, the image situated at the midpoint on the line $\bar{\mathbf{PQ}}$ in pixel space will be the superposition of the two *images* taken with equal weights $\mathbf{M} = 0.5 \cdot \mathbf{P} + 0.5 \cdot \mathbf{Q}$, and not the profile view of the hybrid *shape* of Figure 3.

3 Exploring the representation of similarity relationships among several objects

3.1 Motivation

We first addressed the question whether a complex two-dimensional arrangement of several shapes in the (70-dimensional) parameter space can be recovered by MDS from the table of inter-object similarity ranks collected from the subjects. The classical method for obtaining similarity data for processing with MDS requires that the subject give a numerical estimate of similarity when presented with a pair of stimuli. The major drawback of this method is that many subjects do not feel comfortable with the task of providing an explicit estimate of similarity. We avoided this problem by using a forced choice experimental paradigm, described below.

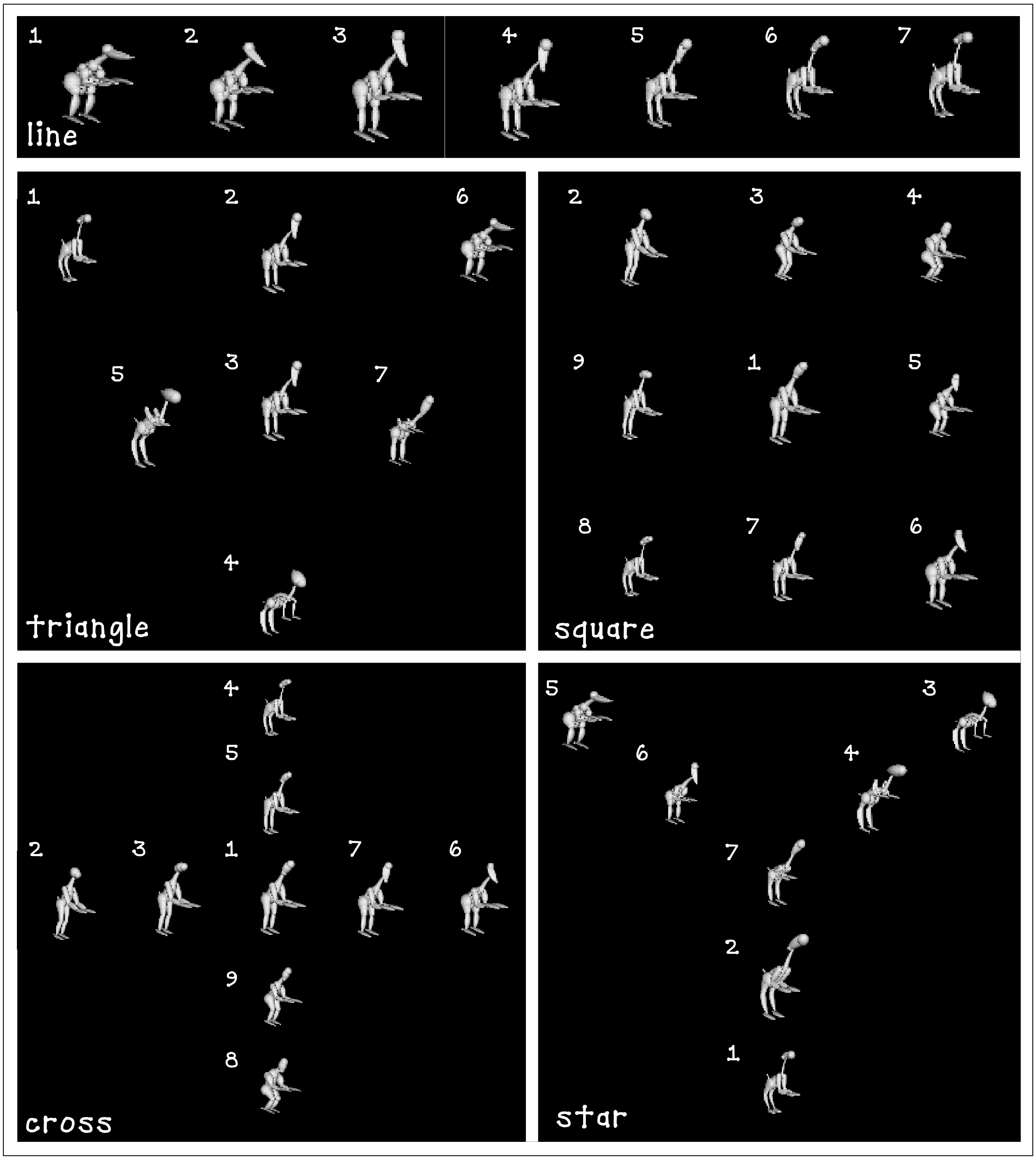


Figure 4: The five parameter-space configurations employed in the generation of the stimuli for the experiments. The numbers near each shape are used as labels to annotate the MDS configurations throughout the report.

3.2 Experimental method: CPP

Each experiment involved a set of N objects, for which we had to measure all the pairwise similarities S_{ij}, $1 \leq i \leq N$, $1 \leq j \leq N$, to be placed in a $N \times N$ proximity table. In each trial of the "compare pairs of pairs" or CPP experiments, 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. The objects were displayed in continuous rotation around their vertical axis. The subject had to select 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. After the subject made the choice by pressing the proper key on the computer keyboard, the next set of four shapes was shown. The choice of the subject was recorded together with the corresponding object names. This method presents a technical problem: 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.

3.3 CPP results

Five series of CPP experiments were conducted, corresponding to the different shape space configurations described in section 2.2. The CPP LINE, 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 experimental object configurations with the appropriate labels are displayed in Figure 4.

Following each experiment, the proximity tables were constructed from the rank data as described in section B.1, and were subjected to a verification of transitivity of ranking. The latter step was necessary because transitivity in the differential similarity judgements 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, the 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).

Because the results of the CPP experiments are of the form "objects (i, j) are more/less similar to each other than objects (p, q)", it is straightforward to check transitivity for triplets of object pairs. Specifically, if objects (i, j) are more similar than (p, q), and (p, q) are more similar than (v, w), then objects (i, j) should be 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 very low $(0 \ldots 4\%)$.

3.3.1 The CPP LINE experiment

Seven subjects were tested with seven objects spread out equidistantly on a line in the parameter space. In the following MDS plots, the labels 1 to 7 corresponding to the 7 objects in parameter space.

An individual example. The MDS plots for subject *Tatiana* are displayed in Figure 5. The 1D MDS solution reveals good recovery of the parameter space configuration, although the spacing is not equal. Adding a second dimension removes all the stress and produces an irregular horseshoe-like shape. The percentage of intransitive triplets was 1.4%.

The pooled data. The configurations displayed in Figure 6 were obtained by running MDS on the pooled data of all subjects. The recovery of the parameter space configuration is almost perfect, even

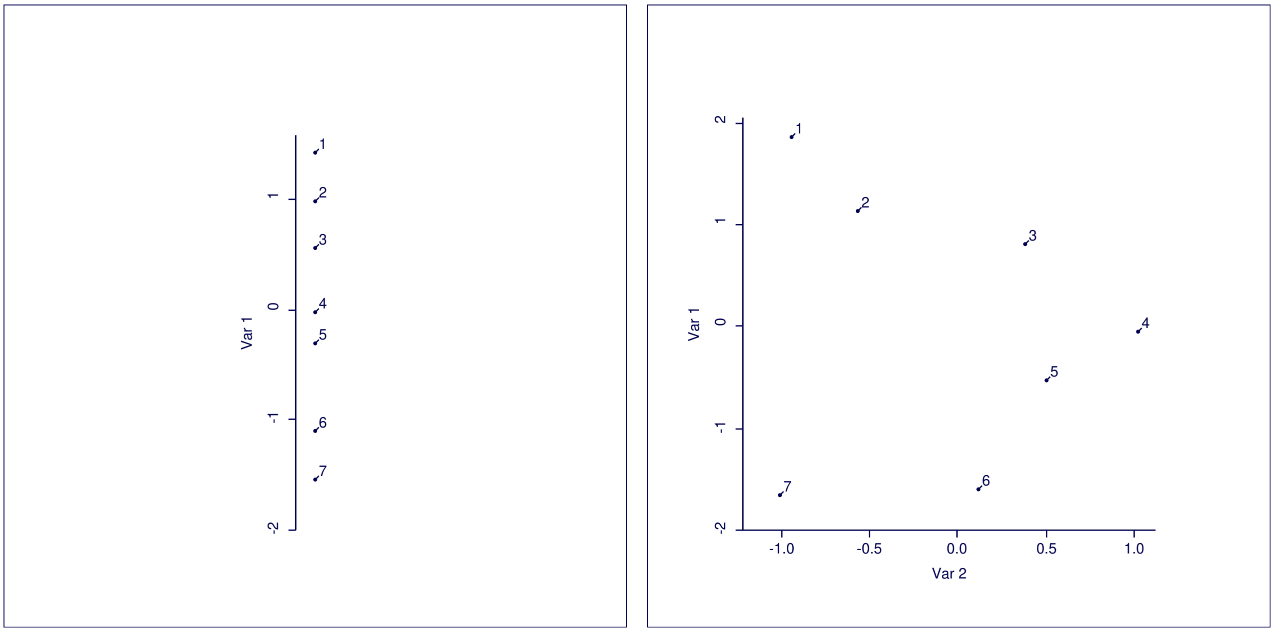


Figure 5: CPP LINE, subject Tatiana. *Left:* the 1D MDS solution (Badness-of-Fit 0.06, Distance Correlation 0.99). The values of Badness-of-Fit and of Distance Correlation here and below should be compared to the chance-level values for 7×7 matrices, stated in section A in the appendix. *Right:* the 2D MDS solution (Badness-of-Fit 0.00, Distance Correlation 1.00).

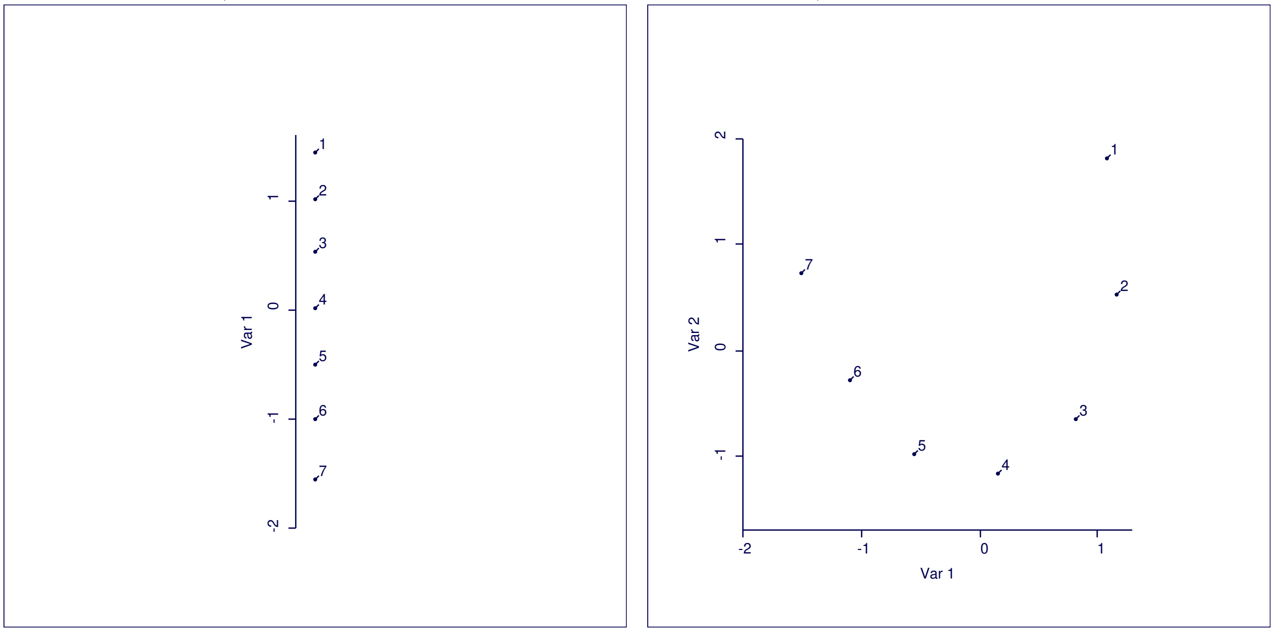


Figure 6: CPP LINE, all subjects. *Left:* the 1D MDS solution (Badness-of-Fit 0.07, Distance Correlation 0.99). *Right:* the 2D MDS solution (Badness-of-Fit 0.04, Distance Correlation 1.00).

though the individual configurations were derformed in various ways. The 2D solution demonstrates the well-known “horseshoe” artifact (Kruskal, 1977), resulting from the excess degree of freedom in fitting the data.

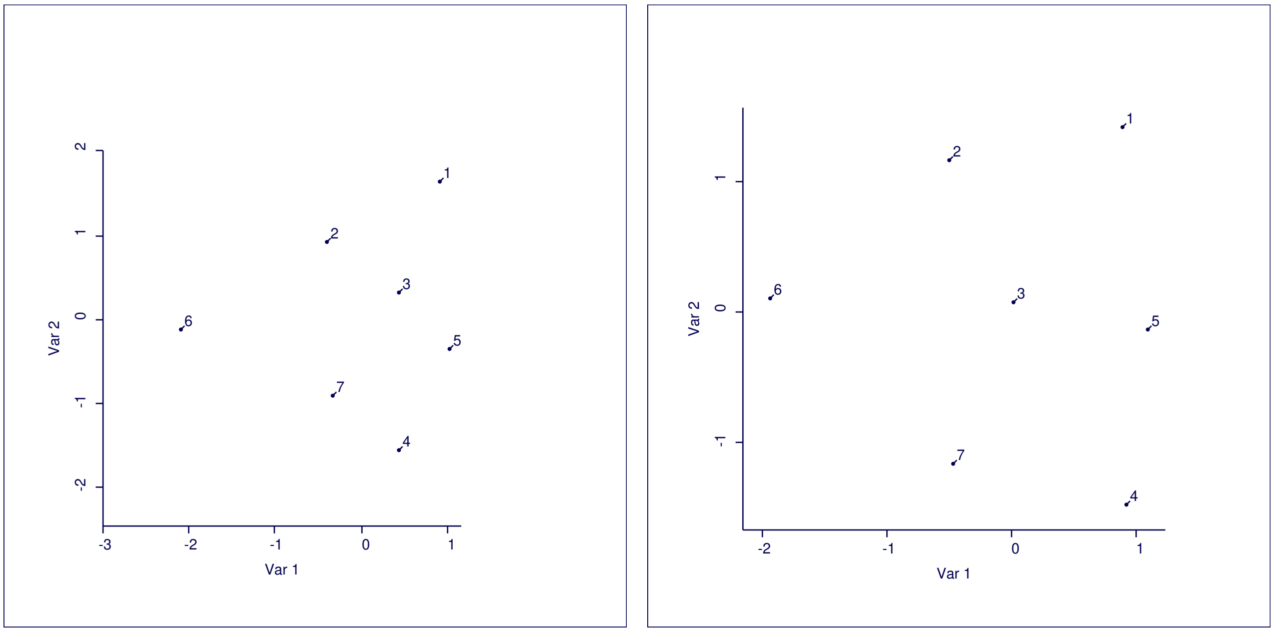


Figure 7: CPP TRIANGLE. *Left:* the 2D MDS solution for subject Rada.(Badness-of-Fit 0.07, Distance Correlation 0.99). *Right:* the 2D MDS solution for all subjects (Badness-of-Fit 0.20, Distance Correlation 0.80).

3.3.2 The CPP TRIANGLE experiments

Seven subjects were tested with seven objects arranged in a planar triangle-like configuration in the parameter space.

An individual example. The 2D MDS plot for subject *Rada* is displayed on the left in Figure 7. The solution reveals a perfect recovery of the parameter space configuration. The percentage of intransitive triplets was 1.4%.

The pooled data. The configuration displayed in Figure 7 was obtained by running MDS on the pooled data of all subjects. The recovery of the parameter space configuration is almost perfect, despite the various deformations that were present in the individual configurations.

3.3.3 The CPP STAR experiments

Six subjects were tested with seven objects arranged in a planar star-like configuration in the parameter space.

An individual example. The 2D MDS plot for subject *Lada* is displayed on the left in Figure 8. The centroid and object number 6 tend to be confused, and the distances between the other midpoints and centroid are overestimated (the midpoints are confused with the extremes). The percentage of intransitive triplets was 0%.

The pooled data. The plot is presented on the right in Figure 8 and was obtained by running MDS

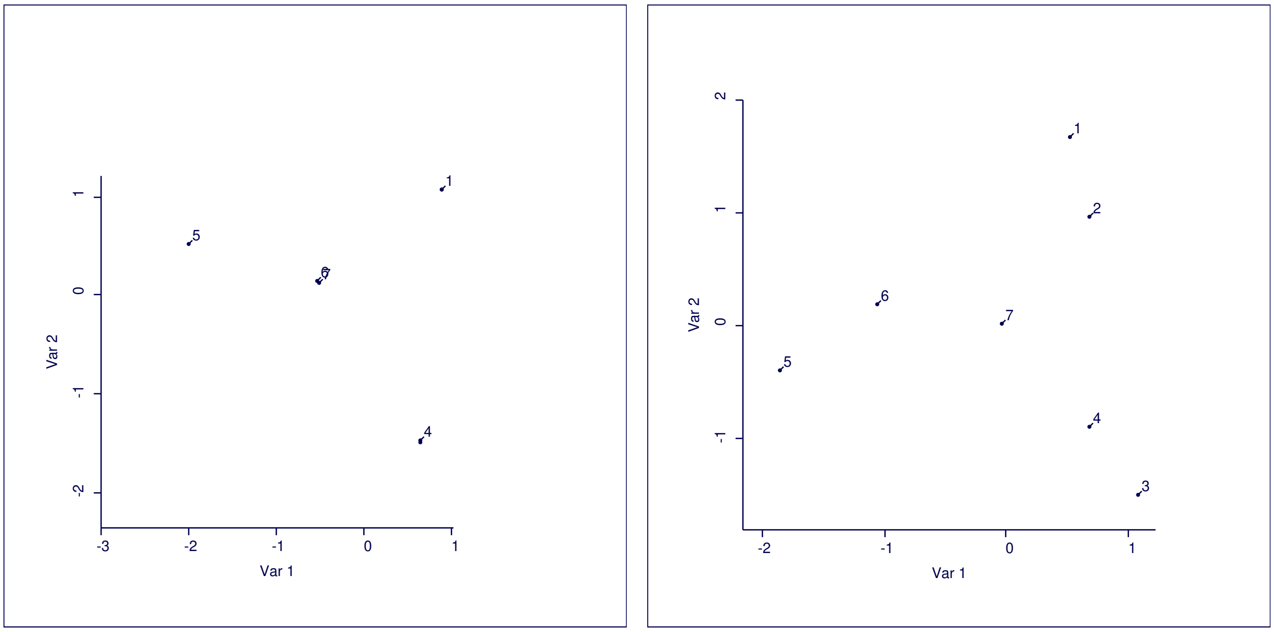


Figure 8: CPP Star. *Left:* the 2D MDS solution for subject Lada (Badness-of-Fit 0.0, Distance Correlation 1.0; labels 1 and 2, 6 and 7, 3 and 4 are overlapped). *Right:* the 2D MDS solution for all subjects (Badness-of-Fit 0.11, Distance Correlation 0.96).

on the proximity data of all the subjects. The recovered configuration is a systematically deformed version of the original configuration: both arms 1–2 and 5–6 are bent to the left.

3.3.4 The CPP Square experiments

Six subjects were tested with nine objects arranged in a rectangular configuration in the parameter space. In the following MDS plots, the labels 1 through 9 correspond to the 9 objects in the parameter space.

An individual example. The 2D MDS plot for subject *Irene* is displayed on the left in Figure 9. The percentage of intransitive triplets was 0%.

The pooled data. The 2D MDS plot appears on the right in Figure 9. The recovery of the parameter space configuration is nearly perfect.

3.3.5 The CPP Cross experiments

Six subjects were tested with nine objects arranged in a cross-like configuration in parameter space.

An individual example. The 2D MDS plot for subject *Valentina* is displayed on the left in Figure 10. Note that the 2 – 3 and 6 – 7 "arms" of the cross are deformed in a similar manner.

The pooled data. The 2D MDS plot is on the right in Figure 10. The recovery of the parameter space configuration is qualitatively correct, even though a degree of deformation is visible.

3.4 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

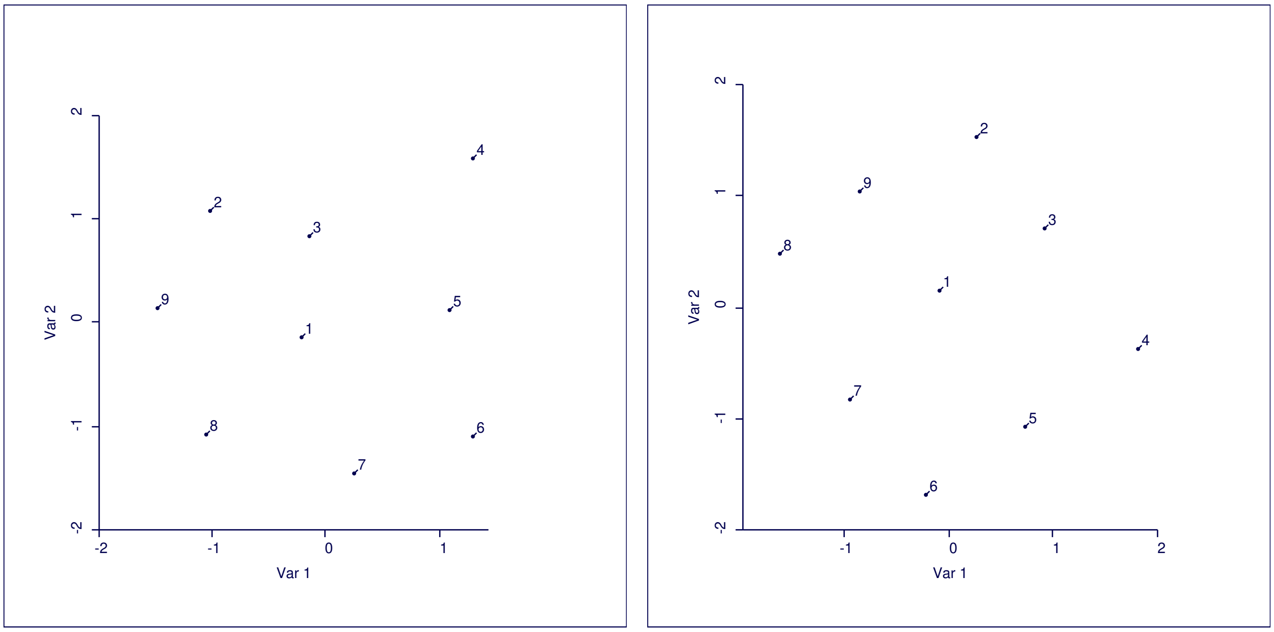


Figure 9: CPP SQUARE. *Left:* the 2D MDS solution for subject Irene (Badness-of-Fit 0.07, Distance Correlation 0.98). *Right:* the 2D MDS solution for all subjects (Badness-of-Fit 0.14, Distance Correlation 0.92).

of basic-level[2] familiarity with the stimuli, we conducted a series of experiments in which the test objects were scrambled versions of the animal shapes used in the experiments described in the preceding section.

3.4.1 Making scrambled objects

The procedure applied for generating nonsense objects was as follows: each animal shape was taken apart by retaining one half of each symmetrical left-right pair of parts, and was 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 — as opposed to the original animals — had no symmetries, and were perceptually more complex than the original animals (cf. (Poggio and Vetter, 1992; Vetter et al., 1992)). Two examples of the scrambled animals used in the experiments are shown in Figure 11.

3.4.2 CPP experiments involving scrambled objects

Two subjects were run on a STAR-like configuration of scrambled animals, and two other subjects were run on a triangular configuration. The subjects did not report any difficulty of principle in making the differential similarity decisions. The only difference from the animal-shape experiments was that the time needed to make a decision was significantly longer. The recovery of the original configuration was as good as for the normal animals. The MDS configuration for subject *Ilana* is displayed on the left in Figure 12. The star is deformed, but the qualitative arrangement is correct: the centroid is mapped

[2] A series of studies (see (Rosch et al., 1976) for a review) provide converging evidence that in the hierarchical structure of object categories there exists a certain level, called basic level, which is the most salient according to a variety of criteria (such as the ease and preference of access). Taking as an example the hierarchy "quadruped, mammal, dog, Dachshund, the basic level is that of "dog".

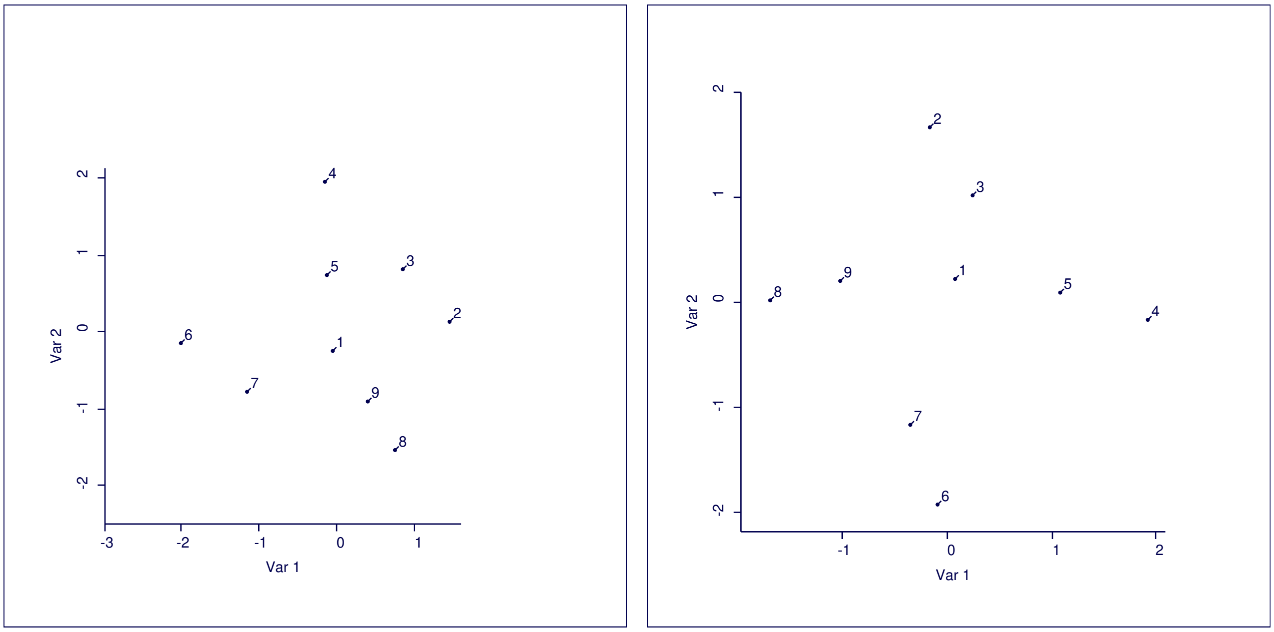


Figure 10: CPP Cross. *Left:* the 2D MDS solution for subject Valentina (Badness-of-Fit 0.04, Distance Correlation 0.99). *Right:* the 2D MDS solution for all subjects (Badness-of-Fit 0.14, Distance Correlation 0.92).

near the center of the configuration, the midpoints of the star arms are positioned between the the centroid and the images of the extremities; the arms of the star are, however, not straight. The MDS configuration for subject *Rada* is displayed on the right in Figure 12. Only point 6 is displaced.

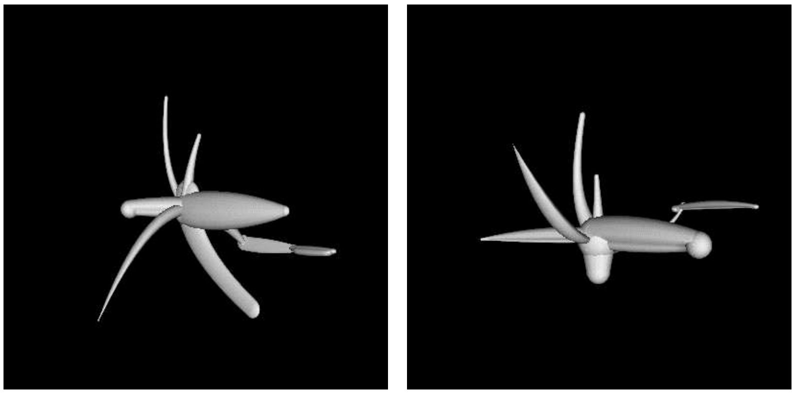


Figure 11: Two scrambled animals (see section 3.4).

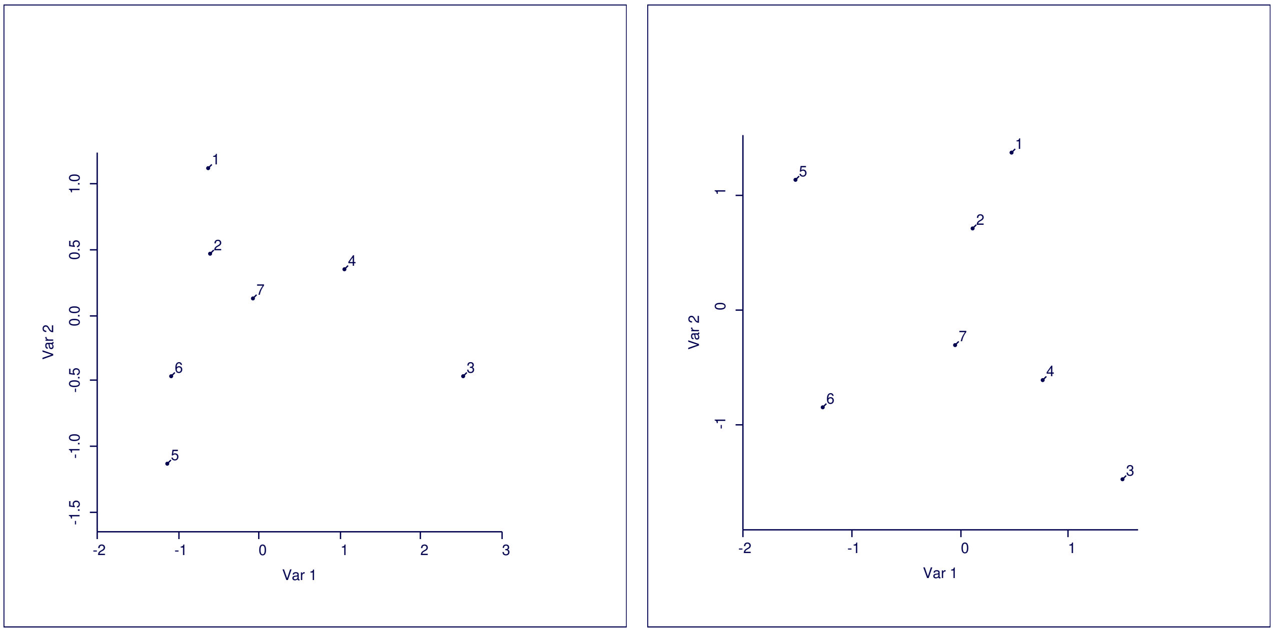


Figure 12: CPP Star, scrambled. *Left:* the MDS configuration for subject *Ilana* (Badness-of-Fit 0.00, Distance Correlation 1.00). *Right:* the MDS configuration for subject *Rada* (Badness-of-Fit 0.01, Distance Correlation 1.00).

4 Exploring the representation of relationships among views of several objects

4.1 Motivation

The CPP experiments described above support the hypothesis of explicit 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 judgments of object similarity ranks. 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 relations between complex 3D objects are reflected in the response times and error rates obtained in a fast discrimination task performed on *views* of these objects.

In the experiments described below, the subjects were briefly 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 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 the experiments was twofold:

- 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;

- 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 2.2, 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.

4.2 Experimental method: DMTS

A variant of the delayed match to sample (DMTS) procedure was used. Pairs of static views of the same object or of different objects were consecutively and briefly flashed on the screen. The subject's task was to decide 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 couple of images of the experimental sequence was shown. The subject received no training, that is, he or she was 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.

Five series of experiments were carried out, corresponding to the LINE, TRIANGLE, STAR, SQUARE, CROSS configurations described in section 2.2 and illustrated in Figure 4. In the experiments which involved seven objects (LINE, TRIANGLE, STAR), four viewpoints per object (corresponding to 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 "same" and "different" decisions more balanced. This brought the proportion of "same" pairs to 39%, out of a total of 546 view pairs. In the experiments which involved 9 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.

4.3 Analysis of results

The proximity tables (see section B.2 in the appendix) were analyzed using multidimensional scaling and discriminant analysis. Nonmetric MDS was applied both to the view and to the object proximity tables, to determine the psychological-space configuration of the stimuli. MDS was applied both on a subject by subject basis, and to the pooled data of all subjects. To estimate classification performance that could be supported by the recovered MDS configuration, the output of the MDS analysis of the view proximity table was submitted to nonparametric discriminant analysis,[3] using the SAS procedure DISCRIM (Sas, 1989). Based on the view configuration obtained from MDS, in which each point represents a view and is labeled with the name of the corresponding animal, the DISCRIM procedure

[3]Since no assumption (such as normality) regarding the distribution within the groups could be justifiably made, nearest neighbor classification, which is nonparametric (Duda and Hart, 1973), was employed.

developed discriminant criteria for classifying the observations (views) into groups (objects). The misclassification rate was then estimated by cross-validation. The error rate estimate produced by cross-validation is stated below in conjunction with each pooled-subjects MDS output.

4.3.1 The DMTS Line experiments

Five subjects were run on stimuli spread out equidistantly on a line connecting a humanoid shape and a biped avian shape in the parameter space. Three other subjects were run on shapes situated between a biped avian and a quadruped dog-like shape.

In the following plots, in the object MDS solution the labels 1 to 7 represent, respectively, the seven objects in parameter space; in the view plots, the labels 1 through 4 represent the four views of object 1, the labels 5 through 8 — the four views of object 2, etc.

An individual example. Subject *Mike* performed at a fair 75% success rate level. The 1D and 2D object plots are presented in Figure 13. The fit for the 1D configuration (on the left in the figure) is poor; moreover, the order of objects 6 and 7 is reversed. Adding a second dimension to the MDS solution removes much of the stress and produces a horseshoe-like configuration. The view configuration for this subject is dislpayed in Figure 14.

The pooled data. The object configurations displayed in Figure 15 were obtained by running MDS on the pooled data of all subjects. The 2D view solution (Figure 16) displays the familiar horseshoe pattern. The views are correctly clustered into objects.

Discriminant analysis of the view configuration. The error rate estimates for the seven shape classes are:

	1	2	3	4	5	6	7
Rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.7500	0.7500
Priors	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429

The global error rate is 0.21, concentrated in object classes 6, 7.

4.3.2 The DMTS Triangle experiments

Eight subjects were tested on shapes arranged in a triangular configuration in parameter space.

An individual example. The MDS plots for subject *Ania* are displayed in Figure 17. The success rate was 76%. The recovered 2D MDS configuration reflects well the original one, save for the centroid A3 which is almost superimposed on A2, and is slightly outside the triangle.

The pooled data. The configurations displayed in Figure 18 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 are deformed in various ways.

Discriminant analysis of the view configuration. The error rate estimates for the seven animal classes are:

	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

There are no classification errors: the global error rate is 0.00.

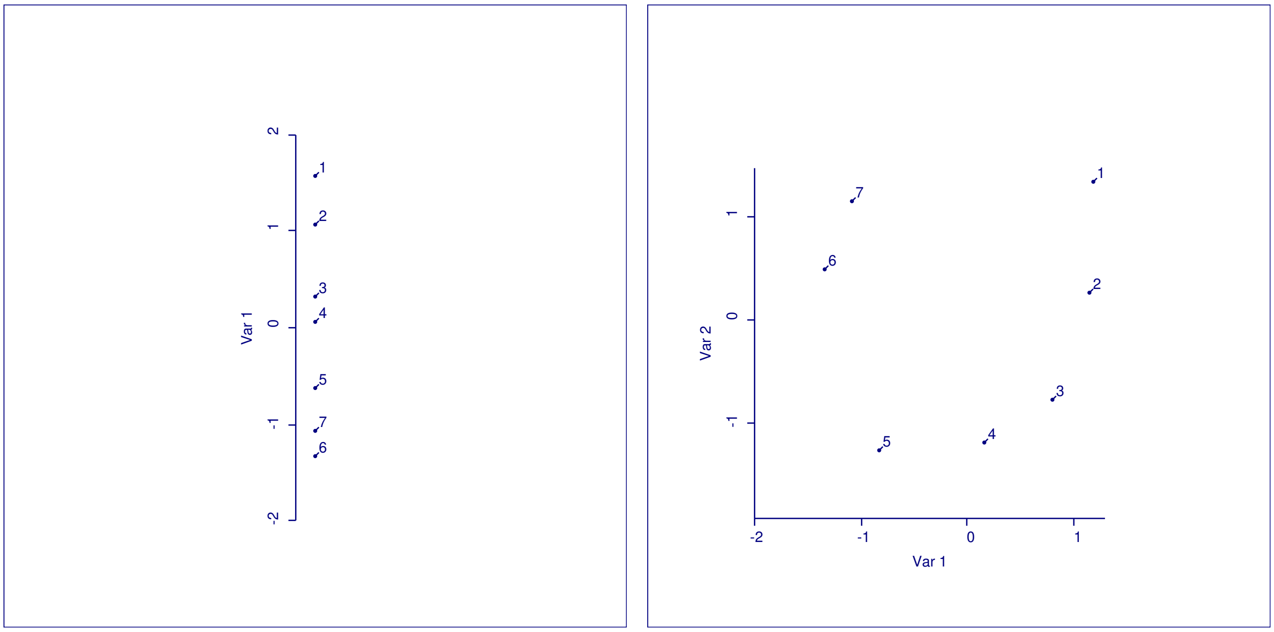


Figure 13: DMTS Line. *Left:* the 1D MDS object solution for subject Mike (Badness-of-Fit 0.23, Distance Correlation 0.89). *Right:* the 2D MDS object solution (Badness-of-Fit 0.07, Distance Correlation 0.98).

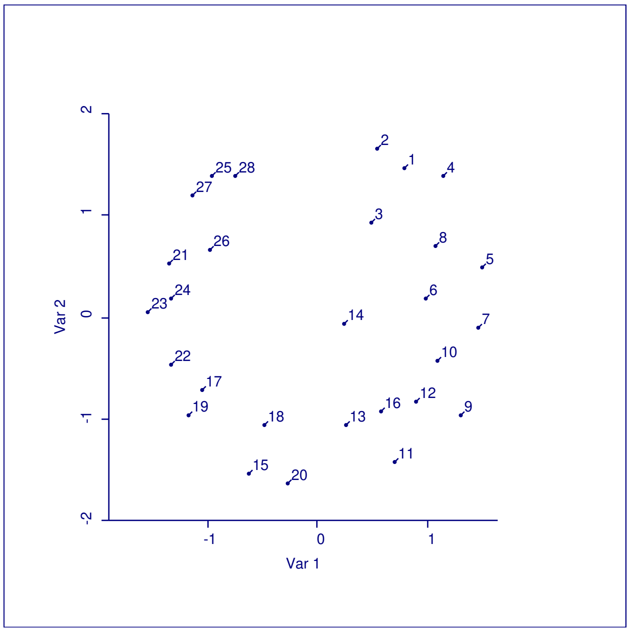


Figure 14: DMTS Line. The view plot for subject Mike (Badness-of-Fit 0.28, Distance Correlation 0.71).

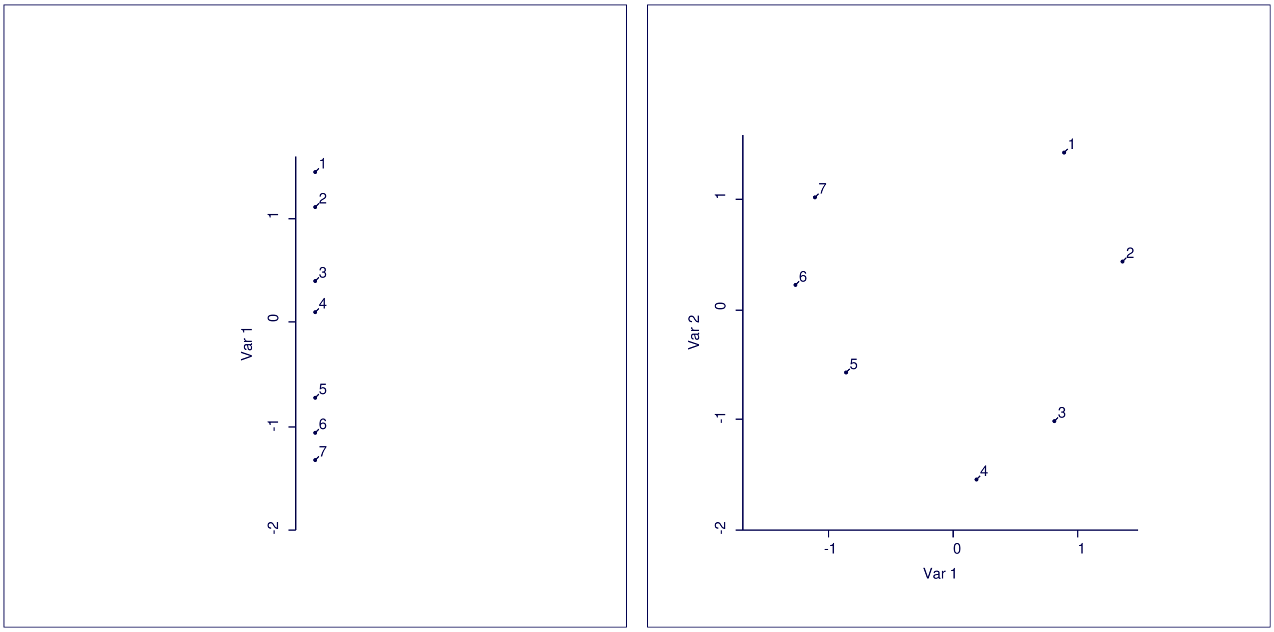


Figure 15: DMTS Line. *Left:* the 1D MDS object solution for the pooled data (Badness-of-Fit 0.26, Distance Correlation 0.85). *Right:* the 2D MDS solution (Badness-of-Fit Criterion 0.15, Distance Correlation 0.91).

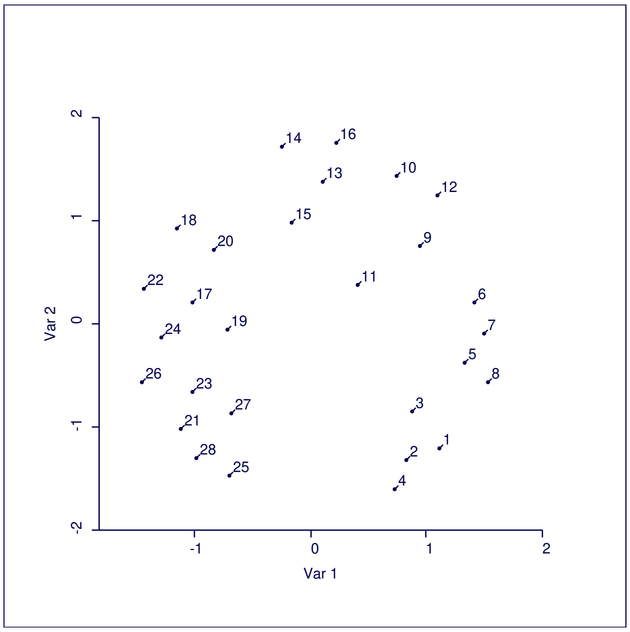


Figure 16: DMTS Line. The view plot for the pooled data (Badness-of-Fit Criterion 0.30, Distance Correlation 0.67).

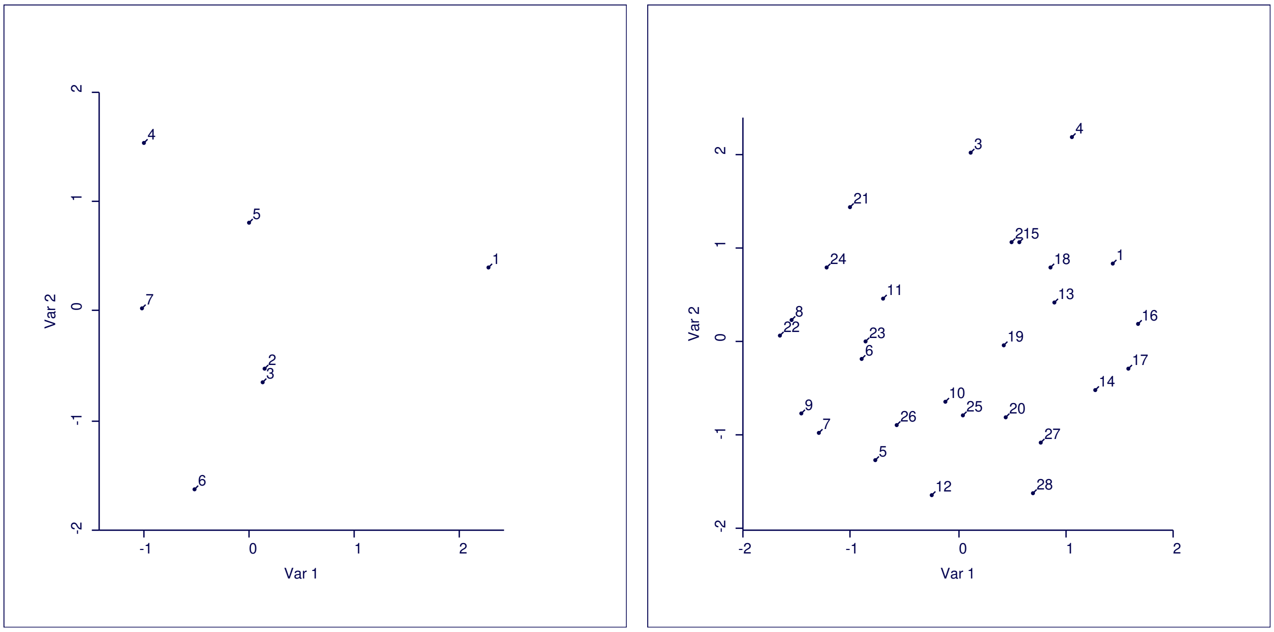


Figure 17: DMTS Triangle. Subject *Ania*. *Left:* the 2D MDS object solution (Badness-of-Fit 0.04, Distance Correlation 0.99). *Right:* the 2D MDS view solution (Badness-of-Fit 0.23, Distance Correlation 0.79).

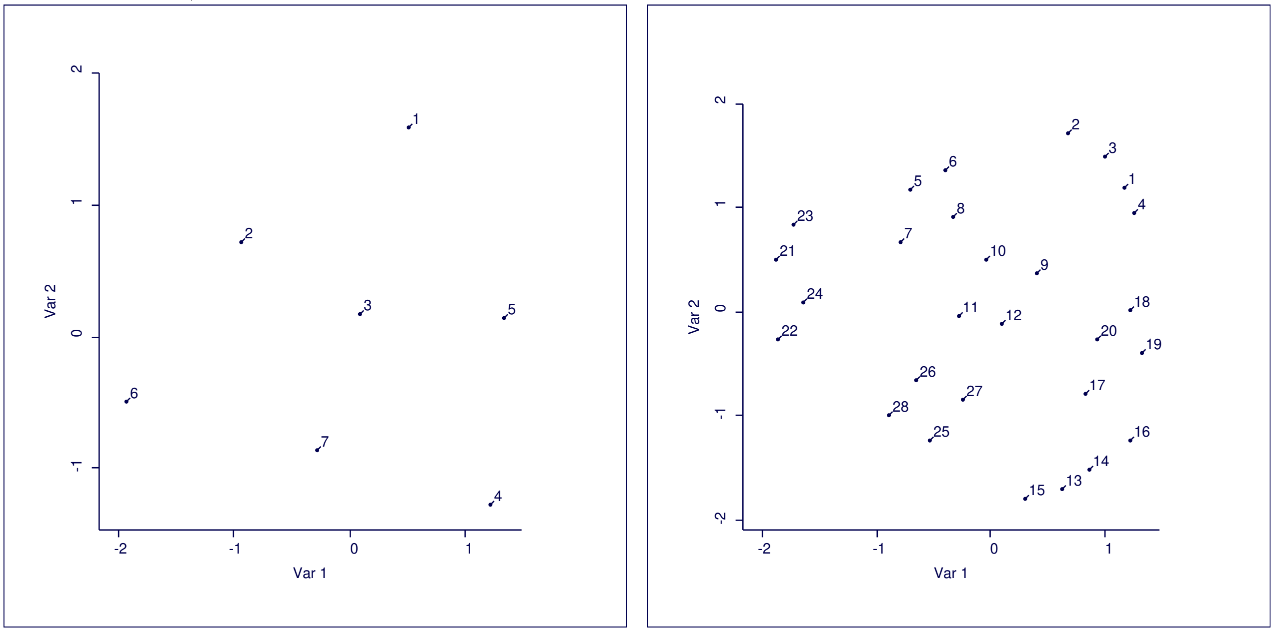


Figure 18: DMTS Triangle. All subjects. *Left:* the 2D MDS object solution (Badness-of-Fit 0.16, Distance Correlation 0.88). *Right:* the 2D MDS view solution (Badness-of-Fit 0.29, Distance Correlation 0.74).

4.3.3 The DMTS Star experiments

Seven subjects were tested on shapes arranged in a star-like configuration in parameter space.
An individual example. The MDS plots for subject *Boris* are displayed in Figure 19. The success rate was 78%. The objects on the arms of the star tend to be confused.
The pooled data. The configurations displayed in Figure 20 were obtained by running MDS on the pooled data of all subjects. The recovery of the parameter space configuration is qualitatively correct.
Discriminant analysis of the views configuration. The error rate estimates for the seven animal classes are:

	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

The global error rate is 0.29.

4.3.4 The DMTS Square experiments

Eight subjects were tested with nine objects positioned at the vertices of a square in parameter space.

In the following MDS plots, in the "object" MDS solution the labels 1 to 9 represent the nine objects in parameter space; in the view plots, the labels 1-3 represent the three views of object 1, labels 4-6 — the four views of object 2, etc.
An individual example. The MDS plots for subject *Tania* are displayed in Figure 21. The success rate was 67%. Objects 2 and 3 are confused; however, the general layout of the stimuli reflects their parameter space distribution.
The pooled data. The configurations displayed in Figure 22 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 are not.
Discriminant analysis of the views configuration. The error rate estimates for the nine object classes are:

	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

	8	9	Total
Rate	0.6667	0.0000	0.4074
Priors	0.1111	0.1111	

Note the relatively high misclassification rate: 0.41.

4.3.5 The DMTS Cross experiments

Seven subjects were tested with nine objects arranged in a cross-like configuration in parameter space.
An individual example. The MDS plots for subject *Elena* are displayed in Figure 23. The success rate was only 67%. The 2D MDS configuration is deformed in a characteristic way, the centroid becoming the tip of a fan-like configuration.

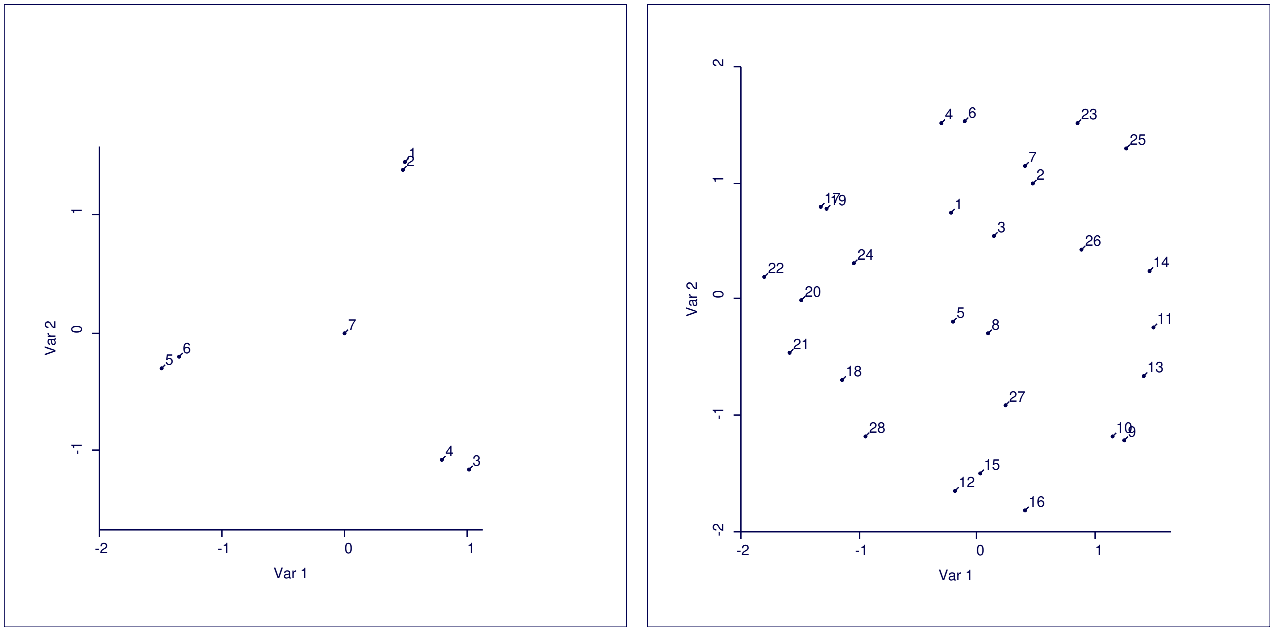


Figure 19: DMTS Star. Subject *Boris*. *Left:* the 2D MDS object solution (Badness-of-Fit 0.00, Distance Correlation 1.00). *Right:* the 2D MDS view solution (Badness-of-Fit 0.28, Distance Correlation 0.70).

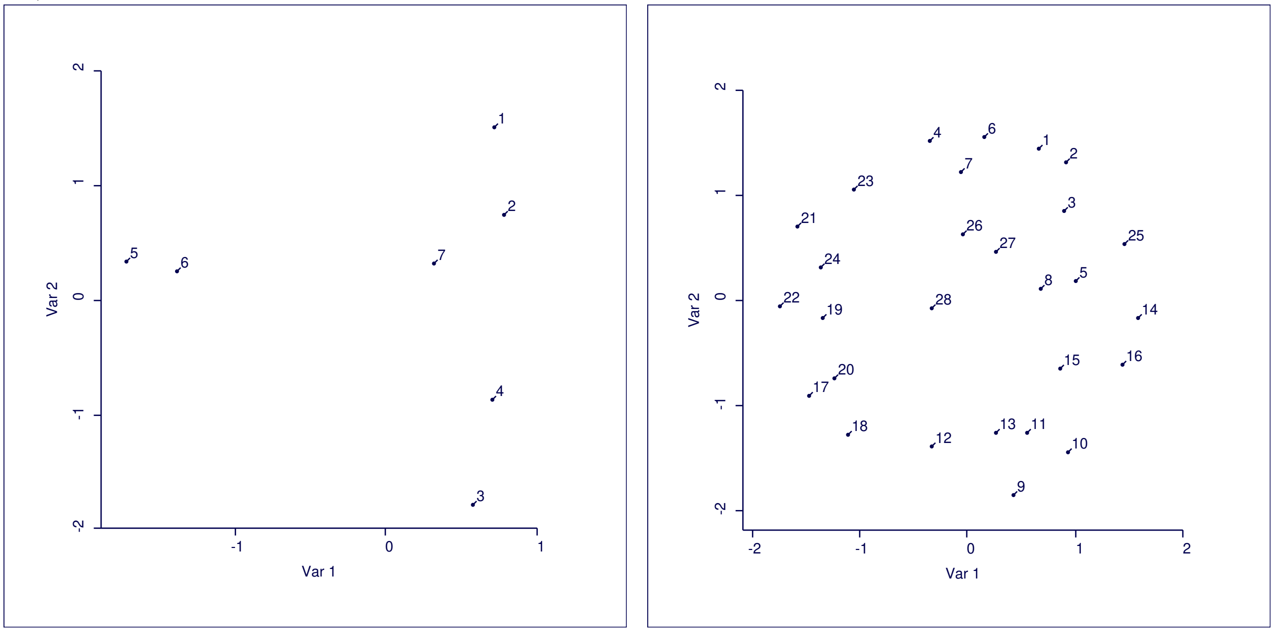


Figure 20: DMTS Star. All subjects. *Left:* the 2D MDS object solution (Badness-of-Fit 0.14, Distance Correlation 0.94). *Right:* the 2D MDS view solution (Badness-of-Fit 0.33, Distance Correlation 0.57).

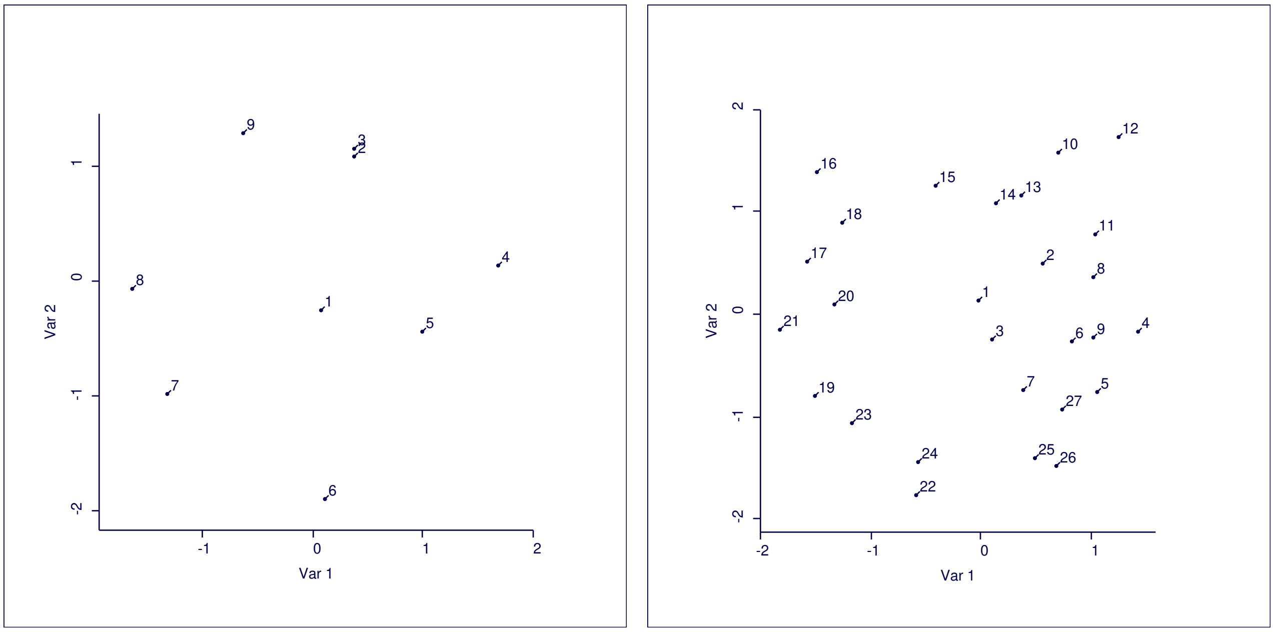


Figure 21: DMTS SQUARE. Subject *Tania*. *Left:* the 2D MDS object solution (Badness-of-Fit 0.02, Distance Correlation 1.00). *Right:* the 2D MDS view solution (Badness-of-Fit 0.15, Distance Correlation 0.93).

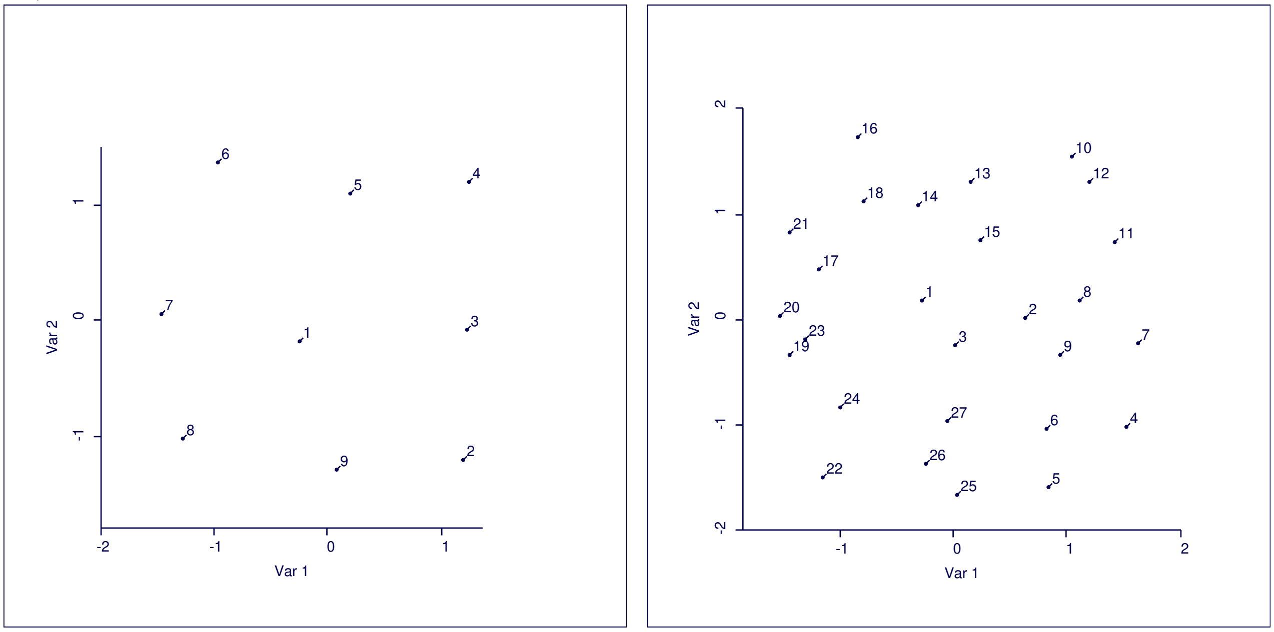


Figure 22: DMTS SQUARE. All subjects. *Left:* the 2D MDS object solution (Badness-of-Fit 0.19, Distance Correlation 0.81). *Right:* the 2D MDS view solution (Badness-of-Fit 0.29, Distance Correlation 0.65).

The pooled data. The configurations displayed in Figure 24 were obtained by running MDS on the pooled data of all subjects.
Discriminant analysis of the views configuration. The error rate estimates for the nine object classes are:

	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

	8	9	Total
Rate	0.0000	0.0000	0.1481
Priors	0.1111	0.1111	

The global error rate is quite low: only 0.15.

4.4 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 nonsense objects. We conducted two DMTS experiments with scrambled animal shapes (the scrambled animals were produced according to the scheme described in section 3.4): one experiment involved a star-like parameter space configuration, and the other one a triangular configuration.

4.4.1 Scrambled animals: configuration Star

Five subjects were used in this series of experiments. The average success rate was 70%, which means that the subjects separated correctly the different input views into objects. However, in order to reach such a success rate, comparable to the performance level in the DMTS experiments with regular animals, the exposure time had to be increased to about 750 *msec*.
An individual example. The plot for subject *Denis* is displayed on the left in Figure 25. The success rate was high, 78%. Even though there is significant deformation of the original configuration, it can be seen that the arms of the star have been altered systematically, all in the same manner, that is, rendered parallel with the preservation of the correct ordering.
The pooled data. The configuration obtained from running MDS on the pooled data of all subjects is depicted on the right in Figure 25; it has practically the same qualitative characteristics.
Discriminant analysis of the views configuration. The error rate estimates for the seven object classes are:

	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
Rate	0.5714

Note the relatively high value (0.57) of the global misclassification rate.

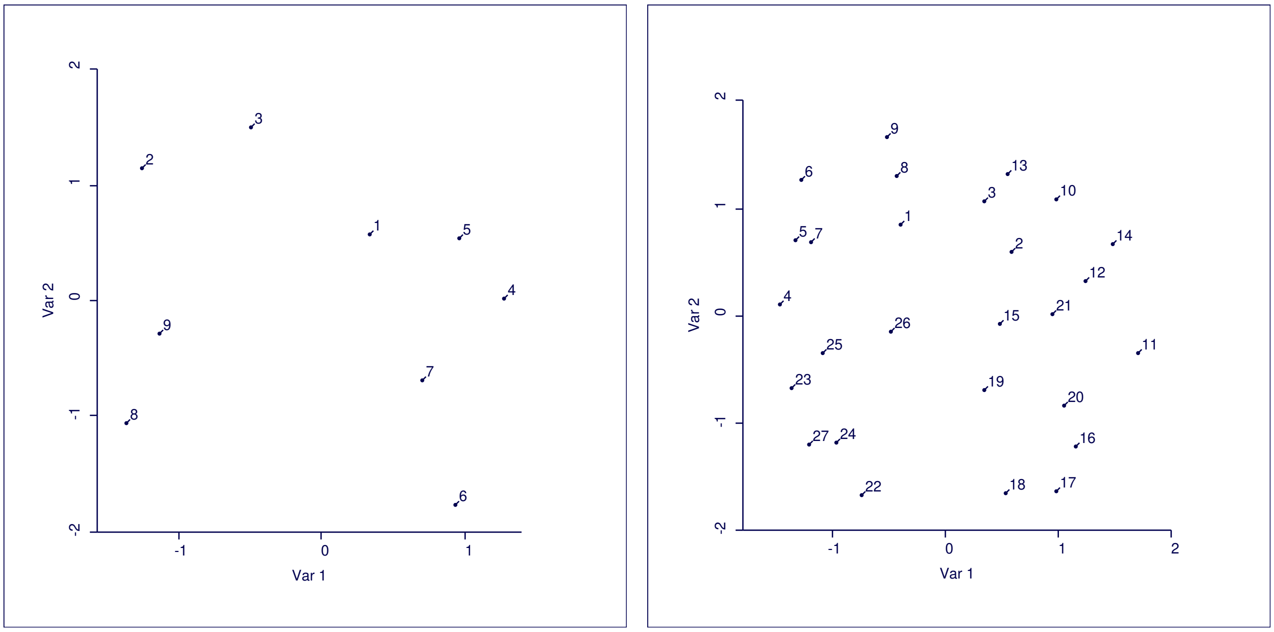


Figure 23: DMTS Cross. Subject *Elena*. *Left:* the 2D MDS object solution (Badness-of-Fit 0.08, Distance Correlation 0.97). *Right:* the 2D MDS view solution (Badness-of-Fit 0.26, Distance Correlation 0.75).

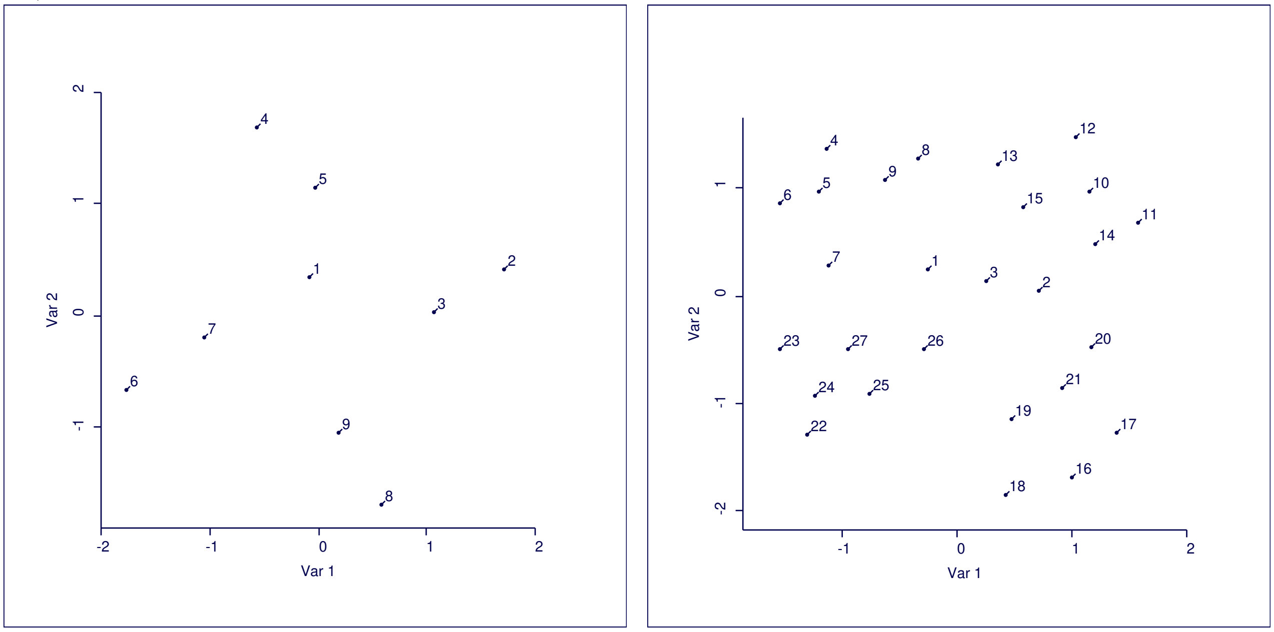


Figure 24: DMTS Cross. All subjects. *Left:* the 2D MDS object solution (Badness-of-Fit 0.18, Distance Correlation 0.87). *Right:* the 2D MDS view solution (Badness-of-Fit 0.29, Distance Correlation 0.70).

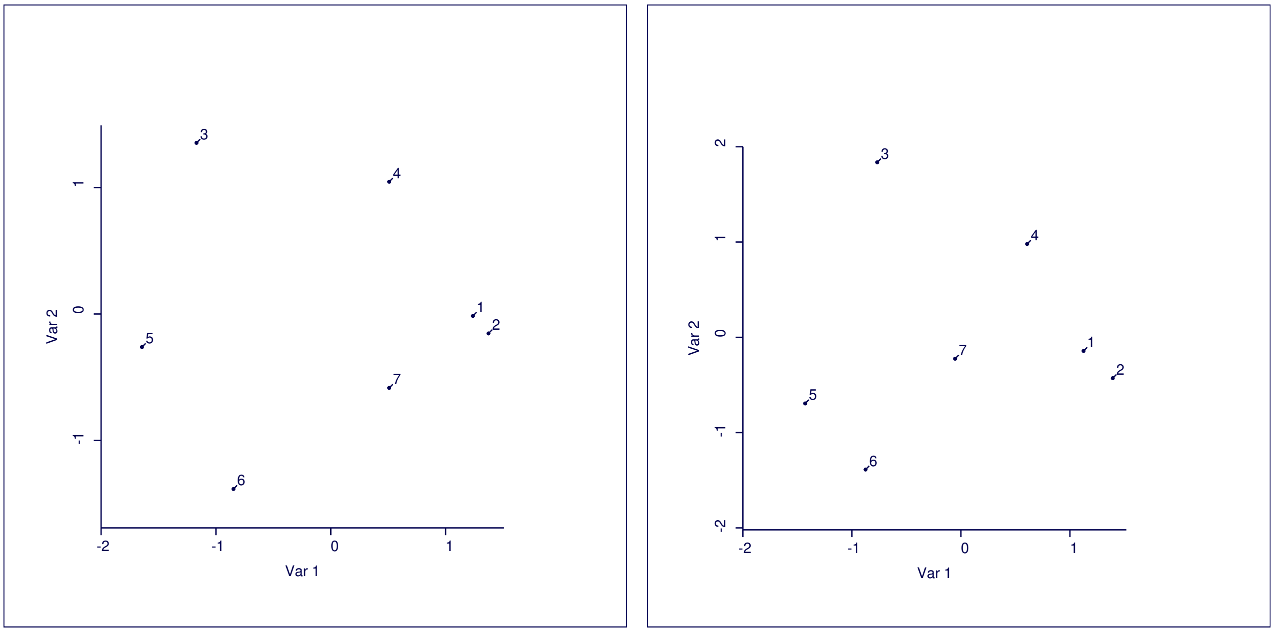


Figure 25: DMTS Scrambled Star. *Left:* the 2D MDS object solution for subject *Denis* (Badness-of-Fit 0.05, Distance Correlation 0.99). *Right:* the 2D MDS solution for the pooled data (Badness-of-Fit 0.13, Distance Correlation 0.94).

4.4.2 Scrambled animals: configuration Triangle

Five subjects were run in this configuration. Again, in order to get a success rate in line with the values obtained for the experiments with animal shapes, (the average was 72%) the exposure time had to be increased to 750 *msec*.

An individual example. The plot for subject *Eliana* is displayed on the left in Figure 26. The success rate was high, 72%, and the subject was the one of the two for which in the recovered configuration the original configuration could be recognized.

The pooled data. The configuration obtained from running MDS on the pooled data of all subjects is depicted on the right in Figure 26; note that the original structure of the parameter space is completely lost.

Discriminant analysis of the views configuration. The error rate estimates for the seven object classes are very high:

	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

	Total
Rate	0.6667

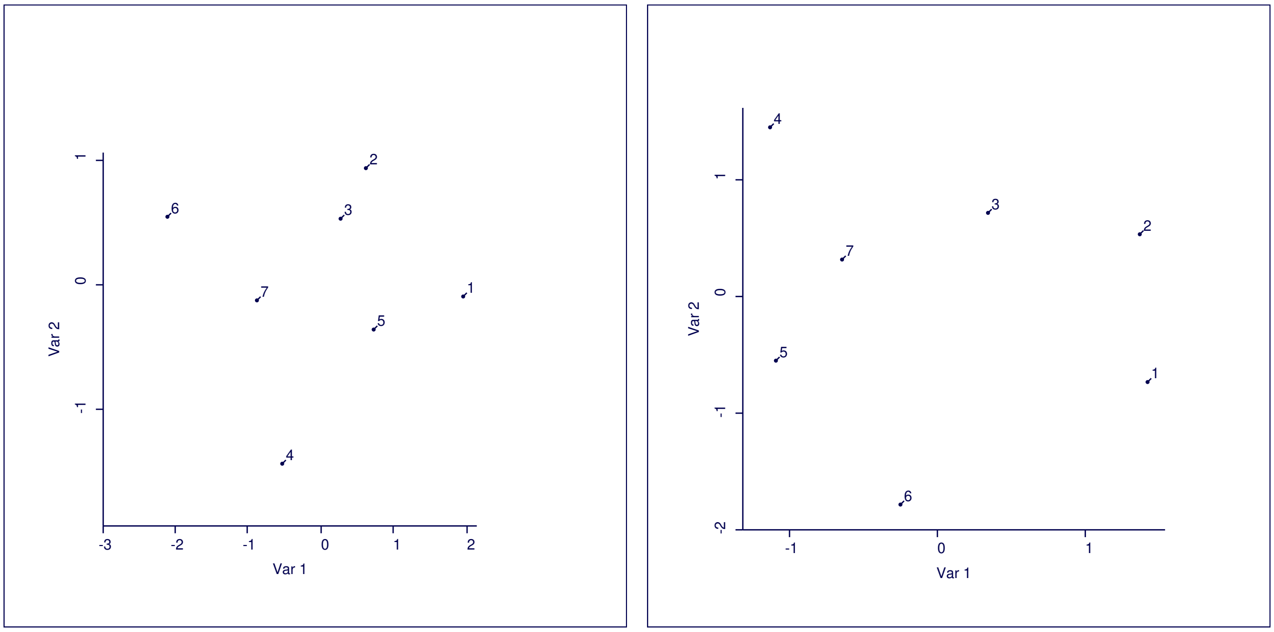


Figure 26: DMTS Scrambled Triangle. *Left:* the 2D MDS object solution for subject *Eliana* (Badness-of-Fit 0.07, Distance Correlation 0.98). *Right:* the 2D MDS solution for the pooled data (Badness-of-Fit 0.21, Distance Correlation 0.79).

5 General discussion

5.1 The CPP experiments

The recovery of the true shape-space configurations of stimuli from the data of subjects who carried out the Compare Pairs of Pairs (CPP) task supports the notion of similarity-based representation, stated in the introduction. The near-veridical internal representation of shape-space similarity patterns implied by our results can be contrasted with the poor sensitivity of subjects to 3D shape information per se (in the context of object discrimination and recognition), as indicated by a number of findings, surveyed in (Jolicoeur and Humphrey, 1994). Thus, the hypothesis that the human visual system represents directly the similarities among the shapes of 3D objects and not their detailed 3D structure (Edelman, 1995c) appears to be both parsimonious (in that it does not postulate an intermediate step of 3D shape recovery prior to the assessment of similarity) and compatible with a range of psychophysical findings.

Remarkably, the subjective similarity ratings in the CPP experiments turned out to be, to a very large extent, transitive. In connection with this finding, it should be noted that transitivity is a property of the metric space (geometrical) models of similarity, and does not necessarily hold 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.

The ability of our subjects to recover the shape-space configuration not only for animal shapes but also for scrambled objects 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).

5.2 The DMTS experiments

To establish that the findings regarding the representation of similarity were not peculiar to the CPP task, we have repeated the entire series of experiments using the delayed match to sample (DMTS) paradigm. Here too, we were able to recover from the subject data the shape-space structures built into the relationships between the stimuli. In the case of the animal shapes, the recovery of the original parameter-space structure was qualitative, sometimes missing the fine metric details of the true configuration. For example, the shape corresponding to the centroid of the shape-space structure was likely to be positioned somewhere inside the recovered configuration, but not always near its true center. These deformations, however, looked systematic and not random (as they would if the MDS procedure were fitting noise rather than data). As a rule, the pooled-subject configurations were much closer to the objective parameter-space patterns. The high reliability of the configurations recovered from the pooled data is indicated also by the estimates of the goodness of fit accompanying each figure (these can be compared to the statistical controls provided in section A in the appendix.

In the scrambled-shape DMTS experiments, a level of performance (proportion of correct responses) comparable to that obtained with animal shapes could only be reached with substantially longer stimulus presentation times. Furthermore, the deformation of the object configurations produced by MDS from scrambled-shape data was much more prominent. This finding is compatible with similarly poor recovery of the true shape-space configuration for wire-like objects in the DMTS experiments reported in (Edelman, 1995b). A possible explanation for the difficulties encountered by subjects who are asked to recognize "unnatural" (wire-like or scrambled) shapes is provided by the hypothesis of representation by similarity to prototypes (Edelman, 1995c). Presumably, shapes that lie in a region of the shape space that is far away from the closest reference (prototypical) objects are misrepresented,[4] leading to difficulties in comparing such shapes and in committing them to memory.

5.3 Conclusions

We saw that, as a rule, the psychological-space configurations recovered by MDS from subject data are either identical to or systematically transformed versions of the objective shape-space configuration built into the stimuli, but, of course, not directly perceivable by the subjects. While the same recovery of objective similarity relationships has been demonstrated in the past in the classical MDS experiments, our results show for the first time that such recovery is possible with complex 3D objects. This finding is especially surprising given the complicated relationship between the parameter values that define the location of an object in the multidimensional shape space, and the image of the object as it appears on the screen.

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

[4]Two possible sources of computational problems in this case are the low response (or no response) from the recognition modules tuned to familiar objects in the face of a radically unfamiliar one, and the low precision of comparing near-equal and large distances between the stimulus and the representations of reference objects.

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.

5.4 Future work

Modeling the results. To appreciate the computational aspects of the problem of faithful representation of similarity, routinely solved by our subjects, one may try to replicate the human data with a model. It may be claimed that the parameterization of the stimuli in our experiments was by construction perceptually significant, and that the best way to replicate the good recovery of the parameter-space configuration would be then to tune the model to those characteristics of the stimuli that are considered traditionally (Biederman, 1987) as perceptually salient: axis curvature, section taper, *et cetera*. However, the results reported in (Edelman, 1995b) indicate that the recovery of the parameter-space pattern can be replicated with a model based on representation by distances to prototypes encoded in the space of activities of simple receptive fields.[5] This finding raises the question whether the perceptually salient features of 3D shapes are inherently involved in any instance of representation of a 3D object (Biederman, 1987), or are merely convenient coalitions of more primitive features (such as the activities of simple receptive fields), which may be dissolved and reformed according to a different roster on the spur of the moment, as suggested by Herrnstein's notion of polymorphic features (Herrnstein, 1984).

Perception of similarity across categories. In all the animal shapes used in our experiments, corresponding parts were always identifiable, despite the extreme variability in the appearance of the animals induced by changing the parameter values. While this feature of the animal-like stimuli facilitates the estimation of differential similarity in tasks such as CPP, it is restrictive in the sense that questions related to superordinate-level classification and to the perception of similarity across natural kinds are not addressed. It seems doubtful that the DMTS paradigm is suitable for exploring cross-category similarity (delayed match to sample involving radically different objects is likely to yield low-variance response times and near-zero error rates). At the same time, the comparison-based CPP paradigm would confront the subject with the need to make differential similarity judgements on pairs of objects belonging to different superordinate categories, encouraging similarity estimation at a conceptual rather than a perceptual level. A new experimental approach must, therefore, be developed to explore the shape space at a superordinate level.

Exploring long-term memory representations. The experiments reported in the preceding sections studied the similarity structure of representations of stimuli that were perceptually available (as in the CPP task), or were retained in short-term memory (as in the DMTS task). Only one study known to us explored the representation of shape similarity in *long-term* memory (Shepard and Chipman, 1970). In that study, MDS-derived configurations obtained from perception data were compared with those obtained from recall data; shapes of 15 of the states of US were used as stimuli. A systematic exploration of the similarity structures induced in long-term memory representation space by controlled sets of 3D shapes is currently under way in our laboratory. In these experiments, the subjects learn a set of labeled animal shapes, and then are tested in pairwise comparisons resembling the CPP task. Preliminary results of this study are encouraging: the fair recovery of the parameter-space patterns

[5]But not with a model based on pixel-space distances among the stimuli.

from recall-based comparison data suggests that the long-term memory representations may indeed be structured according to similarities.

Acknowledgements

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.

References

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

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

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

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.

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

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

Duvdevani-Bar, S., Cutzu, F., and Edelman, S. (1995). A computational model of similarity-based representation of the shape space. Cs-tr, Weizmann Institute of Science. in preparation.

Duvdevani-Bar, S. and Edelman, S. (1995). On the representation by similarity to prototypes. Cs-tr, Weizmann Institute of Science. in preparation.

Edelman, S. (1993). Representation, Similarity, and the Chorus of Prototypes. CS-TR 93-10, Weizmann Institute of Science. to appear in Minds and Machines, 1995.

Edelman, S. (1994). Representation without reconstruction. *Computer Vision, Graphics, and Image Processing*, 60:92–94.

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

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:–.

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

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

Herrnstein, R. J. (1984). Objects, categories, and discriminative stimuli. In Roitblat, H. L., Bever, T. G., and Terrace, H. S., editors, *Animal Cognition*, pages 233–261, Hillsdale, NJ. Erlbaum.

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

Jolicoeur, P. and Humphrey, G. K. (1994). 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. (1977). The relationship between multidimensional scaling and clustering. In Ryzin, J. V., editor, *Classification and clustering*, pages 17–44. Academic Press, New York.

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

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

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

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

Poggio, T. and Vetter, T. (1992). Recognition and structure from one 2d model view: observations on prototypes, object classes, and symmetries. A.I. Memo No. 1347, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

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

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

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

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

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.

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

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

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

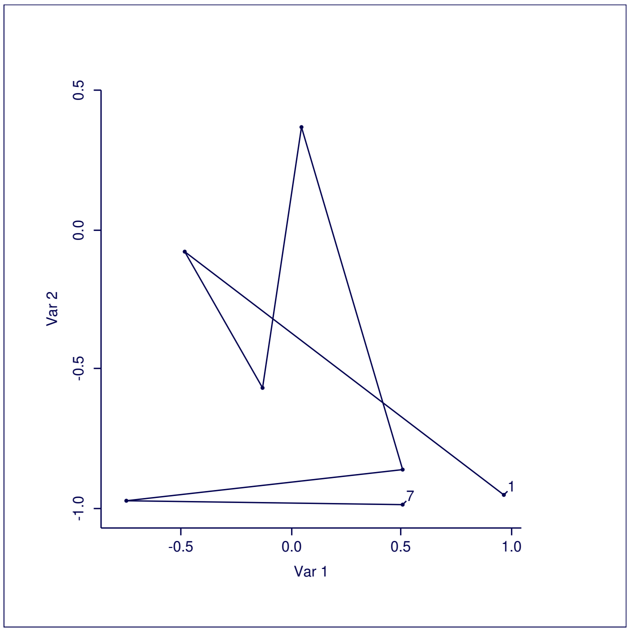
Vetter, T., Poggio, T., and Bülthoff, H. H. (1992). 3D object recognition: Symmetry and virtual views. A.I. Memo No. 1409, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Weiss, Y. and Edelman, S. (1995). Representation of similarity as a goal of early visual processing. *Network*, 6:19–41.

A Evaluating the accuracy of nonmetric multidimensional scaling reconstruction

Nonmetric multidimensional scaling (MDS) is the basic tool used for data analysis in the present research. To show that it is appropriate for our purposes, we have conducted computer simulations to demonstrate that planar, random configurations of seven points (corresponding to the least number of objects used in our experiments) can be successfully recovered by nonmetric MDS from the ranks of their mutual distances.

A graphical example of the power of nonmetric MDS is shown in Figure 27: the original configuration is presented on the left and the recovered configuration on the right. Note that the original configuration of points is almost perfectly recovered.



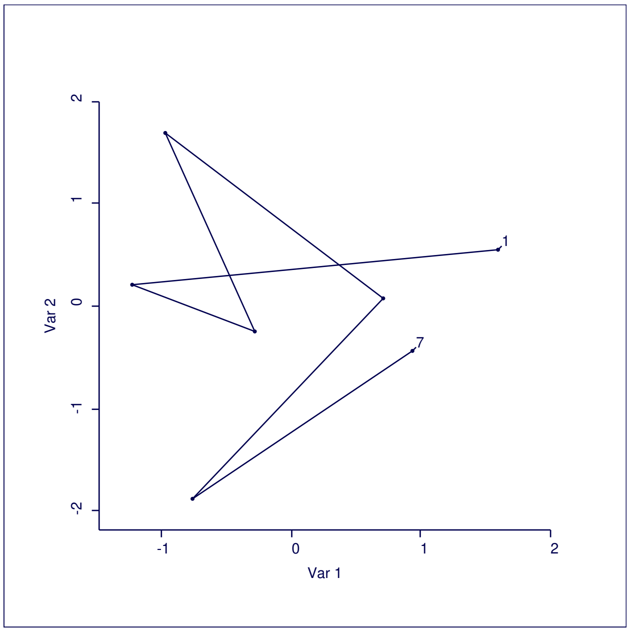


Figure 27: *Left:* a random configuration of seven points in the plane. *Right:* its reconstruction by nonmetric multidimensional scaling.

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 of the corresponding configuration). The Pearson correlation coefficient characterising 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.

The output of the SAS MDS procedure includes, in addition to the recovered configuration, two number 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.

To appreciate the significance of the Badness-of-Fit and the Distance Correlation figures obtained for the experimental data, it is useful to consider the chance-level values of these parameters. To do so, we ran nonmetric MDS on 100 symmetric, random proximity matrices of size 7×7, with entries uniformly distributed between 0 and 1. The Badness-of-Fit was 0.11 ± 0.03 (mean and standard deviation). The Distance Correlation was 0.94 ± 0.03. To obtain the chance-level figures for the pooled-data condition, nonmetric MDS was run on 100 sets of five random-matrix "subjects" each. This resulted in Badness-of-Fit of 0.24 ± 0.01, and Distance Correlation of 0.64 ± 0.05.

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 possibiltiy 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.[6]

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

[6] 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.

View proximity table. The entry (i, j) in the view proximity table was intended to measure the perceived similarity[7] 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 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.

[7] 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.