Representation of similarity in 3D object discrimination

Shimon Edelman

Dept. of Applied Mathematics and Computer Science

The Weizmann Institute of Science

Rehovot 76100

Israel

edelman@wisdom.weizmann.ac.il

Abstract

How does the brain represent visual objects? In simple perceptual generalization tasks, the human visual system performs as if it represents the stimuli in a low-dimensional metric psychological space (Shepard, 1987). In theories of 3D shape recognition, the role of feature-space representations (as opposed to structural (Biederman, 1987) or pictorial (Ullman, 1989) descriptions) has been for a long time a major point of contention. If shapes are indeed represented as points in a feature space, patterns of perceived similarity among different objects must reflect the structure of this space. The feature space hypothesis can then be tested by presenting subjects with complex parameterized 3D shapes, and by relating the similarities among subjective representations, as revealed in the response data by multidimensional scaling (Shepard, 1980), to the objective parameterization of the stimuli. The results of four such tests, accompanied by computational simulations, support the notion that discrimination among 3D objects may rely on a low-dimensional feature space representation, and suggest that this space may be spanned by explicitly encoded class prototypes.

1 Background

A number of recent developments in the computational theory of 3D object recognition indicate that objects may be effectively represented by collections of their 2D images (Poggio and Edelman, 1990; Ullman and Basri, 1991). Indeed, evidence from psychophysical experiments (in particular, repeated findings of viewpoint-dependent recognition performance) suggests that human observers employ a chosen set of stored views of objects in distinguishing among objects from the same basic category (Rock and DiVita, 1987; Tarr and Pinker, 1989; Bülthoff and Edelman, 1992; Edelman and Bülthoff, 1992). The development of a comprehensive multiple-view theory of recognition depends on the resolution of the issue of representation of such stored views. The question of dimensionality of the representation is of particular interest here. At the initial stages of processing in the human visual system, the dimensionality is dictated by certain basic facts of primate neuroanatomy. For example, the information at the output of the retina is embedded in a million-dimensional space, simply because the optic nerve possesses that many fibers. It is clear that subsequent processing of visual information must involve a massive dimensionality reduction. Although some computational models of recognition explicitly address this point (Intrator and Gold, 1993), the nature of the putative low-dimensional representations that support shape processing in human vision is as yet unknown. The present note reports a psychophysical investigation of the hypothesis that visual stimuli are represented as points in a low-dimensional feature space, in which feature values are determined by distances to a few prototypical stored patterns, allowing the resulting representation to reflect closely the objective structure of the stimulus universe.

2 Psychophysical experiments

If shapes are indeed represented as points in a feature space, patterns of perceived similarity among different objects must reflect the structure of this space. The feature space hypothesis can then be tested by presenting subjects with complex parameterized 3D shapes, and by relating the similarities among the internal representations of the stimuli to their objective parameterization. Although subjective internal representations are not directly accessible to the experimenter, it may be possible to recover the relationships between the representations of different stimuli from the response data, using multidimensional scaling, or MDS (Shepard, 1980). To determine the configuration of n stimuli using nonmetric MDS, one starts with an $n \times n$ table, in which the value placed at the (i, j)'th entry is monotonically related to the similarity between stimulus i and stimulus j. In the present study, the objective task configuration was formed by arranging the stimuli in a two-dimensional pattern in a multidimensional parameter space. The interstimulus similarity values were obtained experimentally by measuring response times in a series of delayed matching to sample trials involving all possible stimulus pairs, as explained below.

2.1 Method

Stimuli. Four experiments were carried out, each of which employed a different family of 3D shapes (Figure 1*a*). The stimuli in each experiment belonged to two jointly parameterized computer-generated object classes. In experiment 1 (Figure 2) the objects resembled monkeys or dogs, and were jointly parameterized by a set of 56 variables, which encoded sizes, shapes and placement of the limbs, the ears, the snout, etc. Sixteen test images were obtained from the prototypes by a four-step procedure. First, the equation of the line in R^{56} connecting the two prototypes was computed. Second, for each prototype, a subspace of dimensionality 55 normal to that line was formed. Third, two exemplars from each of these two parallel subspaces were chosen, so that the distance between each exemplar and the corresponding prototype was fixed. Finally, each exemplar was rendered on a graphics workstation, as seen from four different viewpoints. The resulting 16 stimuli images from experiment 1 are shown in Figure 2; the stimuli for the other experiments were generated analogously.

Procedure. The experimental paradigm was delayed matching to sample (see Figure 1*b*). The subjects were told that they will see pairs of images of two objects, taken from a variety of viewpoints. In each trial, the subject was shown an image of one of the objects, then a mask, then again an image of either of the objects. The subject's task was to determine whether or not the two images belonged to the same 3D shape. The response could be made at any time following the display of the second image. Each experimental session began with a series of training trials, which resembled the subsequent testing in everything except that during training the subjects received auditory feedback for incorrect responses. The testing stage, consisting of 256 trials (all pairings of the 16 stimuli images), began when the subject responded correctly in more than 90% of the trailing 30 trials. This procedure resulted in a standardization of the subject's performance level prior to the testing.

The correct-trial response times (RTs) of each subject, in seconds, were entered in a 16×16 table (if the response was erroneous, the trial was marked as yielding a missing value; the mean error rates in the four experiments were 5%, 9%, 25%, and 17%). The table was filled as follows. For images i and j that belonged to the same object class, the value of RT was entered into the (i, j)'th place in the table. For images i and j that belonged to different classes, the quantity $2.5 - \mathrm{RT}$ was entered. The need for such preprocessing of the RT values stems from the requirement that stimulus similarity data submitted to nonmetric MDS vary monotonically with the true distances between the subject's internal representations of the stimuli. Now, the RT in a "same" trial (i.e., when the two images indeed belong to the same object) can be assumed to *increase* monotonically with the distance between the representations of the two images in the subject's psychological space: the more similar the images, the easier it should be to give a correct "same" response. In comparison, the RT in a "different" trial should *decrease* monotonically with the psychological distance between the two images: the more different the images, the easier it should be to produce a correct "different" response. To comply with the overall requirement of monotonicity, "different" trials should thus be made to yield a quantity inversely related to RT, such as a constant minus RT.[1] It should be noted that because these conditions are stated in terms of the subjective psychological space (the same space that one wishes to recover using MDS), they are not amenable to direct verification. Indirect justification for the assumed dependence of RT on psychological distance may be found in certain computational models of decision making (Ratcliff, 1981). A more decisive argument would be the success of MDS in recovering an interpretable configuration, accompanied by a high correlation between the data and the MDS-derived distance table. As we shall see, this argument is indeed upheld by the data.

2.2 Results

The tables of the RT data were processed by a nonmetric individually weighted multidimensional scaling procedure ((Carroll and Chang, 1970); SAS procedure MDS (Sas, 1989)). The main results of MDS are illustrated in Figure 3 (top left plot in each of the four panels). In experiment 1 (animal shapes) the 2D configuration recovered by MDS revealed an astonishingly faithful replica of the main low-dimensional traits of the structure of the parameter space used to generate the stimuli, namely, the distinction between the two object classes, and the within-class variation orthogonal to that distinction (Figure 3*a*). In comparison, in experiment 4 (wire-like shapes) the 2D configuration was not at all interpretable in terms of the objective contrasts built into the stimuli (Figure 3*d*). Notably, the correlations between the recovered and the real distances among data points were markedly different in experiment 1 and experiment 4 (0.80 and 0.29, respectively). The configuration recovered in experiment 2 (scrambled animals; Figure 3*b*) resembled that of experiment 1. The configuration in experiment 3 (geon wires; Figure 3*c*) was somewhat more noisy.

To estimate the likelihood of the high distance correlation being an artifact of the analysis method, the MDS procedure for each experiment was repeated for 100 data sets obtained by replacing each RT value by a number from a normal distribution with the same mean and

[1] The chosen value of the constant, 2.5 *sec*, was determined by computing the histogram of the RTs and by looking for the bend in the tail of the distribution. Less than 1% of the RTs exceeded this threshold in each of the four experiments.

standard deviation as the subjects' RTs for that experiment.[2] This procedure yielded mean distance correlations of 0.29 ± 0.02, 0.28 ± 0.03, 0.34 ± 0.05, and 0.30 ± 0.02, respectively, in the four experiments. It is, therefore, highly unlikely that the distance correlation of 0.80 obtained in experiment 1 is due to chance.

2.3 Discussion

These results indicate that when the subjects do succeed to represent the stimuli, the representations appear to be of the low-dimensional feature-space variety. Although the dimensions of this feature space bear a highly complex relationship to the raw physical dimensions of the stimuli shapes, the configuration formed by the representations of the individual stimuli is readily interpretable in terms of the metric shape contrasts pertinent to the experimental task. The successful recovery of the true parametric distinction among the stimuli from the human response data in experiments 1 through 3, but not in experiment 4, suggests that visual shapes lead to the formation of efficient and useful (that is, low-dimensional and parametrically veridical) representations insofar as they resemble some "natural" object categories.

3 Computational simulations

The notion of resemblance between shapes, invoked in the above discussion, begs the question of computational definition of similarity. What mechanism can support the extraction of low-dimensional features of similarity from image data? One possibility is suggested by the structural theories of recognition (Biederman, 1987), which hold that objects are similar if they share nonaccidental features (which are, presumably, more salient in animal shapes

[2] Methods for estimating directly the variance of the correlation from the subject data, such as the bootstrap (Efron and Tibshirani, 1993), are not applicable in the present case, because the correlation derived from nonmetric MDS is based on the rank order of the data, and not on the means computed over some of its partitions.

than in wire-frame assemblies). According to the structural approach, the nonaccidental shape properties recovered from the image of a scene lead to the formation of a symbolic representation that captures its 3D geometry. The three computational simulations described below indicate that a much simpler mechanism, based on the idea of filter-space similarity to prototypes, may suffice for this purpose.

3.1 Method

Common to all three simulations is a bank of $k = 200$ filters that transduced the images (the same ones seen by the human subjects) into a 200-dimensional Euclidean space, R^{200} (the receptive fields of the filters were radially elongated Gaussians, randomly positioned over the image; see (Edelman, 1992)). Dissimilarity between pairs of images could then be computed as distance in R^{200} (the dissimilarity patterns induced by such representations approach their asymptotic values already for $k \approx 150$ (Weiss and Edelman, 1993)).

In the first simulation (raw filter model), the distances were entered into a table which was then submitted directly to multidimensional scaling. In the second simulation (prototype model), the two points $p_1, p_2 \in R^{200}$ corresponding to the two class prototypes were first computed. The point $p \in R^{200}$ corresponding an arbitrary image was then represented as a vector in R^2 of the form $[d(p, p_1), d(p, p_2)]^T$, where $d(\cdot, \cdot)$ is the metric induced by the Euclidean norm in R^{200}. In the third simulation (random prototype model), the reference points p_1, p_2 were chosen randomly from the hypercube in R^{200} defined by the extreme values of the components of the 16 individual vectors there.

3.2 Results

The performance of the models was assessed by three different methods, and was compared to the human subject data, as described below.

Linear separability of the two classes in the MDS-derived configurations. In experiment 1 (animal shapes), the configuration derived by MDS from the prototype-model data consisted of two linearly separable clusters corresponding to the two object classes (see Figure 3). The clusters resulting from the data of the other two models were not linearly separable. The separation of the clusters in experiment 2 was less pronounced for the prototype model, and vanishing for the other two models, as well as for all three models in experiments 3 and 4. In the human data, linear separability was absent in experiment 4.

Classification based on the MDS-derived configurations. One of the important advantages of dimensionality reduction is the facilitation of learning from examples, which, in turn, allows the system to rely on simple classification methods such as the nearest neighbor (Duda and Hart, 1973). Assuming that the human visual system indeed maintains at some level of processing a low-dimensional representation similar to the 2D pattern obtained by MDS, it would be interesting to assess the utility of this representation for classifying the experimental stimuli. This was done by subjecting the configurations derived by MDS from the human data and from the models to discriminant analysis (SAS procedure DISCRIM (Sas, 1989)). The nearest-neighbor error rate found by this procedure is a non-parametric measure of performance, and is computed by a method related to cross-validation. To obtain the error rates plotted in Figure 4, each of the 16 points in a given configuration was left out successively in a round-robin fashion, and the classification error was averaged over the resulting 16 runs. Note that the dependence of the performance of the prototype model on shape family resembles that of the human subjects, but the error rate is somewhat higher, except in experiment 1. Thus, the 2D configuration recovered from the prototype model data can support classification rate approaching that of the human subjects.

A comparison between the pattern of RT data and the filter-space distances. The third analysis estimated the degree to which the pattern of RTs exhibited by the human subjects resembled the patterns of interstimulus distances produced by each of the models. The resemblance was defined as the canonical correlation between the RT tables of the human subjects (**H**) and the distance tables obtained with the prototype (**P**), random-prototype (**R**) and the raw-filter (**F**) models, and was computed using the SAS procedure CANCORR (Sas, 1989). Given two sets of variables, CANCORR computes a linear combination from each set (called a canonical variable) such that the correlation between the two canonical variables is maximized. This correlation is called the first canonical correlation. The procedure then continues by finding a second set of canonical variables, uncorrelated with the first pair, that produces the second highest canonical correlation, and so on.

This procedure was applied to the experimental data as follows. First, the RT values from each of the five subjects (the **H** variable sets, preprocessed according to the procedure described in section 2.1) and the interstimulus distances from $3 \times 5 = 15$ model runs (the **P**, **R**, and **F** variable sets) in every experiment were arranged in vectors of length 256 (corresponding to the 16×16 stimulus combinations). Next, the canonical correlations between sets **H** and **P**, **H** and **R**, and **H** and **F** were computed, separately for each experiment. In experiment 1, the first canonical correlation between the human and the prototype data (**H**-**P**) was $r = 0.90$ (significant at $p < 0.0001$; regression of **H** on **P** accounted for an average proportion $R^2 = 0.65$ of the data variance); the subsequent **H**-**P** canonical correlations in this experiment were not significant. The **H**-**R** correlation was $r = 0.38$ ($p < 0.0001$; $R^2 = 0.13$). The **H**-**F** correlation was even weaker: $r = 0.23$ ($p < 0.02$; $R^2 = 0.07$). In experiment 4, the **H**-**P** correlation was not significant.[3] Altogether, the canonical correlation analysis showed that interstimulus distances produced by the prototype model predicted well the pattern of response times exhibited by the human subjects in the animal-shapes experiment.

3.3 Discussion

The data produced by the models provide three controls that facilitate the interpretation of the psychophysical findings. First, if the dissimilarity between the raw filter responses in R^{200}

[3] Interestingly, in this experiment the first three canonical correlations for **H**-**F** were strong ($r_{1,2,3} = 0.86, 0.47, 0.43$, all with $p < 0.0004$; $R^2 = 0.71$), and so were the first 3 **H**-**R** correlations ($r_{1,2,3} = 0.76, 0.43, 0.34$, all with $p < 0.001$; $R^2 = 0.53$), as if the subjects had no stored prototypes to rely upon, and had to recourse to less structured and more error-prone representations.

had revealed the true configuration of the objective parameter space, one might conclude that configuration recovery via MDS is trivially easy and thus not informative. The meaningless configuration recovered by the raw-filter model with the animal-like stimuli indicates that this is not so. Second, the level of performance of the true-prototype model suggests that a system that is subject to biologically reasonable constraints on its architecture (i.e., the reliance on filters as the main computational mechanism) can maintain a faithful representational replica of the distal stimulus structure, provided that the representation resides in a sufficiently low-dimensional space. Third, the advantage of the true-prototype approach over the random-prototype one, as revealed in the simulation of experiment 1, underscores the importance of choosing a proper basis for that low-dimensional representation space.

4 Summary

The main conclusions to be drawn from the results of the psychophysical experiments and of the computational simulations may be illustrated with the help of the sketch of the process of dimensionality reduction in the primate visual pathway that appears in Figure 5. First, the presence in the subject response data of the low-dimensional layout of the stimulus space shows that this dimensionality reduction process is efficient and accurate: in a few dozens of trials, the human subjects abstracted the proper low-dimensional task space from the high-dimensional presentation space in which it was embedded, and formed a correct representation of the pattern of similarities among the stimuli.

Second, the computational simulations indicate, somewhat surprisingly, that this dimensionality reduction, which can be considered as detection of features that encode perceptually prominent shape contrasts, may be partially supported by convolution with a bank of filters, followed by comparison with a set of stored prototypical response patterns. The question marks appearing in Figure 5 in the places corresponding to the higher visual areas can, therefore, be replaced by a concrete computational model, which may be subjected to further testing.

The prototype model has a simple and natural implementation in terms of a hierarchical network employing receptive fields as the only required computational mechanism. Specifically, the stored prototypes can be considered as receptive fields selective for complex shapes, or, alternatively, as basis functions for shape, a notion that has parallels in some network theories of visual recognition (Poggio and Edelman, 1990; Poggio, 1990; Edelman, 1993), and is supported by neurobiological data on the structure of object representations in the inferotemporal cortex of the monkey (Fujita et al., 1992; Tanaka, 1992). A more extensive discussion of this approach to representation is beyond the scope of this paper, and may be found in (Edelman, 1993).

Varieties of multidimensional models of perception and cognition in which stimuli are encoded by distances to stored pattern distributions have been explored extensively in a number of recent works (Ashby and Perrin, 1988; Ashby, 1992). Connectionist implementations of categorization in multidimensional feature spaces (Kruschke, 1992; Hanson and Gluck, 1993) tend to resemble in their structure the basis function networks of (Poggio, 1990), as well as the prototype-based approach adopted here. The success of these models in accounting for a range of findings concerning human categorization is, therefore, encouraging.

The surprising performance of the prototype model raises the possibility that the shape space may be effectively spanned not by a small number of generic building blocks or geons, as proposed in (Biederman, 1987), but by filter activities corresponding to entire prototypical objects. It would be interesting to reconsider the experimental results usually quoted in support of the geon theory, to determine the extent of their compatibility with the prototype-based model. Such a comparative study should concentrate on the main differences between the two theories, namely, on the issues of binding and of atomistic vs. holistic representation. The experiments described here were designed to explore the multidimensional shape-space approaches to representation and not to test the geon theory. Their results, therefore, provide only an indirect evidence regarding the value of the geon approach, by showing that a considerably simpler method of representation of structure may be experimentally and computationally viable. Thus, rather than refuting the alternative theories, the present work suggests a synthesis of the three main approaches to shape representation — the pictorial, the structural, and the featural — in which the pictures are replaced by snapshots of filter responses, the structure is encoded by filter-space distance to several similar structures, and the feature values in the representation space reflect the metrics of the viewed shapes.

Acknowledgements

I thank Tomaso Poggio for timely encouragement, Shabtai Barash, Florin Cutzu, Yacov Hel-Or, Nathan Intrator, Shimon Ullman, and Yair Weiss for useful discussions, Edna Schechtman for statistical advice, and two anonymous reviewers for constructive comments. This research is supported by the National Fund for Basic Research, administered by the Israel Academy of Arts and Sciences. SE is an incumbent of the Sir Charles Clore Career Development Chair.

References

Ashby, F. G., editor (1992). *Multidimensional models of perception and cognition.* 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.

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.

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

Edelman, S. (1992). Class similarity and viewpoint invariance in the recognition of 3D objects. CS-TR 92-17, Weizmann Institute of Science.

Edelman, S. (1993). Representation, Similarity, and the Chorus of Prototypes. CS-TR 93-10, Weizmann Institute of Science. to appear in Minds and Machines, 1994.

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.

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

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

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

Intrator, N. and Gold, J. (1993). Three-dimensional object recognition in gray-level images: the usefulness of distinguishing features. *Neural Computation*, 5:61–74.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99(1):22–44.

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.

Ratcliff, R. (1981). Parallel processing mechanisms and processing of organized information in human memory. In Anderson, J. A. and Hinton, G. E., editors, *Parallel models of associative memory*. Erlbaum, Hillsdale, NJ.

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

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

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.

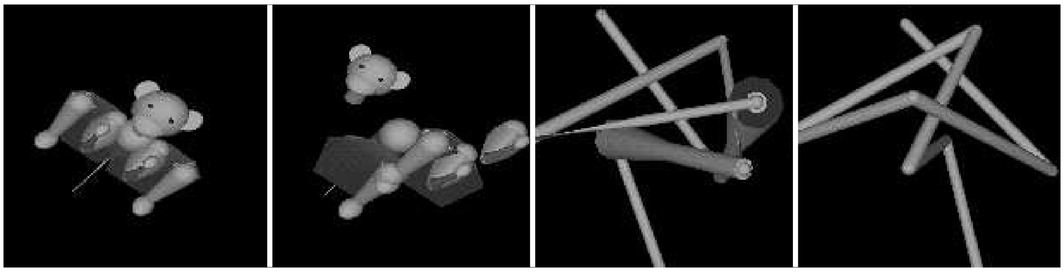
Tanaka, K. (1992). Inferotemporal cortex and higher visual functions. *Current Opinion in Neurobiology*, 2:502–505.

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

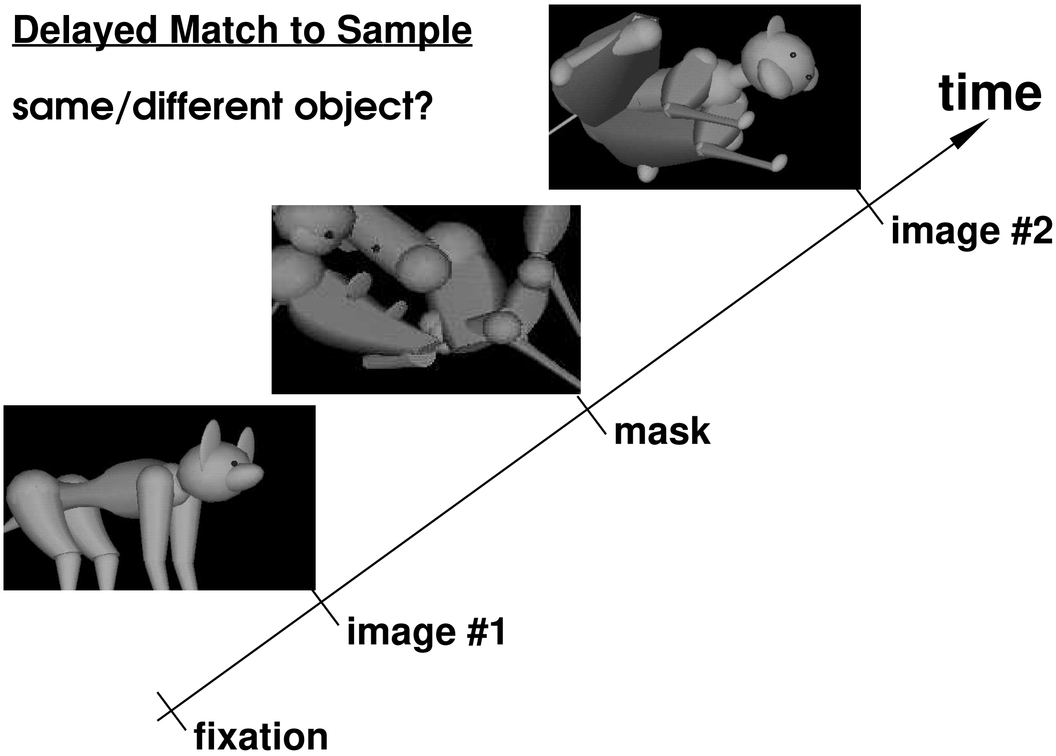
Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition*, 32:193–254.

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

Weiss, Y. and Edelman, S. (1993). Representation with receptive fields: gearing up for recognition. CS-TR 93-09, Weizmann Institute of Science.

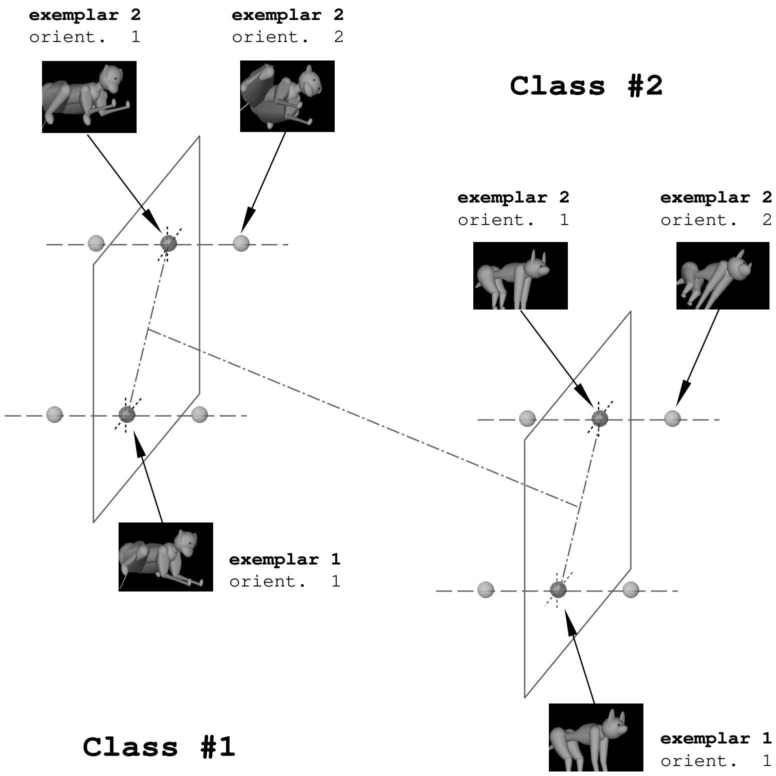


a

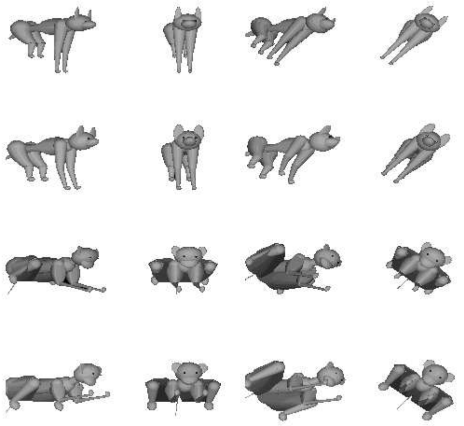


b

Figure 1: *a,* the four shape families used as stimuli in the experiments. Left to right: animals, scrambled animals (obtained by randomly changing the relative location of parts of the animal shapes), geon wires (concatenations of seven elongated distinctive 3D parts), plain wires (concatenations of seven cylindrical parts). In each experiment, the two prototypical shape classes were obtained using the parameter space procedure described in Figure 2. The parameter space was 56-dimensional in the first two cases, and 24-dimensional (3 coordinates for each of the 8 vertices of a wire) in the last two cases. *b,* the time course of a delayed matching to sample trial.



a



b

Figure 2: *a,* The layout of the stimuli in the parameter space (see section 2.1 for details). Here, the stimuli are from experiment 1, in which two computer-generated object classes, one resembling monkeys, and the other dogs, were jointly parameterized by a set of 56 variables. *b,* All 16 images of the animal-like shapes resulting from the above procedure.

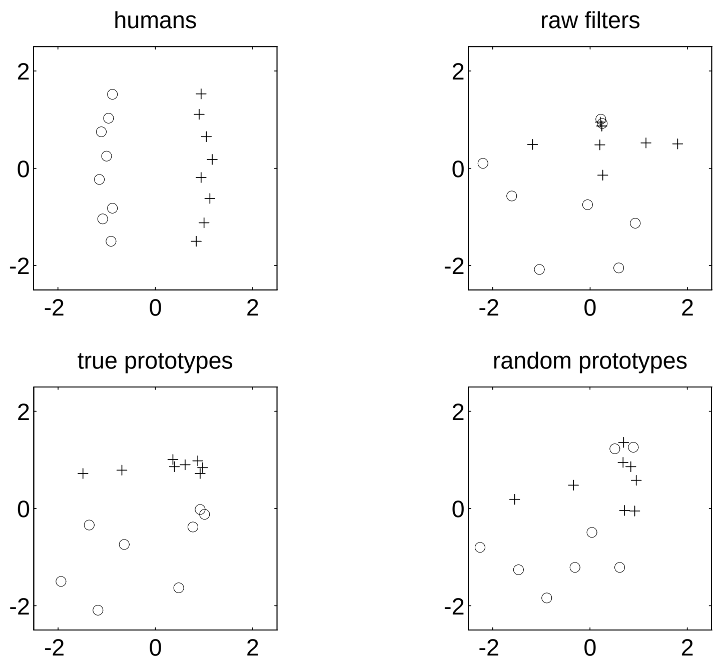


Figure 3a: experiment 1 (animal shapes).

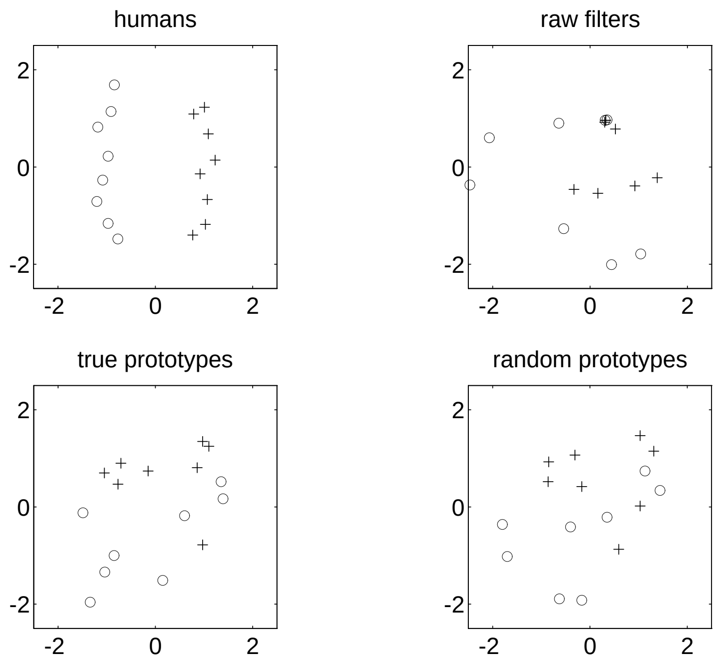


Figure 3b: experiment 2 (scrambled animal shapes).

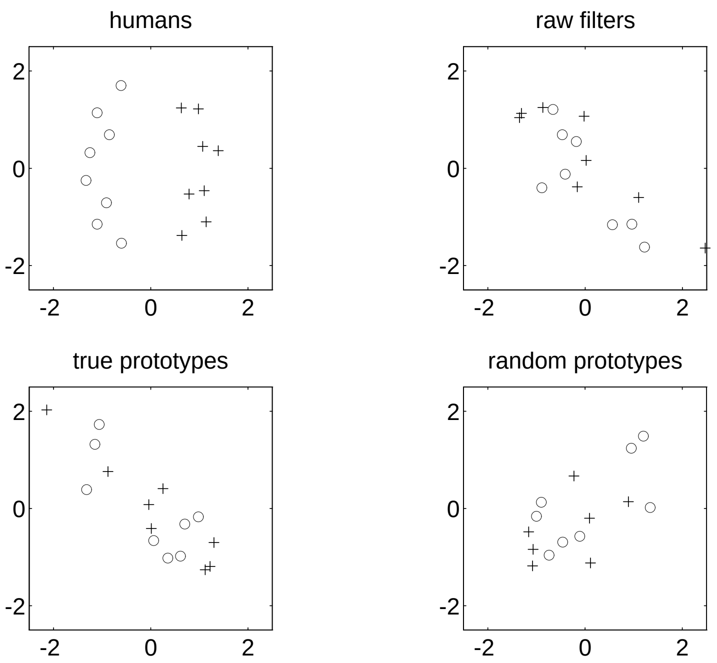


Figure 3c: experiment 3 (geon wire shapes).

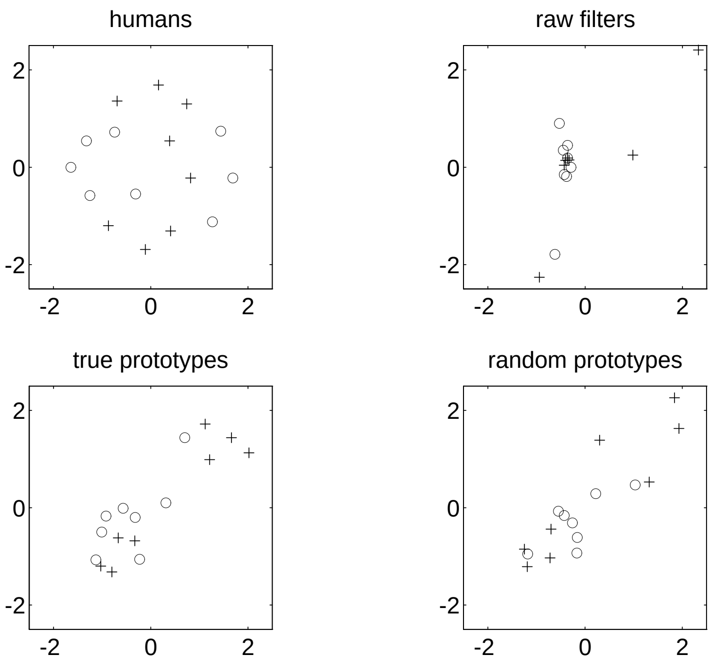


Figure 3d: experiment 4 (plain wire shapes).

Figure 3: (preceding page).

a, Experiment 1 (animal shapes); the 2D configurations recovered by the SAS MDS procedure (Sas, 1989) (a scree test (Kruskal and Wish, 1978) here and in the other experiments showed that using more than two dimensions was not warranted). Points belonging to the two object classes are marked by + and ∘, respectively. The human data are from 5 subjects. Data for each of the three models were obtained by repeating 5 times the basic simulation run in which the receptive fields of the 200 filters were randomly repositioned with respect to the image. The MDS stress and the distance correlations were: for humans 0.30, 0.80; for the prototype model 0.19, 0.93; for the random-prototype model 0.22, 0.90; for the raw-filter model 0.24, 0.88.

b, Experiment 2 (scrambled animal shapes). The 5 human subjects here were different from the participants of experiment 1. The MDS stress and the distance correlations were: for humans 0.27, 0.82; for the prototype model 0.22, 0.89; for the random-prototype model 0.25, 0.82; for the raw-filter model 0.22, 0.91.

c, Experiment 3 (geon wire shapes). The 5 human subjects here were different from the participants in experiment 4. The MDS stress and the distance correlations were: for humans 0.29, 0.67; for the prototype model 0.11, 0.98; for the random-prototype model 0.13, 0.97; for the raw-filter model 0.15, 0.96.

d, Experiment 4 (plain wire shapes). To facilitate the estimation of the effect of shape class, data from the same 5 human subjects who participated in experiment 1, but not in experiment 3, were used here (a subsequent analysis of data from 10 additional subjects, six from experiment 1 and four from experiment 4, yielded in all respects the same pattern of performance). The MDS stress and the distance correlations were: for humans 0.24, 0.29; for the prototype model 0.14, 0.97; for the random-prototype model 0.13, 0.97; for the raw-filter model 0.08, 0.99.

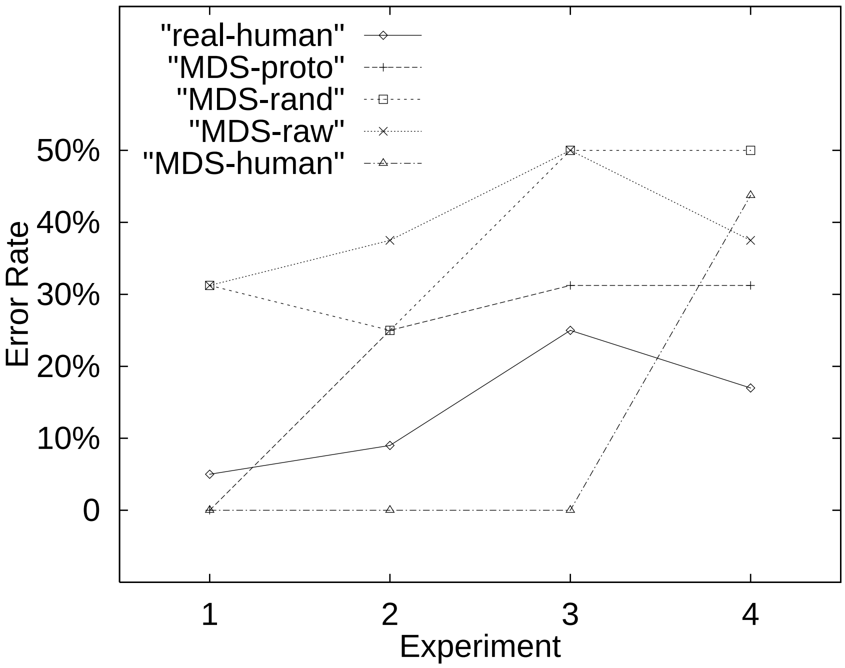


Figure 4: True error rate of the human subjects, plotted along with the performance of a nearest-neighbor classifier applied to the configuration derived by MDS for each of the models, and for the human RT data (see section 3.2).

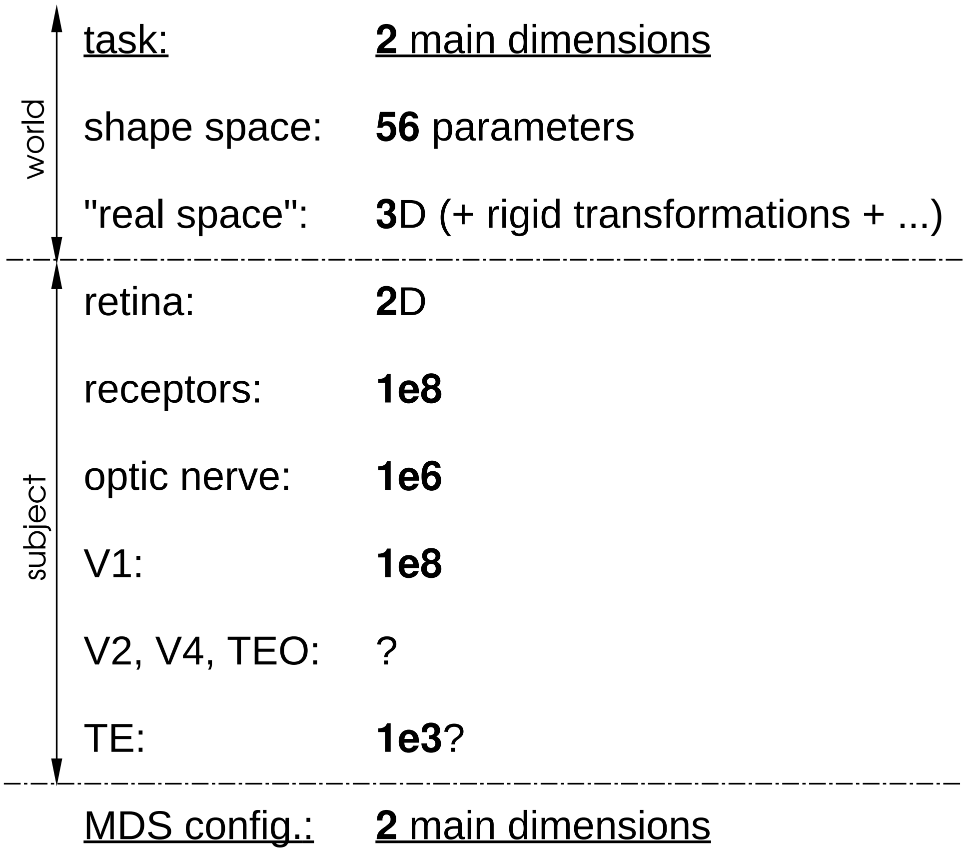


Figure 5: The outcome of experiment 1, from the point of view of dimensionality of the successive representations presumably involved in the processing of the stimuli by the human visual system.