SCISPACE
formerly Typeset

Open access • Journal Article • DOI:10.1038/NN767

Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. — Source link

Hans Op de Beeck, Johan Wagemans, Rufin Vogels

Institutions: Katholieke Universiteit Leuven

Published on: 01 Dec 2001 - Nature Neuroscience (Nature Publishing Group)

Related papers:

- Visual categorization shapes feature selectivity in the primate temporal cortex
- Inferotemporal cortex and object vision.
- Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex
- Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex
- Hierarchical models of object recognition in cortex

Share this paper:

View more about this paper here: https://typeset.io/papers/inferotemporal-neurons-represent-low-dimensional-85crzy5my3

© 2001 Nature Publishing Group http://neurosci.nature.com

Inferotemporal neurons represent low-dimensional configurations of parameterized shapes

Hans Op de Beeck[1,2], Johan Wagemans[2] and Rufin Vogels[1]

[1] *Laboratorium voor Neuro- en Psychofysiologie, K.U. Leuven, Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium*

[2] *Laboratory of Experimental Psychology, K.U. Leuven, Tiensetraat 102, B-3000 Leuven, Belgium*

Correspondence should be addressed to R.V. (rufin.vogels@med.kuleuven.ac.be)

Published online: 19 November 2001, DOI: 10.1038/nn767

Behavioral studies with parameterized shapes have shown that the similarities among these complex stimuli can be represented using a low number of dimensions. Using psychophysical measurements and single-cell recordings in macaque inferotemporal (IT) cortex, we found an agreement between low-dimensional parametric configurations of shapes and the representation of shape similarity at the behavioral and neuronal level. The shape configurations, computed from both the perceived and neuron-based similarities, revealed a low number of dimensions and contained the same stimulus order as the parametric configurations. However, at a metric level, the behavioral and neural representations deviated consistently from the parametric configurations. These findings suggest an ordinally faithful but metrically biased representation of shape similarity in IT.

© 2001 Nature Publishing Group http://neurosci.nature.com

The capacity to categorize stimuli is fundamental to all living organisms[1,2]. Theories of categorization agree upon the importance of the similarity between stimuli to account for many aspects of categorization performance[3–5]. However, it is not straightforward to compute the degree of similarity between stimuli that can vary across a high number of dimensions, like complex shapes. Fortunately, the similarities among a set of complex stimuli can often be described in a more compact way[6–8]. Indeed, stimuli from many behaviorally relevant sets can be represented in a low-dimensional representation space in which the proximity between stimuli is related to their similarity. For example, by presenting the randomly ordered shapes of Fig. 1d in a particular order (**Fig. 1a–c**), the similarities can be easily described by a two-dimensional square-like configuration. Several behavioral studies that have varied complex shape differences parametrically revealed that primates are able to represent the similarities between shapes in a low-dimensional representation space without ever seeing these stimuli in their parametric configuration[9–12].

Here we aim to study directly the neural basis of these low-dimensional representation spaces. Object recognition and categorization in macaques is thought to depend on the inferotemporal cortex (IT)[13,14]. Single IT neurons are selective for moderately complex object features[15], but several studies have found little relationship between the similarities between complex objects and the responses of single IT neurons[16,17]. However, one needs to manipulate shape similarity parametrically to investigate how the responses of IT neurons to complex stimuli are related to the proximity of these stimuli in a low-dimensional space. Thus, we investigated whether the response pattern across a population of IT neurons can reveal a low-dimensional and faithful representation of shape similarity using parameterized shapes.

As the analysis of the visual input in the visual system is highly nonlinear, the neuronal representation space could deviate from the configurations in parameter space in several ways. Previous psychophysical studies found an ordinal agreement between parametric configurations and their perceptual representation (the same number of dimensions and the same stimulus order)[9–12], so we also expected this result at the neuronal level. However, configurations that fit perfectly on an ordinal level can differ on a metric level. Thus, apart from investigating ordinal relationships, we looked for consistent metric anomalies in perceptual and neuronal representation spaces with respect to the parametric similarities.

In this study, we varied complex shape dimensions, that is, radial frequency components (RFCs), that have been used previously with human subjects[10]. These parametric variations affect a high number of perceptually salient dimensions[10,18], but are not coded directly in the macaque visual system[19]. Nevertheless, in agreement with computational work[11], we found that human and monkey subjects were able to represent the similarities between such shapes in low-dimensional representation spaces that agree well with pixel-based configurations (**Fig. 2a–e**). Moreover, the perceived similarities of the monkeys corresponded better with the stimulus similarities when these were computed using IT neuron responses (**Fig. 2f**) than with the pixel-based similarities. Further behavioral experiments in monkeys showed that the low-dimensional stimulus configuration predicts categorization performance.

Results

We investigated the underlying representation space of 24 shapes that were divided into 3 groups (**Fig. 1**). The within-group con-

© 2001 Nature Publishing Group http://neurosci.nature.com

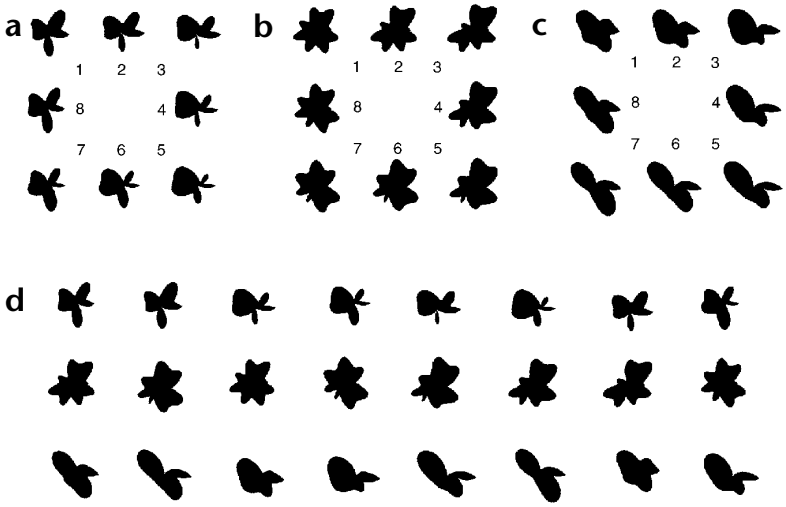

Fig. 1. Visual stimuli. (**a–c**) Three groups of eight shapes were used. Within each stimulus group, the parameter–space configuration of the stimuli is represented by the square arrangement of the stimuli. The top-left stimulus in each square has a low amplitude value for both manipulated radial frequency components. (**d**) Same 24 stimuli, but the 8 shapes from each stimulus group are presented in a random order in a row.

figurations formed two-dimensional squares in the parametric space, and similar but not identical configurations are present when using pixel-based similarities (**Fig. 2a**).

Perceptual representation space: human data

We verified the perceptual representation of the similarities between shapes in two human subjects performing similarity ratings. First, we applied hierarchical cluster analysis to determine whether these similarity ratings reflected the expected clustering of stimuli into the three groups. As expected, all stimuli from a given stimulus group were assigned to the same cluster before they were clustered with stimuli from the other two groups. A substantial proportion of variance (85%) in the similarity judgments was accounted for when reducing the 24 stimuli to these 3 clusters.

Next, we analyzed the similarities among the eight shapes from each group. To determine whether the perceived similarities converge with the pixel-based similarities, we computed the congruence *C* between these two sets of similarities. For all groups of stimuli and for both subjects, significantly high congruences were found (**Table 1**).

We used multidimensional scaling (MDS) to determine whether the shape similarities could be captured by the distance between the stimuli in a low-dimensional space. Two-dimensional MDS-derived configurations were computed for each subject and each group of stimuli separately (**Fig. 2b** and **c**). The more similar the two stimuli, the smaller the distance between the points representing each stimulus. Even with as few as two dimensions, these configurations accounted for most of the variance in the similarity ratings. All within-group similarities were represented in low-dimensional configurations that strongly resembled the expected configurations. First, no qualitative deviations, such as the reversal of a pair of shapes, were seen. We quantified this ordinal agreement between perceived and expected configurations by computing the Spearman rank order correlation (r_s) of the polar angles of the stimulus points with respect to the center of each configuration. If one takes stimulus 1 as the reference with polar angle zero and proceeds clockwise, then one would expect stimulus 2 to have the next-smallest polar angle, followed by stimulus 3. The expected order of the stimuli was preserved perfectly in all stimulus groups ($r_s = 1.0$, $p < 0.01$). Second, details of the expected configurations, such as their 'squareness,' were maintained. In a square, the angle between two lines connecting a point with its two neighboring points is 90° at corner points, and 180° at points along the sides. Other configurations with the same stimulus order, such as eight equidistant points on a circle, do not possess this property. A chi-square test was used to assess the dependence of angle size (smaller or larger than 135°) and type of point (corner or side point in the parametric configuration) for each subject. Angles were greater at side points compared to corner points for each subject ($\chi^2 = 10.74$, $p < 0.01$ for subject 1; $\chi^2 = 20.17$, $p < 0.01$ for subject 2). Thus, the configurations of each subject differentiate between configurations of the same stimulus order, such as a square or a circle.

To determine whether remaining metric deviations from the expected configurations were systematic rather than due to measurement noise, we determined the inter-subject consistency of deviations of the perceived similarities with respect to the pixel-based similarities. For group A and B stimuli, *C* was lower when the perceived similarity ratings were compared between the two subjects (**Table 1**, *C*(human 1, human 2)) than when each subject's data were compared with the pixel-based similarities (*C*(pixel, human 1) and *C*(pixel, human 2)). This suggests that the deviations of the perceived similarities were likely due to measurement noise. However, there was a striking consistency in the way the perceived similarities of the two subjects deviated from the pixel-based similarities for group C stimuli (compare length and direction of red lines in **Fig. 2b** and **c**).

Perceptual representation space: monkey data

Two monkeys (Y and E) were trained in a same-different task. Following previous studies in macaques[12,20], the proportion of correct 'different' responses for a particular stimulus pair was taken as measure of the dissimilarity of that pair. This measure can be subjected to the same analyses as performed on the human data.

First, hierarchical cluster analysis showed that the most similar stimuli belong to the same stimulus group. All stimuli from a given stimulus group were assigned to the same cluster, before they were clustered with stimuli from the other two groups. However, the clustering in monkeys was less than in human subjects. The proportion of variance accounted for by reducing the 24 stimuli to 3 clusters was smaller for the monkey data (averaged across monkeys, 47% of variance explained; humans, 85%).

Second, analysis of the similarity data for each stimulus group separately showed that the congruence with the pixel-based similarities was always highly significant (**Table 1**). Two-dimensional MDS-derived configurations (**Fig. 2d** and **e**) explained most of the variance in the similarity ratings (averaged across monkeys, 96%, 88% and 97% for group A, B and C, respectively). Overall, the within-group similarities were represented in low-dimensional configurations that captured most aspects of the expected configurations. First, the stimulus order matched the expected order ($r_s = 1$, $p < 0.01$) for all stimulus groups in each monkey, except for group B in monkey Y (reversal of stimuli 6

© 2001 Nature Publishing Group http://neurosci.nature.com

© 2001 Nature Publishing Group http://neurosci.nature.com

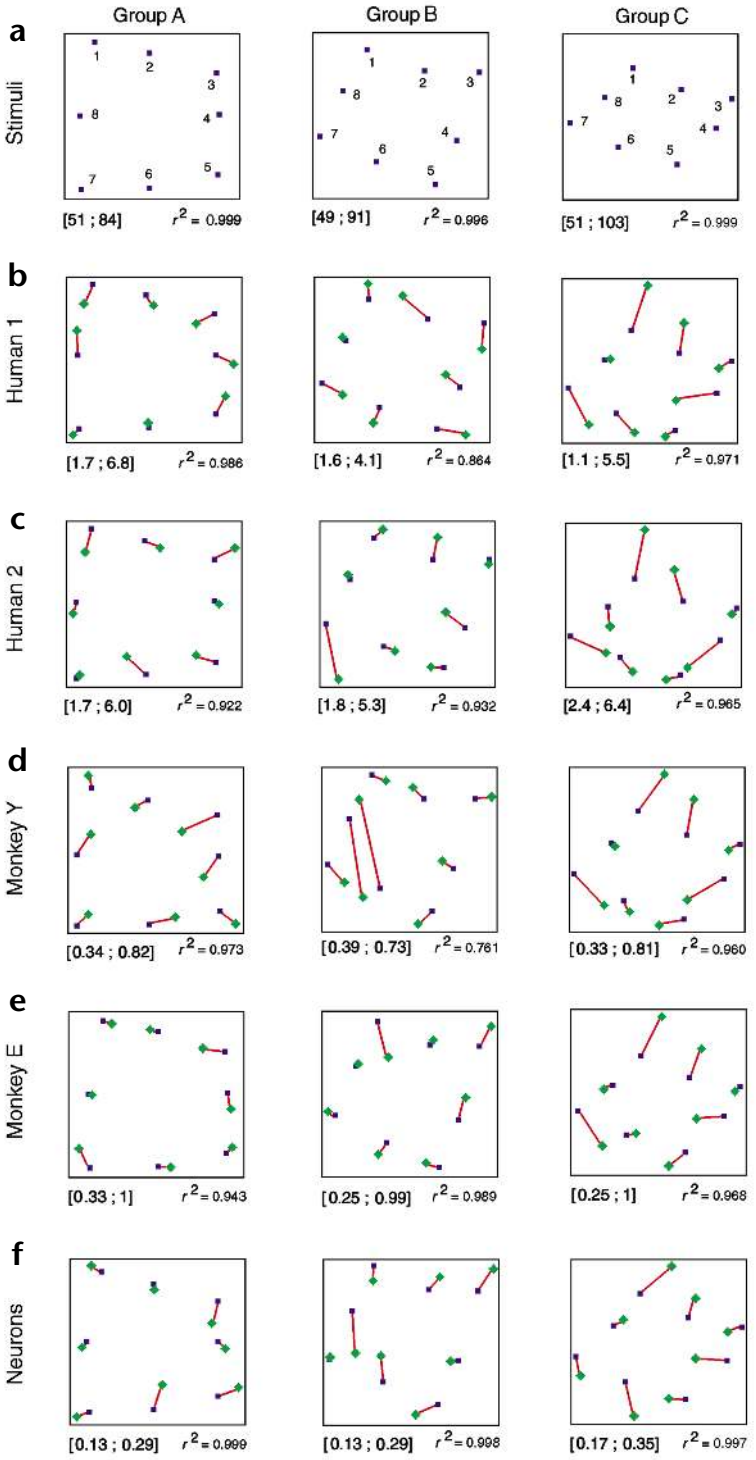

Fig. 2. MDS-derived two-dimensional configurations of stimuli. (**a**) Configurations based on the pixel-based similarity between stimuli. The numbers refer to the stimuli as labeled in Fig. 1. The other panels show a comparison between the configurations (blue squares) of stimuli in (**a**) and the configurations (green diamonds) found for the perceived similarities of the first (**b**) and second (**c**) human subject, the perceived similarities of monkey Y (**d**) and monkey E (**e**), and the neuron-based similarities (**f**). The amount of variance (r^2) of the similarity values that was explained by the two-dimensional configuration is denoted below each panel, as is the range ([minimum; maximum]) of similarity values. These ranges are expressed in different scales for each dataset and can only be compared among stimulus groups within each dataset (pixels, Euclidean distance; human, mean similarity rankings; monkey, proportion different responses; neuronal data, distance in 124-dimensional neuronal space). To aid a visual comparison of the pixel-based with the other configurations, the latter were Procrustes transformed (combination of translation, scaling, rotation and reflection)[11]. Red lines connect corresponding points in both configurations. As the configurations have arbitrary origin, scale and orientation, the labeling and scale on both axes in each graph are omitted.

and 8 ($r_s = 0.93$, $p < 0.01$)). The latter deviation was not a consequence of a poor reliability, as it was replicated. Second, there was a clear difference in the MDS-derived configurations between corner and side points as defined in the expected configurations, with larger angles at side points compared to corner points ($\chi^2 = 8.17$, $p < 0.01$ for monkey Y; $\chi^2 = 10.74$, $p < 0.01$ for monkey E).

We determined whether the remaining deviations of the perceived from the pixel-based similarities were consistent between the two monkeys. There was a significantly high congruence between the data of monkeys Y and E for stimulus groups A and C, although no such inter-subject consistency was present for group B (**Table 1**, *C*(monkey Y, monkey E)). Each monkey tended to represent stimulus 3 of group A more toward the center of the configuration than expected from the pixel-based configurations (**Fig. 2d** and **e**). For group C, the perceived shape configurations showed a bias towards the vertical parametric dimension (**Fig. 1c**), a trend also present in human subjects (**Fig. 2b** and **c**). This consistency between humans and monkeys for group C was assessed using the averaged human and monkey data, and was significant (**Table 1**, *C*(humans, monkeys); **Supplementary Fig. 1**, see supplementary information page of *Nature Neuroscience* on line for the application of MDS on these averaged data).

Neuronal representation space in IT cortex

We recorded from 124 single neurons in area TE of monkeys F ($n = 51$) and M ($n = 73$) while they were learning to categorize the 24 stimuli in 2 classes. Recordings were performed in the lower bank of the superior temporal sulcus and lateral to the anterior middle temporal sulcus. Most neurons (82%) responded twice as strongly to their preferred shape than to the least preferred shape. Many of these neurons combined high within-group selectivity with lower between-group selectivity. For example, the neuron in Fig. 3a responded differently to similar shapes within each stimulus group (for example, shapes C1 and C3), whereas it responded with similar strength to shapes that look dissimilar to human and monkey observers (for example, shapes C1 and A6). Many neurons were very selective within stimulus groups, but stimuli from different groups often elicited similar responses (**Fig. 3b**). Indeed, the median of the distribution of the selectivity index (SE, see Methods; **Fig. 4a**) was 2.4, indicating that most neurons do not respond more strongly to all stimuli of one group compared to the responses to stimuli of the other groups (as can

© 2001 Nature Publishing Group http://neurosci.nature.com

© 2001 Nature Publishing Group http://neurosci.nature.com

Table 1. Congruence between configurations for each group of stimuli.

	Group A	Group B	Group C
Congruence between pixel-based similarities and empirical similarities			
C (pixels, human 1)	0.973+	0.987+	0.960+
C (pixels, human 2)	0.989+	0.987+	0.976+
C (pixels, humans)	0.991+	0.995+	0.974+
C (pixels, monkey Y)	0.979+	0.984+	0.979+
C (pixels, monkey E)	0.988+	0.983+	0.976+
C (pixels, monkeys)	0.989+	0.990+	0.979+
C (pixels, neurons)	0.996+	0.993+	0.991+
Congruence between two sets of empirical data			
C (human 1, human 2)	0.967	0.976	0.981*
C (monkey Y, monkey E)	0.985*	0.972	0.990*
C (humans, monkeys)	0.987	0.988	0.992*
C (monkeys, neurons)	0.990*	0.991*	0.984*

+Probability of random configuration, $p < 0.001$. *Probability that deviations from pixel-based similarities are random, $p < 0.05$.

be illustrated by sorting the stimuli according to response strength, **Supplementary Figs. 1** and 2).

The within-category selectivity of these neurons was regular, meaning that their responses decreased monotonically with increasing parametric distance from the preferred shape of a stimulus group. For example, the neuron in Fig. 3a responded well to shape C1 and was still responsive for the neighboring shapes C2 and C8, but no comparable responses were elicited at any point further in stimulus space. As a consequence, polar plots for each stimulus group revealed regular and unimodal tuning curves (**Fig. 3b**). The ratio index (RR) captures the deviation of this tuning curve with respect to an unimodal sinusoidal modulation (see Methods). The three neurons in Fig. 3b illustrate tuning curves across the whole distribution of RR values (**Fig. 4b**). For all cases with a good within-category selectivity (that is, a response difference between preferred and least preferred shape above 50%, $n = 121$), the median RR index was 1.9. Thus, the responses of most IT neurons were closely related to the parametric variation of shape similarity.

We used the responses of the 124 neurons to compute the neuron-based similarity between each pair of stimuli. Notwithstanding the low between-group selectivity of many neurons, hierarchical cluster analysis of the neuron-based similarities revealed the expected clustering at the population level: all stimuli from a group were

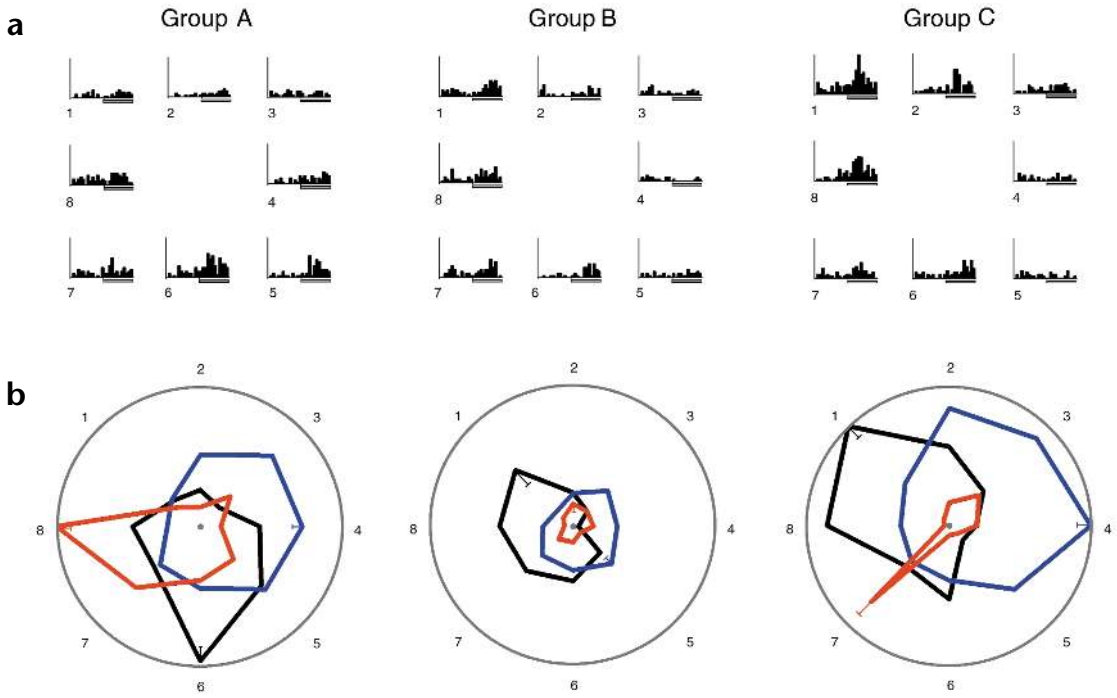

Fig. 3. Responses of single IT neurons. (**a**) Peristimulus time histograms of a single IT neuron from monkey M. The ordering and numbering of the panels is the same as in Fig. 1. Stimulus presentation (300 ms) is indicated by the bar underneath each histogram. The histograms have a bin width of 25 ms and the height of each bin is normalized to the maximum bin across all histograms (94 spikes/s). (**b**) Polar plots of the within-group response pattern, making use of the radial position of each stimulus with respect to the center of the parametric square configuration (numbering of shapes is the same as in **a**). The black tuning curves represent the responses of the neuron of (**a**), whereas the responses of two other single neurons (monkey F) are shown in red and blue. Responses across all three panels are normalized to the maximum response for each neuron separately (maximum responses were 27, 59 and 34 spikes/s for the black, blue and red curves, respectively). The standard error of the mean for the maximum of each tuning curve is indicated. The RR indices for stimulus groups A, B and C were 1.8, 3.6 and 2.8, respectively, for the neuron indicated by black lines (SE = 27); 5.2, 3.0 and 4.8 for the neuron indicated by blue lines (SE = 2.3); and 1.6, 0.68 and 0.33 for the neuron indicated by red lines (SE = 3.8).

© 2001 Nature Publishing Group http://neurosci.nature.com

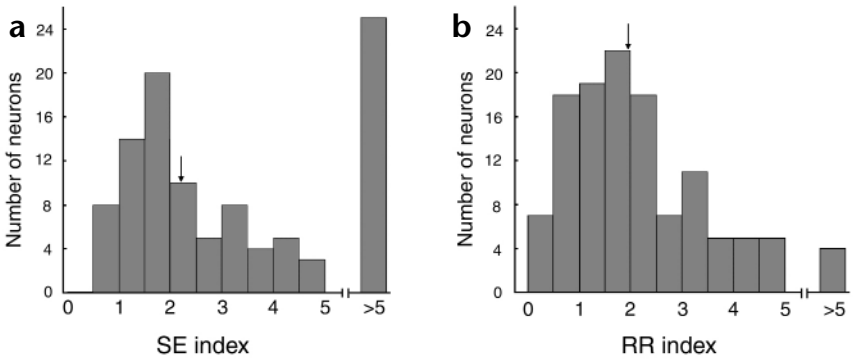

Fig 4. Distribution of SE and RR indices. (**a**) The SE index is calculated for all neurons responding with at least twice as many spikes to their preferred shape than to the least preferred shape (n = 102). (**b**) The RR index is computed for all cases with a good within-category selectivity (that is, a response difference between preferred and least preferred shape above 50%, n = 121). All values above five are collapsed in one column (maximum was 38 and 14 for SE and RR, respectively). The median of each distribution is indicated by an arrow.

assigned to the same cluster, before they were clustered with stimuli from other groups. Because 45% of the variance was explained by assigning the stimuli to 3 groups, the degree of clustering at the neuronal level is comparable to that found in the monkey behavioral data, but less so than in human subjects.

For each stimulus group, the congruence between the neuron-based similarities and the pixel-based similarities was highly significant (**Table 1**). The proportion of variance explained by 2-dimensional MDS-derived configurations (**Fig. 2f**) exceeded 99% for each group. As found for the perceived similarities, the positions of the eight shapes in these configurations captured most aspects of the expected configurations. First, we found the same stimulus order in the neuron-based configurations compared to the expected configurations for stimulus groups A and C ($r_s = 1$, $p < 0.01$), and only a small deviation for group B ($r_s = 0.98$, $p < 0.01$). Second, the MDS-derived configurations were square-like, with larger angles at side points compared to corner points ($\chi^2 = 8.17$, $p < 0.01$).

Reasonable fits could already be obtained with smaller samples of neurons. The mean congruence between neuron- and pixel-based similarities was significant with as few as eight neurons (**Fig. 5a**). The congruence depends also on the tuning properties of the neurons in addition to the effect of sample size (**Fig. 5b**). First, congruence was lower for a sample of neurons with an irregular tuning (RR < median RR) compared to neurons with a regular tuning (RR > median RR). Second, restricting the range of preferred shapes (tuning optima) reduced congruence when combining at least four neurons. Thus, both the regularity of the tuning of single neurons and the range of their preferred shapes contribute to the efficiency of the coding of shape similarity at the population level.

The neuron-based similarities seem to faithfully represent many metric aspects of the perceived similarities of the monkey. Indeed, the congruence between the behavioral and neuronal data was greater than would be expected based on their respective congruences with the pixel-based similarities for all stimulus groups (**Table 1**; C(monkeys, neurons)). The neuron-based similarities were a better predictor of the averaged perceived similarities than were the pixel-based similarities (the same result was obtained with the behavioral data considered for each monkey separately). The neural as well as the perceptual representation of stimulus A3 tended more toward the center of the configuration than expected from the pixel-based configurations. For group B, both the averaged perceived similarities and the neuron-based configurations tend to differentiate the stimuli on the right of the parametric configuration more strongly than those on the left (**Fig. 1b**; although the behavior-based representation of the stimuli on the left is not consistent between monkeys, each animal was more sensitive for differences between stimuli on the right compared to stimuli on the left, as was the population of neurons). The results for group C indicate that the neurons tend to be more sensitive for differences along the vertical parametric dimension, as were the animals and human subjects at the behavioral level.

Effects of categorization on neuronal representation

During the recordings, the monkeys learned to categorize the shapes from each stimulus group into two classes (**Fig. 6**). In some stimulus groups, the division of the stimulus space into specific regions followed a simple linear decision rule (group C and A in monkeys F and M, respectively). In other groups, however, the categorization rule required the stimulus space to be parceled into specific quadrants (group B in monkey M) or into arbitrary decision regions (groups A and B in monkey F and group C in monkey M). In this last category rule, highly similar shapes require a different response.

A possible effect of categorizing the images during the recordings on the neural responses was examined by comparing the neuronal data between monkeys for each stimulus group separately. Indeed, each monkey learned a different rule for a particular group, allowing a comparison between the neural

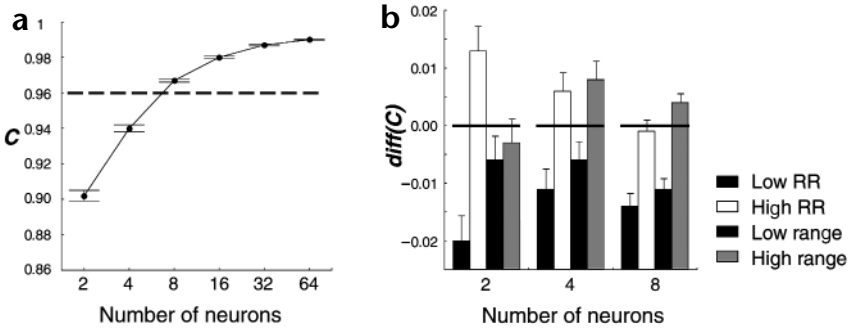

Fig. 5. Congruence (C) between pixel-based and neuron-based similarities as a function of number and type of neurons. Each data point represents the mean congruence of 150 random samples (whiskers indicate the standard error of the mean). (**a**) Congruence as a function of sample size. The ordinate scale starts from the expected value when the pixel-based similarities are compared with the similarities in random configurations (C = 0.86). The congruence of an individual sample is significantly different from this expected value for all values above the dashed line (C = 0.96). (**b**) Difference between the congruence from (**a**) and the congruence obtained when neurons are selected from subsets of the population. A positive difference corresponds to a higher congruence relative to that shown in (**a**). RR, ratio index.

© 2001 Nature Publishing Group http://neurosci.nature.com

© 2001 Nature Publishing Group http://neurosci.nature.com

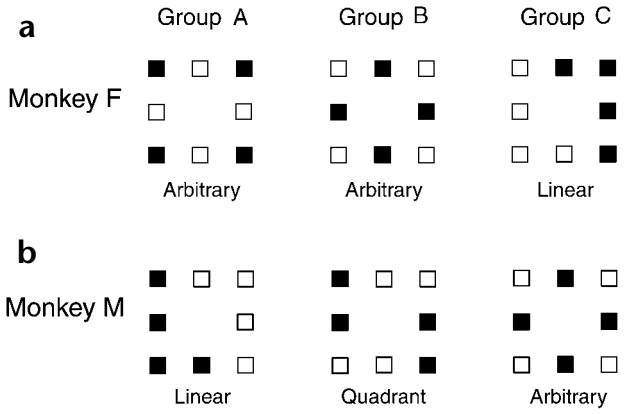

Fig. 6. The division of each stimulus group into two response categories for each monkey. Filled and open squares refer to stimuli that were associated with a leftward or a rightward eye movement, respectively. The stimulus order is the same as in Fig. 1.

similarity spaces for different category rules. We observed no systematic effects of categorization. First, the within-category selectivity of single neurons was regular with respect to the parametric similarity between stimuli and it did not follow the category rule. For example, when a monkey had learned an arbitrary category rule, no neurons responded stronger to all stimuli from one category compared to the responses to the other category. Second, arbitrary rules did not change the dimensionality or the order of shapes within the MDS-derived configurations. Third, even more subtle metric changes, like expansions/contractions of dimensions or parts of the stimulus space, were not induced by the categorization task.

The effect of similarity on categorization performance

If stimulus similarity determines categorization performance, then several predictions can be made regarding the difficulty of learning the different category rules. First, the more tightly stimuli from a given category are clustered (that is, the larger the inter-category distance relative to the intra-category distance), the easier it will be to categorize any given stimulus. Second, the more closely a stimulus is located to a boundary between two categories, the more difficult it will be to assign it to one or the other of those categories.

The errors the monkeys made while learning these categorization problems confirmed both predictions. First, the more tightly that shapes having to be categorized into the same class are clustered, the more easily a particular categorization was learned (**Fig. 7a**). Linear rules were learned more easily than a quadrant rule, but even the latter was more easily learned than the arbitrary rules. Second, in the case of a linear rule, the category assignments of shapes that were close to a category boundary were learned more slowly compared to more distantly located shapes (**Fig. 7b**). Nevertheless, performance on the stimuli close to a linear category border was above 90% after one week of training. This shows that the monkeys were able to discriminate neighboring stimuli in the parametric space. So, the difficulties with the arbitrary rules (performance after several weeks of training lower than 85%, **Fig. 7a**) were not merely due to stimulus discriminability, but indeed reflect the clustering of the stimuli within the representation space.

Discussion

Our data reveal a rather faithful representation of the physical similarities among high-dimensional stimuli in monkeys at the neuronal level. The neuronal representation of a configuration of shapes preserves the dimensionality and stimulus order of the configuration, and even some metric properties like the difference between a square and circle configuration. The response patterns of single neurons were related to the distance between shapes within these low-dimensional representation spaces, a property that contributes to the efficient coding of similarity at the population level. Notwithstanding the close agreement between physical and neuron-based similarities, our results revealed that the neuron-based similarities were a better predictor of perceived similarities than were the physical similarities. Finally, the low-dimensional representation of shape similarity determined the difficulty of learning specific categorization rules, but the reverse was not the case: applying a categorization rule did not change the representation of similarity in IT.

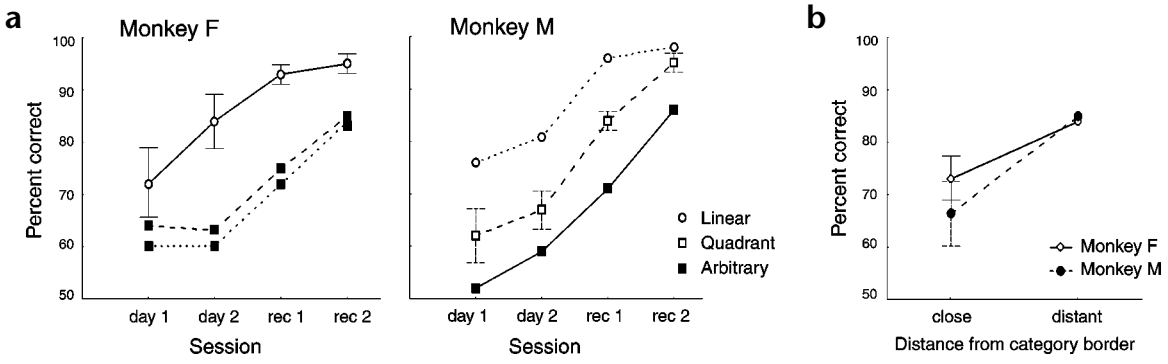

Fig. 7. The performance of the two monkeys in the categorization task for each stimulus group. The stimulus–response associations for each group and monkey are shown in Fig. 4. (**a**) Performance averaged across all stimuli from each stimulus group for different sessions: day 1 and 2 (performance at the end of the first and second session, respectively), and rec 1 and 2 (behavioral performance during the first and second half of the recording sessions, respectively). Point style represents the type of category rule, whereas the line style refers to the different stimulus groups (group A, dotted line; B, dashed line; C, solid line). Vertical bars indicate 95% confidence intervals of one line plot (the number of observations is equal for the other stimulus groups). (**b**) Performance for the linear categorization problems in the first two sessions as a function of the distance between each stimulus and the optimal decision boundary in parametric space. Pairwise close stimuli are C1, C2, and C5, C6 for monkey F, and A1, A2, and A5, A6 for monkey M. Vertical bars indicate 95% confidence intervals for performance with close stimuli in each monkey.

© 2001 Nature Publishing Group http://neurosci.nature.com

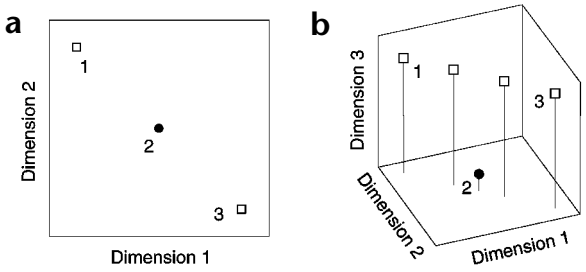

Fig. 8. Possible responses of a hypothetical neuron. The stimuli to which the neuron responds are indicated with open squares; black circles indicate no response. (**a**) A representation of three stimuli within two dimensions. (**b**) Including a parametric variation of the similarity between stimuli 1 and 3 requires the addition of a third dimension for all similarities between stimuli to be captured.

Previous behavioral studies have already shown that humans and monkeys represent parameterized two- and three-dimensional shapes in a manner that preserves relative similarities among the stimuli in parameter space (refs. 9–12 and Sigala *et al.*, *Soc. Neurosci. Abstr.* **26**, 448.10, 2000). Computational work[21] has revealed that such representation of similarity can be simulated by a network composed of units that encode shapes by computing their similarity to reference shapes ('radial basis functions'[22,23]). Our data indicate that shape similarity is implemented at the neuronal level in a similar way. First, we noted responses of single neurons within a single stimulus group that follow the pattern expected of radial basis functions: the responses to other shapes than the most optimal shape of a group (the 'reference shape') decreased gradually with increasing distance from the position of the reference shape in parameter space. The responses of single neurons across different stimulus groups seemed to contradict this conclusion, insofar as many neurons responded with similar strength to dissimilar shapes from different stimulus groups (**Fig. 3**). The same observation has been made in previous studies that revealed that images capable of activating an IT neuron need not be similar to one another[17,18,24]. It is tempting to relate this observation to the fact that the concept of radial basis function networks has been introduced to explain the representation of similar but not distant objects, and that objects that are highly dissimilar need not be embedded into the same low-dimensional space[6]. However, the previous difficulties with finding a correlation between shape similarity and the responses of IT neurons could also be related to the lack of proper stimulus parameterization. Indeed, the high-dimensional nature of these complex stimuli makes it difficult to ascertain what relationship the shapes may have with regard to one another as long as similarity is not controlled for in a parametric way. For instance, consider a realistic pattern of responses to three dissimilar stimuli whose physical similarities can be described within a two-dimensional space (data points 1–3 in **Fig. 8a**). Within this stimulus set, a neuron responding strongly to stimuli 1 and 3 but not to stimulus 2 would exemplify a response pattern bearing no relationship to the similarity between the stimuli. However, including additional stimuli in the stimulus set (**Fig. 8b**) might reveal a more regular response pattern with a systematic unimodal tuning for a third stimulus dimension. Thus, we need a parametric control of similarity to arrive at a more principled understanding of the tuning characteristics of these neurons. Indeed, using parameterized stimuli, we found regular tuning curves of IT neurons for combinations of complex shape dimensions.

A second point of agreement between our data and the computational efforts is the demonstration that a population of neurons tuned to a set of reference shapes can underlie a low-dimensional and ordinally faithful representation of shape configurations. In addition, our results confirm that both assumptions of the model, a regular tuning on one hand, and different optimal shapes for different neurons on the other hand, contribute to the efficiency of the coding of shape similarity. However, by examining consistent metric deviations from the expected representation spaces we found that the representation of similarity at the behavioral and neuronal level is not completely faithful at the metric level. Indeed, the neuron-based similarities were always a better predictor of the monkey-perceived similarities than were the pixel-based similarities. With the idea of a network using radial basis functions, one can account for most aspects of our data, but cannot predict these metric biases (although they could be explained *ad hoc*). We applied measures of physical similarity that make no assumptions about how the image is analyzed in the visual system (in contrast to, for example, a wavelet analysis with oriented filters). Of course, physical similarity can be quantified by an almost infinite number of measures beyond the limited set we have used, and some of these alternative measures could do better. But, even in the latter case, we need a model that can explain why some measures are better than others although they all agree in an ordinal sense. Many theoretical models have tried to characterize the nonlinear processing steps in the hierarchically organized visual system (for example, see refs. 25–28). They make different predictions about the occurrence of systematic biases toward a higher sensitivity for some stimulus differences than for others, but our study was not designed to differentiate between these models.

In agreement with categorization models using the concept of similarity[3–5], we showed that the representation of similarity has a profound influence on categorization performance. However, computing similarity is only a first step. Learning a particular categorization rule implies that the low-dimensional representation space is parceled into regions that require the same response. Neuropsychological theories of categorization localize the representation of stimulus similarity within extrastriate cortex, but the stimulus–response mapping is presumed to be done in other areas such as prefrontal cortex, hippocampus and basal ganglia[29–33]. In line with this strict segregation of stimulus–response associations from the representation of stimulus similarity, the representation spaces we found at the neuronal level in visual cortex were not altered by learning a specific categorization rule. Further research is needed to compare the visual responses in different brain regions using parameterized stimuli to see how these visual representation spaces are transformed into category or 'response spaces.'

Methods

Subjects. Four monkeys (*Macaca mulatta*) and two naive humans were subjects. All procedures[34,35] were approved by the K.U. Leuven Ethical Committee for animal experiments and followed NIH guidelines.

Stimuli. The stimuli were 24 closed contours defined by 7 RFCs[36], and were divided into 3 groups with distinct RFCs. Within each group, differences between shapes were induced by independently varying the amplitude of two RFCs, creating eight shapes arranged in a square-like manner in the two-dimensional amplitude space (**Fig. 1**). The shapes (maximum size, 6°) were presented on a gray background (luminance,

© 2001 Nature Publishing Group http://neurosci.nature.com

© 2001 Nature Publishing Group http://neurosci.nature.com

© 2001 Nature Publishing Group http://neurosci.nature.com

17 candelas/m^2 (cd/m^2)) and filled with noise (pixel luminances, either 0.05 or 34 cd/m^2). The mean luminance of the shapes and background was equal. Shape position was randomized within a square region ranging from 2° to 4° centered on the fixation spot, except during the single-cell recordings where all shapes were presented foveally.

As a first physical dissimilarity measure, we computed the distance between two stimuli in a two-dimensional configuration equal to the parametric space. Second, the Euclidean and the city-block distance between two shapes, i and j, within the space of 256 × 256 pixels[11,37], was computed:

$$(\Sigma_{x=1:256}\Sigma_{y=1:256}\,|\mathrm{pix_i}(x, y) - \mathrm{pix_j}(x, y)|^r)^{1/r} \quad (1)$$

For shape and background pixels, pix(x, y) is 1 and 0, respectively; for city-block and Euclidean metrics, r is 1 and 2, respectively. Finally, we calculated the average information that a pixel of one image provides about the corresponding pixel in another image[38]. The congruence between these pixel-based dissimilarity measures was high. The Euclidean distances provided the best fit with the behavioral and neuron-based similarity measures and are used in the Results.

Behavioral tasks. Monkeys Y and E were trained in a same-different task with fixation control[39]. After 700-ms fixation (window, 2°), two shapes were shown successively for 300 ms (interstimulus interval, 500 ms). A leftward and rightward saccade was the correct response in trials with identical (50% of the trials) or different shapes, respectively. Aborted trials (for example, responses before the end of the second stimulus) were not included in the analyses. Correct responses were rewarded by juice. Each monkey was trained for several months in this task using a wide variety of images. Their performance level for highly dissimilar novel stimuli was 95% correct. During the experiment, all pairs of shapes were presented for 16 trials each, whereas additional testing (72 trials/pair) was done for pairs of shapes within a stimulus group. One additional session was run for monkey Y with group B stimuli (34 trials/pair). The performance level was 79% (Y) and 88% (E) correct when all shapes were paired and 69% (Y) and 80% (E) correct when only within-group comparisons were shown, which is as good as in other studies[12,20].

The 2 other monkeys learned to categorize the 24 shapes into 2 response categories (**Fig. 6**). The procedure was the same as for monkey Y, except that only one shape was shown, after which the monkeys had to make either a leftward or a rightward eye movement. Each monkey had several months of experience in this task with other stimuli.

Similarity judgments from the human subjects were obtained with a rating method. The rating and the same-different technique produce equivalent results[11,40]. The subjects were asked to fixate a spot while two stimuli were shown for 300 ms (interstimulus interval, 500 ms). They had to rate the similarity between the two stimuli on a scale from one (very similar) to nine (very dissimilar). All possible pairs of shapes were presented four times for each subject, and additional testing (10 trials/pair) was done for pairs of shapes within a stimulus group.

Recordings. IT recordings[34,35] were performed in monkeys F and M during the categorization task. Recording sites were verified using CT images with the guiding tube *in situ* in monkey M and *post mortem* in monkey F. We searched for responsive neurons by presenting all 24 shapes. Responsive neurons were investigated further by presenting all shapes for at least 6 trials (median 12 trials). All further analyses were done using the mean number of spikes in the 50–350-ms interval after stimulus onset after normalization to the maximum response. All neurons were shape-selective (one-way ANOVA[41]). The selectivity index (SE) compared the selectivity within the group to which the optimal stimulus belonged (maxR and minR(within) being the responses to the best and the worst stimulus within that group), with the maximum response elicited by a stimulus from the other groups (maxR(between)):

$$\mathrm{SE} = (\mathrm{maxR} - \mathrm{minR(within)})/(\mathrm{maxR} - \mathrm{maxR(between)}) \quad (2)$$

SE ranges between 0 (no within-group selectivity) and infinity (no between-group selectivity).

The within-group response pattern was represented on a polar plot using the radial position of each stimulus with respect to the center of the parametric configuration, and the unimodality of the selectivity was determined by a fast Fourier transform of these plots[42]. Unimodal selectivities will be reflected in a Fourier spectrum dominated by the first-order component. The ratio (RR) of the size of this first-order component to the largest of all other components was taken as a measure of the dominance of the first-order component. A high RR value reflects an unimodal tuning, but a low RR does not necessarily correspond to an irregular tuning. (Many factors contribute to a low RR, such as an asymmetrical but unimodal tuning.) So, the RR index underestimates the relationship between shape similarity and IT selectivity.

To analyze the representation of the shape similarities at the population level, we computed the distance between a pair of stimuli i and j in the multidimensional space spanned by the responses of all neurons[35,43,44]:

$$[(\Sigma_{n\,=\,1:124}\,|\mathrm{Resp_i}(n) - \mathrm{Resp_j}(n)|^2)/124]^{1/2} \quad (3)$$

Here, n is the cell number. The inverse of these distances measures the similarity of the neural representations of two shapes.

The effect of tuning properties on the coding of similarity at the population level was assessed as follows. First, we drew a random sample of neurons (with replacement) from the population. The sample size ranged from 2 to 64, and 150 samples of each size were analyzed (50 for each stimulus group). For each stimulus group, only neurons that responded to at least one shape were included, giving a population size of 105, 109 and 117 for groups A, B and C, respectively. Because we found no consistent differences among stimulus groups, the results were pooled. Second, we selected neurons from four different subsets. For each subset, we selected 150 samples for each of 3 sizes (2, 4 or 8). In the first and second subset, we selected only neurons with RR lower and higher, respectively, than the median RR of the population. In the third subset, the neurons in a sample had similar preferred shapes. In each sample, the first neuron was randomly drawn but the other neurons were required to have tuning curves with an optimum at the same or a neighboring stimulus. In the fourth subset, the optima of the neurons were distributed (optima were 180° apart for two neurons, 90° apart for four neurons, and were at all positions for eight neurons).

Analysis of similarity data. The different sets of similarity data were analyzed using Statistica software (StatSoft, Tulsa, Oklahoma). First, we checked whether the 24 shapes showed the expected clustering in three groups by performing hierarchical cluster analysis[45]. This algorithm starts from a configuration with as many clusters as stimuli, and groups similar stimuli in several steps (starting with the most similar stimuli) until all stimuli form one cluster. We describe the results obtained by using a weighted pair-group average, but other rules such as single and complete linkage revealed the same pattern of results. We have summarized the results by describing at which level stimuli are clustered with stimuli from the same and different stimulus groups. The proportion of variance in the original data that could be accounted for by three clusters of stimuli measures the degree of clustering. The dissimilarity between stimuli within a cluster was set to zero, whereas the 'linkage distance' at the level at which two clusters were grouped was taken as the dissimilarity between stimuli from different clusters. Second, we analyzed the similarity data from each stimulus group separately with nonmetric MultiDimensional Scaling (MDS)[43,45].

As a statistical criterion to decide whether the perceived and neuron-based similarities converged with the similarities from the parametric or pixel-based configurations, we computed the congruence coefficient C between different sets A and B of similarity data[46]:

$$C = (\Sigma_{n\,=\,1:28}\,\mathrm{d_A}(n) \times \mathrm{d_B}(n))/(\Sigma_n \mathrm{d_A}^2(n) \times \Sigma_n\,\mathrm{d_B}^2(n))^{1/2} \quad (4)$$

The index n provides a summation across the 28 pairs of distances. The closer C approximates 1, the better the fit. The significance of a C value was evaluated using Monte Carlo simulations (n = 1000) to determine the expected congruence when a square configuration is compared with a random two-dimensional configuration of eight points. This distribution approximated a normal distribution with mean 0.865 and a standard deviation of 0.029, meaning that a value above 0.960 is highly significant ($p < 0.001$), which is similar to the result found for a range

© 2001 Nature Publishing Group http://neurosci.nature.com

of configurations[11]. As this index is skewed to unity, small differences in C in the range of 0.95–1 are meaningful.

Measurement noise would introduce deviations of a measured configuration with respect to the actual configuration that are unrelated from one set of measurements to the other. However, if the deviations are consistent across datasets, then one can conclude that the actual configuration is distorted with respect to the expected configuration. This is analogous to determining whether a correlation between two variables (the two measured configurations) can be explained by their correlations with a third variable (the expected configuration).To simulate the effect of measurement noise, we added uniform noise to the similarities expected in a square configuration. We used several noise magnitudes to cover the range of congruences found in our study. The fit between two noisy datasets was almost never higher than the best of both fits found by comparing each noisy dataset with the perfect square similarities ($p < 0.001$). Even a fit between two noisy sets that is larger than the worst but lower than the best of these is significant ($p < 0.05$). If a configuration is measured twice, then the difference between the measurements will be larger than the difference between one measurement and the configuration. If the difference between the two measurements is smaller, then probability is high that a systematic factor determines the way in which the measurements deviate from the expected configuration (**Table 1**).

Note: Supplementary information is available on the Nature Neuroscience web site (http://neuroscience.nature.com/web_specials).

Acknowledgements

This work was supported by the Geneeskundige Stichting Koningin Elizabeth, Interuniversitaire Attractiepolen (IUAP P4/22) and the University Research Council (IDO/98/002). We thank M. De Paep, P. Kayenbergh, G. Meulemans, G. Vanparrys, S. Van Wetter and G. Kayaert for technical assistance, and S. Raiguel and F. Wichmann for their comments on the manuscript. H.O.d.B. is a research assistant of the Fund for Scientific Research (FWO) Flanders.

Received 6 September; Accepted 25 October 2001

1. Edelman, G. M. *Neural Darwinism* (Basic Books, New York, New York, 1987).
2. Smith, E. E. & Medin, D. L. *Categories and Concepts* (Harvard Univ. Press, Cambridge, Massachusetts, 1981).
3. Nosofsky, R. M. Choice, similarity, and the context theory of classification. *J. Exp. Psychol. Learn. Mem. Cogn.* **10**, 104–114 (1984).
4. Ashby, F. G. & Perrin, N. A. Toward a unified theory of similarity and recognition. *Psychol. Rev.* **95**, 124–150 (1988).
5. Shepard, R. N. Toward a universal law of generalization for psychological science. *Science* **237**, 1317–1323 (1987).
6. Edelman, S. *Representation and Recognition in Vision* (MIT Press, Cambridge, Massachusetts, 1999).
7. Roweis, S. T. & Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323–2326 (2000).
8. Tenenbaum, J. B., de Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319–2323 (2000).
9. Shepard, R. N. & Cermak, G. W. Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cogn. Psychol.* **4**, 351–377 (1973).
10. Cortese, J. M. & Dyre, B. P. Perceived similarity of shapes generated from fourier descriptors. *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 133–143 (1996).
11. Cutzu, F. & Edelman, S. Representation of object similarity in human vision: psychophysics and a computational model. *Vision Res.* **38**, 2229–2257 (1998).
12. Sugihara, T., Edelman, S. & Tanaka, K. Representation of objective similarity among three-dimensional shapes in the monkey. *Biol. Cybern.* **78**, 1–7 (1998).
13. Dean, P. Effects of inferotemporal lesions on the behavior of monkeys. *Psychol. Bull.* **83**, 41–71 (1976).
14. Logothetis, N. K. & Sheinberg, D. L. Visual object recognition. *Annu. Rev. Neurosci.* **19**, 577–621 (1996).
15. Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
16. Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **4**, 2051–2062 (1984).
17. Gochin, P. M., Colombo, M., Dorfman, G. A., Gerstein, G. L. & Gross, C. G. Neural ensemble coding in inferior temporal cortex. *J. Neurophysiol.* **71**, 2325–2337 (1994).
18. Wilkinson, F., Wilson, H. R. & Habak, C. Detection and recognition of radial frequency patterns. *Vision Res.* **38**, 3555–3568 (1998).
19. Albright, T. D. & Gross, C. G. Do inferior temporal cortex neurons encode shape by acting as Fourier Descriptor filters? *Proc. Int. Conf. Fuzzy Logic & Neural Networks, Izuka, Japan*, 375–378 (1990).
20. Sands, S. F., Lincoln, C. E. & Wright, A. A. Pictorial similarity judgements and the organization of visual memory in the rhesus monkey. *J. Exp. Psychol.* **111**, 369–389 (1982).
21. Edelman, S. Representation is representation of similarities. *Behav. Brain Sci.* **21**, 449–467 (1998).
22. Poggio, T. & Girosi, F. Regularization algorithms for learning that are equivalent to multilayer networks. *Science* **247**, 978–982 (1990).
23. Poggio, T. & Edelman, S. A network that learns to recognize three-dimensional objects. *Nature* **343**, 263–266 (1990).
24. Vogels, R. Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. *Eur. J. Neurosci.* **11**, 1239–1255 (1999).
25. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, San Francisco, California, 1982).
26. Riesenhuber, M. & Poggio, T. Models of object recognition. *Nat. Neurosci.* **3**, 1199–1204 (2000).
27. Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychol Rev.* **94**, 115–147 (1987).
28. Wallis, G. & Rolls, E. T. Invariant face and object recognition in the visual system. *Prog. Neurobiol.* **51**, 167–194 (1997).
29. Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U. & Waldron, E. M. A neuropsychological theory of multiple systems in category learning. *Psychol. Rev.* **105**, 441–481 (1998).
30. Murray, E. A., Bussey, T. J. & Wise, S. P. Role of prefrontal cortex in a network for arbitrary visuomotor mapping. *Exp. Brain Res.* **133**, 114–129 (2000).
31. Wise, S. P. & Murray, E. A. Role of the hippocampal system in conditional motor learning: mapping antecedents to action. *Hippocampus* **9**, 101–117 (1999).
32. Asaad, W. F., Rainer, G. & Miller, E. K. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* **21**, 1399–1407 (2000).
33. Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316 (2001).
34. Op de Beeck, H. & Vogels, R. Spatial sensitivity of macaque inferior temporal neurons. *J. Comp. Neurol.* **426**, 505–518 (2000).
35. Vogels, R., Biederman, I., Bar, M. & Lorincz, A. Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *J. Cogn. Neurosci.* **13**, 444–453 (2001).
36. Zahn, C. T. & Roskies, R. Z. Fourier descriptors for plane closed curves. *IEEE Trans. Comp.* **21**, 269–281 (1972).
37. Grill-Spector, K. *et al.* Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* **24**, 187–203 (1999).
38. Rieke, F., Warland, D., de Ruyter van Steveninck, R. R. & Bialek, W. *Spikes: Exploring the Neural Code Vision* (MIT Press, Cambridge, Massachusetts, 1997).
39. Robinson, D. A. A method of measuring eye movements using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Eng.* **101**, 131–145 (1963).
40. Cortese, J. M. *Perceptual Similarity of Closed Contours.* Thesis, Univ. of Illinois (1992).
41. Kirk, R. E. *Experimental Design: Procedure for the Behavioral Sciences* (Brooks-Cole, Belmont, California, 1968).
42. Worgotter, F. & Eysel, U. T. Quantitative determination of orientational and directional components in the response of visual cortical cells to moving stimuli. *Biol. Cybern.* **57**, 349–355 (1987).
43. Young, M. P. & Yamane, S. Sparse population coding of faces in the inferotemporal cortex. *Science* **256**, 1327–1331 (1992).
44. Kobatake, E., Wang, G. & Tanaka, K. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J. Neurophysiol.* **80**, 324–330 (1998).
45. Shepard, R. N. Multidimensional scaling, tree-fitting, and clustering. *Science* **210**, 390–398 (1980).
46. Borg, I. & Leutner, D. Measuring the similarity of MDS configurations. *Multivar. Behav. Res.* **20**, 325–334 (1985).

© 2001 Nature Publishing Group http://neurosci.nature.com