The representation of perceived shape similarity and its role for category learning in monkeys: A modeling study

Hans P. Op de Beeck [a,*], Johan Wagemans [a], Rufin Vogels [b]

[a] *Laboratory of Experimental Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium*
[b] *Laboratory for Neuro- and Psychophysiology, K.U.Leuven, Belgium*

Abstract

Categorization models often assume an intermediate stimulus representation by units implementing "distance functions", that is, units that are activated according to the distance or similarity among stimuli. Here we show that a popular example of these models, ALCOVE, is able to account for the performance of monkeys during category learning when it takes the perceived similarity among stimuli into account. Similar results were obtained with a slightly different model (ITCOVE) that included experimentally measured tuning curves of neurons in inferior temporal (IT) cortex. These results show the intimate link between category learning and perceived similarity as represented in IT cortex.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Shape perception; Object recognition; Extrastriate cortex; Monkey

1. Introduction

The term 'visual categorization' refers to the process by which a possibly infinite number of visual input patterns are grouped into a relatively small number of output categories. Many models of visual categorization have been proposed (e.g., Ashby & Waldron, 1999; Kruschke, 1992; Kruschke, 2005; Medin & Schaffer, 1978; Nosofsky, 1984). Despite major differences in how stimuli and response selection are modeled in each of these theories, most quantitative models of visual categorization assume some sort of spatial stimulus representation. In these models stimuli are represented as points or regions in a stimulus space so that the discriminability among any two stimuli is related to their distance in stimulus space.

Spatial categorization models have been implemented in neural networks (Ashby & Waldron, 1999; Kruschke, 1992). In these networks, stimuli are represented by means of units that implement 'distance functions'. Each of these units has a maximum value when the input corresponds to one particular position in the input space (the 'preferred' stimulus) and it falls off with increasing distance between the input and the preferred position. Here we investigate whether the concept of distance functions and the related models are consistent with (i) the behavior of rhesus monkeys in a categorization task and (ii) the observed tuning properties of single neurons in inferior temporal (IT) cortex.

1.1. Distance functions and category learning

Probably the most frequently used neural network implementation of spatial categorization models is ALCOVE, a three-layer network for categorization (Kruschke, 1992; see Fig. 1). The inputs to the ALCOVE network are the values of exemplar stimuli along each stimulus dimension. The
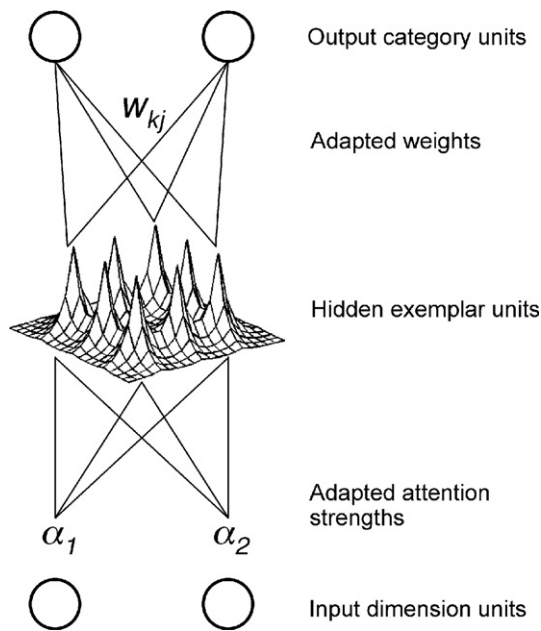
Fig. 1. The standard architecture of the ALCOVE network (see Sections 1.1 and 2.5; figure based on Kruschke, 1992).

hidden units implement distance functions centered at the locations in stimulus space that correspond to the learned exemplars. The tuning function of each hidden unit is characterized by its position in stimulus space. The output layer contains as many units as there are categories to learn. The network learns to associate each input pattern with the correct output unit by adapting two sets of weights with error-driven learning. First, the overall weight of each stimulus dimension is gated by an attention strength factor (α_1 and α_2 in Fig. 1) that reflects the learned relevance of that dimension for the categorization task. Second, the weights between hidden units and output units (w_{kj} in Fig. 1) are adapted during learning.

The standard ALCOVE is a neural network implementation of one specific class of categorization models, namely exemplar models (e.g., Medin & Schaffer, 1978; Nosofsky, 1984). Exemplar models provide a very good account of human and monkey behavior in categorization tasks (e.g., Kruschke, 1992; Nosofsky & Johansen, 2000; Sigala, Gabbiani, & Logothetis, 2002). The success of these models and of the ALCOVE model in particular suggests that humans and monkeys represent visual stimuli in terms of distance functions.

One reason for the popularity of the ALCOVE architecture is that some other spatial categorization models can be implemented with only slight modifications in architecture (Johansen & Palmeri, 2002). For example, ALCOVE turns into an implementation of prototype models of categorization if the hidden layer contains one hidden unit for each category (a distance function tuned to the prototype of a category). ALCOVE has also been adapted to include hybrid forms of spatial representations and feature-based representations (e.g., Lee & Navarro, 2002; Verguts, Ameel, & Storms, 2004).

1.2. Distance functions in the brain

A likely candidate site for the implementation of the intermediate representations that are involved in visual shape categorization in macaque monkeys is the inferotemporal (IT) cortex in the ventral visual stream (Logothetis & Sheinberg, 1996; Riesenhuber & Poggio, 1999). Several theoretical analyses have suggested that stimuli might be represented in the brain by a recoding of the visual input into distance functions (Edelman, 1999; from a theoretically point of view, these distance functions form a set of basis functions, see Poggio & Girosi, 1990; Pouget & Snyder, 2000). Single-cell recordings in IT cortex have demonstrated the usefulness of the concept of 'distance functions' in understanding the representation of shape in high-level visual cortex (Logothetis, Pauls, & Poggio, 1995; Op de Beeck, Wagemans, & Vogels, 2001). In our study (Op de Beeck et al., 2001), we used a stimulus set created by a parametric manipulation of shape contour. This set was composed of three groups of 8 shapes with the two-dimensional configurations in parameter space as shown in Fig. 2A. Behavioral tests showed that the parametric configurations are, to a large degree, faithfully represented by human and monkey subjects, as shown in Fig. 2B for each stimulus group. The stimulus order in the representational space and its dimensionality agreed with the stimulus order and dimensionality of the parametric configurations. Nevertheless, some metric deviations were also noted, resulting in small but consistent deformations of the representational space with respect to parametric space.

The single-cell recordings in macaque IT cortex revealed a close correspondence between pair-wise shape distances in this behavioral representational space, and pair-wise differences in the IT population responses. Even the small deformations of the behavioral representational space compared to parametric space were present in the population responses. At the level of individual neurons, most IT neurons responded maximally to the presentation of one particular shape in a stimulus group, and responses diminished gradually for shapes located more distantly in stimulus space, exactly as expected when implementing 'distance functions'.

Although these results indicate that IT neurons may implement distance functions, the neuronal tuning curves showed a high level of diversity. Computational models such as the models introduced by Kruschke (1992) and Edelman (1999) contain units with identical tuning properties that differ only in their preferred position in stimulus space. The IT response profiles, however, displayed great diversity in their tuning properties. There was a wide range of tuning widths with some neurons responding to only one stimulus but other neurons responding to almost all stimuli. Furthermore, although the tuning profiles of many neurons showed a close correspondence with shape similarity, a minority deviated from this general picture.
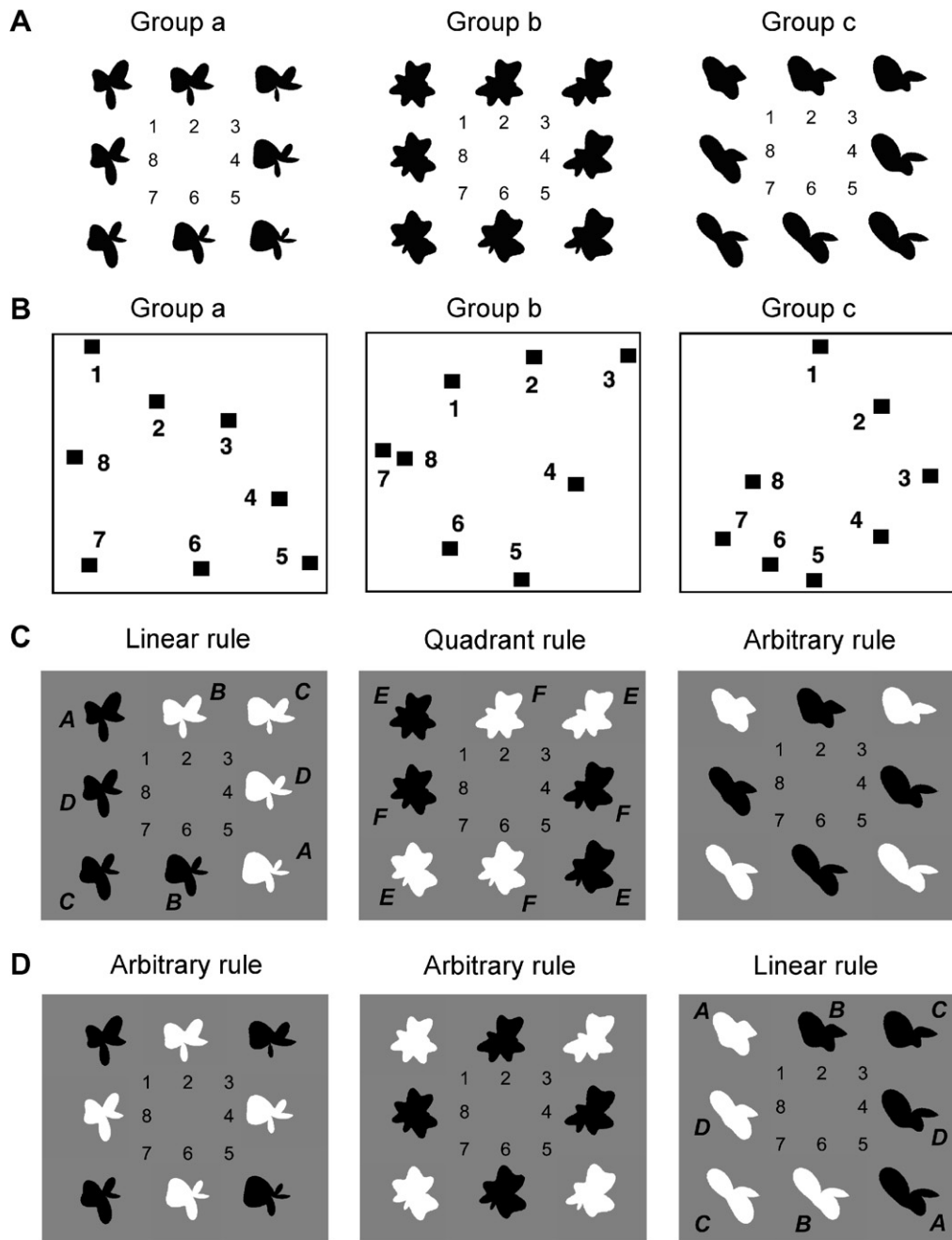
Fig. 2. Stimuli used in the behavioral and neurophysiological experiments. (A) The three stimulus groups (a–c), with stimulus position in each group representing the position of stimuli in parametric space. (B) MDS-derived configurations with stimulus position representing the position of stimuli according to perceived similarity (as described by Op de Beeck et al., 2001). (C) Illustration of the three category rules. The black or white color of the stimuli represents their category membership according to each rule. Capital letters denote stimuli with a particular position in stimulus space with respect to the category border (A–D for the linear rule, and E–F for the quadrant rule). The pairing of rule and stimulus group as shown in panel C (linear, quadrant, and arbitrary rule for stimulus group a, b, and c, respectively) corresponds to the learning scheme of monkey 1, the pairing in monkey 2 is shown in panel D.

1.3. The present investigation

Up to now, two lines of evidence were used to support the hypothesis that the visual system recodes its input into a set of distance functions for categorization: (i) the correspondence between the performance of a distance function network such as ALCOVE and behavior and (ii) neuronal tuning curves in IT cortex. None of the previous studies

combined these two lines of evidence by showing that a network with empirical neuronal tuning curves can account for categorization behavior. In particular, previous implementations of ALCOVE have not considered the effect of the diverse nature of neuronal tuning curves. Not only might such diversity directly affect network performance and modeling capacity, but it also evokes questions about how particular tuning functions in the intermediate repre-

sentation are selected while the network learns the input/output transformation needed for categorization: Which units will be selected by which category rules? How is this selection related to other parameters of the model such as attention weights? Would the diversity and the neuronal selection influence performance of the network? These questions will be addressed in the present work.

2. Methods

2.1. Stimuli

The stimulus set included 3 groups of 8 stimuli (Fig. 2A). The square arrangement of the stimuli within each stimulus group represents the parameter-space configuration of the stimuli. Op de Beeck et al. (2001) measured the perceptually weighted similarity space of these stimuli (shown in Fig. 2B) from the performance of two monkeys in a sequential matching task.

2.2. Category rules

The category learning tasks included three different category rules, referred to as the linear rule, the quadrant rule, and the arbitrary rule (Fig. 2C–D). In these rules, exemplar stimuli could differ in the average similarity of an exemplar to all other exemplars in its category (within-category similarity) and its average similarity to exemplars of other categories (between-category similarity). For example, stimulus 1 in the left panel of Fig. 2C belongs to one category together with stimuli 6, 7, and 8, but stimulus 1 is relatively different from these other exemplars of its category (relatively low average within-category similarity) and relatively similar to some exemplars of the other category (especially stimulus 2). In contrast, stimulus 7 in the same panel is more representative for its category and very different from all exemplars of the other category. Despite these stimulus differences, some stimuli are equivalent in within-category and between-category similarity in parametric space due to the construction of the stimulus space and category rules. For example, stimuli 1 and 5 in the left panel of Fig. 2C are equivalent according to these metrics, only

their category differs, e.g. the within-category similarity of stimulus 1 for the category of stimuli shown in black is the same as the within-category similarity of stimulus 5 for all white stimuli. For our analyses we have pooled "equivalent" exemplars with the same within-category and between-category similarity in parametric space, and the result is shown in Fig. 2C–D by capital letters. This distinction is only made for the linear and quadrant rule, since for the arbitrary rule, all exemplars of a category have the same within-category and between-category similarity.

2.3. Behavioral data during category learning

The behavioral data of this paper have been published by Op de Beeck et al. (2001), and were obtained from different monkeys than the monkeys used to measure the perceptual similarity space. Two monkeys (in the present paper referred to as monkey 1 and monkey 2) learned to categorize the stimuli of each stimulus group into two categories. Monkey 1 learned the linear, quadrant and arbitrary rule for stimulus group a, b, and c, respectively (Fig. 2C). His performance during training is shown in Fig. 3. Monkey 2 learned the linear rule for stimulus group c and the arbitrary rule for groups a and b (Fig. 2D).

2.4. Neuronal recordings

As described previously (Op de Beeck et al., 2001), we recorded the responses of 124 neurons in the IT cortex of two monkeys (monkey 1 and 2) while they performed a categorization task. The responses of each neuron were normalized by dividing all responses by the response to the preferred stimulus.

2.5. ALCOVE implementations

The network architecture as used in this paper followed Kruschke (1992) and included three layers (Fig. 1). The activation of hidden unit j for the exemplar stimulus e is given by

$$a_{je}^{\text{hid}} = \exp[-c\Sigma_i\alpha_i|h_{ji} - a_{ie}^{\text{in}}|] \quad (1)$$

where c is a scaling parameter determining the specificity of the unit (large values imply a narrow activation profile), α_i is the weight that is given to dimension i (attention strength), h_{ji} is the position of hidden unit j along dimension i, and a_{ie}^{in} is the value of the input stimulus e along dimension i.

Two output nodes correspond to the possible response categories and their activation is given by:

$$a_{ke}^{\text{out}} = \Sigma_j w_{kj}a_{je}^{\text{hid}} \quad (2)$$

where w_{kj} is the association weight between hidden unit j and output node k. These weights and the attention strengths are adjusted by gradient descent error learning (the 'humble teacher' version of Kruschke, 1992). In order to compare model performance with behavioral performance the network output is transformed to response probabilities as given by

$$P_e(K) = \exp(\phi a_{ke}^{\text{out}})/\Sigma_k\exp(\phi a_{ke}^{\text{out}}) \quad (3)$$

where $P_e(K)$ is the probability that exemplar e is categorized in category K and ϕ is a response mapping parameter (Kruschke, 1992). The reported results were obtained with a response mapping parameter of $\phi = 2.0$, a learning rate for the weights w_{kj} of $\lambda_w = .01$, and a learning rate for the attention strengths α_I of $\lambda_a = .0011$. All the reported results were consistent across different parameter settings (except for parameter c, of which the effect is reported in Section 3).

2.6. ITCOVE implementations

We constructed a variant of ALCOVE, referred to as ITCOVE, in which the output of the hidden layer was based on the responses of the recorded IT neurons. This layer contains 124 units, one for each recorded
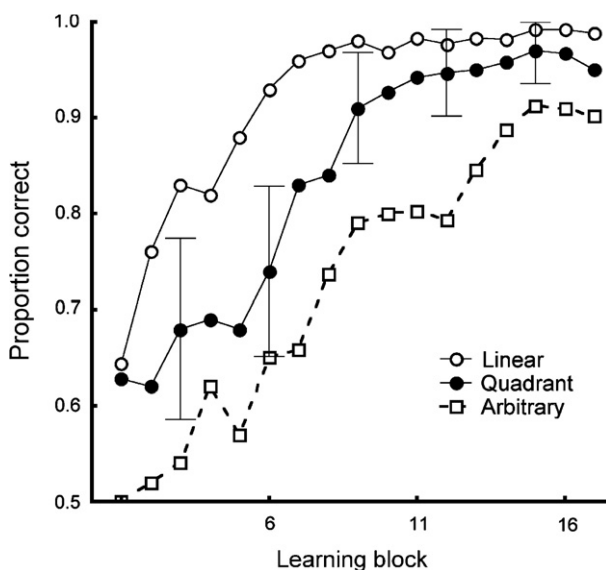

Fig. 3. Performance of monkey 1 on each category rule as a function of discretized learning blocks of 600 trials each. Whiskers indicate 95% confidence intervals. These data have been published in another format by Op de Beeck et al. (2001).

neuron. As the responses of each neuron were recorded for the three groups of eight stimuli, we actually constructed three ITCOVE networks, one for each stimulus group.

The output of the hidden layer in ALCOVE is deterministic, that is, the output is always the same for a particular input pattern. For ITCOVE, we constructed not only a deterministic variant, but also a variant that included trial-by-trial variability in the output of the network units. The response of each unit to each presentation of an exemplar was selected from a normal distribution with mean and variance estimated using the observed mean response strength and response variance during the recordings. The inclusion of trial-by-trial variance of the neuronal responses leads to some variability in network performance between simulations, so all simulations of ITCOVE were repeated ten times and averaged for each stimulus group (the mean result averaged over stimulus groups is presented here, except where the results deviated between networks).

Several versions of ALCOVE and ITCOVE are compared in the Results section, with the number of hidden units or neurons varying from 8 (ALCOVE) to 124 (ITCOVE), and with variations in the selectivity of these units. These two parameter variations have strong implications for the overall speed of learning in the network (faster learning with more units and with more selective units), but they do not change the relative order of learning of rules or specific stimuli. Nevertheless, the effects of these two parameters on overall network performance make it difficult to compare different networks, for example in Fig. 4. We solved this problem by selecting the reported learning blocks so that the average performance within each block was comparable to the average behavioral performance (a "learning block" includes one presentation of each stimulus to the network). This average performance is calculated across all stimulus groups, learning rules, and in the case of ALCOVE across the implementations with tuning in parametric and perceptual space. In every case, the rank order in the performance for different rules or different stimuli in the selected blocks proved to be consistent with those in the other learning blocks. In this report we will focus on the order of learning different rules, and we will investigate in addition which specific stimuli are learned better, and the conclusions were not affected by the specific learning block that is used when comparing rules or stimuli.

2.7. Category selectivity (CS)

The category selectivity of a unit or neuron, CS, is given by

$$CS_{jk} = \Sigma_{e \in K} a_{je} - \Sigma_{e \notin K} a_{je} \qquad (4)$$

where $\Sigma_{e \in K}$ denotes a summation across all exemplars that belong to category K. This index compares the average response to exemplars of category K with the average response to the other stimuli.

All correlations between neural tuning properties and weights presented in this paper compare sets of values with $N = 124$ (the number of recorded neurons or units in ITCOVE). Correlations greater than 0.18 are statistically significant ($p < 0.05$, $N = 124$).

2.8. Analyses of neuronal response profiles

In addition to CS as an index for category selectivity (formula (4)), two indices were used to describe the tuning of each neuron within each stimulus group. The first index, the depth of selectivity (DOS), was introduced by Rainer and Miller (2000) and is given by

$$DOS = [S - (\Sigma_{i=1:S} R_i) / \max(R)] / (S - 1) \qquad (5)$$

With S the number of stimuli in each group (8).

The second index (RR), described by Op de Beeck et al. (2001) (see also Worgötter & Eysel, 1987) was based on a Fourier transform of a polar plot constructed using the radial position of each stimulus with respect to the center of the parametric configuration. A high RR value indicates that the relationship between radial position and response strengths can be captured by a unimodal sinusoidal modulation, and it will decrease as a consequence of any deviation from this highly regular modulation (e.g., bimodal tuning).

3. Results

We compared the performance of different versions of ALCOVE with the previously reported behavior of monkeys that learned three category rules (Fig. 2; see Op de Beeck et al., 2001): the linear, quadrant, and arbitrary rule. First, we will describe some important aspects of the behavioral data. Second, we will describe the performance of the ALCOVE model and how it matches monkey performance. Third, we describe the results obtained with a network, labeled ITCOVE, in which we have substituted the
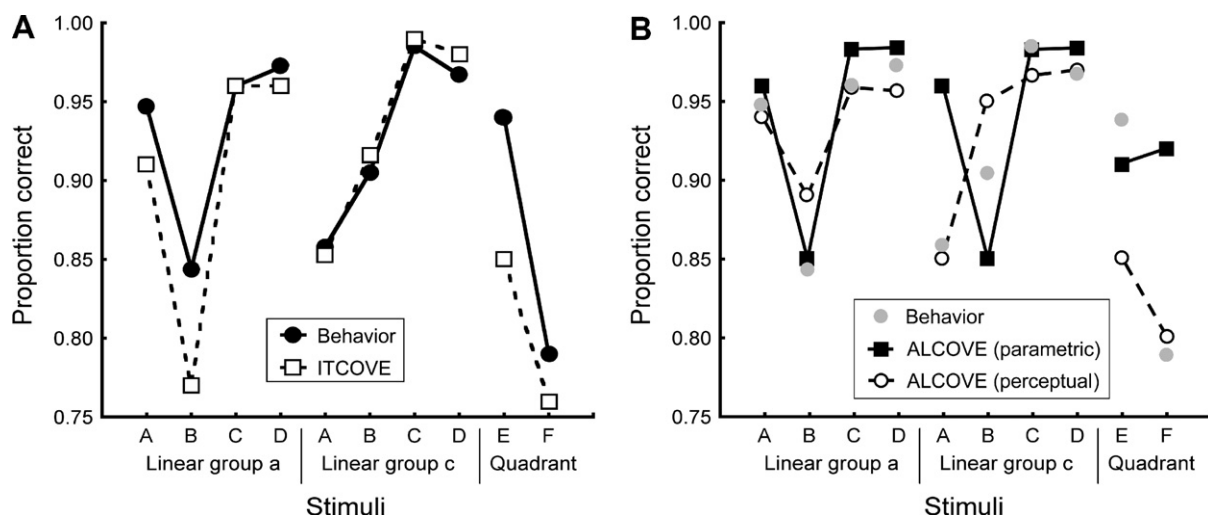

Fig. 4. Performance of models and monkeys for stimuli with a particular position with respect to the linear and quadrant category rules. (A) Performance of the monkeys and the ITCOVE model for stimuli with a particular position with respect to the linear and quadrant category rules. (B) The predictions of the ALCOVE model with units tuned for parametric or for perceptual similarity. "Linear group a" refers to the linear category rule applied to stimulus group a, "linear group c" is the linear category rule applied to stimulus group c, and "quadrant" refers to the quadrant rule (which was only applied to stimulus group b). The behavioral data for linear group a and the quadrant rule were obtained from monkey 1, and the behavioral data for linear group c from monkey 2 (see Fig. 2).

responses of real IT neurons for the mathematically defined hidden units. We will describe the performance of ITCOVE and finally we will discuss the importance of the diversity of tuning profiles in ITCOVE's stimulus representation.

3.1. Behavioral performance of the monkeys

Monkey 1 learned the linear, quadrant and arbitrary rule for stimulus group a, b, and c, respectively (Fig. 2C). His performance during learning is shown in Fig. 3. Monkey 2 learned the linear rule for stimulus group c and the arbitrary rule for groups a and b (Fig. 2D). The level of performance in these two subjects depended on the category rule. Monkey 1 performed better for the linear rule than for the quadrant rule, and better for the quadrant rule than for the arbitrary rule. Monkey 2 displayed a similar difference in performance between linear and arbitrary rules, that is, showing better performance with the linear rule.

These differences between learning rules have been described in Op de Beeck et al. (2001), but in addition there were difference in performance among the stimuli within a stimulus group, as shown in Fig. 4. Exemplar models (Medin & Schaffer, 1978) assume that the performance for a given exemplar depends on the average similarity of that exemplar to all other exemplars in its category (within-category similarity) and its average similarity to exemplars of other categories (between-category similarity). The pooling of exemplars with the same within-category and between-category similarity is represented in Fig. 2C–D by capital letters (see Section 2 for details). Within-category and between-category similarity was defined in parametric space. This parametric space is the same for all stimulus groups, but Op de Beeck et al. (2001) showed that the perceived stimulus differences are not equal to parametric or physical stimulus differences (Fig. 2B; see Introduction). For this reason, different stimulus groups are not pooled in Fig. 4. Monkey 1 learned the linear rule for stimulus group a ("linear group a" in Fig. 4) and monkey 2 for stimulus group c ("linear group c" in Fig. 4). The data for the two groups (or monkeys) are consistent in that performance for exemplar stimuli 'C' and 'D' is better than performance for exemplars 'A' and 'B'. However, a marked difference was found in that for linear group a performance was better for exemplar stimuli 'A' than for exemplars 'B' (95% versus 84%, $p < .01$) while for linear group c performance was worse for exemplars 'A' than for exemplars 'B' (86% versus 91%, $p < .01$; see Fig. 4A). In the next section we will use ALCOVE to show that these differences between stimulus spaces are due to the deviations between the parametric stimulus space and the perceived stimulus space that were found by Op de Beeck et al. (2001).

3.2. ALCOVE implementations

First, we implemented the ALCOVE network in its standard format. There were two input units, one for each of the two input dimensions. The values assigned to these units ranged from zero to four and represented the values of stimuli on the two parametric dimensions (e.g., stimuli 1–3 of Fig. 2 were associated with value zero of one input unit, and stimuli 5–7 with value four of the same unit). There was one hidden unit for each exemplar (8 hidden units), and one output unit for each category (2 units). The network learned a category rule by changing the attention weight that is given to each input dimension and by changing the weights between hidden units and output units (see Section 2 for details).

The performance of ALCOVE for the three category rules (linear, quadrant, and arbitrary) depended on parameter c, the selectivity of the hidden units. Fig. 5 displays the performance for each rule during learning when the hidden units were relatively selective ($c = 7$). In this case, the hidden unit that prefers exemplar 1 will respond to exemplar 2 with only 25% of the response to exemplar 1. This network learned the three category rules in the same order as the monkeys did: first the linear rule, and last the arbitrary rule. Although the relative order of rule learning was consistent across a wide range of c values, there were some important quantitative effects of this parameter. With less selectivity (small c values), the linear rule was learned much faster than the two other rules. In fact, a decrease in selectivity slowed learning for each rule, but less so for the linear rule than for the other two rules. For example, with $c = 1.2$, the quadrant and the arbitrary rule were associated with a performance of only 61% and 54%, respectively, in the same learning block (block 100) for which performance reached 90% correct for the linear rule. The difference in performance between the three rules became progressively smaller with larger c values.
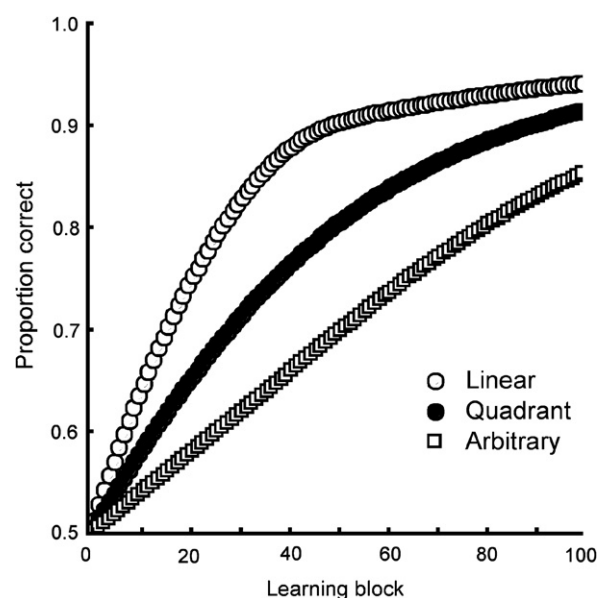

Fig. 5. Performance of the standard ALCOVE network with $c = 7$. The line plots show the performance for each category rule as a function of time.

These effects of parameter c suggest that a stimulus representation having units with a low selectivity is less suited for learning non-linear rules than for learning a linear rule (the term 'less suited' applies to the speed of learning; asymptotic performance of the ALCOVE model after an infinite number of learning blocks is always nearly 100% correct, no matter which rule has to be learned). A similar conclusion was drawn by Baldi and Heiligenberg (1988). The adapted attention strengths in ALCOVE support this conclusion. Increasing both attention strengths results in hidden units with a higher selectivity along both dimensions, and thus we would expect a larger increase of attention strengths during learning of a non-linear rule than during learning of a linear rule. Indeed, the initial attention strength of 1 for each dimension was increased in the last learning block to 1.3 and 1.6 for the quadrant and the arbitrary rule, respectively. For the linear rule, the attention strength also increased to 1.2 for the horizontal dimension in Fig. 2 (the most relevant dimension for this rule), but it decreased to .5 for the vertical dimension. Thus, the fact that the attention strengths can be adapted was useful for each learning rule, for the linear rule because the hidden units became sensitive to the difference in the relevance of the two input dimensions, and for the non-linear rules in order to increase the general selectivity of hidden units. As a consequence, the same network architecture, but lacking the ability to change attention strengths, learned each rule more slowly although the order in which it learned the different rules was identical to that of the network with adaptive attention strengths.

This standard ALCOVE network failed to model subjects' performance on individual stimuli within stimulus groups (Fig. 4B). The network's performance in the last learning block for stimuli 'A', 'B', 'C', and 'D' in the linear rule was 96%, 85%, 98%, and 98%, respectively. As noted before, monkey performance showed a comparable rank order when the linear rule was learned for stimulus group a, but the ranking of the behavioral performance was opposite to that of the network for stimuli 'A' and 'B' of group c (Fig. 4). For the quadrant rule, ALCOVE showed the same levels of performance for stimuli 'E' and 'F' (91% and 92%, respectively), while monkey behavior showed far better performance for stimuli 'E' than for stimuli 'F' ($p < 0.01$).

A likely explanation of the discrepancy between network performance and behavioral data is that the parametric stimulus space is not represented faithfully in perceptual space. Some stimulus differences are perceived as being larger or smaller than would be expected from the distance between stimuli in parametric space. If this explanation is correct, then ALCOVE should be able to model monkey behavior provided that its input represents the value of stimuli along the two dimensions that form the perceptual space, rather than using the parametric dimensions. The three sets of data points in Fig. 4 were obtained with different stimulus groups, and so each dataset should be compared with the performance of ALCOVE implemented

with the perceptual configuration of the corresponding stimulus group, instead of using the parametric configuration (which was the same for all stimulus groups). The results revealed that this adapted version of ALCOVE is able to model the aforementioned significant differences between stimuli in how well they are learned by monkeys (Fig. 4B). The performance of ALCOVE with perceived similarities in the last learning block (as well as in previous learning blocks) was better for 'A' exemplars than for 'B' exemplars for stimulus group a (A: 94% correct, B: 89%), and better for 'B' exemplars than for 'A' exemplars for group c (A: 85%, B: 95%). When ALCOVE with perceived similarities learned the quadrant rule for stimulus group b, its performance showed the expected difference between 'E' and 'F' exemplars (85% and 80%, respectively).

Across all 10 datapoints in Fig. 4, the correlation between behavioral performance and the performance of ALCOVE with perceived similarities was 0.82, compared to a correlation of 0.57 between behavioral performance and the performance of ALCOVE with parametric similarities.

Thus, overall, the power of ALCOVE in predicting which stimuli will be learned best is greatly enhanced when the network input reflects the perceptual stimulus space, although small quantitative differences remain between the behavioral results and the ALCOVE predictions. Note that this fit of ALCOVE with categorization behaviour is all the more striking since the perceptual stimulus space and the categorization learning performance was obtained in different groups of animals.

3.3. ITCOVE implementations

The hidden units with mathematically defined distance functions in ALCOVE were replaced in ITCOVE by the response profiles of 124 neurons recorded in the inferotemporal cortex of two macaque monkeys as described by Op de Beeck et al. (2001). Three different networks were constructed since the responses of each neuron were recorded with three stimulus groups. Results were usually consistent between stimulus groups and, unless otherwise noted, the average performance is shown here.

Technically, ITCOVE is no longer a three-layer network since only one set of weights is applied, to the connections between each neuron and the two output units. A version of ALCOVE without adaptable attention strengths would also be a two-layer instead of a three-layer network. The contribution of attention strengths to category learning in the present context is doubtful for several reasons. First, the fit between ALCOVE and monkey behavior was the same whether or not attention strengths were adaptable in the model. Second, our previous study (Op de Beeck et al., 2001) showed that the tuning of IT neurons was not influenced by learning a linear, quadrant, or arbitrary rule. As such, there was no indication that attention strengths were adapted during learning. Third, as we will argue in more detail later, some of the findings in the liter-

ature that seem to confirm the importance of attention strengths might be accounted for without adaptive attention strengths once one accepts that stimuli are represented with a diversity of tuning functions.

3.4. ITCOVE and monkey behavior

In a first implementation of ITCOVE, we used the mean response of each neuron across trials. The order in which ITCOVE learned the three category rules (Fig. 6A) was similar to the ranking found behaviorally. However, unlike the monkey performance, the difference in ITCOVE performance between the quadrant and the arbitrary rule was very small. As such, the performance of ITCOVE resembles the findings of ALCOVE when hidden units were broadly tuned (small c value). Thus, one way to interpret the ITCOVE performance is that it was caused by a relatively low selectivity of the recorded neurons, and that ITCOVE provided no acceptable fit with monkey behavior.

However, we implemented ITCOVE with the mean response of each neuron across trials, and as such we neglected the trial-by-trial variance in neuronal responses. Some of the averaged neuronal tuning curves could be relatively unreliable, and if ITCOVE were to rely on such a subset of neurons to learn one of the rules (e.g., the arbitrary rule), then this could enhance performance in an artificial way. This problem was addressed by simulating the trial-by-trial variance of the neuronal responses (see Section 2). The introduction of trial-by-trial response variance increased the correspondence between model and monkey performance (Fig. 6B), with a larger difference between the quadrant and arbitrary rule than observed in the implementation with no response variance. The learning of the linear rule was hardly affected by the presence of response

variability (overall reduction in percentage correct of 3%), but more pronounced effects were found for the quadrant and arbitrary rule (reductions of 11% and 18%, respectively).

As found for monkey behavior and ALCOVE implemented using perceived similarities, ITCOVE learned exemplars 'A' better than exemplars 'B' with stimulus group a, but exemplars 'B' better than exemplars 'A' with stimulus group c (Fig. 4A). Furthermore, ITCOVE showed better performance for stimuli E than for stimuli F. Across all 10 datapoints in Fig. 4A, the correlation between behavioral performance and the performance of ITCOVE was 0.92. We suggested previously (Op de Beeck et al., 2001) that the deviations in the perceptual space of stimuli compared to the parametric space were also present in the neuronal representation of the stimuli. The correspondence between ITCOVE performance, monkey behavior, and ALCOVE based on perceived but not parametric similarity indicates that such deviations between neuronal and parametric space are also an important factor for predicting category learning.

3.5. ITCOVE and neuronal tuning properties

As with previous implementations of ALCOVE described in the literature, our implementation involved few hidden units (one for each exemplar) with a regular exponential tuning in stimulus space. Learning affected which output unit a hidden unit is connected to by a positive and a negative weight, but the absolute values of these weights did not differ between hidden units. So, in ALCOVE, the weights to the two output units of different hidden units had nearly the same absolute value, only the sign differs. As a consequence, there was no substantial
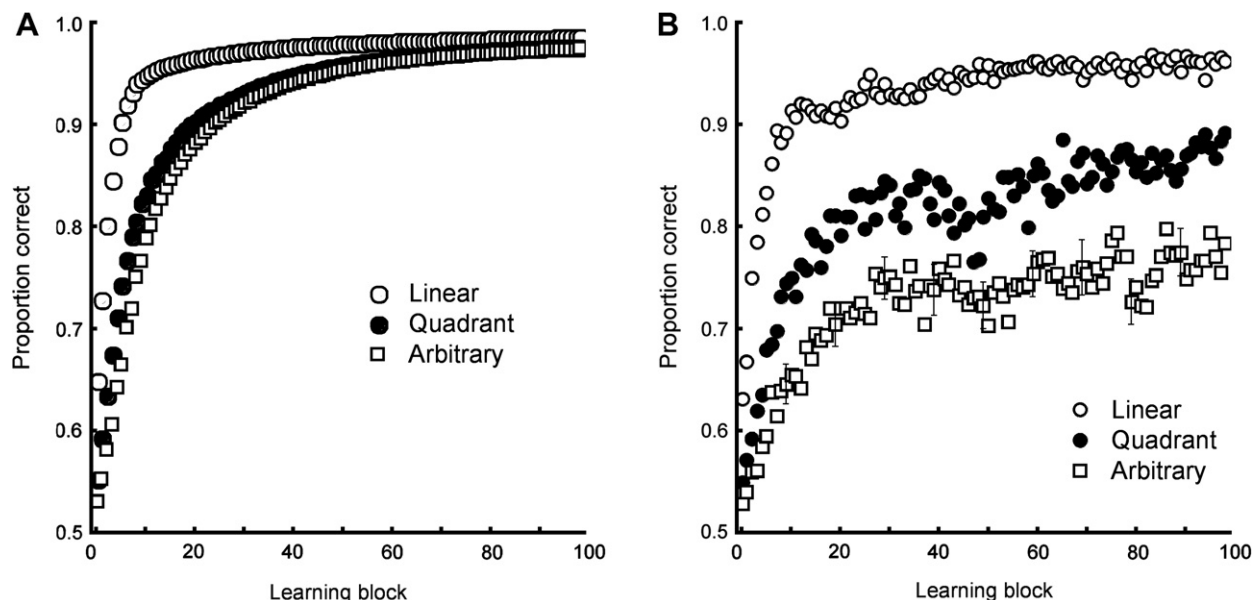

Fig. 6. Performance of ITCOVE for the three category rules as a function of time, (A) without trial-by-trial variance and (B) with trial-by-trial variance. Error bars in (B) reflect the typical range in standard error of the mean across simulations.

selection of hidden units in that there was almost no difference between hidden units in their connection strength to the output layer—*connection strength* being defined here as the absolute value of the weights to the two output units. This indicates that all hidden units were almost equally useful for learning each of the three category rules. However, we might have a completely different situation in the ITCOVE network as this network included many units with a wide diversity of tuning properties. So we wondered whether and how the tuning properties of a neuron influence its usefulness during category learning.

One such tuning property is the tuning width or selectivity of a neuron, which we measured by computing the 'depth of selectivity' or 'DOS' (Rainer & Miller, 2000). This index varies between 0 (the same response to all stimuli) and 1 (response to only one stimulus). The mean DOS of the neurons was .38 (for comparison, the DOS for the hidden units in ALCOVE was .38 with $c = 1.2$). We computed the correlation between DOS and the connection strength of each neuron in ITCOVE after the last learning block with each category rule. The results revealed a higher connection strength for more selective units when ITCOVE learned a linear or a quadrant rule (resulting in significant correlations between connection strength and DOS, $r = .24$ and .21, respectively), but not significantly so in the case of an arbitrary rule ($r = .07$, ns). These correlations are consistent with the fact that the standard version of ALCOVE performed best with very selective hidden units.

A second tuning property is regularity, that is, the degree to which the tuning of a neuron can be described by a function with a single optimum in stimulus space and a gradual fall-off towards other positions in stimulus space. Op de Beeck et al. (2001) used RR as an index for tuning regularity, a value that quantifies to what degree the polar plots of neuronal tuning deviate from a unimodal sinusoidal modulation. A large RR value means that the ordinates given by the responses to the 8 stimuli can be well fitted by a sinus when the responses are plotted as a function of the stimulus numbers 1–8 (in this way, the two-dimensional circular stimulus configuration is converted to one dimension as is often done for orientation tuning). The mean RR for the hidden units in ALCOVE with $c = 1.2$ (which are units with a very regular tuning) was 3.8. The median RR value in the neuronal population and in ITCOVE was 1.7. There was a rule-dependent correlation between RR and connection strength in ITCOVE after learning. In the case of a linear category rule, there was a positive correlation (.41), no significant correlation was found for the quadrant rule (−.14,

ns), and a significant negative correlation (−.22) was present for the arbitrary rule.

Thus, the tuning width and the tuning regularity of a neuron have implications for the neuron's usefulness for learning different category rules. This is supported by a comparison of the performance of networks that are composed of units with selected tuning properties. Networks were constructed that consisted of 40 units randomly selected from the population of 124 neurons but restricted to either high selectivity (DOS>.4), low selectivity (DOS < .4), regular tuning (RR>1.7), or irregular tuning (RR < 1.7), or with no restrictions at all. Table 1 shows the performance of these networks after 60 learning blocks. Compared to the unrestricted control condition, a restricted selection resulted in higher or lower performance in a way that agreed with the correlations, described above, between tuning properties and connection strength. A high DOS network always outperformed a low DOS network, showing that a high selectivity is beneficial for category learning. Learning a linear category rule was improved with a high tuning regularity, while the other two category rules were learned best with irregular tuning properties.

Tuning width and regularity are two properties of a neuronal tuning curve that are defined in stimulus space in a way that is independent of the input/output mapping (category rule) that has to be learned. Most probably, an index that takes into account how well a neuronal tuning curve separates all exemplars of a given category from all other stimuli would be a better choice for predicting the connection strength of a neuron. A straightforward index for category selectivity of a neuron is CS (see formula (4) in Section 2), which is the difference between the sum of the responses of a neuron for all exemplars of one category minus the sum of the responses for all other stimuli. The CS predicted the connection strength after learning between a neuron and an output unit almost perfectly (mean $r = .95$). Thus, the basic factor determining the connection strengths in ITCOVE is the degree to which a neuron responds more to all exemplars from one category than to all exemplars from the other category.

4. Discussion

We showed that the performance of the ALCOVE model fits monkey categorization behavior only when the input to the model is based on the perceived similarities which can differ from the parametric similarities. The fact that discrepancies between perceived and parametric shape

Table 1
Model performance (proportion correct and SEM across simulations) with various types of tuning properties as a function of rule

Rule	High selectivity	Low selectivity	High regularity	Low regularity	No selection
Linear	.90 (.01)	.81 (.02)	.91 (.01)	.73 (.02)	.86 (.01)
Quadrant	.73 (.02)	.66 (.02)	.67 (.01)	.73 (.02)	.70 (.01)
Arbitrary	.68 (.02)	.63 (.02)	.58 (.01)	.67 (.02)	.63 (.02)

changes needed to be incorporated into ALCOVE for it to reproduce categorization behavior confirms the importance of these differences for behavioral performance. First of all, this result confirms the pervasiveness of these incongruencies across individuals, since the categorization task and the same/different task involved different subjects. Second, the incongruencies influenced behavioral performance in two fundamentally different tasks: same/different and categorization. These results illustrate that perceived shape differences need to be quantified since they are not necessarily equivalent to parametric or physical shape differences (for some stimulus sets it has been suggested that such incongruencies do not exist, see Peters, Gabbiani, & Koch, 2003).

We showed in addition that the correspondence between network performance and behavior is also high when the intermediate stimulus representation in the ALCOVE model is replaced with the responses of real IT neurons ('ITCOVE' model), at least when the trial-by-trial variance of these responses is taken into account. This correspondence is not trivial, as the behavioral and neuronal data were obtained in different groups of animals, and given that our initial ALCOVE implementation (based on parametric differences) failed to predict several aspects of behavior. This result provides evidence for a link between category learning and neuronal responses in IT cortex. A general link between behavior in tasks that require shape and object recognition and neural processing in IT cortex is also supported by recent evidence that the trial-by-trial variation in neural responses is related to behavioral choices (Uka, Tanabe, Watanabe, & Fujita, 2005), and by behavioral effects of microstimulation of face cells (Afraz, Kiani, & Esteky, 2006).

The successful combination of a computational network with neurophysiological data supports the idea that distance functions constitute a type of basis function set that can account for behavioral data and that is biologically plausible. However, the neurons of which the responses are incorporated in ITCOVE displayed a high diversity in tuning properties, more so than the units in the ALCOVE model. We showed that the diversity in the properties of actual neuronal tuning profiles is associated with a strong selection of which neurons influence the output of the network. A simple index for category selectivity (CS) predicted almost perfectly whether or not a particular unit would acquire a high weight during supervised category learning by ITCOVE. There is a close relationship between this CS index and the ordinal difference index used by Thomas, Van Hulle, and Vogels (2001). Their difference index predicted the importance of a unit for an unsupervised clustering of stimuli into natural categories such as trees versus non-trees. Thus, the importance of a given neuron for learning a particular category distinction appears to depend upon the same tuning properties in both supervised and unsupervised categorization. The degree to which a category distinction corresponds to an intrinsic clustering in stimulus space will be related to the relative frequency

of neurons with a high CS value. These neurons are relatively abundant for natural categories such as trees versus non-trees or animate versus inanimate (Kiani, Esteky, Mirpour, & Tanaka, 2007), probably somewhat less so for artificially defined linear category rules, and they are rare for more complex rules such as the quadrant and arbitrary rules. Nevertheless, the diversity in real tuning profiles increases the odds that neurons with a high CS index exist even for complex category rules.

4.1. Neuronal selection: At which stage and for which stimuli does it happen?

Is the remarkably high diversity in neuronal tuning properties and the associated neuronal selection relevant, that is, should we care about the observation that the stimulus representation in the real brain is more diverse and complex than the representations included in computational models? In our data, we found a close agreement in the output performances of ALCOVE (without such diversity) and ITCOVE (with such diversity), so maybe we should conclude that the tuning diversity and the related selection/weighting of neurons might be irrelevant after all. However, our behavioral paradigm was not designed to assess the behavioral effects of this diversity, nor to assess the effects of attention strengths. In fact, evidence in the literature for the importance of attention strengths (e.g., Kruschke, 1992; Nosofsky, 1984) could be used to argue in favor of the behavioral implications of tuning diversity. Ashby and Lee (1991) have already argued that behavioral data do not allow one to distinguish selection mechanisms lying between input and intermediate representations (attention strengths) from later selection mechanisms (the neuronal selection we have discussed in the present paper). A similar discussion applies to the interpretation of learning effects in other behavioral paradigms (e.g., Dosher & Liu, 1999). Neurophysiological data have suggested that attentional mechanisms influence the coding of stimuli in visual cortex (for a review, see Treue, 2001). Most relevant is a study by Sigala and Logothetis (2002) arguing that category learning enhances the selectivity of IT neurons for dimensions that are relevant for the learned category rule. Although these studies suggest that adaptable attention strengths do play a role in some circumstances, we cannot exclude the possibility that part of the learning-induced increase in behavioral sensitivity for relevant stimulus properties would be caused by selection at later stages in the information processing hierarchy (Ghose, Yang, & Maunsell, 2002; Mirabella et al., 2007). Most likely the two types of selection (early and late selection) should both be taken into consideration.

One possible explanation of the fact that some but not all relevant studies show attention effects is related to the use of different sorts of stimulus dimensions. Sigala and Logothetis (2002) used category rules for which the relevant stimulus differences were located in other parts of the stimuli (e.g., the eyes of a face) compared to the irrele-

vant stimulus differences (e.g., changes in the mouth and the nose). As a consequence, spatial attention could explain these results. Op de Beeck et al. (2001) used parametric dimensions that changed the *global* contour of the shapes, and spatial attention could not enhance the processing of relevant compared to irrelevant stimulus differences. So, the degree to which dimensions can be selectively attended to is determined by the specific combination of dimensions. In fact, this proposal was made long ago by Garner (1974), who labeled dimensions that can be selectively attended to as 'separable' dimensions, and other dimensions as 'integral'. The stimuli of Sigala and Logothetis (2002) vary along separable dimensions, whereas we showed in another study that our stimuli vary along integral dimensions (Op de Beeck, Wagemans, & Vogels, 2003). Thus, it is conceivable that the discrepancy between the results of these two studies is related to the distinction between separable and integral dimensions. With integral dimensions, attention strengths might not be adapted and neuronal selection might be restricted to a later stage, that is, the stage at which visual responses are mapped onto the output categories (the connection between hidden layer and output layer in ALCOVE).

4.2. How useful is the concept of "distance functions"?

Is the concept of distance functions still valid given the diversity in IT tuning profiles? One source of diversity was tuning selectivity. This factor is easily incorporated into the framework of distance functions by changing the variance of the exponential tuning between units. Furthermore, neurons can vary in the dimensions that they are most sensitive to. As such, neuronal selection is able to increase the sensitivity for a relevant stimulus dimension relative to an irrelevant dimension. Op de Beeck et al. (2003) (see also Kayaert, Biederman, Op de Beeck, & Vogels, 2005; Wagemans, Wichmann, & Op de Beeck, 2005) suggested how the distinction between integral and separable dimensions could be related to the way in which different units differ in their sensitivity for individual dimensions.

A second source of diversity was found in tuning regularity, which is more difficult to incorporate into the framework of distance functions. Can irregular tuning curves still be regarded as distance functions? An important point of consideration here is that the responses of these neurons were recorded within a restricted stimulus space. Op de Beeck et al. (2001) argued that a parametric variation of similarity in stimulus space is necessary to determine whether the responses of neurons are related to stimulus distance within this space. However, even with this sort of parametric manipulation, we know only what a neuron's responses are within that limited stimulus space. Given this restricted knowledge, it is possible that a neuron that seems to be a counterexample of distance functions in our restricted stimulus space is in fact tuned to stimulus similarity in another stimulus space characterized by other dimen-

sions. In the present case, the parametric dimensions characterize global shape changes, but there are undoubtedly some neurons in our sample that are sensitive to changes in only a restricted part of the stimuli, or to higher-order shape characteristics. The tuning profiles of such neurons could be irregular in the stimulus space that is characterized by global shape changes, while at the same time being regular in the broader stimulus space that also takes into account other shape dimensions.

This account of the diversity of tuning profiles is biologically plausible. It has been known for several decades that the primate brain processes visual input in a number of hierarchical steps and that neurons at each level are tuned to a wide variety of stimulus dimensions and features. Given this neurophysiological background, we introduce no new concepts here. However, spatial models of categorization simplify how stimuli are represented, and this simplification explains some of the failures of these models in the cognitive literature (e.g., Erickson & Kruschke, 1998; Lee & Navarro, 2002; Nosofsky & Johansen, 2000). A spatial representation of a set of stimuli contains an explicit representation of only a low number of dimensions (as many as the dimensionality of the stimulus space), and as such these models are not able to fit behavioral data in cases where subjects solve a category rule by focusing upon a dimension not explicitly included in the spatial representation. Spatial representations in models of visual categorization have to become richer and more biologically plausible in order to overcome these restrictions. For example, they should incorporate the fact that visual stimuli are processed in a hierarchical manner and that they can be represented at multiple levels. As in the study of Peters et al. (2003), future computational work should combine finer-grained models of how successive stimulus representations are constructed from the visual input (e.g., Riesenhuber & Poggio, 1999; Steyvers & Busey, 2001) with the learning rules and stimulus/response mapping proposed in more abstract models of categorization.

Recent studies have provided evidence for yet another difference between the implementation of distance functions in the brain and in computational models. These studies showed that the preferred stimuli for which IT neurons are tuned tend to be located at extreme positions in the stimulus space that is shown during the experiments (De Baene, Premereur, & Vogels, 2007; Kayaert et al., 2005; Leopold, Bondar, & Giese, 2006). As a result, most tuning curves are monotonic in this stimulus space. While we defined the distance functions in ALCOVE as Gaussian functions, centered at the 8 shapes in each stimulus group, these functions are also monotonic because all 8 shapes were located at the borders of the stimulus space. Thus, this spatial arrangement does not allow to distinguish bell-shaped from monotonic tuning functions – and the two types of functions are "distance functions" by definition (response declines as a function of distance to the preferred stimulus). However, many category learning experiments include exemplars that are located at more central posi-

tions in stimulus space, and in such cases exemplar models assume a stimulus representation that includes the same number of units for the central and for the more extreme exemplars, while prototype models assume units centered at the category prototypes. Clearly, the distribution of preferred stimuli for IT neurons does not fit with either of these models.

4.3. Other stages involved in category learning

We focused on the representation of stimuli as found in IT cortex and its relation to how stimuli are represented in computational models. Of course, it is just as important to investigate how and where the category rule is represented within the brain and whether the properties of this rule representation can be accounted for by computational models (e.g., the output layer in the ALCOVE model). After training has been completed, each category unit in the output layer of ALCOVE responds maximally to exemplars of one category and minimally to exemplars of other categories. This corresponds to results obtained in prefrontal cortex with single-cell recordings in macaque prefrontal cortex (Freedman, Riesenhuber, Poggio, & Miller, 2001; Jiang et al., 2007). For example, neurons in monkey dorsolateral prefrontal cortex are very sensitive for the category border, with a low sensitivity for intra-category differences combined with a high sensitivity for inter-categorical differences (at least at the population level; individual units showed less consistent category specificity). In addition to prefrontal cortex, other brain areas (e.g. basal ganglia) have been implicated in the stimulus/response mapping in category learning (e.g., Ashby & Waldron, 1999; Freedman & Assad, 2006; Vogels, Sary, Dupont, & Orban, 2002).

Acknowledgments

References

Afraz, S. R., Kiani, R., & Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature, 442*, 692–695.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General, 120*, 150–172.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review, 3*, 363–378.

Baldi, P., & Heiligenberg, W. (1988). How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers. *Biological Cybernetics, 59*, 313–318.

De Baene, W., Premereur, E., & Vogels, R. (2007). Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. *Journal of Neurophysiology, 97*, 2900–2916.

Dosher, B., & Liu, Z. (1999). Mechanisms of perceptual learning. *Vision Research, 39*, 3197–3221.

Edelman, S. (1999). *Representation and recognition in vision*. Cambridge, MA: MIT Press.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*, 107–140.

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature, 443*, 85–88.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*, 312–316.

Garner, W. R. (1974). *The processing of information and structure*. New York: Wiley.

Ghose, G. M., Yang, T., & Maunsell, J. H. (2002). Physiological correlates of perceptual learning in monkey V1 and V2. *Journal of Neurophysiology, 87*, 1867–1888.

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron, 53*, 891–903.

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology, 45*, 482–553.

Kayaert, G., Biederman, I., Op de Beeck, H., & Vogels, R. (2005). Tuning for shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience, 22*, 212–224.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology, 97*, 4296–4309.

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

Kruschke, J. K. (2005). Category learning. In K. Lamberts & R. L. Goldstone (Eds.), *The handbook of cognition* (pp. 183–201). London: Sage.

Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review, 9*, 43–58.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature, 442*, 572–575.

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

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19*, 577–621.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238.

Mirabella, G., Bertini, G., Samengo, I., Kilavik, B. E., Frili, D., Della Libera, C., & Chelazzi, L. (2007). Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. *Neuron, 54*, 303–318.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 10*, 104–114.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7*, 375–402.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Macaque inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience, 4*, 1244–1252.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2003). The effect of category learning on the representation of shape: Dimensions can be biased but not differentiated. *Journal of Experimental Psychology: General, 4*, 491–511.

Peters, R. J., Gabbiani, F., & Koch, C. (2003). Human visual object categorization can be described by models with low memory capacity. *Vision Research, 43*, 2265–2280.

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

Pouget, A., & Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience, 3*, 1192–1198.

Rainer, G., & Miller, E. K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron, 27*, 179–189.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Sigala, N., Gabbiani, F., & Logothetis, N. K. (2002). Visual categorization and object representation in monkeys and humans. *Journal of Cognitive Neuroscience, 14*, 187–198.

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature, 415*, 318–320.

Steyvers, M., & Busey, T. (2001). Predicting similarity ratings to faces using physical descriptions. In M. Wenger & J. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. Hillsdale, NJ: Erlbaum.

Thomas, E., Van Hulle, M. M., & Vogels, R. (2001). Encoding of categories by noncategory-specific neurons in the inferior temporal cortex. *Journal of Cognitive Neuroscience, 13*, 190–200.

Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences, 24*, 295–300.

Uka, T., Tanabe, S., Watanabe, M., & Fujita, I. (2005). Neural correlates of fine depth discrimination in monkey inferior temporal cortex. *Journal of Neuroscience, 25*, 10796–10802.

Verguts, T., Ameel, E., & Storms, G. (2004). Measures of similarity in models of categorization. *Memory & Cognition, 32*, 379–389.

Vogels, R., Sary, G., Dupont, P., & Orban, G. A. (2002). Human brain regions involved in visual categorization. *Neuroimage, 16*, 401–414.

Wagemans, J., Wichmann, F. A., & Op de Beeck, H. (2005). Visual perception I: Basic Principles. In K. Lamberts & R. L. Goldstone (Eds.), *The handbook of cognition* (pp. 3–47). London: Sage.

Worgötter, F., & Eysel, U. T. (1987). Quantitative determination of orientational and directional components in the response of visual cortical cells to moving stimuli. *Biological Cybernetics, 57*, 349–355.