MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL AND COMPUTATIONAL LEARNING
DEPARTMENT OF BRAIN AND COGNITIVE SCIENCES

A.I. Memo No. 1679
C.B.C.L. Paper No. 183
December 14, 1999

A note on object class representation and categorical perception

Maximilian Riesenhuber and Tomaso Poggio

Abstract

We present a novel scheme ("Categorical Basis Functions", CBF) for object class representation in the brain and contrast it to the "Chorus of Prototypes" scheme recently proposed by Edelman [4]. The power and flexibility of CBF is demonstrated in two examples. CBF is then applied to investigate the phenomenon of Categorical Perception, in particular the finding by Bülthoff *et al.* [2] of categorization of faces by gender without corresponding Categorical Perception. Here, CBF makes predictions that can be tested in a psychophysical experiment. Finally, experiments are suggested to further test CBF.

This report describes research done within the Center for Biological and Computational Learning in the Department of Brain and Cognitive Sciences and in the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology. This research is sponsored by a grant from Office of Naval Research under contract No. N00014-93-1-3085, Office of Naval Research under contract No. N00014-95-1-0600, National Science Foundation under contract No. IIS-9800032, and National Science Foundation under contract No. DMS-9872936. Additional support is provided by: AT&T, Central Research Institute of Electric Power Industry, Eastman Kodak Company, Daimler-Benz AG, Digital Equipment Corporation, Honda R&D Co., Ltd., NEC Fund, Nippon Telegraph & Telephone, and Siemens Corporate Research, Inc. M.R. is supported by a Merck/MIT Fellowship in Bioinformatics.

1 Introduction

Object categorization is a central yet computationally difficult cognitive task. For instance, visually similar objects can belong to different classes, and conversely, objects that appear rather different can belong to the same class. Categorization schemes may be based on shape similarity (*e.g.,* "human faces"), on conceptual similarity (*e.g.,* "chairs"), or on more abstract features (*e.g.,* "Japanese cars", "green cars"). What are possible computational mechanisms underlying categorization in the brain?

Edelman has recently presented an object representation scheme called "Chorus of Prototypes" (COP) [4] where objects are categorized by their similarities to reference shapes, or "prototypes". While this categorization scheme is of appealing simplicity, the reliance on a single metric in a global *shape space* imposes severe limitations on the kinds of categories that can be represented. We will discuss these shortcomings and present a more general model of object categorization along with a computational implementation that demonstrates the scheme's capabilities, relate the model to recent psychophysical observations on categorical perception (CP), and discuss some of the model's predictions.

2 Chorus of Prototypes (COP)

In COP, "the stimulus is first projected into a high-dimensional measurement space, spanned by a bank of [Gaussian] receptive fields. Second, it is represented by its similarities to reference shapes" ([4], p. 112, caption to Fig. 5.1).

The categorization of novel objects in COP proceeds as follows (ibid., p. 118):

1. A category label is assigned to each of the training objects ("reference objects"), for each of which an RBF network is trained to respond to the object from every viewpoint;

2. a test object is represented by the activity pattern it evokes over all the output units of the reference object RBF networks (i.e., the "similarity to reference shapes" above);

3. categorization is performed using the activity pattern and the *labels associated with the output units of the reference object RBF networks.* Categorization procedures explored were winner-take-all, and k-nearest-neighbor using the training *views* (this time taking the prototypes to be not the objects but the object views), *i.e.,* the centers of individual RBF units in each network, with the class label in this cased based on the label of the majority of the k closest stored views to the test stimulus.

The appealingly simple design of COP also seems to be its most serious limitation: While a representation based solely on shape similarities seems to be suited for the taxonomy of some novel objects (cf. Edelman's example of the description of a giraffe as a "cameleopard" [4]), such a representation appears too impoverished when confronted with objects that can be described on a variety of levels: A car, for instance, can look like several other cars (and also unlike many other objects), but it could also be described as a "cheap" car, a "green"* car, a "Japanese" car, an "old" car, etc. — different qualities that are not simply or naturally summarized by shape similarities to individual prototypes but nevertheless provide useful information to classify or discriminate the object in question from other objects of similar shape. The fact that an object can be described in such abstract categories, and that this information appears to be used in recognition and discrimination, as indicated by the findings on categorical perception (see below), calls for an extension of Chorus to permit the use of several categorization schemes in parallel, to allow the representation of an object within the framework of a whole dictionary of categorization schemes that offers a more natural description of an object than one global shape space.

While Edelman ([4], p. 244) suggests a refinement of Chorus where weights are assigned to different dimensions driven by task demands, it is not clear how this can happen in one global shape space if two objects can be judged as very similar under one categorization scheme but as rather different under another (as, for instance, a chili pepper and a candy apple in terms of color and taste, resp.). Use of different categorization schemes appears to require reversible temporary warping of shape space depending on which categorization scheme is to be used, which runs counter to the notion of one general representational space.

3 A Novel Scheme: Categorical Basis Functions (CBF)

In CBF, the receptive fields of stimulus-coding units in measurement space are not constrained to lie in any specific class — unlike in COP, there are no class labels associated with these units. The input ensemble drives the *unsupervised*, *i.e.,* task-independent learning of receptive fields. The only requirement is that the receptive fields of these *stimulus space-coding units (SSCUs)* cover the stimulus space sufficiently to allow the definition of arbitrary classification schemes on the stimulus space (in the simplest version, "learning" just consists in storing all the training examples by allocating an SSCU to each training stimulus).

These SSCUs in turn serve as inputs to units that are trained on categorization tasks in a *supervised* way — in fact, if each training stimulus is represented by one SSCU, then the network would be identical to a standard radial basis function (RBF) network. Figure 1 illustrates the CBF scheme.

Novel stimuli in this framework evoke a characteristic activation pattern over the existing categorization units (as well as over the SSCUs). In fact, CBF can be seen as an extension of COP: instead of a representation based on similarity in a global shape space alone (as in "the object looks like *xyz*",

*The idea of representing color through similarities to prototype objects seems especially awkward considering that it first requires the build-up of a library of objects of a certain color with the sole purpose of allowing to "average out" object shape.

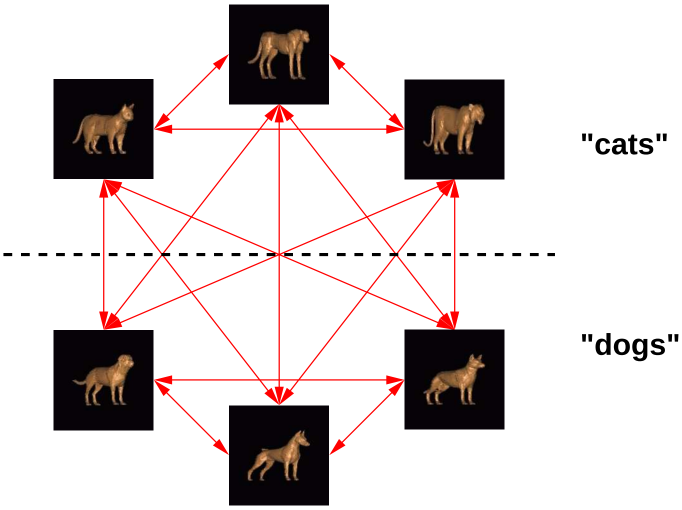

Figure 2: Illustration of the cat/dog stimulus space. The stimulus space is spanned by six objects, three "cats" and three "dogs". Our morphing software [13] allows us to generate 3D objects that are arbitrary combinations of the six prototypes. The lines show possible morph directions between two prototypes each, as used in the test set.

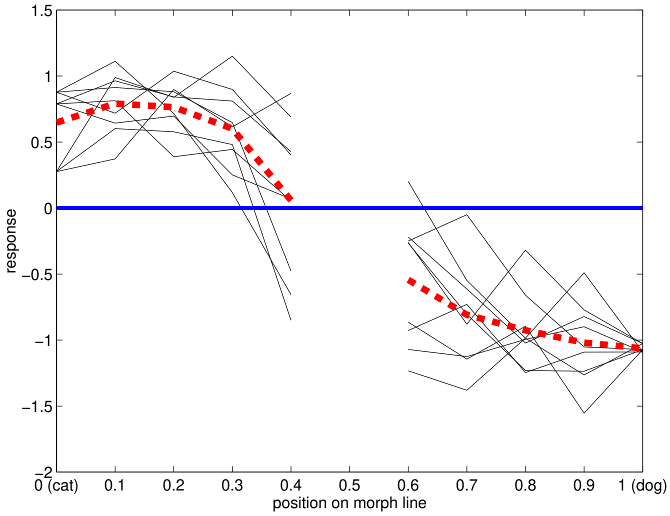

Figure 3: Response of the categorization unit (based on 144 SSCU, 256 afferents to each SSCU, $\sigma_{SSCU} = 0.7$) along the nine class boundary-crossing morph lines. All stimuli in the left half of the plot are "cat" stimuli, all on the right-hand side are "dogs" (the class boundary is at 0.5). The network was trained to output 1 for a cat and -1 for a dog stimulus. The thick dashed line shows the average over all morph lines. The solid horizontal line shows the class boundary in response space.

where *x,y,z* can be objects for which individual units have been learned), abstract features, which are the result of prior category learning, are equally valid for the description of an object (as in "the object looks expensive/old/pink"). Hence, an object is not only represented by expressing its similarity to learned shapes but also by its membership to learned categories, providing a natural basis for object description.

In the proof-of-concept implementation described in the following, SSCUs are identical to the view-tuned units from the model by Riesenhuber and Poggio [12] (in reality, when objects can appear from different views, they could also be view-invariant — note that the view-tuned units are already invariant to changes in scale and position [12]). For simplicity, the unsupervised learning step is done using k-means, or just by storing all the training exemplars, but more refined unsupervised learning schemes, which better reflect the structure of the input space, such as mixture-of-Gaussians or other probability density estimation schemes, or learning rules that provide invariance to object transformations [15] are likely to improve performance. Similarly, the supervised learning scheme used (Gaussian RBF) can be replaced by more biologically plausible or more sophisticated algorithms (see discussion).

3.1 An Example: Cat/Dog Classification

To illustrate the capabilities of CBF, the following simulation was performed: We presented the hierarchical object recognition system (up to the C2 layer) of Riesenhuber & Poggio [12] with 144 randomly selected morphed animal stimuli, as used in a very recent monkey physiology experiment [6] (see Fig. 2).

A view-tuned model unit was allocated for each training stimulus, yielding 144 view-tuned units (results were similar if the 144 stimuli were clustered into 30 units using k-means, see appendix). The activity patterns over the 144 units to each of the 144 stimuli were used as inputs to train a gaussian RBF output unit, using the class labels 1 for cat and -1 for dog as the desired outputs. The categorization performance of this unit was then tested with the same test stimuli as in the physiology experiment (which were not part of the training set). More precisely, the testing set consisted of the 15 lines through morph space connecting each of the prototypes, each subdivided into 10 intervals, with the exclusion of the stimulus at the mid-points (which in the case of lines crossing the class boundary would lie right on the class boundary, with an undefined label), yielding a total of 126 stimuli. Figure 3 shows the response of the categorization unit to the stimuli on the category boundary-crossing morph lines, together with the desired label. A categorization was counted as correct if the sign of the network output was identical to the sign of the class label.

Performance on the training set was 100% correct, performance on the test set was 97%, comparable to monkey performance, which was over 90% [6]. The four categorization errors the model makes lie right at the class boundary.

3.2 Introduction of parallel categorization schemes

To demonstrate how different classification schemes can be used in parallel within CBF, we also trained a second network to perform a different categorization task on the same stimuli. The stimuli were resorted into three classes, each based on one cat and one dog prototype. For this categorization task, three category units were trained (on a training set of 180 animal morphs, taken from training sets of an ongoing

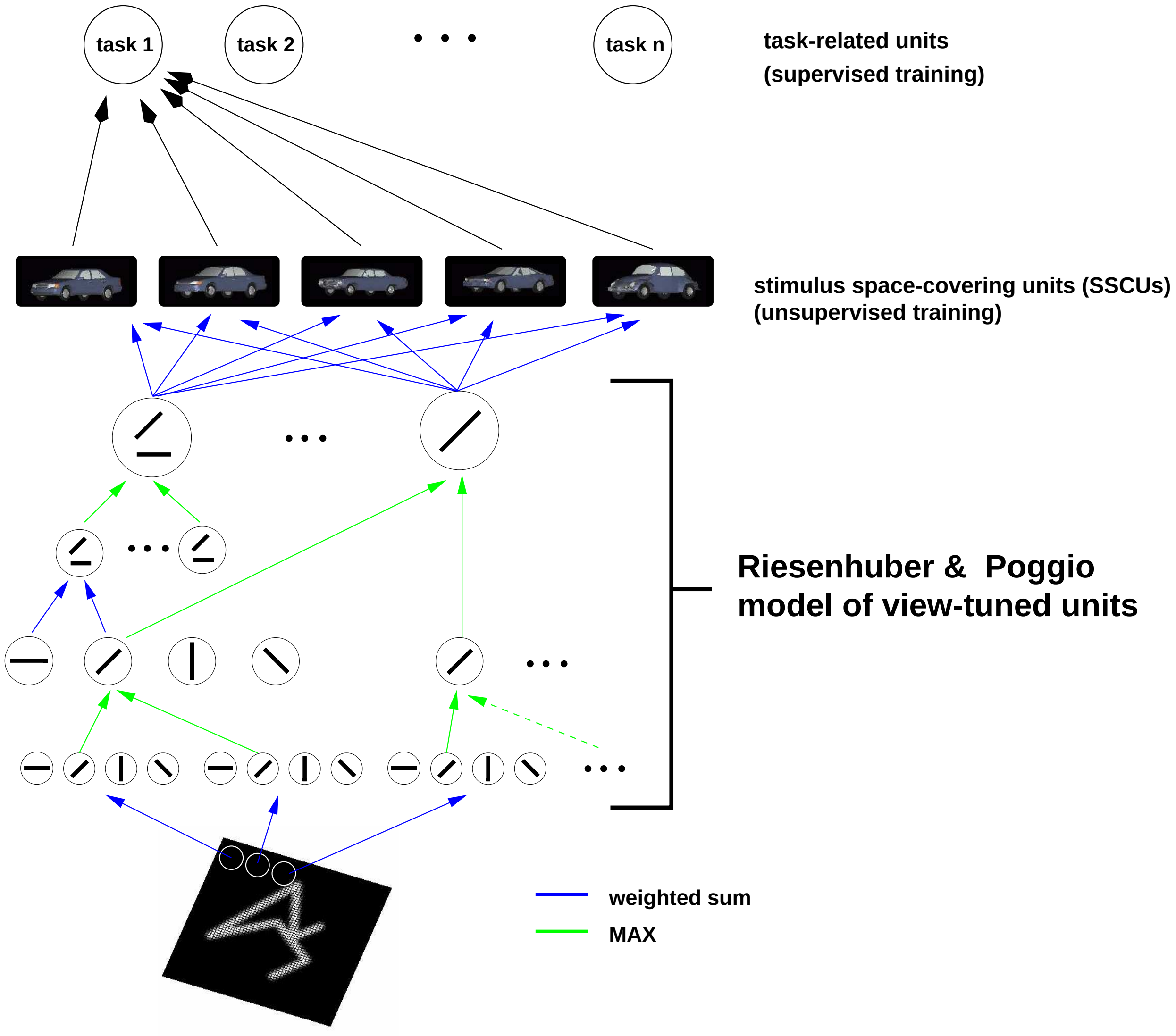

Figure 1: Cartoon of the CBF categorization scheme, illustrated with the example domain of cars. Stimulus space-covering units (SSCUs) are the view-tuned units from the model by Riesenhuber & Poggio [12]. They self-organize to respond to representatives of the stimulus space so that they "cover" the whole input space, with no explicit information about class boundaries. These units then serve as inputs to task-related units that are trained in a supervised way to perform the categorization task (*e.g.,* to distinguish American-built cars from imports, or compacts from sedans etc.). In the proof-of-concept implementation described in this paper, the unsupervised learning stage is done via k-means clustering, or just by storing all the training exemplars, and the supervised stage consists of an RBF network.

physiology project), each one to respond at a level of 1 for stimuli belonging to "its" class and a level of -1 for stimuli from the other two classes.[†] Each category unit received input from the same 144 SSCUs as the cat/dog category unit described above.

As mentioned, it is an open question how to best perform multi-class classification. We evaluated two strategies: i) categorization is said to be correct if the maximally activated category unit corresponds to the true class ("max" case); ii) categorization is correct if the signs of the three category units are equal to the correct answer ("sign" case).

Performance on the training set in the "max" as well as in the "sign" case was 100% correct. On the testing set, performance using the "max" rule was 74%, whereas the performance for the "sign" rule was 61% correct, the lower numbers on the test set as compared to the cat/dog task reflecting the increased difficulty of the three-way categorization. We are currently training a monkey on the same categorization task, and it will be very interesting to compare the animal's performance on the test set to the model's performance.

4 Interactions between categorization and discrimination: Categorical Perception

When discriminating objects, we commonly do not only rely on simple shape cues but also take more complex features into account. For example, we can describe a face in terms of its expression, its age, gender etc. to provide additional information that can be used to discriminate this face from other faces. This suggests that training on *categorization* tasks could be of use also for object *discrimination*.

The influence of categories on perception is expected to be especially strong for stimuli in the vicinity of a class boundary: In the cat/dog categorization task described in the previous paragraph, the goal was to classify all members of one class the same way, irrespective of their shape. Hence, when presented with two stimuli from the same class, the categorization result will ideally not allow to discriminate between the two stimuli. On the other hand, two stimuli from different classes are labelled differently. Thus, one would expect greater accuracy in discriminating stimulus pairs from different classes than pairs belonging to the same class (note that in this paper we are not dealing with the discrimination process itself — while several mechanisms have been proposed, such as a representation based directly on the SSCU activation pattern, or one based on the activity pattern over prototypes such as view-invariant RBF units [4, 9], we in this section only discuss how prior training on categorization tasks can provide additional information to the discrimination process, without regard to how the latter might be implemented computationally).

This phenomenon, called *Categorical Perception* [8], where linear changes in a stimulus dimension are associated with nonlinear perceptual effects, has been observed in numerous experiments, for instance in color or phoneme discrimination.

A recent experiment by Goldstone [7] investigated Categorical Perception (CP) in a task involving training subjects on a *novel* categorization. In particular, subjects were trained on a combined task that first required them to *categorize* stimuli (rectangles) according to size or brightness or both and then to *discriminate* stimuli from the same set in a same-different design.

The study found evidence for *acquired distinctiveness*, *i.e.,* cues (size and brightness, resp.) that were task-relevant became perceptually salient even during other tasks. The task-relevant interval of the task-relevant dimension became selectively sensitized, *i.e.,* discrimination of stimuli in this range improved (*local sensitization* at the class-boundary — the classical Categorical Perception effect), but dimension-wide sensitization was, to a lesser degree, also found (*global sensitization*). Less sensitization occured when subjects had to categorize according to size *and* brightness, indicating competition between those dimensions.

4.1 Categorical Perception in CBF

The CBF scheme suggests a simple explanation for category-related influences on perception: When confronted with two stimuli differing along the stimulus dimension relevant for categorization, the different respective activation levels of the categorization unit provide additional information to base the discrimination on, and thus discrimination across the category boundary is facilitated, as compared to the case where no categorization network has been trained. Fig. 4 illustrates this idea: The (continuous) output of the categorization unit(s) provides additional input to the discrimination network in a discrimination task. In a categorization task, the output of the category unit is thresholded to arrive at a binary decision, as is the output of the discrimination network in a yes/no discrimination task.

In particular, *global* sensitization would be expected as a side effect of training the categorization unit if its response is not constant within the classes, which is just what was observed in the simulations shown above (Fig. 3): The "catness" response level of the categorization unit decreases as stimuli are morphed from the cat prototypes to cats at the class boundary and beyond. Its output is then thresholded to arrive at the categorization rule, which determines the class by the sign of the response (cf. above). *Local* sensitization (Categorical Perception) occurs as a result of a stronger response difference of the categorization unit for stimulus pairs crossing the class boundary than for pairs where both members belong to the same class.

In agreement with the experiment by Goldstone [7], we would expect competition between different dimensions in CBF when class boundaries run along more than one dimension (*e.g.,* two, as in the experiment), as compared to a class boundary along one dimension only: For the same physical change in one stimulus property (one dimension), the re-

[†]Multi-class classification is a challenging and yet unsolved computational problem — the scheme employed here was chosen for its simplicity.

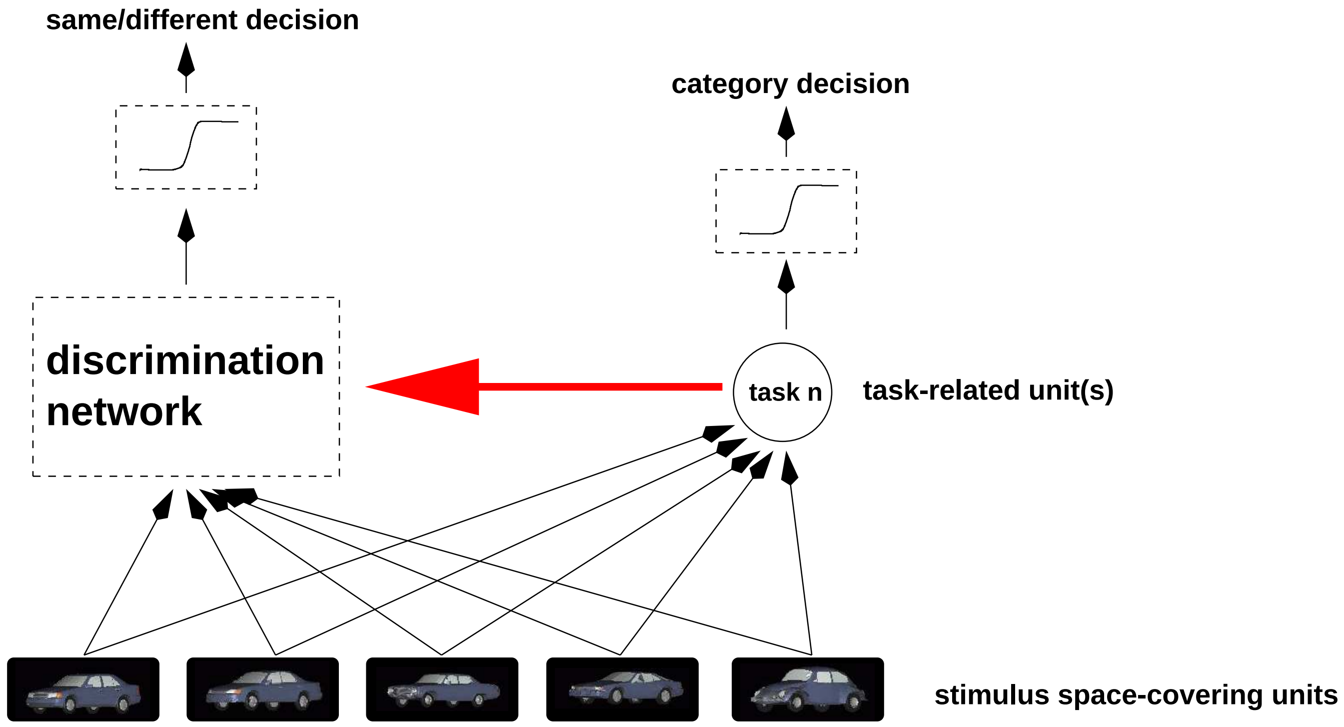

Figure 4: Sketch of the model to explain the influence of experience with categorization tasks on object discrimination, leading to global and local (Categorical Perception) sensitization. Key is the input of the category-tuned unit(s) to the discrimination network (which is shown here for illustrative purposes as receiving input from the SSCU layer, but this is just one of several alternatives), shown by the thick horizontal arrow.

sponse of the categorization unit should change more in the one-dimensional than in the two-dimensional case since in the latter case crossing the class boundary requires change of the input in both dimensions.

4.2 Categorization with and without Categorical Perception

Bülthoff *et al.* have recently reported [2] that discrimination between faces is *not* better near the male/female boundary, *i.e.,* they did not find evidence for CP in their study, even though subjects could clearly categorize face images by gender.

Such categorization without CP can be understood within CBF: Following the simulations described above, CP in CBF is expected if the response of the category unit shows a stronger drop across the class boundary than within a class, for the same distance in morph space. Suppose now the slope of the categorization unit's response is uniform across the stimulus space, from the prototypical examplars for one class (*e.g.,* the "masculine men") to the prototypical exemplars of the other class (*e.g.,* the "feminine women"). If the subject is forced to make a category decision, *e.g.,* using the sign of the category unit's response, as above, the stimulus ensemble would be clearly divided into two classes (noise in the category unit's response would lead to a smoothed out sigmoidal categorization curve). However, in a discrimination task, the difference of response values of the category unit for two stimuli across the boundary would not be different from the difference for two stimuli within the same class (if the within-pair distance for both pairs with respect to the category-relevant dimension is the same). Hence, no Categorical Perception, or, more precisely, any local sensitization would be expected.

In CBF, the slope of a category unit's response curve is influenced by the extent of the training set with respect to the class boundary. To demonstrate this, we trained a cat/dog category unit as described above using four training sets differing in how close the representatives of each class were allowed to get to the class boundary (which was again defined by an equality in the sum over the cat and dog coefficients). Introducing the "crossbreed coefficent", c, of a stimulus belonging to a certain class (cat or dog) as the coefficient sum of its corresponding vector in morph space over all prototypes of the *other* class (dog or cat, resp.), training sets differed in the maximum value of c, ranging from 0.1 to 0.4 in steps of 0.4 (c values of stimuli in each training set were chosen uniformly within the permissible interval, and training sets contained an equal number of stimuli, *i.e.,* 200). The first case, $c = 0.1$, thus contained stimuli that were very close to the prototypical representatives of each class, whereas the $c = 0.4$ set contained cats with strong dog components and dogs with strong cat components, resp.

Fig. 5 shows how the average response along the morph lines differs for the two cases $c = 0.1$ and $c = 0.4$. The legend shows in parentheses the performance on the training set and on the test set, resp.; the number after the colon shows the

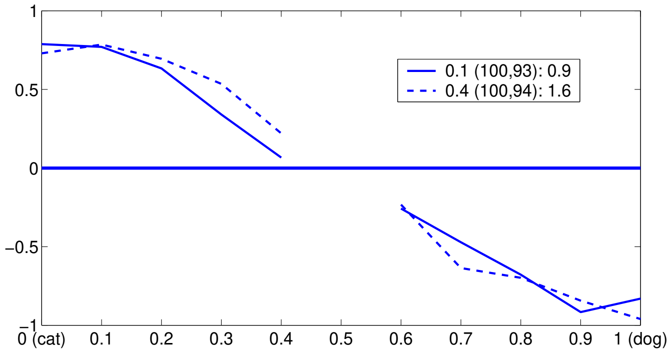

Figure 5: Average responses over all morph lines for the two networks (parameters as in Fig. 3) trained on data sets with $c = 0.1$ and $c = 0.4$, respectively. The legend shows in parentheses the performance (on the training set and on the test set, resp.); the number after the colon shows the average change of response across the morph line (absolute value of response difference at positions 0.4 and 0.6) divided by the response difference for that morph line averaged over all other stimulus pairs 0.2 units apart.

average change of response across the morph line (absolute value of response difference at positions 0.4 and 0.6) relative to the response difference for that morph line averaged over all other stimulus pairs 0.2 units apart. While categorization performance in both cases is very similar (93% vs. 94% correct on the test set), the relative change across the class border is much greater for the $c = 0.4$ case than in the $c = 0.1$ case, where the response drops almost linearly from position 0.2 to position 0.9 on the morph line (incidentally, the relative drop of 1.6 in the $c = 0.4$ case is very similar to the drop observed in prefrontal cortical neurons of a monkey trained on the same task [6] with the same maximum c value).

Thus, CBF predicts that the amount of categorical perception is related to the extent of the training set with respect to the class boundary: If the training set for a categorization task is sparse around the class boundary (as is the case for face gender classification where usually most of the training examplars clearly belong to one or the other category with a comparatively lower number of androgynous faces), a lower degree of CP would be expected than in the case of a training set that extends to the class boundary.

It will be interesting to test this hypothesis experimentally by training subjects on a categorization task where different groups of subjects are exposed to subsets of the stimulus space differing in how close the training stimuli come to the boundary. Category judgment can then be tested for (randomly chosen) stimuli lying on lines in morph space passing through the class boundary. In a second step, subjects would be switched to a discrimination task to look for evidence of CP. The prediction would be that while subjects in all groups would divide the stimulus space into categories (not necessarily in the same way or with the same degree of certainty, as there would be uncertainty regarding the exact location of the class boundary that increases for groups that were only trained on stimuli far away from the boundary), the degree of CP should increase with the closeness of the training stimuli to the true class boundary. Naturally, the categorization scheme used in this task should be novel for the subjects to avoid confounding influences of prior experience. Hence, a possible alternative to the cat/dog categorization task described above would be to group car prototypes (randomly) into two classes and then train subjects on this categorization task.

One issue to be addressed is whether the fact that subjects are trained on different stimulus sets will influence discrimination performance (even in the absence of any categorization task). For the present case, simulations indicate only a small effect of the different training sets on discrimination performance (see Fig. 6), but it is unclear whether this transfers to other stimulus sets. However, while the different training groups might differ in their performance on the untrained part of the stimulus space due to the different SSCUs learned, the prediction is still that the area of improved discriminability should coincide with the subjects' location of the class boundary rather than with the extent of the training set. To avoid range and anchor effects [3] (see footnote below), stimuli should be chosen from a continuum in morph space, *e.g.,* a loop.

Why has no CP been found for gender classification while other studies have found evidence for CP in emotion classification using line drawings [5] as well as photographic images of faces [3]?‡ For the case of emotions, subjects are likely to have had experience with not just the "prototypical" facial expression of an emotion but also with varying combinations and degrees of expressions and have learned to categorize them appropriately, corresponding to the case of high c values in the cat/dog case described above, where CP would be expected.

5 COP or CBF? — Suggestion for Experimental Tests

It appears straightforward to design a physiological experiment to elucidate whether COP or CBF better model actual category learning: A monkey is trained on two different categorization tasks using the same stimuli (for example, the cat/dog stimuli used in the simulations above). The responses of prefrontal cortical neurons (which have been shown in a preliminary study using these stimuli [6] to carry category information) to the test stimuli are then recorded from while the monkey is passively viewing the test stimuli (*e.g.,* during a fixation task). In CBF, we would expect to find neurons showing tuning to either categorization scheme, whereas COP would predict that cell tuning reflects a single metric in shape space. In the former case, it will be interesting to compare neural responses to the same stimuli while the monkey is performing the two different categorization tasks to look

‡CP has also been claimed to occur for facial identity [1], but the experimental design appears flawed as stimuli in the middle of the continuum were presented more often than the ones at the extremes, and prototypes were easily extracted from the discrimination task, biasing subjects discrimination responses towards the middle of the continuum [11].

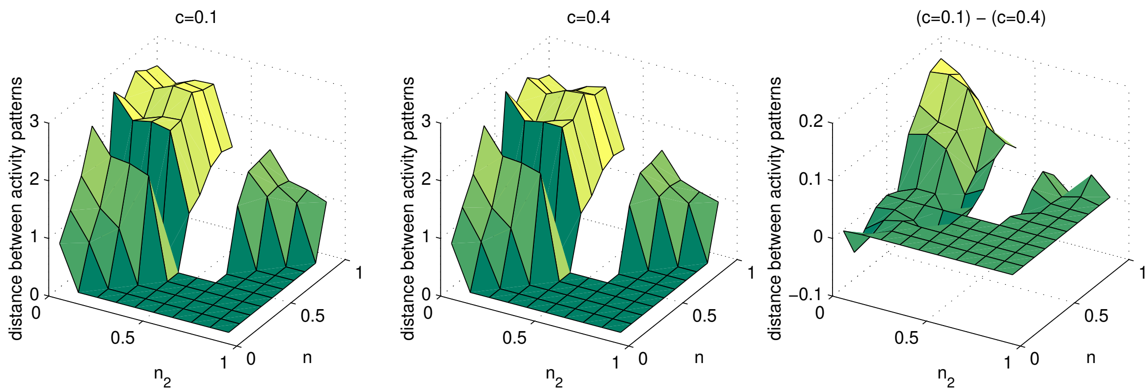

Figure 6: Comparison of Euclidean distances of activation patterns (over 144 SSCU, as used in the previous simulations) for stimuli lying at two different positions on morph lines for the cases of $c = 0.1$ and $c = 0.4$. The left panel shows the average euclidean distance between the activity pattern for a stimulus at position n (y-axis) and a stimulus on the same morph line at position n_2 (x-axis), for the network trained on the data set with $c = 0.1$ (note that there were no stimuli at the 0.5 position). The middle panel shows the corresponding plot for the network trained on $c = 0.4$, while the right panel shows the difference between the two plots: Differences between the two networks are usually quite low in magnitude (note the different scaling on the z-axes), suggesting that discrimination performance in the $c = 0.1$ case should be close to the $c = 0.4$ case.

at response enhancement/suppression of neurons involved in the different categorization tasks.

6 Conclusions

We have described a novel model of object representation that is based on the concurrent use of different categorization schemes using arbitrary class definitions This scheme provides a more natural basis for classification than the "Chorus of Prototypes" with its notion of one global shape space. In our framework, called "Categorical Basis Functions" (CBF), the stimulus space is represented by units whose receptive fields self-organize without regard to any class boundary. In a second, supervised stage, categorization units receiving input from the stimulus space-covering units (SSCUs) come to learn different categorization task(s). Note that this just describes the basic framework — one could imagine, for instance, the addition of slow time-scale top-down feedback to the SSCU layer, analogous to the GRBF networks of Poggio and Girosi [10], that could enhance categorization performance by optimizing the receptive fields of SSCUs. Similarly, the algorithms used to learn SSCUs (k-means clustering or simple storage of all training examples) and the categorization units (RBF) should just be taken as examples. For instance, (a less biological version of) CBF could also be implemented using Support Vector Machines [14]. In this case, a categorization unit would only be connected to a sparse subset of SSCUs, paralleling the sparse connectivity observed in cortex.

A final note concerns the advantages of CBF for the learning and representation of class hierarchies: While the simulations presented in this paper limited themselves to one level of categorization, it is easily possible to add additional layers of sub- or superordinate level units receiving inputs from other categorization units. For instance, a unit learning to classify a certain breed of dog could receive input not only from the SSCUs but also from a "generic dog" unit, or a "quadruped" unit could be trained receiving inputs from units selective for different classes of four-legged animals, in both cases greatly simplifying the overall learning task.

Acknowledgements

Thanks to Christian Shelton for k-means and RBF MATLAB code, and for the morphing programs used to generate the stimuli [13]. Special thanks to Prof. Eric Grimson for help with taming the "AI Lab Publications Approval Form".

References

[1] Beale, J. and Keil, F. (1995). Categorical effects in the perception of faces. *Cognition* **57**, 217–239.

[2] Bülthoff, I., Newell, F., Vetter, T., and Bülthoff, H. (1998). Is the gender of a face categorically perceived? *Invest. Ophthal. and Vis. Sci.* **39**(4), 812.

[3] Calder, A., Young, A., Perrett, D., Etcoff, N., and Rowland, D. (1996). Categorical perception of morphed facial expressions. *Vis. Cognition* **3**, 81–117.

[4] Edelman, S. (1999). *Representation and Recognition in Vision*. MIT Press, Cambridge, MA.

[5] Etcoff, N. and Magee, J. (1992). Categorical perception of facial expressions. *Cognition* , 227–240.

[6] Freedman, D., Riesenhuber, M., Shelton, C., Poggio, T., and Miller, E. (1999). Categorical representation of visual stimuli in the monkey prefrontal (PF) cortex. In *Soc. Neurosci. Abs.*, volume 29, 884.

[7] Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *J. Exp. Psych.: General* **123**, 178–200.

[8] Harnad, S. (1987). *Categorical perception: The groundwork of cognition*. Cambridge University Press, Cambridge.

[9] Poggio, T. and Edelman, S. (1990). A network that learns to recognize 3D objects. *Nature* **343**, 263–266.

[10] Poggio, T. and Girosi, F. (1989). A theory of networks for approximation and learning. Technical Report AI Memo 1140, CBIP paper 31, MIT AI Lab and CBIP, Cambridge, MA.

[11] Poulton, E. (1975). Range effects in experiments on people. *Am. J. Psychol.* **88**, 3–32.

[12] Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neurosci.* **2**, 1019–1025.

[13] Shelton, C. Three-dimensional correspondence. Master's thesis, MIT, (1996).

[14] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.

[15] Wallis, G. and Rolls, E. (1997). A model of invariant object recognition in the visual system. *Prog. Neurobiol.* **51**, 167–194.

Appendix: Parameter Dependence of Categorization Performance for the Cat/Dog Task

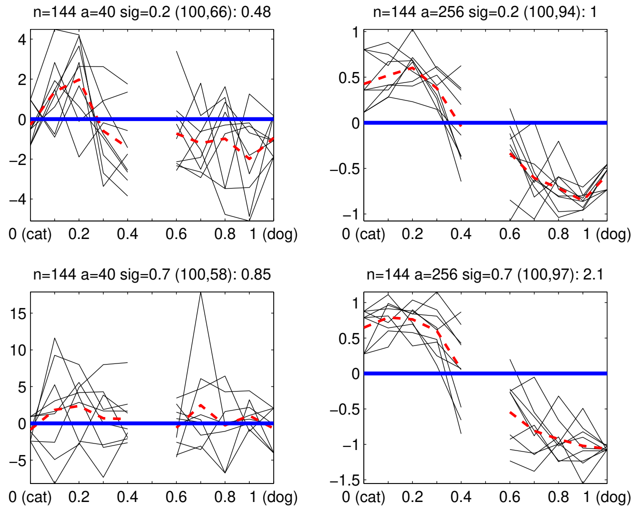

Figure 7: Output of the categorization unit trained on the cat/dog categorization task from section 3.1, for 144 SSCUs (where each SSCU was centered at a training example) and two different values for the σ of the SSCU and the number of afferents to each SSCU (choosing either all 256 C2 units or just the 40 strongest afferents, cf. [12]). The numbers in parentheses in each plot title refer to the unit's categorization performance on the training and on the test set, resp. The number on the right-hand side is the average response drop over the category boundary relative to the average drop over the same distance in morph space within each class (cf. section 4.2). Note the poor performance on the test set for a low number of afferents to each unit, which is due to overtraining. The plot in the lower right shows the unit from Fig. 3.

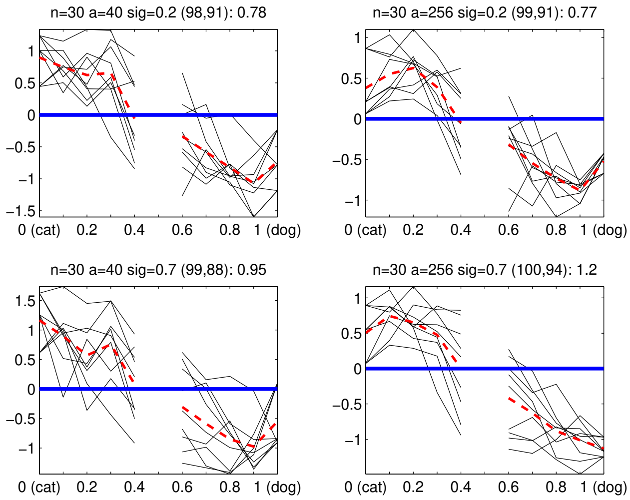

Figure 8: Same as the above figure, but for a SSCU representation based on just 30 units, chosen by a k-means algorithm from the 144 centers in the previous example.