Categorical Perception of Novel Dimensions

Robert L. Goldstone, Mark Steyvers, and **Kenneth Larimer**
Indiana University
Department of Psychology/Program in Cognitive Science
Bloomington, IN. 47405
rgoldsto@indiana.edu, msteyver@indiana.edu, klarimer@indiana.edu

Abstract

Categorical perception is a phenomenon in which people are better able to distinguish between stimuli along a physical continuum when the stimuli come from different categories than when they come from the same category. In a laboratory experiment with human subjects, we find evidence for categorical perception along a novel dimension that is created by interpolating (i.e. morphing) between two randomly selected bezier curves. A neural network qualitatively models the empirical results with the following assumptions: 1) hidden "detector" units become specialized for particular stimulus regions with a topologically structured competitive learning algorithm, 2) simultaneously, associations between detectors and category units are learned, and 3) feedback from the category units to the detectors causes the detectors to become concentrated near category boundaries. The particular feedback used, implemented in an "S.O.S. network," operates by increasing the learning rate of weights connecting inputs to detectors that are neighbors to a detector that produces an improper categorization.

Introduction

Models of category learning typically assume that the stimuli to be categorized can be described in terms of perceptual features or dimensions, and that concept learning involves linking these perceptual descriptions to categories (e.g. Kruschke, 1992). As such, in these "feed-forward" models, processing starts with a perceptual input, and output is in the form of a categorization.

Although categorization is clearly dependent on perceptual input, many researchers have also argued for a reciprocal influence of concept learning on the development of percepts (e.g. Goldstone, 1995). The notion that concepts can influence perception can be traced back at least as far as the Sapir-Whorf hypothesis (Whorf, 1941). The current work explores a version of this hypothesis, and provides a computational mechanism for simultaneous, reciprocal influences between perceptual inputs and acquired concepts. The particular variety of conceptual influence on perception explored here concerns whether specific regions of a novel perceptual dimension can become perceptually sensitized if the region is important for a learned categorization. Our modelling approach for accounting for the observed effects is to develop topologically ordered "detectors" that tend to be densely clustered at the boundary between categories.

Categorical Perception

The most relevant empirical support for categorical influences on perceptual sensitivity comes from work on "categorical perception" (for a review, see Harnad, 1987). According to this phenomenon, people are better able to distinguish between physically different stimuli when the stimuli come from different categories than when they come from the same category. For example, Liberman, Harris, Hoffman, and Griffith (1957) generated a set of vowel-consonant syllables going from /be/ to /de/ to /ge/ by varying a particular physical value along a dimension. Results showed that when the physical difference between speech sounds was equated, subjects were better able to discriminate between two sounds that belonged to different phonemic categories such as /be/ and /de/ than they were able to discriminate between two sounds that belonged within the /be/ category.

Research in our laboratory has explored the development of categorical perception during an experimental session (Goldstone, 1994). Goldstone first trained subjects in one of several categorization conditions in which one physical dimension (e.g. size or brightness) was relevant and another was irrelevant. Subjects were then transferred to same/different judgments ("Are these two squares physically identical?"). Ability to discriminate between squares in the same/different judgment task, measured by Signal Detection Theory's d', was greater when the squares varied along dimensions that were relevant during categorization training. More relevant to categorical perception effects, regions within a dimension were selectively sensitized if they occurred at the boundary between categories. For example, if objects less than 2.5 cm belonged to Category A and objects greater than 2.5 cm belonged to Category B during training, then transfer results indicated heightened sensitivity to this particular region of the size dimension relative to other size values.

Sensitization versus Construction of Dimensions

The above experiments indicate that laboratory experience can perceptually sensitize dimensions and local regions within a dimension. The experimentally explored dimensions that display categorical perception have been pre-existing dimensions. For example, although laboratory training can sensitize size or regions of the size dimension, nobody doubts that our subjects have a notion of size as a dimension by the time they participate in the experiment. Although Goldstone (1994) found categorization-dependent sensitization within the integral dimensions of color brightness and saturation, categorical perception for truly arbitrary dimensions has not yet been found. Such a demonstration would argue for two levels of perceptual learning. In the first, particular values of existing dimensions are sensitized due to categorization demands. In the second, new dimensions are developed for describing stimuli because of their diagnosticity, or ability to cover the range of stimuli. Some researchers have speculated that this second type of

learning has been severely underestimated by the use of laboratory stimuli that are clearly delineated into preexisting dimensions such as orientation, number, and size (e.g. Schyns, Goldstone, & Thibaut, 1995). The current experiment explores whether learned categories can cause sensitization of specific values along novel dimensions.

Experiment in Concept Learning Along an Arbitrary Dimension

In this experiment, a categorization is created that depends on the value of a stimulus along a new dimension. The new dimension is created by selecting two similar, arbitrarily curved objects, and treating these objects as endpoints on a continuum. Intermediate objects are then created by blending these endpoints in varying proportions. Thus, a negative contingency between the proportion of two shapes is formed: the greater the percentage of Shape A in an object, the less Shape B will be present. The arbitrary dimension can be considered as "the proportion of A relative to B" dimension, although subjects may attend to a small region of the shapes during categorization. Subjects learn one of two categorizations based on different cut-off values along this dimension, and then are transferred to a task that measures their perceptual sensitivity at various points along this dimension.

Method

Subjects. One hundred and forty undergraduate students from Indiana University served as participants in order to fulfill a course requirement, not including 12 subjects whose data was excluded for failing to meet a learning criterion of 70% correct categorizations. Forty-nine students were in the left split categorization condition, 45 students were in the right split condition, and 46 students were in the irrelevant categorization condition.

Materials. Stimuli were bezier curves based on 9 control points. Bezier curves are constructed by smoothly passing curves through or near an ordered set of control points. Two random bezier curves were constructed, and 60 intermediate curves were generated by linearly interpolating between the two random endpoints. From these 60 curves, the central 7 curves were selected as the stimuli to be displayed during categorization. An additional set of 7 other curves, to be used in the control categorization condition, were created in the same manner from two different randomly chosen random curves. In this manner, the 7 curves within a dimension can be considered as intermediate frames from a movie that morphs from one arbitrary shape to another. The 7 stimuli used are shown in Figure 1. By choosing only the central 7 stimuli from the A-to-B continuum, the categorization and perceptual discrimination tasks are set to a reasonably high level of difficulty. Each stimulus was approximately 9 cm wide by 7 cm tall, and was displayed at a distance of 25 cm from the subject.

Procedure. There were two tasks in the hour-long experiment - category learning followed by same-different judgments. There were three categorization conditions: left split, right split, and irrelevant categorization. As shown in Figure 1, for the left split group, the first three curves to the left belonged to Category 1, and the last four curves belonged to Category 2. For the right split group, the first four curves to the left belonged to Category 1, and the remaining curves belonged to Category 2. For the irrelevant categorization group, the first three curves from a dimension with completely different endpoint shapes belonged to Category 1, and the remaining curves belonged to Category 2.

During the categorization training, 40 repetitions of the seven curves were shown in random order. On an individual trial, a curve was shown in a randomly generated location on the screen. The curve remained on the screen until the subject pressed a key corresponding to their guess as to the curve's category. Category responses were made by pressing the keys "1" and "2." After a response was made, feedback was given as to the correctness of the response, and the correct category label was displayed. After 1.5 sec, the screen was erased, and after another 1 sec, the next trial began.

All three categorization training groups received the identical subsequent discrimination experiment, using the seven curves shown in Figure 1. Subjects were shown pairs of adjacent curves as ordered in Figure 1, or the identical curves repeated twice, and responded either "same" or "different." Subjects were instructed to press the "S" key on the keyboard if they believed the two curves to be physically identical, and to press the "D" key if they believed the two curves to differ in any way except location. The interval between trials was 1500 msec. Subjects made 150 same/different judgments in all.

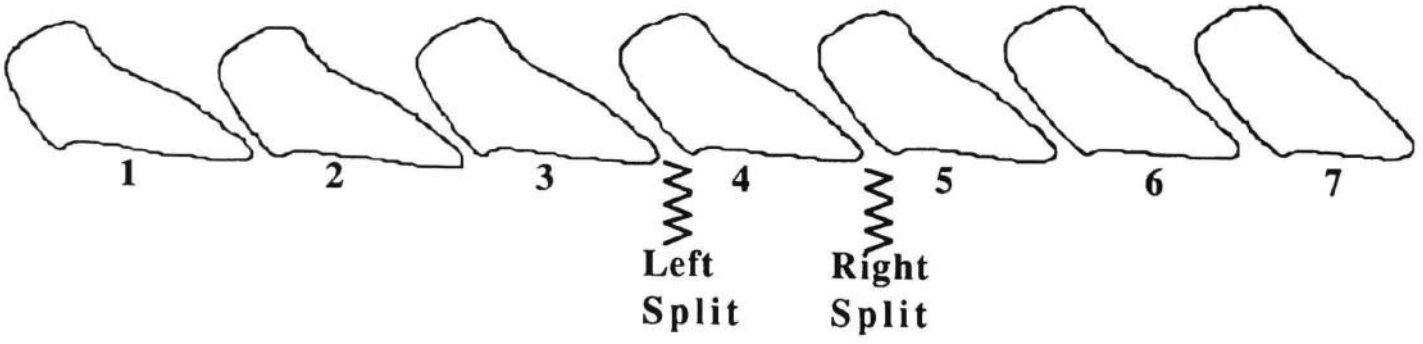

Figure 1

Results

The data of principle interest are subjects' sensitivities at discriminating between various pairs of curves, broken down as a function of their categorization condition. A d' measure of sensitivity was calculated. A d' of 0 indicates a complete lack of sensitivity in distinguishing "Same" from "Different" trials; d' values increase as sensitivity increases.

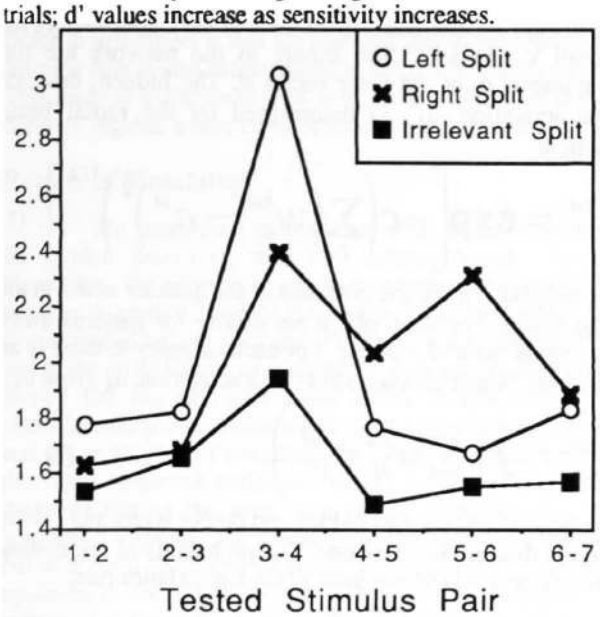

Figure 2

With 7 curves there are 6 pairs of adjacent curves. The d' for each of these 6 pairs in each categorization condition is shown in Figure 2. Overall, there were main effects due to both tested pair, $F(5, 122) = 4.3$, $p < .01$, and categorization condition, $F(2, 122) = 6.5$, $p < .01$. The former effect seems to be attributable to subjects' ability to discriminate between stimuli 3 and 4 (from Figure 1) more easily than other pairs. The latter effect is due to subjects in the left and right split groups having elevated sensitivity relative to the control groups. This effect is consistent with a large literature showing that preexposure to stimuli leads to their heightened discriminability (Hall, 1991).

Most relevant to learned categorical perception, a significant interaction between categorization condition and tested pair was found, $F(10, 122) = 2.9$, $p < .01$. As such, the categorization training in the first stage of the experiment altered the discriminabilities of stimuli in the experiment's second stage. To better visualize the exact effect of this influence, Figure 3 plots the sensitivity (d') obtained from the right split group minus the sensitivity from the left split group, for each of the six pairs of adjacent curves. As such, positive values signify greater sensitivity for the right split group than for the left split group. Although the effects of the splits are not symmetric, the general effect of categorization training seems to be that discriminability is relatively high for stimuli that fall near the category boundary. Even if we restrict our attention just to the 3-4 and 4-5 pairs, significantly higher d's are found when the pair rests on a boundary that was influential for categorization (d' = 2.54) than when it does not (d' =2.08), $F(1,122)=2.5$, $p < .01$.

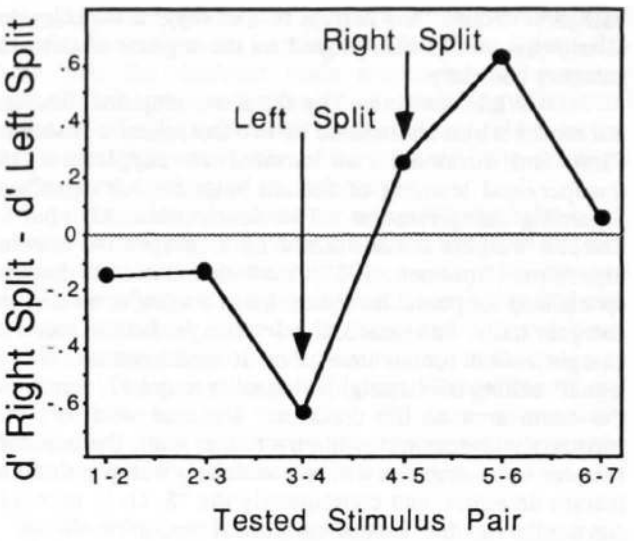

Figure 3

Discussion

The pattern of sensitivity to regions of the continuum formed by interpolating between two randomly selected bezier curves varies across the three categorization groups. This pattern is most accurately described as follows: ability to discriminate between two physically different curves is relatively high when the boundary between laboratory-trained categories falls relatively close to the pair of curves. Although there is only a substantial difference between the categorization groups for three pairs, this generalization applies to all six pairs in Figure 3 - sensitivity is higher for the right split categorization group when, and only when, the tested pair is closer to the right split than the left split.

Modelling Learned Categorical Perception Along a Novel Dimension

One useful property of neural networks for modeling learned categorical perception is that hidden units that intervene between input and output representations are capable of creating internal representations that capture important regularities in the inputted materials. Several models develop hidden units that can be interpreted as learned feature detectors (e.g. Schyns, 1992). Our model will use this technique in order to create topologically ordered feature detectors that tend to respond to specific values along a arbitrary continuum.

Empirical and Theoretical Constraints on Modeling

There are several empirical results, with respect to both categorization and subsequent perceptual sensitivity, that a model of learned categorical perception should show (Harnad, 1987). First, categorization judgments should rapidly change as the boundary between categories is crossed. Second, categorization of "caricatured" items (displaced away from a category's central tendency in the direction opposite to the boundary between the categories) should be at least as good, and often times better, than categorization of the central tendencies of the categories (Goldstone, in press). Third, sensitivity for discriminating physically different stimuli should be higher when the items straddle two categories than when they fall in a single

category. Fourth, the current results suggest that elevated sensitivity should also extend to the regions next to the category boundary.

While constrained by the above empirical findings, our model is also constrained by two theoretical motivations. First and foremost, we wanted to supplement the unsupervised learning of feature detectors with feedback regarding categorization. The development of input-to-detector weights is constrained by a competitive learning algorithm (Kohonen, 1982) such that detectors become specialized for particular inputs, but is also influenced by the category units. In essence, if a detector predicts an incorrect categorization for an item, then it sends out an "S.O.S. signal" calling for its neighboring units to quickly move into the same area as the detector. Because detectors that incorrectly categorized will attract other units, the boundary between two categories will be particularly well populated by feature detectors, and consequently the "S. O. S. network" can predict flexible, learned categorical perception effects.

The second theoretical motivation for our model is to develop categorical perception starting from relatively raw, perceptual inputs. As such, the first stage of our network converts gray-scale two-dimensional drawings of curves to Gabor filter representations that describe the inputs in terms of spatially organized line segments. The detectors are trained upon these Gabor filter representations.

Details of the S.O.S. Network

The classification part of the model is a neural network similar to ALCOVE (Kruschke, 1992). The hidden layer of detectors are radial basis exemplar nodes maximally sensitive to stimuli at the position of these exemplar nodes. The output layer consists of nodes that classify the activation pattern of these exemplar nodes. The crucial differences with ALCOVE are that the exemplar nodes are topologically arranged in a one-dimensional lattice, and exemplar nodes can move their position in input space through competitive learning. These features allow the model to self-organize the exemplar nodes along the input dimensions. More importantly, because the learning rate is set proportional to the classification error, greater sensitivity near the category boundary can be predicted.

We used materials of the same type as those used in the experiment. A morphing sequence of 28 bezier curves was created, with each picture having 128x128 pixels. Each stimulus was filtered through Gabor filters (Daugman, 1985) with overlapping receptive fields to extract local features. Gabor filters with orientations of 0, 45, 90 and 135 degrees operated on 6 x 6 overlapping receptive fields that were regularly spaced over the input picture. In total then, the Gabor filter output vector a, has 144 components. In figure 4a, one bezier curve is shown. Figures 4b, c, d and e, show the filtered activations over the receptive fields in the four orientations for this bezier curve. The transformation from the stimuli in pixel space to the Gabor filter space preserved the local similarity relations of the stimulus sequence; the distance between the Gabor vectors for stimuli k and k+1 was always smaller than the distance of the Gabor vectors for stimuli k and k+2. The inputs to the network are the components a_i of the filter vector a. The hidden, detector node activation, a_j, is determined by the radial basis function:

$$a_j^{hid} = \exp\left[-c\left(\sum_i \left(w_{ji}^{hid} - a_i^{in}\right)^2\right)^{1/2}\right]. \quad (1)$$

The weights, w_{ji} are the positions of the detector nodes in the input space. The drop off in sensitivity for patterns away from w_{ji} is determined by c. For each category k, there is an associated classification node k, with activation, a_k given by:

$$a_k^{out} = f\left(\sum_j w_{kj}^{out} a_j^{hid}\right). \quad (2)$$

The weights w_{kj} connect hidden and output layers and f is the sigmoid discriminant function. The probability of responding with category k is determined by the Luce choice rule,

$$P_{resp}(k) = \frac{a_k^{out}}{\sum_K a_K^{out}}. \quad (3)$$

The sum of squared error,

$$E = \sum_j \left(t_k - a_k^{out}\right)^2, \quad (4)$$

is based on the teacher signal t_k for node k which is 1 if the input stimulus belongs to category k, and 0 otherwise. Gradient descent is used to update the weights. The weights, w_{kj}, from hidden to output nodes are determined by:

$$\Delta w_{kj}^{out} = \lambda a_k^{out}\left(1 - a_k^{out}\right)\left(t_k - a_k^{out}\right)a_j^{hid} \quad (5)$$

where λ is the learning rate. The position weights, w_{ji}, of the hidden nodes are updated with a competitive learning rule,

$$\Delta w_{ji}^{hid} = E\eta\Lambda_{(j,j^*)}\left(a_i^{in} - w_{ji}^{hid}\right), \quad (6)$$

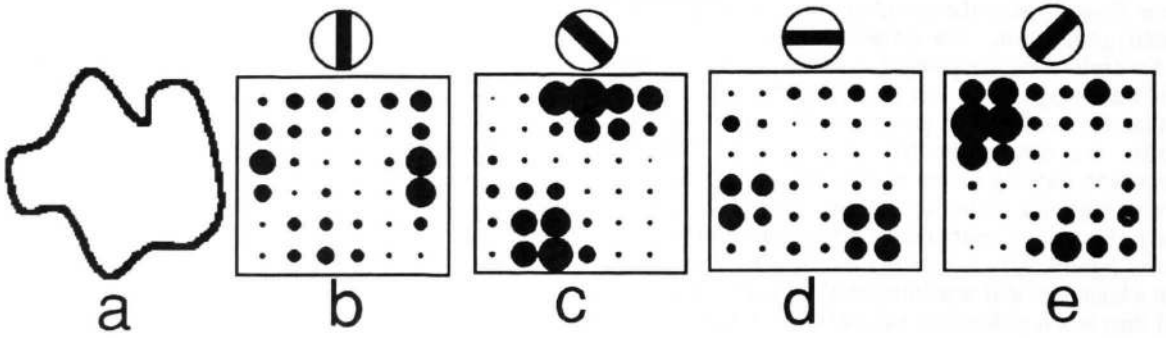

Figure 4

where the learning rate is proportional to a constant h and two terms: the neighborhood function $\Lambda(j,j^*)$, and the classification error E. In the function $\Lambda(j,j^*)$, j^* is the hidden detector that has the smallest Euclidian distance to the input pattern. $\Lambda(j,j^*)$ is 1 for $j=j^*$ and falls off as a power function of distance $|j-j^*|$. This learning rule typically leads to a partial or complete topological ordering of the position weights w_{ji} in input space (Kohonen, 1982). The important factor in this model is that the learning rate in the competitive learning rule is also proportional to the classification error E. This leads to a distribution of detector positions that is more dense in regions where classification error is greatest.

Results of Simulation

We performed simulations with 28 input patterns, 14 hidden detectors, and two output nodes. In one simulation, the split between the two categories was placed between patterns 10 and 11 (left split). In another simulation, it was placed between 18 and 19 (right split). In figures 5a and b, the probabilities of responding with either category are shown for the left and right splits, respectively. The classification probabilities are highest between the extremes and the prototypes of the categories; thus, the model exhibits the often observed caricature effect whereby response is maximal not at the prototype of the category (e.g. the stimulus 5.5 for the left category in the left split condition) but at a point displaced from the prototype in the direction opposite to the other category. In figures 5c and d, the responses a_j for each of the 14 detector nodes are shown for each stimulus value. Each curve corresponds to the response profile for one detector. These figures give some insight into the distributions of the position weights w_{ji} of the hidden nodes in input space, because activation a_j is maximal for input at the position weight w_{ji}. The figures show that the hidden nodes are more densely distributed around the categorization boundary as a result of the feedback of classification error in the learning rule (6). These figures also show that the detector node responses for patterns surrounding the maximally responding detector monotonically decrease with distance from this detector. This reflects the preservation of the local similarity relations by the spatial topology of the detectors.

A sensitivity measure for same/different judgements in the model was constructed by taking the Euclidian distance between the hidden node activation patterns for the two patterns to be judged. In figure 5e and f, sensitivity is shown for comparisons of patterns 1 and 3, 3 and 5, 5 and 7 etc. The peak sensitivity occurs approximately at the category boundary. This occurs because slightly different stimuli that occur near the category boundary will cause substantially different activation patterns on the detector units, given the dense concentration of detectors in this region.

Discussion

The experiment and computer simulation support the possibility that category learning can entail not only the sensitization of regions of a preexisting dimensions, but can also sensitize regions of new dimensions. The dimensions are unlikely to have existed before the experiment because they were created by interpolating between arbitrary curves. The dimension is either interpretable as "Proportion of Shape A relative to Shape B" or in terms of some smaller sub-component that continuously changes from Shape A to B. As with standard categorical perception effects, sensitization relative to the control condition is greatest for stimuli at the boundary between the categories.

The simulation provides insights into the phenomenon of categorical perception along new dimensions. First, Kohonen's self-organizing feature map

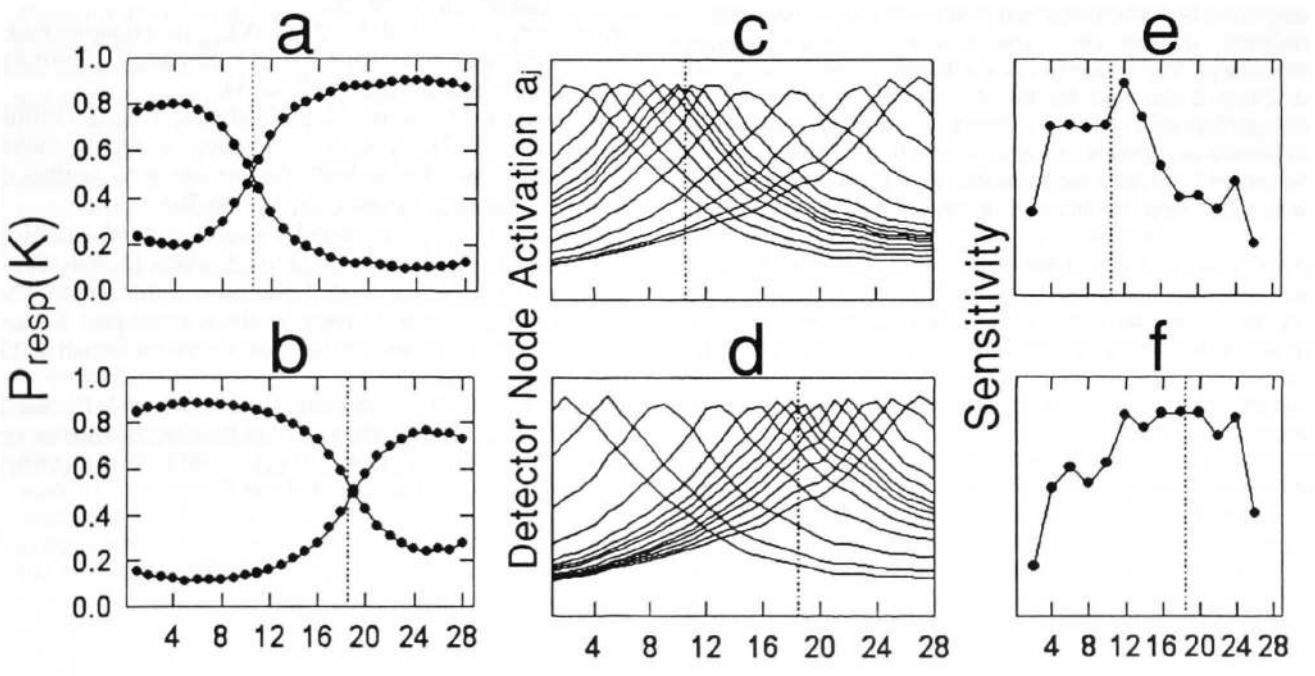

Figure 5

algorithm is typically understood as developing detectors for specific stimuli. Although this is certainly one way to understand our network's behavior, it can also be understood as creating detectors for regions along a dimension. Second, the network shows how the structure implicit in stimuli that fall along a new dimension can be captured by the topological positions of detectors units. The natural similarity relations between adjacent stimuli (in Figure 1) leads, without supervision, to the construction of a locally and globally well-ordered sets of detector units. Once the network has settled, the detectors on the left and right ends will be specialized for the two extreme curves, and the detectors in between will handle the intermediate curves in proper order.

The final major insight of the network's treatment of categorical perception effects, embodied by the S.O.S. principle, is that these effects can be modelled by creating relatively dense representations of items at the border between categories. This treatment of categorical perception differs from other neural network implementations (Anderson, Silverstein, Ritz, & Jones, 1977; Harnad, Hanson, & Lubin, 1994). In these other approaches, each category has its own attractor, and the stimuli that fall into one category will all be propelled toward the category's attractor. Categorical perception occurs because inputs that are very close but fall into different categories will be driven to highly separated attractors. In contrast, in our S.O.S. network, categorical perception emerges because many detectors will congregate at the category boundary, and thus small differences at this boundary will be reflected by different patterns of detector activity. There are two potential advantages of our account. First, categorical perception effects can arise even when there is no demand to categorize the stimuli, once the detectors have moved toward the boundary. This fits the requirements of the same/different task well because physical identity, not category identity, is the basis for these judgments. Second, our account explains how stimuli falling on the same side of a category boundary may also become more discriminable after categorization training, if they are sufficiently close to the category boundary. The results from the human experiment suggest that this is the case for people. In networks that explain categorical perception by creating different attractors for different categories, unique items that are close to the boundary but fall in the same category become more similar with processing, not more distinctive.

In conclusion, category learning can lead to the development of new dimensions. Once developed, regions within these dimensions can be selectively sensitized if they are important for determining category boundaries. The qualitative effect of category learning on perceptual sensitivity can be modeled by a neural network that simultaneously develops detectors for dimension values and associations between detectors and categories. Within this framework, there is a top-down influence of categorization that gives rise to categorical perception - when a detector produces an improper categorization, then learning rates for its neighboring detectors are momentarily increased. In this manner, the difficult-to-categorize regions of a dimension will garner a high density of detectors, thereby permitting sensitive discriminations at the category boundaries.

Acknowledgements

We would like to thank John Kruschke and Philippe Schyns for many helpful comments. This work was funded by National Science Foundation grant SBR-9409232 awarded to Robert Goldstone

References

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. Psychological Review, 84, 413-451.

Daugman, J.G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America A, 2, 1160-1169.

Garner, W. R. (1974). The processing of information and structure. Hillsdale, NJ: Erlbaum.

Goldstone, R. L. (in press). Isolated and Interrelated Concepts. Memory and Cognition.

Goldstone, R. L. (1994). influences of categorization on perceptual discrimination. Journal of Experimental Psychology: General, 123, 178-200.

Goldstone, R. L. (1995). Effects of categorization on color perception. Psychological Science, 6, 298-304.Hall, G. (1991). Perceptual and associative learning. Clarendon Press: Oxford.

Harnad, S. (1987). Categorical perception. Cambridge University Press: Cambridge.

Harnad, S., Hanson, S. J., & Lubin, J. (1994). Learned categorical perception in neural nets: Implications for symbol grounding. in V. Honavar & L. Uhr (Eds.) Artificial intelligence and neural networks: Steps toward principled integration. Academic Press: Boston

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological cybernetics, 43, 59-69.

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

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 54, 358-368.

Schyns, P. (1992). A modular neural network model of concept acquisition. Cognitive Science, 15, 461-508.

Schyns, P. G., Goldstone, R. L., & Thibaut, J-P (1995). The development of features in object concepts. Indiana University Cognitive Science Technical Report #133. Bloomington, IN.

Whorf, B. L. (1941). Languages and logic. in J. B. Carroll (ed.) Language, Thought, and Reality: Selected papers of Benjamin Lee Whorf. MIT Press (1956), Cambridge, Mass. (pp. 233-245).