Simulating Conceptually-Guided Perceptual Learning

Alexander Gerganov (agerganov@cogs.nbu.bg)
Department of Cognitive Science and Psychology, New Bulgarian University
21 Montevideo St., 1618 Sofia, Bulgaria

Maurice Grinberg (mgrinberg@nbu.bg)
Department of Cognitive Science and Psychology, New Bulgarian University
21 Montevideo St., 1618 Sofia, Bulgaria

Paul C. Quinn (pquinn@udel.edu)
Department of Psychology, University of Delaware
Newark, DE 19716 USA

Robert L. Goldstone (rgoldsto@indiana.edu)
Department of Psychological and Brain Sciences, Indiana University
1101 East Tenth Street, Bloomington, IN 47405 USA

Abstract

Traditional models of perceptual learning usually assume that learning occurs through changes of weights to fixed primitive features or dimensions. A new model for perceptual learning is presented which relies on simple and physiologically plausible mechanisms. The model suggests how flexible features can be dynamically derived from input characteristics in the course of learning and how diagnostic shape representations could be formed due to conceptual influences.

Keywords: perceptual learning, neural networks, categorization, concept learning.

Introduction

Perceptual learning refers to performance improvement in different sensory tasks as a result of practice, training, or simple exposure. In the domain of visual perception, these tasks range from simple detection and discrimination of geometric shapes to more complex tasks like face recognition and object categorization. One important question concerns the nature of the processes that lead to perceptual learning. Evidence has been provided for a wide range of changes – from input based representation modifications to influences of expectation, attention, or task. Because of the highly complex and intertwined interactions of different processes, a deliberate blurring of the boundary between concepts and percepts has been proposed (Goldstone & Barsalou, 1998). There is a need for theories and models that account for conceptual influences on perceptual learning.

Computational modeling is often used to simulate perceptual learning processes (e.g., Mozer, Zemel, Behrmann, & Williams, 1992; Petrov, Dosher, & Lu, 2005; Poggio, Fahle, & Edelman, 1992). Modeling places important constraints on explanations about perceptual learning and pushes theoretical accounts to be more quantitative and concrete. Testable behavioral predictions are often derived from simulations. Models of perceptual learning, however, rarely try to account for performance in different tasks at the same time. They should be able to operate in the absence as well as in the presence of reward feedback. In addition, many of the models rely on a finite number of fixed representations (primitives) as the elementary building blocks for composing concepts. Such accounts fall short of capturing the richness of visual pattern learning phenomena. There is experimental evidence suggesting that perceivers not only learn to selectively weight existing dimensions, but also learn to isolate dimensions that were originally psychologically fused together (Goldstone & Steyvers, 2001), and reorganize visual inputs into new units (Behrmann, Zemel, & Mozer, 1998; Goldstone, 2000).

In the present article, a neural network model is described which relies on the physiologically plausible learning mechanisms of competitive and Hebbian learning. The model focuses on simulating results from task-dependent perceptual learning, which may involve both a higher-level conceptual influence and a lower-level perceptual reorganization. Studies with adults show that perceptual learning is influenced by the feedback presented to learners (Shiu & Pashler, 1992) but can also take place without feedback (Watanabe, Náñez, & Sasaki, 2001). Experimental data from infants show also that perceptual learning can occur without feedback (Quinn, Schyns, & Goldstone, 2006). Accordingly both supervised and unsupervised learning should be incorporated into a full model of environmentally induced perceptual plasticity. The model for perceptual learning presented below is able to simulate both influences.

Several simulations are reported that correspond to empirical results from behavioral studies. Finally, conclusions are put forward about the way statistics from visual patterns can lead to the building of flexible primitive features and how higher-level conceptual tasks can influence the formation of complex shape representations.

The Model

The model for perceptual learning consists of two main layers and an artificial input retina (Figure 1). The first layer is based on the competitive learning paradigm (Rumelhart & Zipser, 1985). However units compete only for a small part of the input—that is, each unit has a receptive field and competes only with other units with the same receptive field. In the current implementation of the model there is no overlap between receptive fields. Competing units are organized in inhibitory clusters—two units with the same receptive field cannot be active at the same time. Only the winner for this receptive field is active. A competitive unit is connected with horizontal Hebbian weights to all units from the other inhibitory clusters. The horizontal Hebbian connections link the parts of an input pattern in terms of coactivation of the competitive units that are specialized to those parts. The activation of a competitive unit is computed in two time-steps according to the following equations:

$$A^d_{i,k}(t) = \sum_{j=1}^{n} I^d_{j,k} W^d_{i,j}$$

$$A^d_{i,k}(t+1) = A^d_{i,k}(t) + \eta \sum_{\substack{p=1 \\ p \neq d}}^{c} \sum_{l=1}^{s} W^{d,p}_{i,l} A^p_{l,k}(t),$$

where $A^d_{i,k}(t)$ is the activation of unit i from cluster d in moment t when input pattern k is presented, $I^d_{j,k}$ is the activation of input pixel j from receptive field d for pattern k, $W^d_{i,j}$ is the weight of the connection between unit i and pixel j, $A^p_{l,k}(t)$ is the activation in moment t of competitive unit l from cluster p for pattern k, $W^{d,p}_{i,l}$ is the weight of the horizontal connection between unit i from cluster d and unit l from cluster p, n is the number of pixels in receptive field d, s is the number of competitive units from cluster p, and c is the number of clusters. In the following simulations, s is the same for all clusters, that is, the number of competitive units in the different clusters is constant. The parameter η is set to 0.1 and represents the smaller contribution of the horizontal connections compared to the bottom-up activation. The winner from each cluster is determined as the most active unit inside the cluster. The output units have sigmoid activation functions.

Learning for the connections between an input receptive field and the competitive units from the corresponding inhibitory cluster follows the classical formula:

$$\Delta W^d_{i,j} = \begin{cases} M(I^d_{j,k} - W^d_{i,j}) & \text{if unit } i \text{ loses on stimulus } k \\ L(I^d_{j,k} - W^d_{i,j}) & \text{if unit } i \text{ wins on stimulus } k \end{cases},$$

where L is the learning rate for the winning unit (0.1 for all simulations), M is the learning rate for the losing unit – it is set to 0.001 for all simulations. $I^d_{j,k}$ is the activation of the retina pixel j from receptive field d when input k is presented, and $W^d_{i,j}$ is the weight between pixel j from receptive field d and competitive unit i. The stimuli are presented as activation patterns on the retina, where each pixel is either 1 (active) or 0. Activation of competitive units is normalized so that the winning unit's activation is 1 and all the losing units from the cluster sharing the same receptive field are inhibited to have zero activation. The horizontal Hebbian weights learn according to the Hebbian rule:

$$\Delta W^{d,p}_{i,l} = \alpha A^d_i A^p_l - D,$$

where α is the learning rate, A^d_i is the activation of unit i from cluster d, A^p_l is the activation of unit l from cluster p, and D is the decay rate of the weights.

The competitive layer is fully connected to the output layer with Hebbian weights that learn according to the same rule as the horizontal connections, with the exception that they have different decay and learning rates. All Hebbian weights were set to zero in the beginning of a simulation.

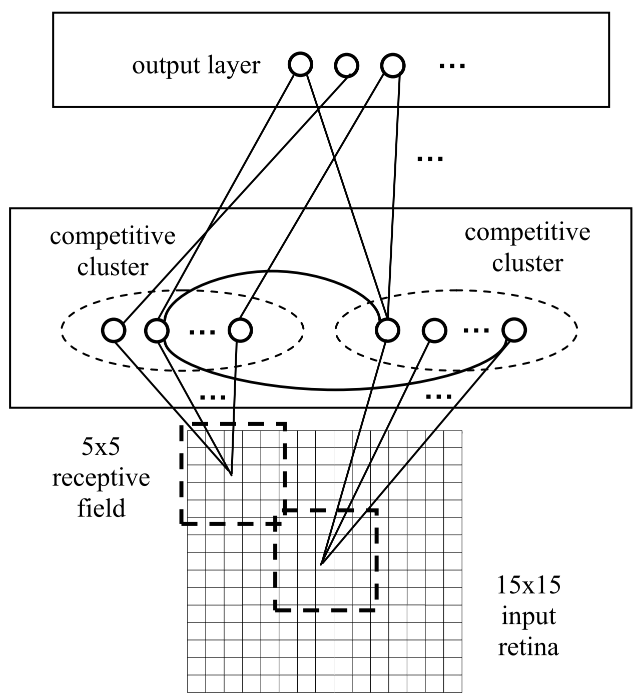

Figure 1: The model for perceptual learning. Only some of the connections are shown for visualization purposes. See the text for full details.

The network learns after each pattern. The competitive layer corresponds to lower-level cells with small receptive fields that cover only small parts of an input, while the output units correspond to more complex structures that are thought to participate in higher-level cognitive tasks

Simulations and Results

Two types of simulations are possible with the described model. The first type corresponds to learning without feedback. In this operational mode, the output layer is activated at random since no teacher signal is available. In other words, this is unsupervised learning of the competitive layer, based only on the characteristics of the input space. When feedback is available, a particular pattern of activation appears on the output layer as a teacher signal. This signal represents the influence of higher-level conceptual processes on learning.

Unsupervised Learning

The unsupervised learning of the competitive layer alone was simulated with stimuli close to those used in Quinn and Schyns (2003) and Quinn et al. (2006). Using an unsupervised model to simulate empirical results from infants seems like a natural correspondence given that infants in the first few months of life do not receive instruction on how to organize their visual experiences. A series of experiments were conducted to determine whether infants, like adults, can perceive visual patterns in terms of parts extracted through category learning rather than parts that would be derived from adherence to gestalt organizational principles.

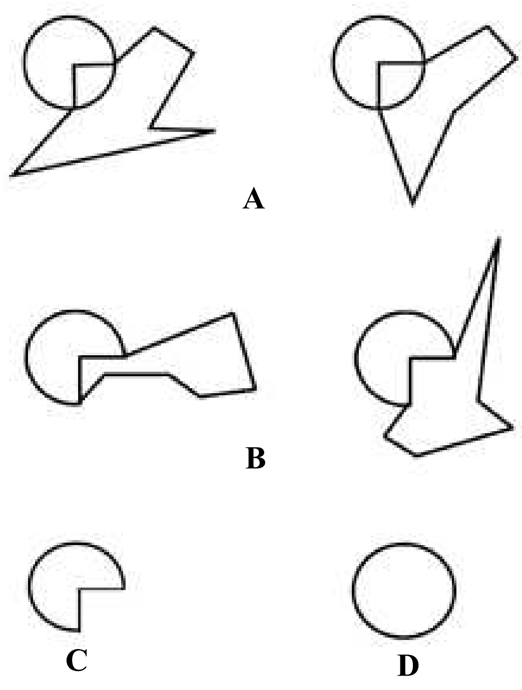

Figure 2: Stimuli from Quinn and colleagues, 2006.

When 3- to 4-month-olds were presented with visual patterns consisting of overlapping circle and polygon shapes (Figure 2A), the infants tended to interpret these forms in terms of a polygon and circle, consistent with a good continuation principle. This was evidenced by infants being more surprised (looking longer) by a subsequently presented pacman shape (Figure 2C) than a circle (Figure 2D). However, when a separate group of 3- to 4-month-olds was first presented with a series of patterns containing the three-quarter "pacman" shapes (Figure 2B), and then subsequently with the patterns shown in Figure 2A, the infants interpreted the ambiguous patterns in Figure 2A as containing a pacman instead of a circle, as evidenced by their greater looking times for the circle than the pacman. These experimental results strongly suggest that unsupervised learning is capable of overriding gestalt laws of organization such as good continuation if the prior learning history supports an alternative organization.

The model can provide a computational account for these empirical findings. The competitive layer is capable of extracting elements and statistical dependencies from the input structure even if no feedback is available. Thus the gestalt law of continuity was simulated with presentation of simple forms at different positions on the retina. Ten such patterns (three vertical lines, three horizontal lines, and four circles) were presented in random order for 2000 cycles. This pre-training phase simulated the infant's perceptual experience prior to arrival at the laboratory and conceivably corresponds with the experiences of young infants as they encounter visual patterns in the environment. We were interested in the ability of the model to acquire perceptual constraints from commonly occurring patterns instead of explicitly building in the good continuation principle. This could also be interpreted as the evolved representation of naturally occurring statistics in visual patterns (Olshausen & Field, 1996).

The input retina consisted of 225 pixels organized in a 15x15 square matrix. There were 9 non-overlapping square 5x5 receptive fields with 8 units in an inhibitory cluster competing over each of the receptive fields, which makes for a total of 72 nodes in the competitive layer. The learning rate of the horizontal Hebbian weights was 0.05 and the decay rate was set to 0.009. After the pre-training phase, some of the competitive units specialized for parts of lines, while others specialized for arcs of a circle. Then an ambiguous pattern (Figure 3A) was presented. This portion of network training and testing corresponded to the first familiarization test phase in the study with infants, when similar patterns each consisting of an overlapping circle and a polygon were presented which led to the segmentation of the circle and the polygon by infants. The ambiguous pattern given to the model activated four "arc" and two "line" nodes from the competitive layer, thus forming a good, continuous circle and some parts of a polygon which was consistent with the infants' behavior. The activation pattern over the competitive layer is visualized in Figure 3B with the following algorithm – each pixel represents the

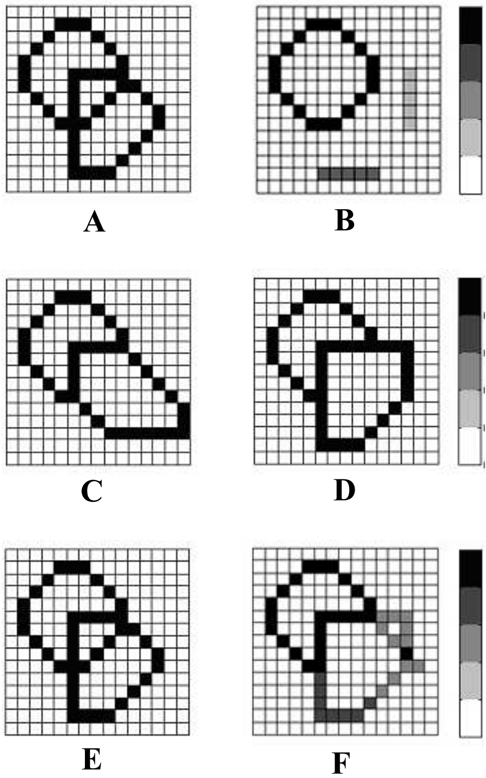

Figure 3: Unsupervised learning simulation

weight between this pixel and the competitive unit multiplied by the competitive unit's activation. This visualization is intended to show that the competitive units were not activated accidentally but represented both the structure of the presented pattern and a learned continuity principle for a circle shape. The polygon shape triggered only the activation of two separate line segments, because the network had never been exposed to any polygon shape and thus did not have the chance to acquire any polygon representation during its pre-training. This result shows that the network does not simply imitate the presented pattern but is affected by its previous knowledge about perceptual grouping that has been stored in the horizontal connections.

The same network was fed for 200 cycles with two patterns containing pacman shapes (Figure 3C, 3D) and again was presented with the ambiguous pattern 3E. This corresponded to the two-part procedure in which the infants were first presented with pacman shapes and subsequently with circle shapes (2B followed by 2A). Once again the model behavior was very similar to what the experimental results suggested. This time the pacman shape was strongly active and some polygon segmentation appeared but was less active than the pacman (Figure 3F). The pacman shape actually was represented by three competitive units specialized for arcs and one specialized for an angle. The "arc" units were initially connected to the fourth arc unit which completed the active circle from Figure 3B; however, after the patterns containing the pacman shapes were repeatedly shown to the network, the angle unit became more active than the arc unit over the same receptive field, which led to the angle unit winning for this receptive field. This could be interpreted as a spontaneous formation of a virtual pacman shape detector that is constructed from smaller low-level representations of three arcs and one angle segment.

Supervised Learning

Supervised learning is often used in studies of adult perceptual learning and can influence the course of learning. Previous experiments (Pevtzow & Goldstone, 1994) have suggested that observers seem to develop perceptual detectors for stimulus elements that are diagnostic of task-critical categorization while they learn to categorize simple patterns. The same patterns, when they receive different categorizations, result in different psychological features being constructed. The nature of the detectors depends not only on the input patterns as in the previous simulation, but on the categorization task as well. As an example, the ambiguous scene in Figure 3A was more likely to be segmented into a circle and polygon when the circle was previously relevant for categorization, and more likely to be segmented into a pacman when the pacman was relevant.

The experimental results from Pevtzow and Goldstone (1994) have been simulated with a model similar to the one presented here (Goldstone, 2000). The previous model however relied on built-in perceptual constraints and input patterns competing to be accommodated by a competitive unit. The present model adds plausible Hebbian learning to the competitive learning mechanism used in Goldstone (2000). The present model also uses more local competition for small parts of an input inside a receptive field instead of competition for the whole input. This leads to a somewhat different interpretation of a detector – in the present model a detector is composed of several smaller competitive units from different receptive fields that form together a coherent shape detector over the whole input retina.

In the following simulations the formation of such detectors was influenced not only by the input properties as in the unsupervised learning but also by a conceptual teacher signal that led to the formation of categorization-relevant detectors at the output layer of the network. A teacher signal was directly presented as a pattern of activation on the output layer during the supervised training. This was done for simplicity since the influence of higher-level categorization or judgment structures can be simulated in different ways – one possible mechanism that was used by Goldstone (2000) was top-down influence from a categorization layer to the detector layer.

A 256 square 16x16 pixel retina was used; competitive units' receptive fields were square 8x8 non-overlapping matrices, which yielded a total of four receptive fields. Each inhibitory cluster consisted of 4 units competing with one another. The output layer had two units. Learning rate for the output Hebbian weights was set to 0.1 and the decay rate was 0.04. The horizontal Hebbian connections had the same learning and decay rates as in the previous simulation.

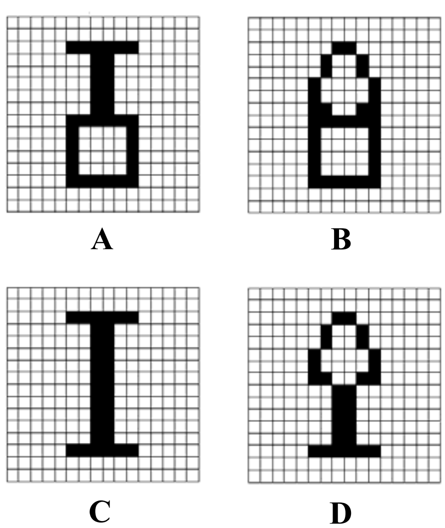

Figure 4: Inputs for the categorization task simulation

Four input patterns were presented to the network (Figure 4). First, feedback was given to the network that 4A and 4B belong to one category (1, 0) and 4C, 4D belong to another (0, 1). With this horizontal categorization rule, 50 cycles were run with the four input patterns presented in a random order during each cycle. The mean squared error of the output units displayed a rapid decrease (Figure 5B). The network learned to distinguish 4A and 4B as members of one category from 4C and 4D belonging to another. That is, when 4A or 4B were presented, output unit 1 was active and unit 2 was not. On the contrary, when 4C or 4D were presented, output unit 2 was active and unit 1 was off. The two output units can be considered detectors for the two categories. The learned weights of the connections between the competitive layer and each of the two output units are shown on Figure 5A. Only two of the competitive units had positive weights to output unit 1 and the other two had positive weights to output unit 2. Thus the output units had learned to ignore the responses of those lower-level nodes that were not relevant for categorization and combined together those parts which were relevant, forming diagnostic shape detectors (Figure 5C, 5D). The formation of the detectors was not influenced by the number of lower-level competitive units that participated in the shape representation. The result was the same with smaller 4x4 receptive fields. This change led only to the same diagnostic shape detectors being composed of four instead of two competitive units. The competitive units participating in a detector's representation were specialized for small input patterns contained within their receptive fields. The global representation activated by the whole input pattern, however, was a continuous shape honoring the Gestalt principle of Good Continuation.

In a second simulation, a vertical categorization rule was applied to a network with identical parameters. This time patterns 4A and 4C were from the same category (1, 0) while patterns 4B and 4D were from the other (0, 1).

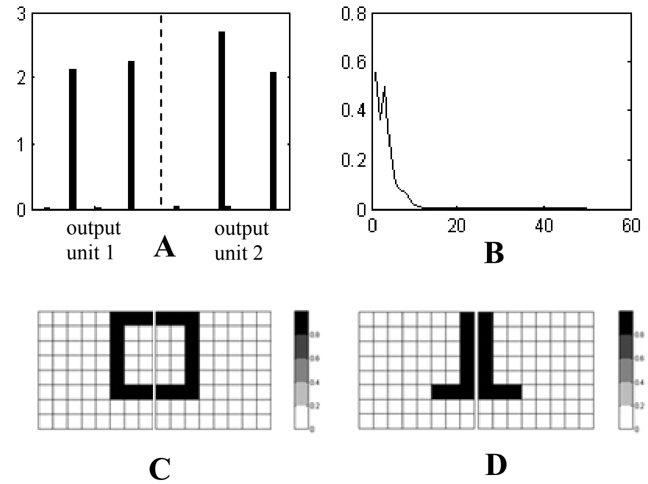

Figure 5: Panel A – weights between the competitive layer and the two output nodes. Panel B – mean square error for the output nodes. Panel C – the pixel-to-unit weights for the two competitive units with positive weights to output unit 1. Panel D – the pixel-to-unit weights for the two competitive units with positive weights to output unit 2.

The results from the second simulation are compared to the outcomes of the first simulation in Figure 6. For visualization purposes the output layer weights are multiplied by the competitive layer weights, which represent the participation of each pixel in the diagnostic shape detectors that were formed at the output layer. The same patterns led to the formation of different detectors when the vertical categorization rule was applied. This result was very stable over simulations and replicated the type of results reported by Pevtzow and Goldstone (1994).

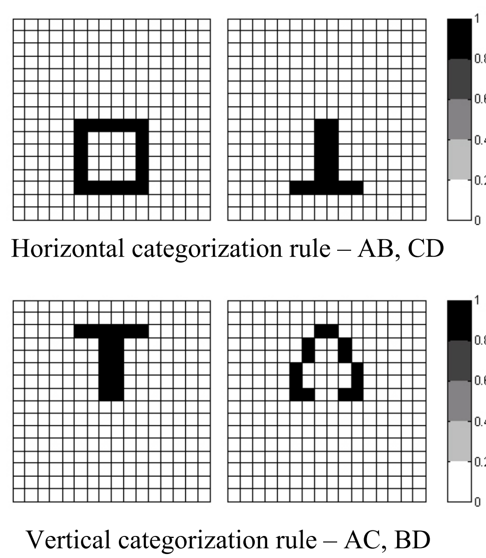

Figure 6: Detectors built according to a horizontal and vertical categorization rule.

Inspection of all specialized competitive units showed that there was no difference in their representation after the vertical and horizontal rule simulations. This means that the general structure of the input space was captured every time by the competitive units. Correct categorization was due to the formation of a diagnostic shape detector at the output layer.

General Discussion

The model shows a reliable ability to replicate at least two empirical results with minimal changes in parameters. Both unsupervised and supervised learning is possible. A general conclusion from the simulation results is that there are automatic low-level changes that capture the structure of visual stimuli irrespective of the given task. However when feedback is available, a more complex shape representation is constructed at a higher-level to accommodate the task requirements.

Another interesting conclusion comes from the unsupervised behavior of the network. The simple and plausible mechanism of competitive learning, reinforced by the horizontal Hebbian connections, is able to extract perceptual categories that are statistically present in the input space. This strongly supports empirical findings that Gestalt principles of perceptual organization can at times be overruled by category learning. The model also suggests a way in which even certain Gestalt principles like continuity can be learned, rather than built-in, as a consequence of experience with a learning environment that includes visual patterned stimulation (Quinn & Bhatt, 2005; Spelke, 1982).

The presented simulations have shown that it is computationally possible to account for both supervised and unsupervised perceptual learning without using built-in primitive features at the level that is eventually diagnostic for categorization. This was achieved by a fairly simple structure and by plausible mechanisms. The suggested model for perceptual learning is a first step toward a more global approach to learning that intends to bring together concepts and perception.

Acknowledgments

This research was funded by NIH Grants HD-42451 and HD-46526 (to the third author), and Department of Education, Institute of Education Sciences grant R305H050116 and National Science Foundation grant 0527920 (to the fourth author).

References

Behrmann, M., Zemel, R. S., & Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1011-1036.

Goldstone, R. L. (2000). A neural network model of concept-influenced segmentation. *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society* (pp. 172-177). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 86-112.

Goldstone, R. L., & Barsalou, L. (1998). Reuniting perception and conception. *Cognition*, *65,* 231-262

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*, 116-139.

Mozer, M. C., Zemel, R. S., Behrmann, M., & Williams, C. K. I. (1992). Learning to segment images using dynamic feature binding. *Neural Computation*, *4,* 650-665

Olshausen B. A., & Field D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381,* 607-609

Petrov, A., Dosher, B., & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, *112,* 715-743.

Pevtzow, R., & Goldstone, R. L. (1994). Categorization and the parsing of objects. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 717-722). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. Science, 256, 1018–1021.

Quinn, P. C., & Bhatt, R. S. (2005). Learning perceptual organization in infancy. *Psychological Science, 16,* 515-519.

Quinn, P. C., & Schyns, P. G. (2003). What goes up may come down: Perceptual process and knowledge access in the organization of complex visual patterns by young infants. *Cognitive Science, 27,* 923-935.

Quinn, P. C., Schyns, P. G., & Goldstone, R. L. (2006). The interplay between perceptual organization and categorization in the representation of complex visual patterns by young infants. *Journal of Experimental Child Psychology*, *95,* 117-127.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 9,* 75-112.

Shiu, L., & Pashler H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics*, 52, 582-88.

Spelke, E. S. (1982). Perceptual knowledge of objects in infancy. In J. Mehler, M. Garrett, & E. Walker (Eds.), *Perspectives on mental representation* (pp. 409-430). Hillsdale, NJ: Erlbaum.

Watanabe, T., Náñez, J., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*, 844-848.