(or anyone) to make a choice (as between a toaster and new Importantly, alignable differences require representations. ?vel cognitive processes like those described in this section are ?plained by using representations. Dynamics are usually less theories of higher level cognitive processing than in theories ?el cognitive processing. Thus, the chapters in this section are because they show that the dynamics that seem so central in and action also apply to more complex cognitive processes.

ES

Simon, H. A. (1976). Computer science as empirical inquiry: Symbols Search.
'ions of the ACM, 19, 113-126.

8

Interactions Between Perceptual and Conceptual Learning

Robert L. Goldstone
Mark Steyvers
Jesse Spencer-Smith
Alan Kersten
Indiana University, Bloomington

Confusions arise when *stable* is equated with *foundational*. Spurred on by the image of a house's foundation, we find it tempting to think that something provides effective support to the extent that it is rigid and stable. We argue that when considering the role of perception in grounding our concepts, exactly the opposite is true. Our perceptual system supports our ability to acquire new concepts by being flexibly tuned to these concepts. Whereas the concepts that we learn are certainly dependent on our perceptual representations, we argue that these perceptual representations are also influenced by the learned concepts. In keeping with one of the central themes of this book, behavioral adaptability is completely consistent with representationalism. In fact, the most straightforward account of our experimental results is that concept learning can produce changes in perceptual representations, the "vocabulary" of perceptual features used by subsequent tasks.

This chapter reviews theoretical and empirical evidence that perceptual vocabularies used to describe visual objects are flexibly adapted to the demands of their user. We extend arguments made elsewhere for adaptive perceptual representations (Goldstone, Schyns, & Medin, 1997; Schyns, Goldstone, & Thibaut, 1998) and discuss research from our laboratory illustrating specific interactions between perceptual and conceptual learning. We describe computer simulations that provide accounts of these interactions by using neural network models. These models have detectors that become increasingly tuned to the set of perceptual features that support concept learning. The bulk of the chapter is organized around

mechanisms of human perceptual learning and computer simulations of these mechanisms.

FIXED AND FLEXIBLE FEATURE SETS

A dominant notion in cognitive science is the idea that cognition involves operations on a fixed set of hardwired primitive features. The fixed set of features provides the building blocks for representing objects. This idea has been highly productive because of its parsimony; a wide variety of objects can be represented by combining a small number of existing features in different arrangements. The field of linguistics saw one of the first feature set theories, in which phonemes are represented by the presence or absence of fewer than 12 features such as *voiced, nasal,* and *strident* (Jakobson, Fant, & Halle, 1963). Ascending in complexity, Schank (1972) proposed representing entire situations, such as ordering food in a restaurant, in terms of a set of 23 primitive concepts such as *physical transfer, propel, grasp,* and *ingest.* In the field of object recognition, Bieder-man (1987) proposed a set of 36 geometric shapes such as *wedge* and *cylinder* to be used for representing objects such as telephones and flash-lights. Wierzbicka (1992) proposed a set of 30 semantic primitives includ-ing *good, want, big,* and *time* to be composed together in structured phrases to generate all other words. Much of the work in concept learning, in-cluding the seminal work of Bruner, Goodnow, and Austin (1956), as-sumes that experimental participants come into the psychology laboratory already possessing the primitive features that they need to learn how to categorize objects. For example, participants know, before the categoriza-tion experiment begins, that a set of objects is to be described in terms of color, number, and shape. In short, a prevalent assumption in cognitive science is that cognition consists of combining the elements of a fixed set of *a priori* features.

Although appropriate and mandatory for answering many questions, this "fixed features" approach systematically overlooks situations in which the concept learner, in addition to learning an association between a set of features and a category, must also learn what counts as a feature. Perceptual features that are currently available constrain the concepts that are acquired (as in traditional concept learning systems), but we argue that the concepts to be acquired also influence the features that are developed. Thus, in contrast to theories that posit fixed features, the alternative pursued here is that the building blocks of cognition are neither fixed nor finite, but rather adapt to the requirements of the tasks for which they are employed (Schyns, Goldstone, & Thibaut, 1998; Schyns & Mur-phy, 1994). As argued by Gibson (1969), the perceptual interpretation of an entity depends on the observer's history, training, and acculturation.

These factors, together with psychophysical constraints, mold one's set of building blocks. There may be no single, unique set of perceptual primitives because the building blocks themselves are adaptive.

One of the notorious difficulties with representations based on a limited set of elements is that it is hard to choose exactly the right set of elements that suffices to accommodate all future entities that need to be represented.[1] On the one hand, if a small set of primitive elements is chosen, then it is likely that two entities that must be distinguished with any likely combination of available primitives eventually arise. On the other hand, if a set of primitives is sufficiently large to construct all entities that might occur, then it likely includes many elements that lie unused, waiting for their moment of need to possibly arise (Schyns et al., 1998). By developing new elements as needed, newly important discriminations can cause the construction of building blocks that are tailored for the discrimination.

There is substantial neurological evidence for perceptual learning via imprinting of specific features in a stimulus. Weinberger (1993) reviewed evidence that cells in the auditory cortex become tuned to the frequency of often-repeated tones. Deeper in the cortex, cells in the inferior temporal area can be tuned by extended experience (over 600,000 trials) to particular views of three-dimensional (3D) objects (Logothetis, Pauls, & Poggio, 1995). Cells in the inferotemporal cortex can also be selective for particular faces, and this specificity is at least partially acquired given that it is especially pronounced for familiar faces (Perrett et al., 1984). There is neurological evidence that cortical areas involved with early perceptual processes are flexible, context sensitive, and tuned by training. For exam-ple, practice in discriminating small motions in different directions sig-nificantly alters electrical brain potentials that occur within 100 millisec-onds of the stimulus onset (Fahle & Morgan, 1996). These electrical changes are centered over the primary visual cortex, which suggests plas-ticity in early visual processing. Karni and Sagi (1991) found evidence, based on the specificity of training to eye (interocular transfer does not occur) and retinal location, that is consistent with early, primary visual cortex adaptation in simple discrimination tasks. In fact, training in an auditory selective attention task may produce differential responses as early as the cochlea—the neural structure that is connected directly to the eardrum via three small bones (Hillyard & Kutas, 1983). In sum, there is an impressive amount of converging evidence that experimental training leads to changes in very early stages of information processing.

These theoretical and neurophysiological sources of evidence for experience producing perceptual changes parallel evidence from expert-novice differences. In many fields, including radiology (Myles-Worsley, Johnston, & Simons, 1988), gender discrimination of day-old chicks (Biederman & Shiffrar, 1987), and beer tasting (Peron & Allen, 1988) experts organize or parse the world differently than do novices. In these fields, part of what it means to be an expert is to have developed perceptual tools for analyzing the stimuli in a domain. In what follows, we explore some potential laboratory analogs of the development of perceptual expertise, albeit on a much shorter course of training. The experiments are organized in terms of particular mechanisms of interaction between perception and concept learning: sensitization of existing perceptual dimensions, sensitization of novel perceptual dimensions, perceptual reorganization, and unitization.

DIMENSION SENSITIZATION

One way in which perception becomes adapted to tasks and environments is by increasing the attention paid to perceptual dimensions that are important, by decreasing attention to irrelevant dimensions, or by both. Attention can be selectively directed toward important stimulus aspects at several different stages in information processing. Attention may be applied relatively late in information processing to strategically emphasize important dimensions (Nosofsky, 1986). Alternatively, attentional shifts may be perceptual, rather than strategic or judgmental, in nature. One source of evidence that shifts are not completely voluntary is that attentional highlighting of information occurs even if it is to the detriment of the observer. When a letter consistently serves as the target in a detection task and then later becomes a distracter—a stimulus to be ignored—it still automatically captures attention (Shiffrin & Schneider, 1977). The converse of this effect, negative priming, also occurs: Targets that were once distracters are responded to more slowly than never-before-seen targets (Tipper, 1992). Although most research has investigated the sensitization of relevant dimensions, perceptual learning can also involve the loss of an ability to discriminate along irrelevant dimensions. For example, Myles-Worsley et al. (1988) showed that expert radiologists have poorer recognition memory for x-rays that do not show disease than do less expert medical professionals. Also, Werker and Tees (1984) showed that adults have poorer discrimination abilities for certain non-native sounds than do infants.

In addition to entire dimensions becoming sensitized if relevant, particularly important regions in a dimension can also be sensitized. The

largest body of empirical work showing an influence of categories on perception comes from work on categorical perception. 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 (Harnad, 1987). For example, Liberman, Harris, Hoffman, and Griffith (1957) generated a continuum of equally spaced consonant-vowel syllables changing continuously from /be/ to /de/. At a certain point on this continuum, people rather abruptly shift from identifying the sound as a /be/ phoneme to identifying it as a /de/. Moreover, people are better able to discriminate between two sounds that belong to different phonemic categories such as /be/ and /de/ than they are able to discriminate between two sounds that belong in the /be/ category, even when the physical differences between the pairs of sounds are equated. As such, perceptual sensitivity is at a peak at the boundary between phonemic categories.

There is an ongoing controversy about whether categorical perception effects are due to innate or learned categories. On the side of innateness, infants as young as 4 months show categorical perception for speech sounds (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Furthermore, chinchillas show categorical perception effects for speech sounds akin to those produced by people (Kuhl & Miller, 1978), even though chinchillas presumably have little exposure to human language. On the side of experience, categorical perception in humans is modulated by the listener's native language and extended training. In general, a sound difference that crosses the boundary between phonemes in a language is more discriminable to speakers of that language than to speakers of a language in which the sound difference does not cross a phonemic boundary (Repp & Liberman, 1987). Laboratory training on the sound categories of a language can produce categorical perception among speakers of a language that does not intrinsically have these categories (Pisoni, Aslin, Perey, & Hennessy, 1982).

Work in our laboratory has found visual analogs to the trained categorical perception effects observed with speech. In Goldstone (1994), participants were first given categorization training involving the sizes or brightnesses of squares. On each trial of categorization training, a square appeared on the screen and participants were asked to categorize it into Category A or B. The "size categorizers" group received feedback indicating that the squares in the left and right two columns of Fig. 8.1 belonged to Category A and Category B, respectively. The "brightness categorizers" group received categorization training in which the squares in the upper and lower two rows of Fig. 8.1 belonged to Category A and Category B, respectively. The squares were calibrated so that the differences between adjacent squares were just barely detectable. Subsequent

to 1.5 hours of categorization training, participants were transferred to a same/different judgment task in which horizontally or vertically adjacent squares from Fig. 8.1 were presented or the same square was repeated twice. Participants were required to respond as to whether the two squares were exactly identical on both their size and brightness or differed even slightly on either dimension. When a dimension was relevant for categorization, participants' same/different judgments along this dimension were more accurate (based on the d' measure from signal detection theory) than those from participants for whom the dimension was irrelevant and from control participants who did not undergo categorization training. This trend, found for both categorization groups, is shown in Figs. 8.2 and 8.3. The greatest sensitization of the categorization-relevant dimension was found along those particular dimension values that were the boundaries between the learned categories. However, the sensitization of the relevant dimen..sion also extended to other values along the dimension even though these other values were originally placed in the same category. Thus, entire relevant dimensions are sensitized, but critical regions in those dimensions are also sensitized. In addition, Fig. 8.3 shows the one case of *acquired equivalence* that was found, in which a dimension that was *irrelevant* for categorization became desensitized relative to control participants. Compared with the control group of participants who were

Brightness / **Size**

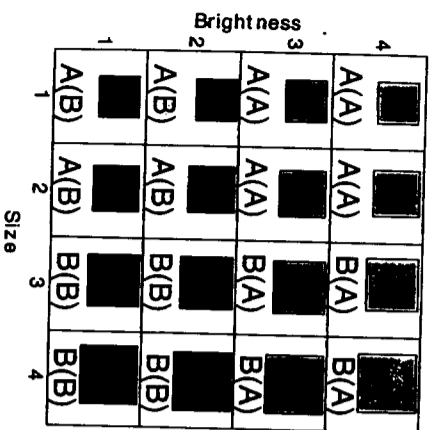|	Size 1	2	3	4
Brightness 1	A(B)	A(B)	B(B)	B(B)
2	A(A)	A(A)	B(A)	B(B)
3	A(A)	A(A)	B(A)	B(A)
4	A(B)	A(B)	B(A)	B(A)

FIG. 8.1. Stimuli used by Goldstone (1994). Sixteen squares were constructed by factorially combining four values of brightness with four values of size. The letters outside the parentheses show the categorizations of the squares when size was relevant. The letters in the parentheses show the categorizations of the squares when brightness was relevant. Adapted from "Influences of Categorization on Perceptual Discrimination," by R. L. Goldstone, 1994, *Journal of Experimental Psychology: General, 123,* p. 183. Adapted with permission.

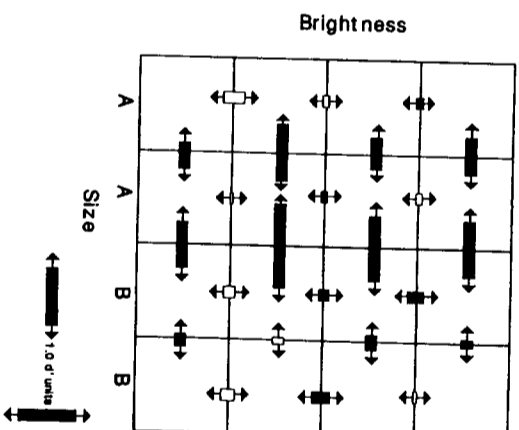### Size Relevant During Categorization

Brightness / **Size**

FIG. 8.2. This figure shows the change in perceptual sensitivity (measured in d' units) that is due to size categorization training. A black rectangle indicates a positive difference when the control groups' sensitivity is subtracted from the size categorizers' sensitivity. A white rectangle indicates a negative difference. The size of the rectangle indicates the magnitude of the difference. Rectangles are placed between the two squares that are being discriminated. The greatest sensitization occurs at the boundary between the two size categories.

1.0 d' units

Brightness Relevant During Categorization
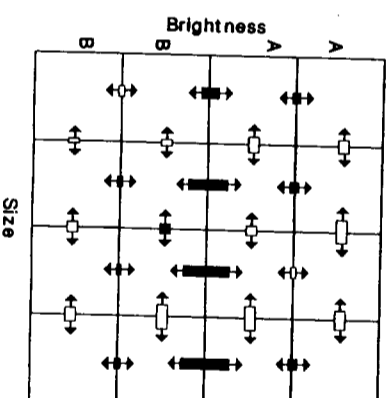
Brightness / **Size**

FIG. 8.3. This figure shows the change in perceptual sensitivity (measured in d' units) that is due to brightness categorization training. Each rectangle reflects the difference between the d' for brightness categorizers and the control group. The predominantly white, horizontal rectangles reflect a significant case of acquired equivalence whereby brightness categorizers are less adept than controls at making size discriminations.

given no categorization training, the desensitization that occurred for relevant dimensions was larger and more reliable than the desensitization that occurred for irrelevant dimensions.

SENSITIZATION OF NOVEL DIMENSIONS

Dimension sensitization following training provides evidence that not only do our perceptual encodings guide our categorizations, but that our categorizations also guide our perceptual encodings. However, shifts in dimensional attention do not necessarily require the postulation of new perceptual vocabulary elements. Existing elements may simply be emphasized or de-emphasized. From the same paradigm, we believe that we are also getting evidence for a second type of perceptual learning involving dimensionalization—the development of new dimensions. Size and brightness are easily distinguishable and are likely to be psychological dimensions for our participants before categorization training. We replicated the experiment just reported by using dimensions that people are less likely to register *as* dimensions (Goldstone, 1994). We used the brightness and saturation of colors, two dimensions that are often cited as the classic examples of integral dimensions (Garner, 1974). Dimensions are considered integral if it is difficult to attend to one dimension without also attending to the other dimension. However, after prolonged categorization experience in which brightness was relevant and saturation was irrelevant (or vice versa), we found that the relevant dimension became selectively sensitized. Thus, dimensions that were once fused can become more isolated with the proper categorization training. This result is consistent with evidence that color experts (art students and vision scientists) are better able to selectively attend to dimensions (e.g. hue, chroma, and value) that make up color than can nonexperts (Burns & Shepp, 1988). A large developmental literature suggests that people often shift from perceiving stimuli in terms of holistic, overall aspects to analytically decomposing objects into separate dimensions (Smith, 1989). This trend can be described as the construction of new perceptual vocabulary elements that are used to build object descriptions.

Sensitization of Entire Novel Dimensions

It has been argued that saturation and brightness, although they are integral dimensions for most people, are not genuinely arbitrary dimensions (Grau & Kemler-Nelson, 1988). To show the dimensionalization process for genuinely arbitrary dimensions, we have recently begun to explore situations where dimensions are generated by morphing between two pairs of arbitrarily chosen faces. One dimension is created by morphing between the top two faces shown in Fig. 8.4, and a second dimension
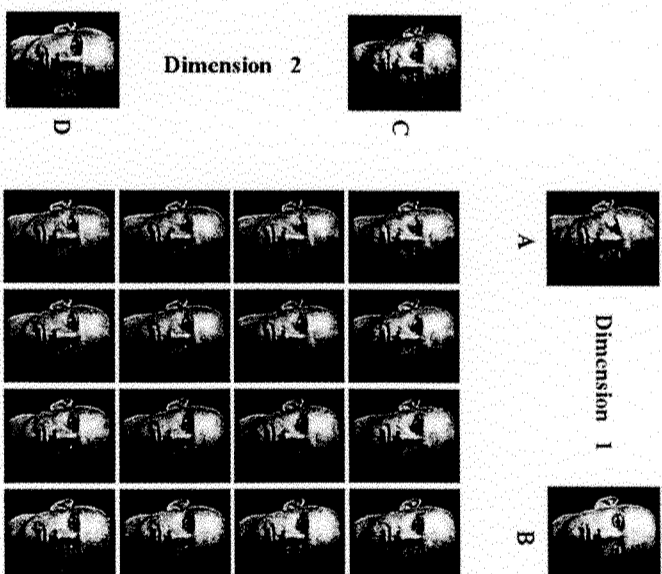
FIG. 8.4. Arbitrary dimensions can be created by generating a series of morphs between two randomly chosen faces. Dimension 1, running horizontally, morphs between Faces A and B. Dimension 2, running vertically, morphs between Faces C and D. Each of the faces in the 4 by 4 array possesses a unique set of coordinates on Dimensions 1 and 2.

is created by morphing between the two faces on the left. Using a technique described by Steyvers (1999), a four by four matrix of faces can be created from these two dimensions such that each face is defined one half by its value on Dimension 1 and one half by its value on Dimension 2. Arbitrary dimensions are thus generated by creating negative contingencies between two faces—the more of Face A that is present in a particular morphed face, the less of Face B there is. The horizontal dimension, Dimension 1, might be called the "The proportion of Face A relative to Face B" dimension. We refer to the vertical dimension as Dimension 2.

Just as in the previously described experiments, participants were initially given a categorization rule to learn that divided the four by four stimulus array of Fig. 8.4 either vertically or horizontally into equal halves. On each trial, participants saw a face and categorized it into one of two categories, with feedback from the computer indicating whether or not the participant was correct. Whereas Goldstone's (1994) participants were transferred to a same/different task, Goldstone and Steyvers (1999) par-

ticipants were transferred to another categorization task. The initial and transfer categorizations were related to each other by one of the seven ways shown in Table 8.1. In the representation used in Table 8.1 and Fig. 8.4, the dimension above the line is relevant, and the dimension below the line is irrelevant. Figure 8.4 would be represented as 1/2; Dimension 1 (morphing from Face A to B) is relevant, and Dimension 2 is irrelevant. Different faces were used as the anchoring end points for each of the four dimensions (one through four). Dimensions that were relevant during the first categorization could continue to be relevant during the second categorization, could become irrelevant, or could become absent altogether, and the same was true for irrelevant dimensions. For example, if the original categorization was 2/3 (Dimension 2 was relevant, and Dimension 3 was irrelevant) and the subsequent categorization was 1/2, then the dimension that was originally relevant becomes irrelevant and a new dimension becomes relevant for the final categorization.

Suggestive evidence of dimensionalization with these materials is that participants become increasingly adept at attending to one dimension while ignoring variation on irrelevant dimensions during the initial category learning. A more important measure is the categorization accuracy during the final categorization phase of the experiment, which was identical for all seven groups and involves the categorization 1/2 (Dimension 1 = relevant, 2 = irrelevant). As such, any systematic differences between conditions on final categorization performance must be due to differences in how the initial categorization prepared them for this final categoriza-

TABLE 8.1
Seven Conditions of an Experiment by
Goldstone and Steyvers (in preparation)

Initial Training	Subsequent Transfer	Relation Between Training and Transfer
1 Relevant / 2 Irrelevant	1 Relevant / 2 Irrelevant	Relevant and irrelevant dimensions are both preserved.
1 Relevant / 3 Irrelevant	" "	Relevant dimension is preserved.
3 Relevant / 2 Irrelevant	" "	Irrelevant dimension is preserved.
3 Relevant / 1 Irrelevant	" "	Irrelevant dimension becomes relevant.
2 Relevant / 1 Irrelevant	" "	Irrelevant dimensions become relevant. Relevant dimension becomes irrelevant.
2 Relevant / 3 Irrelevant	" "	Relevant dimension becomes irrelevant.
3 Relevant / 4 Irrelevant	" "	Control—completely new dimensions.

% Correct Transfer to $\frac{1}{2}$
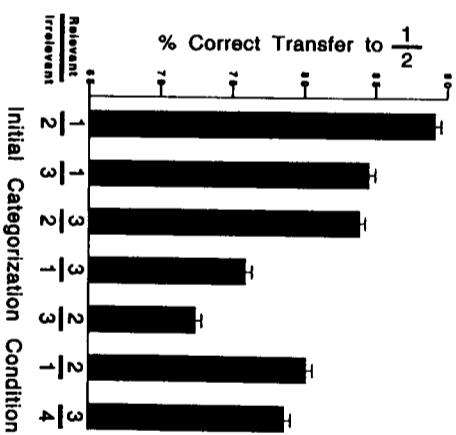
Initial Categorization Condition

FIG. 8.5. Results from the transfer experiment. Each dimension is represented by the two faces that function as its end points. The dimension in the "numerator" was relevant during the initial categorization, and the dimension in the "denominator" was irrelevant. The bars show the overall percentage correct when each of seven initial categorization conditions was transferred to a 1/2 categorization, wherein Dimension 1 was relevant and 2 was irrelevant. The last bar shows the results from the control condition in which the initial and transfer categorization rules used completely different dimensions.

tion. The results, in Fig. 8.5, show several types of transfer based on the initial categorizations. The degree of transfer in a condition is best appraised by comparing it to the 3/4 control condition in which the initial and final categorizations involve completely different faces and dimensions. The categorization advantage of the first three conditions, 1/2, 1/3, and 3/2, over the control condition suggests that participants learn to selectively emphasize relevant dimensions and to de-emphasize irrelevant dimensions. That is, when initial and final categorizations share relevant or irrelevant dimensions, performance is better than in the control condition. This transfer is impressive because these conditions use completely new sets of faces in the final categorization. For example, none of the faces belonging to the 1/2 set is the same as faces from the 3/2 set. The only similarity between these sets is that Dimension 2 is irrelevant for both sets, and this similarity has a large beneficial effect on transfer.

The next two conditions of Fig. 8.5 demonstrate negative transfer effects owing to shared dimensions. Relative to the control condition (3/4), when irrelevant dimensions become relevant (3/1) and when relevant dimensions become irrelevant (2/3), performance suffers. The latter effect is particularly strong and is reminiscent of Shiffrin and Schneider's (1977) results that when participants are trained to respond to a particular letter

as a target, performance is quite poor when that letter later becomes a distracter to be ignored. The results from the 2/1 condition seem surprising at first. In this condition, the relevant dimension becomes irrelevant *and* the irrelevant dimension becomes relevant, and yet performance is better than for the control condition. Our explanation for the beneficial transfer from 2/1 to 1/2 categorizations rests on the observation that both involve the same set of 16 faces. The categorization rules are orthogonal (separated by 90 degrees), splitting the stimuli horizontally or vertically. As such, both rules depend on separating the horizontal dimension from the vertical dimension to selectively attend to only one of these dimensions. Effective performance on the 2/1 categorization requires isolating Dimension 1 from Dimension 2. Once accomplished, this isolation may be useful in acquiring the 1/2 categorization because this categorization also requires the same differentiation of dimensions, albeit for opposite purposes.

This account is supported by additional experiments from our laboratory showing that transfer involving 90-degree rotations of a categorization boundary results in better final performance than does transfer involving 45-degree rotations. Although the 45-degree rotation might be expected to produce better performance because dimensions that are relevant in the final categorization are semirelevant during the initial categorization, this advantage is apparently overwhelmed by the incompatible, cross-cutting dimensionalizations that are required. In contrast, categorization boundaries that are separated by 90 degrees are compatible in promoting the same differentiation of one dimension from another. These results, in conjunction with the results shown in Fig 8.5, suggest that categorization learning involves not only allocating attention to existing dimensions, but also in isolating dimensions in the first place. In fact, it is possible to allocate attention to a dimension only if that dimension has previously been isolated. In some cases, early childhood experience or innate perceptual devices suffice to isolate dimensions, and categorization learning requires only attention weighting. However, in situations involving objects with dimensions that are initially integral for people, category learning also requires the construction of dimensional representations. Once constructed, these representations are employed for learning subsequent categorizations (for further evidence, see Schyns & Rodet, 1997).

Sensitization of Regions of Novel Dimensions

In addition to sensitizing entire dimensions, regions in novel dimensions can become sensitized, giving rise to categorical perception effects. Goldstone, Steyvers, and Larimer (1996) generated an arbitrary dimension by generating two random bezier curves and treating these objects as end

points on a continuum. The values along the dimension were created by morphing from one random bezier to the other, similar to how the face dimensions were created. The stimuli were made by creating 60 linearly interpolated morphs between two random curves and selecting the 7 central curves as stimuli. During categorization training, participants learned one of two categorizations based on different cutoff values along this dimension. For the left-split group, the first three objects in Fig. 8.6 belonged to Category A, and the last four objects belonged to Category B. For the right-split group, the boundary between Categories A and B occurred between the fourth and fifth curves. A third control group learned a comparable categorization, but involving curves that were irrelevant for the subsequent task.

After categorization training, participants were transferred to a same/different judgment task. Participants were shown pairs of highly similar curves or the identical curve repeated twice and were instructed to say whether the curves had exactly the same shape or differed in any way.

The data of principal interest, shown in Fig. 8.7a, were participants' sensitivities at discriminating between pairs of adjacent curves, broken down as a function of their categorization condition. A d' measure of sensitivity was calculated based on participants' ability to correctly respond "same" and "different." Specifically, it is a function of the probability of responding "different" given that the curves were indeed different minus a function of the probability of incorrectly responding "different" given that the curves were the same. The d' values increase as participants' ability to correctly make discriminations increases.

One result of the experiment is that sensitivity is higher for the left-and right-split groups than for the group that was trained on an irrelevant categorization. This effect is consistent with previous work showing that pre-exposure to stimuli leads to their heightened discriminability (Gibson & Walk, 1956). More relevant to categorical perception, there was a significant difference between the pattern of sensitization for the left- and right-split groups. Although the effects of the two groups were not symmetric, the general effect of categorization training is that discriminability is relatively high for stimuli that fall near the category boundary. To
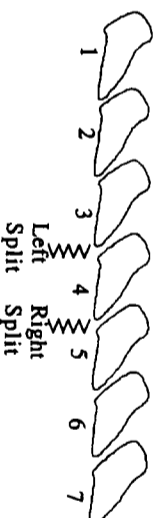

FIG. 8.6. These seven curves were constructed by morphing between arbitrary curves. The left-split and right-split groups saw the same seven curves, but the middle curve was categorized differently by the two groups.
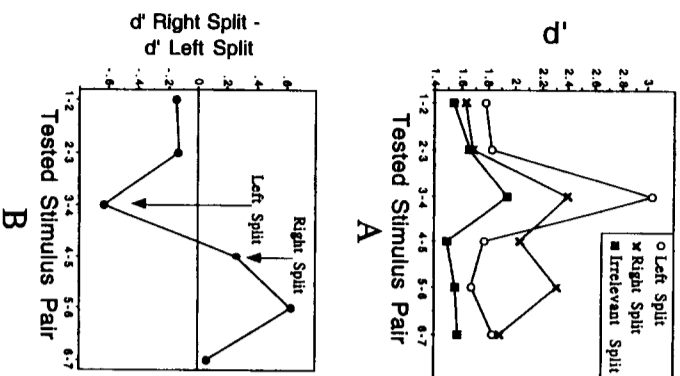
FIG. 8.7. The numbers on the horizontal axis reflect the numbers associated with the compared curves from Fig. 8.6. Fig. 8.7A shows participants' sensitivity (measured in d' units) at discriminating between adjacent curves. Fig. 8.7B plots the same data, but using a derived measure that is the difference between the right and left categorization groups. In general, the categorization condition with the categorization boundary closest to the tested pair had the highest sensitivity at discriminating the pair.

visualize this effect, Fig 8.7B plots a new measure derived from the data shown in Fig. 8.7A. In this figure, the sensitivity (d') of the left-split group is subtracted from the sensitivity of the right-split group. Thus, this measure is positive when the right-split group shows a greater sensitivity than the left-split group for a pair of curves. Figure 8.7B shows that the left-split group does relatively well when and only when the pair of tested curves lies closer to the left boundary than to the right boundary.

A Neural Network Model of Dimensional Sensitization

In developing a computational model for the observed categorical perception effects, we were drawn to neural networks that possess hidden units that intervene between inputs and outputs and are capable of creating internal representations. For our purposes, these hidden units can

be interpreted as learned feature detectors and represent the organism's acquired perceptual vocabulary. The model consists of three processing stages, shown in Fig. 8.8. In the first stage, the input images are processed by a set of Gabor filters. In the second stage, a layer of hidden units learns to represent the perceptual dimensions along which the continuum of stimuli falls. The representation of the hidden units is changed by an unsupervised learning algorithm similar to Kohonen's self-organizing maps (e.g. Kohonen, 1995). In the last stage, a layer of category units classifies the input image based on the activity in the hidden unit layer. The weights from the hidden layer to the category units are learned in a supervised manner (Kruschke, 1992). The critical assumption of the model is that the input-to-hidden weights are influenced by the hidden-to-category weights. By unsupervised learning, the topology of the hidden detector units comes to reflect the morph-based dimension that underlies the experimentally created stimuli. By the category level supervision, the distribution of detectors is biased by the demands of the categorization.

The input patterns to the network are gray-scale, two-dimensional pictures of curves, and the categorization of the curve is supplied as a teacher signal for the category units. Twenty-eight curves are created by using the same technique of morphing between two arbitrary curves used in the experiment. The first stage of the network preprocesses these pictures by a set of Gabor filters (Daugman, 1985) with maximal sensitivities to line segments oriented at 0, 45, 90, or 135 degrees. The receptive fields of the filters are positioned at overlapping local regions of the image. The Gabor filters reduce the information contained in the original images to a manageable amount and capture some of the higher order shape invariants associated with a curve not captured by pixel-based representations. Figure 8.9 shows an example of the transduction of an image into Gabor

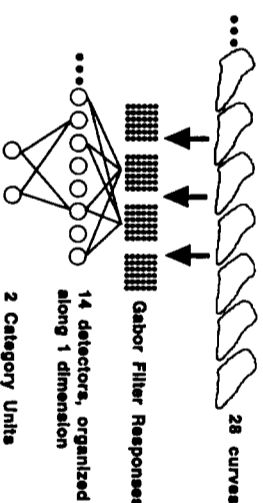FIG. 8.8. An overview of the SOS network. The bezier curve images are passed through Gabor filters, and the resulting response patterns are presented to a one-dimensional set of detector units. These detectors adapt toward the filtered inputs, but are also influenced by the categorization of the inputs. Representative bezier curves, detector units, and connections are shown in this illustration.
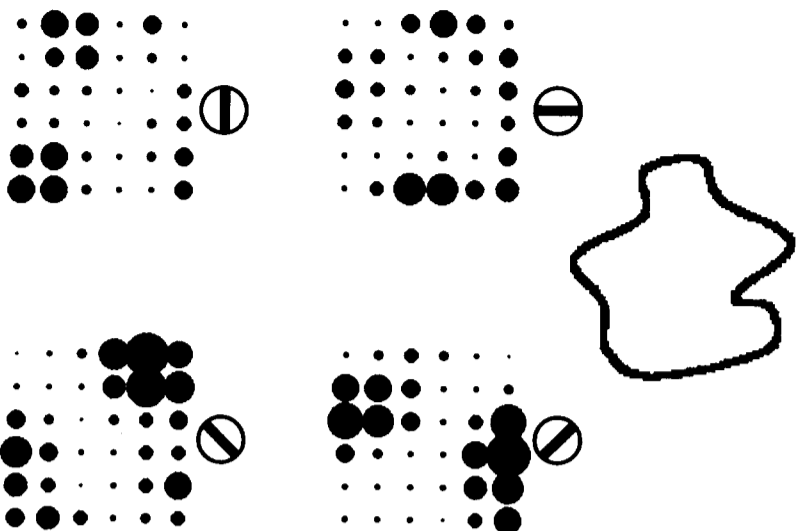
FIG. 8.9. This figure shows the Gabor filter responses for the bezier curve at the top. The activity of a filter is indicated by the magnitude of the black circle. For example, the large circles in the upper-left portion of the 45-degree-angle filter (the lower-right filter) indicate a strong 45-degree-angle component in the upper-left portion of the bezier curve.

filter responses. As a result of this process, each image is represented by 144 real numbers (36 image locations for each of four orientations).

The next processing stage involves the hidden layer consisting of 14 detector units. An individual detector becomes most active when an image leads to a Gabor filter activation pattern that matches the weights from the Gabor filters to the detector. The input-to-detector weight pattern can also be thought of as locations of the detector node in the space determined by the Gabor filters. At the start of learning, the detectors represent random activation patterns of these Gabor filters. By using a competitive learning rule, the hidden detector units become specialized for activation patterns caused by particular images or properties of these images. When a bezier curve is presented, the "winning" detector adjusts its weight

vector toward the curve's Gabor filter representation. A detector wins by having input-to-detector weight values that are closest to the Gabor filter activation. The extent to which the nonwinning detector nodes update their weights is restricted by the topology that is imposed on the feature detectors; we used a one-dimensional lattice such that each detector (except at the two end points) has two neighbors. Far neighbors update their weights less than close neighbors of the winning detector unit. This imposed topology creates a dimensional representation such that neighboring detectors respond to similar images or images having similar properties. More globally, the positions of the 14 detectors come to reflect the arbitrary morph-based dimension.

For the purposes of this chapter, we only want to mention the learning equation for adjusting the weights from the Gabor filter responses to the detector units:

$$\Delta w_{ji}^{det} = ELN_{(j,winner)}(a_i^{in} - w_{ji}^{det}),$$

where Δw_{ji}^{det} is the weight from the Gabor filter Response i to Detector j. L is a constant learning rate. The Function N is dependent on how far Detector j is from the winning detector; far neighbors to a winning detector give values close to 0, and close neighbors give values closer to 1, so that the most learning occurs for close neighbors of the winning unit. The $(a_i^{in} - w_{ji}^{det})$ factor adapts the detectors' weights toward the input activations; if the Gabor response is larger than the weight, then the input-to-detector weight increases to match it. So far, the description of the network conforms to a standard self-organizing map. A new factor is introduced with the term E, which is the total amount of error at the category units. The category unit activations depend on the weighted activation in the previous layer of hidden detector units. The category units learn in a supervised manner so that the errors in predicting category membership are used to update the weights to these category units. The term E is introduced in the learning equation of the hidden nodes to influence the rate of learning; if a stimulus leads to a miscategorization or a relatively undifferentiated response by the categorization units, the term E is high, which leads to an increased rate at which the winning detector unit and its neighbors move toward the current input activation. Stated more metaphorically, the network sends out an SOS to neighboring detector nodes to help handle the current miscategorized input. Because undifferentiated categorization responses or miscategorizations occur most frequently at or near the category boundary, the hidden detector units tend to migrate toward the categorization boundary. As a result, the region near the boundary is more densely populated by detector nodes.

In two separate runs of the SOS network, we chose two different locations for the category split, corresponding to left- and right-split groups.

Figure 8.10 shows the influence of category training on discrimination sensitivity. The activations for each of the 14 detector nodes are shown for each of the 28 curves presented to the network. As such, each of the 14 curves in Panel A (left split) and Panel C (right split) shows a response profile for one detector. The detectors are densely distributed around the categorization boundary as a result of the classification feedback in the learning rule for detectors. Importantly, the detectors are arranged topologically. As we move from left to right along the bank of detectors, we move along the arbitrary dimension that we experimentally formed. As such, the network has implicitly represented an abstract and arbitrary stimulus dimension through the topology of its detectors.

A sensitivity measure for same/different judgments was constructed by taking the Euclidean distance between the detector unit activation patterns for the two curves to be judged. Thus, the model tends to respond "different" to the extent that the two presented input patterns activate different detectors. As shown in Panels B and D of Fig. 8.10, the peak sensitivity occurs approximately at the category boundaries. This occurs because slightly different stimuli that occur near the category boundary cause quite different activation patterns on the detector units, given the dense concentration of detectors in this region.

The SOS network models categorical perception effects by creating relatively dense representations of items at the border between categories.

Detector Node Activations

Sensitivity

Stimulus Number   4 8 12 16 20 24 28
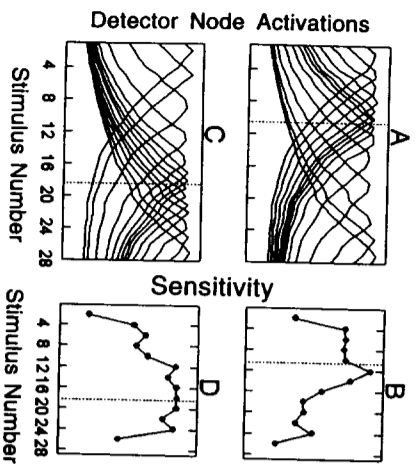
Stimulus Number   4 8 12 16 20 24 28

FIG. 8.10. The SOS network's simulation of the results shown in Fig. 8.7. The category boundaries are shown by dashed, vertical lines. The left panels show activation profiles for the 14 detector units when the categorization boundary was on the left (Panel A) or right (Panel C). In both cases, the categorization boundaries are densely populated with detectors, giving rise to the increase in sensitivity at the category boundaries shown in Panels B (left-split) and D (right-split).

This treatment of categorical perception differs from other neural network implementations. In Anderson, Silverstein, Ritz, and Jones' (1977) model, each category has its own attractor, and the stimuli that fall into one category are all propelled toward the category's attractor. In Harnad, Hanson, and Lubin (1994), the stimuli that fall into one category are repelled from the categorization boundary. One potential advantage of our account is that it 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 our experiment suggest that this is the case for our participants, and this effect has been shown more persuasively by other researchers (Iverson & Kuhl, 1995). 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 general, developing detectors that both cover input patterns and are tailored to categorization requirements is a promising avenue for modeling perceptual learning phenomena related to categorical perception, stimulus pre-exposure, and discrimination learning.

THE SEGMENTATION OF OBJECTS INTO PARTS

People organize objects into parts, not simply by carving nature at the joints, but by carving joints into nature. It is more natural for us to think of an X as being broken down into a left slash and a right slash than as being composed of a V and an upside-down V intersecting at a point, even though both are possible decompositions (McGraw, Rehling, & Goldstone, 1994). Palmer conducted several studies on the naturalness of parts in whole objects, exploring factors that make certain parts more natural than others (Palmer, 1977, 1978). Palmer also developed a quantitative part goodness model that used a number of objective factors about the parts and whole: how close the line segments in a part were to each other, whether they formed closed objects, whether they had similar orientations, and whether the line segments of a part were similar to line segments in other parts. Palmer found that this objective measure of part naturalness correlated highly with empirical methods for assessing subjective part goodness, such as requesting people to rate the naturalness of a part or measuring participants' response times to confirm that a particular part is contained in the whole.

Pevzow and Goldstone (1994) were interested in whether the naturalness of a part in a whole depends on not just the objective physical properties that Palmer considered, but also a person's subjective experience. In particular, we thought that how natural a part is might depend

on whether it has been useful for categorization. In the same way that the world looks like a nail to the person who has a hammer, to the person who has learned that a particular feature is diagnostic for needed categorizations, the world may look like it is built from this feature. To test this conjecture, we gave participants a categorization task, followed by part/whole judgments. During categorization, participants were shown distortions of the four objects A, B, C, and D shown in Fig. 8.11. The objects were distorted by adding a random line segment that was connected to the segments already present. Using an experimental design that should now be familiar to the reader, we gave participants extended training with either a vertical or horizontal categorization rule. For participants who learned that A and C were in one category and B and D were in another, the two component parts at the bottom of Fig. 8.11 were diagnostic. For participants who learned that A and B belonged in one category and C and D belonged to the other, the components on the right were diagnostic. During part/whole judgments, participants were shown a whole and then a part and were asked whether the part was contained in the whole. As with Palmer's studies, it is assumed that the faster a person can correctly confirm the presence of a part, the more natural the part is. Participants were given both present and absent judgments. Participants were given trials with parts that were previously diagnostic or nondiagnostic and with complements of these category parts. A complement is what remains in a whole when a category part (one of the components shown in Fig. 8.11) is removed.

The major result to note from Fig. 8.12 is that participants were faster to correctly respond "present" when the part was diagnostic than when it was nondiagnostic. To the extent that one can find response time analogs of signal detection theory sensitivity and bias, this effect seems to be a sensitivity difference rather than a bias difference, because absent judgments also tended to be faster for diagnostic than nondiagnostic parts. Given that a category part that was diagnostic for one group was nondiagnostic for the other group, it is not simply the physical stimulus properties that determine how readily a person can segment an object into a particular set of parts; segmentation is also influenced by the learned categorical diagnosticity of the parts. The results for complements were unexpected; if a part was relevant during categorization, then participants were relatively slow to respond that the complement was present. One may have predicted the opposite because the part and its complement are consistent with the same segmentation of an object. However, the result is predicted if category parts attract attention to themselves when they are diagnostic, to the detriment of other parts in the display.

We have begun modeling the result from this experiment by using a competitive learning network. As with the experiment, the network is
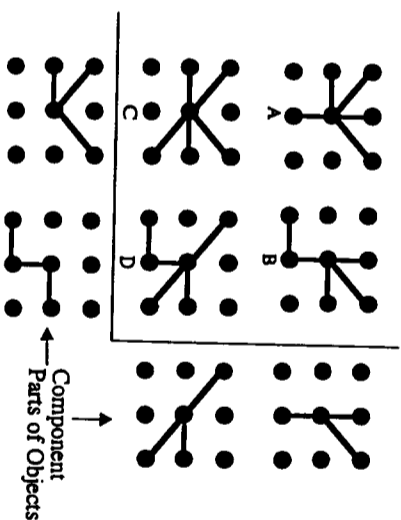

FIG. 8.11. Materials used by Pevtzow and Goldstone (1994). The four Objects A, B, C, and D were categorized into two groups. When A and B were placed in one group and C and D were categorized into two groups. When A and B were placed in one group and C and D were diagnostic. When A and C were placed in one group and B and D were placed in the other, then the parts on the bottom were diagnostic. Adapted from "Categorization and the Parsing of Objects," by R. Pevtzow and R. L. Goldstone, 1994, in *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (Hillsdale, NJ: Lawrence Erlbaum Associates), p. 719. Copyright 1994. Adapted with permission.
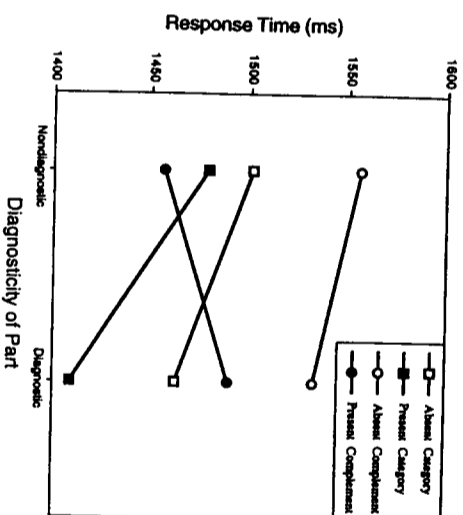

FIG. 8.12. It was easier to detect the presence or absence of a part in a whole object when the part was previously diagnostic for a categorization. The complement (the remaining line segments of the whole once the part has been removed) of a part was harder to find in a whole when the part was previously diagnostic for a categorization.

first given categorization training and then a subsequent segmentation task, using the same network weights. Similar to the simulation of acquired categorical perception, the network involves three layers: one representing the input patterns, one representing a bank of learned detectors, and one reflecting the category assignments of the inputs. Both the weights from the input patterns to the detectors and the weights from the detectors to categories are learned. The same network is used for categorizing and segmenting patterns, but the category units have an impact only during categorization. A schematic illustration of the network for the two tasks is shown in Fig. 8.13. The categorization task uses a standard unsupervised competitive learning algorithm (Rumelhart & Zipser, 1985), but includes a top-down influence of category labels incorporating supervised learning. The network begins with random weights from a 2D input array to a pair of detector units. When an input pattern is presented, the unit with the weight vector that is closest to the input pattern is the winner and selectively adjusts its weights to become even more specialized toward the input. By this mechanism, the originally homogenous detectors become differentiated over time, splitting the input patterns into two categories according to which detector is specialized for each pattern. Abstractly, the competitive learning algorithm, if supplied with jazz and classical pieces of music, automatically learns to group the pieces into these categories without feedback, because the pieces naturally cluster into these two groups. However, given that we want the detectors to reflect the experiment-supplied categories, we need to modify the standard unsupervised algorithm. This is done by including a mechanism

Detectors

Category Units

Categorization Network
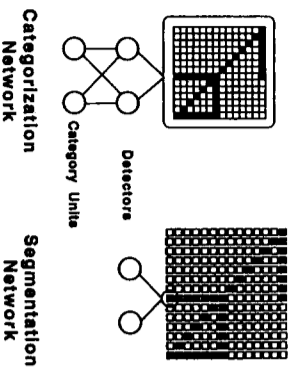
Segmentation Network

FIG. 8.13. The categorization and segmentation network used to simulate the results shown in Fig. 8.10. The same input-to-detector weights are used for both the categorization and segmentation tasks. During categorization, an entire pattern is fed in as an input at the same time, and one detector becomes specialized for the pattern. Effectively, detectors arise that sort the inputs into two categories according to their diagnostic line segments. During segmentation, a pattern is fed in one pixel at a time, and detectors learn to become specialized for pixels. Now, detectors sort the parts of one pattern into two segments.

such that detectors that are useful for categorizing an input pattern become more likely to win the competition to learn the pattern. The usefulness of a detector is assumed to be directly proportional to the weight from the detector to the presented category, which is provided as a label representing the input patterns. The input-to-detector weights do not have to be set before the weights from detectors to categories are learned.

With this modified competitive learning algorithm, if we present the same four pictures but with different categorizations, then different detectors develop. Detectors emerge that tend to selectively represent the diagnostic, shared components of input patterns. If A and B of Fig. 8.11 are assigned to the same category, as are C and D, then detectors tend to emerge that respond preferentially to the component parts on the right side of Fig. 8.11. However, if we change the categorization, then detectors for the lower components are created.

Thus far, the category learning network has been described. The basic insight is that segmentation tasks can also be modeled by using competitive learning, and thus the two tasks can share the same network weights and consequently influence each other. Competitive learning for categorization sorts complete, whole input patterns into separate groups. Competitive learning for segmentation takes a single input pattern and sorts the pieces of the pattern into separate groups. For segmentation, instead of providing a whole pattern at once, we feed in the pattern one pixel at a time. Instead of grouping patterns, the network groups pixels together. With this technique, if the "original pattern" in Fig. 8.14 is presented to the network, the network might segment it in the fashion shown in the top decomposition. This figure shows the weights from the 2D input array to each of two detectors and reflects the specializations of the two detectors. The two segments are complements of each other—if one detector becomes specialized for a pixel, the other detector does not. This stems from the basic operation of the competitive learning algorithm by which the winning detector indirectly inhibits the other detector from learning to adapt to the input. Unfortunately, this segmentation is psychologically absurd; nobody would decompose the original figure into these parts. To create psychologically plausible segmentations, we modify the determination of winners. Topological constraints on detector creation are incorporated by two mechanisms: Input-to-detector weights "leak" to their neighbors in an amount proportional to a Gaussian function of their distance, and input-to-detector weights also spread to each other as a function of their orientation similarity, defined by the inner product of four Gabor filter responses. The first mechanism produces detectors that tend to respond to cohesive, contiguous regions of an input. The second mechanism produces detectors that follow the principle of good continuation, dividing X into two crossing lines rather than two kissing sideways

Original pattern

Segmentation by competitive learning

Segmentation with perceptual constraints

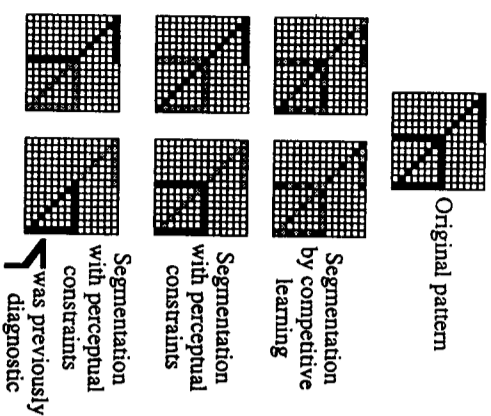Segmentation with perceptual constraints

↙ was previously diagnostic

FIG. 8.14. When a competitive learning network is supplied the original pattern, the detectors specialize for different pixels. When the network is supplemented with perceptual biases to develop spatially coherent, smooth features, segmentations such as the one shown by the middle pair of detectors are obtained. When this latter network is run after it has already acquired a detector for a particular component during categorization training, then the segmentation shown by the last pair of detectors is typically found.

Vs, because the two halves of a diagonal line are linked by their common orientation. Thus, if a detector wins for Pixel X (meaning that the detector receives the most activation when Pixel X is on), then the detector also tends to handle pixels that are close to, and have similar orientations to, Pixel X. With this modification and added dynamics that allow the network to escape local minima,[2] segmentations such as the middle decomposition occur. The segmentations now tend to break the object into whole position occur. The segmentations now tend to break the object into whole line segments that are connected to each other. These two mechanisms are too simplistic to do justice to human perceptual biases on segmentation, but even in this simplistic form, they lead to segmentations that tend to obey the Gestalt laws of good continuation and closure. The segmentation network is a process-model alternative to Palmer's model of segmentation and produces roughly comparable results.

However, the segmentation network has a notable advantage over Palmer's model in explaining Pevtzow and Goldstone's results. The seg-

[2]Local minima were avoided by adding noise to input-to-detector weights and basing the magnitude of this noise on the strength of the input-to-detector weight.

mentation network, because it shares the same input-to-detector weights that were used for the categorization network, can be influenced by previous category learning. Detectors that were diagnostic for categorization are more likely used to segment a pattern because they are already primed. Thus, if a particular shape is diagnostic, the network segments the whole into this shape most of the time, as shown by the bottom decomposition in Fig 8.14. In short, category learning can alter the perceived organization of an object. By establishing multisegment features along a bank of detectors, the segmentation network is biased to parse objects in terms of these features. This application shows that two separate cognitive tasks can be viewed as mutually constraining self-organization processes. Categorization can be understood in terms of the specialization of perceptual detectors for particular input patterns, where specialization is influenced by categorization diagnosticity. Object segmentation can be viewed as specialization of detectors for particular parts in a single input pattern. Object segmentation can isolate an input pattern's single parts that are potentially useful for categorization, and categorization can suggest possible ways of parsing an object that would not otherwise have been considered.

THE UNITIZATION OF COMPONENTS FOR CATEGORIZATION

Thus far, we have described the influence of category learning on the sensitization of pre-existing and novel dimensions and the organization of objects into dimensions and parts. One final mechanism of perceptual learning is unitization. In unitization, a single functional unit is created for a complex pattern, and this functional unit can be identified without an analytic process of breaking it down into components and identifying the components. The letter A may originally be perceived by assembling evidence from independent feature detectors for oriented lines, but with prolonged practice, a single unitized chunk for the entire A image seems to emerge (LaBerge, 1973). Czerwinski, Lightfoot, and Shiffrin (1992; Shiffrin & Lightfoot, 1997) obtained evidence for such a unitization process, by finding large improvements in the speed and efficiency of detecting conjunctively defined targets in a feature search task. The current experiments, reported by Goldstone (in press), similarly explore unitization, but from a complementary perspective. First, our experiments are primarily concerned with the influence of category learning on unitization, under the hypothesis that a unit tends to be created if the parts that make up the unit frequently co-occur and if the unit is useful for determining a

categorization. Second, we use a new method for analyzing response-time distributions to assess the presence of unitization.

Whenever the claim for the construction of new units is made, two objections must be addressed. First, perhaps the unit existed in people's vocabulary before categorization training. Our stimuli are designed to make this explanation unlikely. Each unit to be sensitized is constructed by connecting 5 randomly chosen curves. With 10 curves that can be sampled, there are 5^{10} possible different units. If it can be shown that any randomly selected unit can be sensitized, then an implausibly large number of vocabulary items are required under the constraint that all vocabulary items are fixed and *a priori*. The second objection is that no units need be formed; instead, people analytically integrate evidence from the five separate curves to make their categorizations. However, this objection is untenable if participants, at the end of extended training, are faster at categorizing the units than expected by the analytic approach. Quantifying what "faster than expected" means is the main task at hand.

The categorization task was designed so that evidence for five components must be received before certain categorization responses are made. As such, it was a conjunctive categorization task. The stimuli and their category memberships are shown in Fig. 8.15. Each of the letters refers to a particular segment of a "doodle." Each doodle was composed of five segments, with a semicircle below the segments added to create a closed figure. To correctly place the doodle labeled ABCDE into Category 1, all five components, A, B, C, D, and E, must be processed. For example, if the rightmost component was not attended, then ABCDE could not be distinguished from ABCDZ, which belongs in Category 2. Not only does no single component suffice for accurate categorization of ABCDE, but two-way, three-way, and four-way conjunctions of components also do

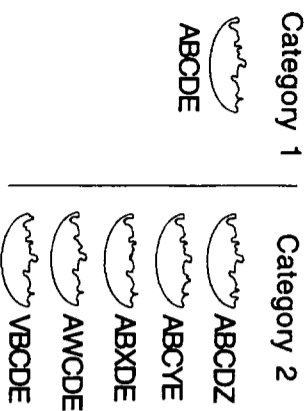| Category 1	Category 2
ABCDE	ABCDZ
	ABCYE
	ABXDE
	AWCDE
	VBCDE

FIG. 8.15. Stimuli used by Goldstone (in press). Each letter represents a particular stimulus segment, and each stimulus is composed of five segments. To categorize the item represented by ABCDE as belonging to Category 1, it is necessary to process information associated with each of the segments.

not suffice. For example, the three-way conjunction C and D and E is possessed by the stimulus ABCDE, but this conjunction does not discriminate ABCDE from AWCDE or VBCDE. Only the complete five-way conjunction suffices to accurately categorize ABCDE.

If unitization occurs during categorization, then the stimulus ABCDE may become treated functionally like a single component with training. If this occurs, then participants should be able to quickly respond that this stimulus belongs to Category 1. A pronounced decrease in the time required to categorize the conjunctively defined stimulus ABCDE is taken as initial evidence of unitization.

For improvement in the conjunctive task to be taken as evidence for unitization, two important control conditions are necessary. First, it is important to show that tasks that do not require unitization do not show comparable speedups. To this end, a control task was included that allows participants to categorize the item ABCDE by attending to only a single component rather than a five-way conjunction. This was done by having Category 2 contain only one of the five Category 2 doodles shown in Fig. 8.15, randomly selected for each participant. This "One" (component) condition should not result in the same speedup as the "All" (components) condition where five components must be attended. If it does, then the speedup can be attributed to a simple practice effect rather than to uniti-zation. Second, it is important to show that stimuli that cannot be unitized also do not show comparable speedups. For this control condition, it was necessary to attend to a five-way conjunction of components, but the ordering of the components in the stimulus was randomized. That is, ABCDE and CEBDA were treated as equivalent. In this "random" condition, a single template cannot serve to categorize the ABCDE stimulus, and unitization should therefore not be possible.

The results of the experiment were suggestive of unitization. The results in Fig. 8.16 reflect only the correct responses to the Category 1 doodle ABCDE because this is the only stimulus that requires the full five-component conjunction to be identified. The horizontal axis shows the amount of practice over a 2-hour experiment. The condition where all components were necessary for categorization, and where they were combined in a consistent manner to create a coherent image, showed far greater practice effects than did the others. This dramatic improvement suggests that the components are joined to create a single functional unit to serve categorization. Particularly impressive speedups were found when and only when unitization was possible and advantageous.

It is possible to get stronger evidence for unitization with this paradigm. The alternative to the unitization hypothesis is that responses in the All task are obtained by integrating evidence from five separate judgments of the type required in the One task. In arguing against this analytic

Response Time (msec)

Blocks of Training

□ All, Random
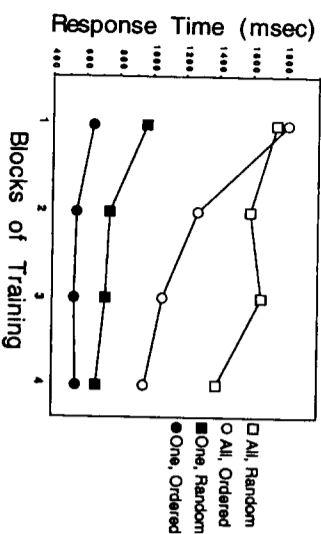○ All, Ordered
■ One, Random
● One, Ordered

FIG. 8.16. Results from Goldstone (in press). The most pronounced improvement was observed when all components were required for a categorization, and the components were always in the same positions.

account, a highly efficient version of the analytic account was devised to observe whether it still predicted response times that were too slow. The first advantage given to the analytic model was fully parallel processing; All responses were made by combining five One responses, but evidence for these five One responses was assumed to be obtained simultaneously. Second, the model was given unlimited capacity; identifying one component was not slowed by the need to identify another component. In obtaining predictions from this charitably interpreted analytic model, it is important to remember that the All task is a conjunctive task. To categorize ABCDE as a Category 1 item with the required 95% categorization accuracy, all five components must be identified. Second, there is intrinsic variability in response times, even in the simple task where only one component must be identified. An analytic model of response times can be developed that predicts what the All task response-time distribution should be, based on the One task distribution. After training, a distribution of response times in the One task can be determined. To derive the analytic model's predictions, we can randomly sample five response times from this distribution. The *maximum* of these five times, rather than the average, is selected because no response can be made to the conjunction until all components have been recognized. We can repeat this selection process several times to create a distribution of the maximums, and this yields the predicted response-time distribution for the All task according to the analytic model. Fortunately, there is an easier way of obtaining the predicted distribution. The One task response-time distribution is converted to a cumulative response-time distribution, and each point on this distribution is raised to the fifth power. If the probability of one component's being recognized in less than 400 milliseconds

Cummulative Probability

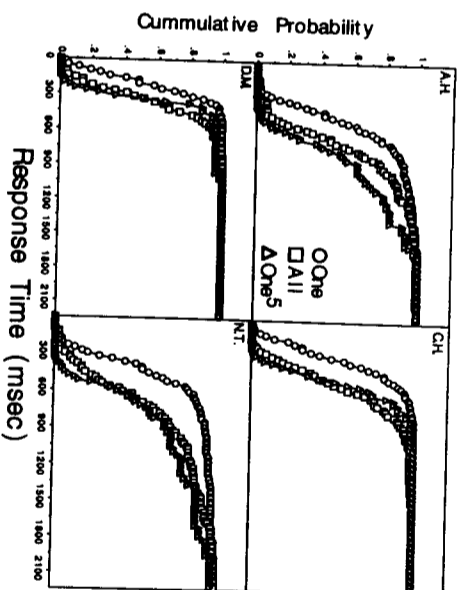○ One
□ All
△ One^5

Response Time (msec)

FIG. 8.17. The cumulative response time distributions for the four participants taken from the last session. The One and All distributions were empirically obtained. The One^5 distribution was obtained by raising each point along the One distribution to the fifth power and represents the analytic model's predicted cumulative distribution for the All task. Violations of this analytic model occur when the All task's distribution is shifted to the left of the analytic model's distribution. Such violations occur for the fastest half of response times for all four participants (significantly so for all participants except C. H.).

is .2, then the probability of all five components' being recognized in less than 400 milliseconds is .2 raised to the fifth power, assuming sampling independence.

A replication of the experiment shown in Fig. 8.16 was conducted that included the ordered All and One tasks. Only four research assistants participated as participants, but unlike the 2-hour experiment described previously, each participant was given eight 2-hour training sessions. The results, shown in Fig. 8.17, are only for Category 1 responses on the final day of the experiment. These results indicate violations of the analytic model. The cumulative response-time distributions show that the One task was naturally the fastest (most shifted to the left). The analytic model's predictions are shown by the curve labeled One^5, which is obtained simply by raising each point on the One curve to the fifth power. For two of the four participants, the fastest All distribution was faster than the analytic models' predictions for all regions of the distribution. For all four participants, the fastest 30% of response times for the All task was faster than predicted by the analytic model, even though all participants were achieving accuracies greater than 95%. Although the advantage of the All over the One^5 distributions may not look impressive, they

were significant by a Kolmogorov–Smirnoff test of distributions for all participants except C_1H_3[3] Thus, by the end of extended training, responses to the conjunctively defined ABCDE curve are faster than predicted by the analytic model, despite its charitable interpretation.

The conclusion we draw from these results is that category learning probably created new perceptual units. Large practice effects are found if and only if stimuli were unitizable (the first experiment), and responses after 14 hours of training were faster to conjunctively defined categories than predicted by a charitably interpreted analytic model. The results shown in Fig. 8.17 violate the analytic model only if negative dependencies or independence is assumed between the five sampled response times that make up one All judgment. Although it is beyond the scope of this chapter, we also have evidence for violations of the analytic model for classes of positive dependencies, using Fourier transformations to deconvolve shared input-output processes from the One task response-time distribution (Goldstone, in press; P. L. Smith, 1990).

One question still remains: Exactly how do people become so fast at categorizing ABCDE in the All task? Two qualitatively different mechanisms can account for the pronounced speedup of the conjunctive categorization: a genuinely holistic match process to a constructed unit, or an analytic model that incorporates interactive facilitation among the component detectors. In a holistic match process, a conjunctive categorization is made by comparing the image of the presented item with an image that has been stored over prolonged practice. The stored image may have parts, but these parts are either arbitrarily small or do not play a functional role in the recognition of the image. There is evidence supporting the gradual development of configural features. Neurophysiological findings suggest that some individual neurons represent familiar conjunctions of features (Perrett & Oram, 1993) and that prolonged training can produce neurons that respond to configural patterns (Logothetis et al., 1995). However, the results can also arise if detecting one component of ABCDE facilitates detection of other components. In either case, the process is appropriately labeled "unitization" in that the percepts associated with different components are closely coupled as a result of training. In fact, an interactive facilitation mechanism can be seen as the mechanism

---

[3]One may ask why the violations of the analytic model are restricted to, or at least maximized at, the fast response times. Most likely, a range of strategies was used for placing ABCDE into Category 1 in the All task. On some trials, an analytic strategy of combining evidence from separately detected components may have been used. On other trials, participants may have detected a single constructed unit. On average, the unit-based trials will be faster than the analytic trials. If the fast and slow response times tend to be based on single units and analytic integration, respectively, then we predict violations of the analytic model to be limited to, or more pronounced for, the fast response times.

that implements holistic unit detection at a higher functional level of description.

GENERAL REMARKS ON ADAPTIVE PERCEPTUAL VOCABULARIES

The general conclusions can be divided into empirical and theoretical ones. Empirically, evidence was found for several types of perceptual learning that accompanies concept learning. Concept learning can cause the perceptual sensitization of existing dimensions such as size and brightness and can cause sensitization of regions in a dimension, a finding suggesting that categorical perception can be due to learned categories. Not only can category learning lead to stretching and shrinking of existing dimensions, but it can also lead to selective sensitization along novel dimensions. These new dimensions may be created by breaking a fused dimension into subdimensions, as was the case with saturation and brightness, or by creating dimensions by morphing between arbitrary end points, as was the case with the bezier curves and bald heads. Finally, the object segmentation and unitization experiments indicate that concept learning can lead to the addition of new elements in a person's perceptual vocabulary. These new vocabulary elements change how objects are organized and can lead to responses that are more efficient than predicted by analytic models that do not develop new vocabularies.

The major theoretical contribution of the research has been to specify some possible ways in which perceptual and conceptual learning might interact. In both the neural networks described, feature detectors are developed that represent the network's set of acquired vocabulary elements. The networks begin with homogenous, undifferentiated detectors that become specialized for different inputs over time. Furthermore, both models have mechanisms by which detector-to-category associations modify the nature of the detectors. It is unnecessary to first develop detectors and then build associations between detectors and categories. These two types of learning can and should go on simultaneously.

Staking Out the Territory

It is worthwhile to step back and ask exactly what is entailed by the claim that perceptual vocabularies adapt to the demands of concept learning. What we mean by a perceptual vocabulary is the set of functional features used for describing objects. A functional feature, in turn, is defined as any object property that can be selectively attended to (for a similar claim, see Smith, Gasser, & Sandhofer, 1997). An organism shows evidence of using Feature X to describe an object if there is behavioral evidence that

X can be considered in isolation from other aspects of the object. Thus, a feature is a "chunk of object stuff" that has been individuated from the rest of the object. This definition explicitly denies that psychologically relevant features are objective properties of the external world. Even if a physicist can measure the illuminance of an object or a chemist can measure the tannin content of a Bordeaux wine, these stimulus properties are not psychological features unless the perceiving organism can isolate them as well. Tying featurehood to selective attention conforms to many empirical techniques for investigating features. Garner (1974) considered two features or dimensions to be separable if categorizations on the basis of one of the features are not slowed by irrelevant variation on the other. Treisman (e.g., Treisman & Gelade, 1980) argued that features are registered separately on different feature maps, which gives rise to efficient and parallel searches for individual features and the automatic splitting of different features that occupy the same object. Given this characterization of a feature, the claim for vocabulary creation seems less controversial than might be thought in view of the prevalence of fixed-vocabulary approaches. A substantial body of evidence from development (Smith, 1989) and expertise (Burns & Shepp, 1988) indicates that children and novices have a harder time selectively attending to stimulus aspects than do adults and experts. Claiming vocabulary creation does not necessitate that features are created *de novo* or that our perceptual system provides us with information that was not present *in any form* in the early stages of sensory processing. It is unclear whether these are even logically tenable positions. All that is required is that the organism shows behavioral evidence that stimulus elements come to be isolated with experience.

In some respects, our claim is similar to those made by theorists of dynamical systems in which an object is recognized if its processing follows the same trajectory as an object presented earlier, without requiring any decomposition of the objects into part representations (see, for example, Thelen & Smith, 1994). Both approaches stress the flexibility and plasticity of perceptual processing, and argue for powerful top-down and contextual influences on perception. However, a fundamental difference between the approaches is that we do posit a set of features that are used for describing objects. Radical versions of the dynamical systems approach have argued that objects are not represented by a set of elementary features at all. Thus, our approach is more closely tied to the traditional "fixed feature set" approach to cognition than it may initially appear. Our approach and the fixed feature set approach both assume that objects are represented in terms of a set of building blocks; the theories simply differ on whether this set is expandable.

In advocating building block representations over complete fluidity, we may be criticized on the same grounds of inflexibility that we used

to criticize fixed feature set theories. In our approach, objects must be described in terms of a finite vocabulary of features that have been previously acquired. Still, we believe that the traditional advantages of building block theories compensate for this inflexibility. By building object descriptions from elements, we have a generative method for creating and accommodating novel objects. The advantages of propositional representations are accrued by establishing explicit relations between represented elements. Systematicities in the appearance, function, and influence of objects can be accounted for in terms of systematicities between their featural representations. Perhaps most important, decomposing objects into features provides an efficient and compressed representational code. Instead of coding objects in a raw, uncompressed manner, short codes can be used to token discrete features that can be associated with complex configurations. If a feature that represents the doodle ABCDE is built, then the large amount of information present in the rich doodle is compressed into a single feature. In the same way that a large data set can be reduced to a few major components (in principal component analysis) or dimensions (in multidimensional scaling), a marked information reduction can be achieved by establishing features that underlie systematic variation among a set of objects. This reduction is accomplished by identifying stimulus aspects that are highly correlated, isolating (differentiating) them from other aspects, and grouping (unitizing) them.

The empirical evidence that people execute such a feature-extraction process comes from transfer experiments. Features that were useful for an earlier categorization are more likely to be applied to a later categorization (see Fig. 8.5), are sensitized for subsequent same/different judgments (Figs. 8.2 and 8.3), are used as features for decomposing subsequent whole objects (Fig. 8.12), and can be detected without analytically composing them from smaller components (Fig. 8.17). The sum of this evidence suggests a feature development process that has a lasting impact on perceptual processing.[4] In dynamic models that recognize objects by pulling the raw object description toward an attractor state caused by a previous episode, there is little reason to expect prominent transfer on the basis of component, rather than overall, similarities.

In light of our commitment to (adaptive) building block representations of objects, our approach is perfectly consistent with evidence that people use particular sets of primitive elements. The geons, textons, or conceptual primitives of fixed feature set theories may be the end product of a general

---

[4]The influence on perception is lasting in the sense that it persists from training to transfer. However, it remains to be seen how permanent these influences are. In most cases, the endurance of a perceptual change is probably positively related to the amount of training required to produce the change.

perceptual learning strategy. Recent research in computer science has shown that sets of primitives, including Gabor filters, and size detectors can be created by a system provided naturalistic scenes (e.g. Miikkulainen, Bednar, Choe, & Sirosh, 1997). The advantages of learning, rather than simply positing, elements are that mechanisms are in place for acquiring slightly different primitives if the environment is modified, and special-ized domains in the environment can have tailored sets of primitives designed for them (Edelman & Intrator, 1997).

Constraints on Computational Models of Perceptual and Conceptual Learning

We have been advocating systems that develop new perceptual vocabu-laries instead of relying on a fixed set of features. Some may object to this on grounds of computational plausibility. The argument is: With such a flexible system it takes too long to learn any category. Even with a fixed set of features, there may be a combinatorial explosion of complex rules involving these features, if we allow rules such as "white and (square or triangle)" (Bruner et al., 1956; Nosofsky, Palmeri, & McKinley, 1994). Moreover, the picture is much more grim if we allow the possibility of creating new features and entering these into Boolean expressions. In this case, there is a combinatorial explosion of potential features combined combinatorially with a combinatorial explosion of logical expressions. Our solution to the difficulties associated with finding good solutions in such an immense search space is to provide two sources of constraints. *Psychophysical constraints* enter in because not anything can be made into a feature. There is a heavy bias for features to be contiguous and coherent and to follow Gestalt laws of organization. In the segmentation network (see Fig. 8.14), these constraints were needed to create psychologically realistic segmentations. Mechanisms biased the acquired features to in-volve similarly oriented and positioned segments and served to constrain the number of features contemplated. Via *categorical constraints*, there is a bias to develop features that are diagnostic for relevant categories.

In the described networks, these two constraints act in parallel. There are problems with the flexibility and efficiency of either serial approach—starting with the set of candidate features admissible by psychophysical constraints such as topological coherence and then choosing the ones from this set that obey the categorical constraints, or vice versa (see also Wis-niewski & Medin, 1994). In a serial approach that uses psychophysical constraints as a first filter on the feature selection, features that should be created if useful for a categorization are excluded. For example, seg-ments that are separated by a pixel are probably eventually formed given enough training. These features can be accommodated by weakening the psychophysical constraint on connectedness, but only at the considerable

cost of failing to sufficiently limit the search space of features generally. Parallel constraints allow the individual constraints to both strongly limit the search space of features but also to be relaxed if required by other constraints simultaneously being satisfied.

Finally, one possibility is that people may actually be quite poor at combining separate and distinct features into logical expressions. Creating categories such as "Large and (square or triangle)" may be rather unnatu-ral, four decades of concept learning research not withstanding. At the same time, people seem to be adept at integrating components to create a single, coherent feature. Humans seem to be much more adept at creating coherent, useful features than they are at simultaneously attend-ing to several unrelated sources of information. By providing mechanisms for the development of novel features, much of the need for searching through the space of logical rules is removed. In many cases, a single feature suffices if it can integrate many stimulus aspects.

Building Perceptual Vocabularies: A Reprise

Cognitive science researchers who have proposed particular fixed sets of primitives have cleverly designed primitives that are genuinely useful for representing words, objects, and events. Our point is simply that ordinary people may be almost as clever as these researchers and may come up with their own sets of tailored elements (Schyns et al., 1998). The advan-tage, of course, is that the elements can be tuned to the particular cate-gories that are important. Fixed feature sets, no matter how cleverly constructed, cannot be perfectly tuned to the individual. Either the fixed feature sets have specific, special-purpose features, in which case the set is efficient at representing some things but incapable of representing other things, or the fixed feature set has a large number of general purpose, universal features, in which case it can represent everything, but not efficiently, taking advantage of the systematicities particular to a domain. Instead, if perceptual vocabularies are created, they are at least diagnostic for the category that they were created to accommodate. In sum, concepts certainly depend on perceptual encodings, but it is not viciously circular to claim that the perceptual encodings also depend on our concepts. In fact, our concepts seem to be able to "reach down" and influence the very features that compose the concepts.

ACKNOWLEDGMENTS

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.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115–147.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 640–645.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Burns, B., & Shepp, B. E. (1988). Dimensional interactions and the structure of psychological space: The representation of hue, saturation, and brightness. *Perception and Psychophysics, 43*, 494–507.

Czerwinski, M., Lightfoot, N., & Shiffrin, R. M. (1992). Automatization and training in visual search. *American Journal of Psychology, 105*, 271–315.

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

Edelman, S. & Intrator, N. (in press). Learning as extraction of low-dimensional representations. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.), *Psychology of learning and motivation* (Vol. 36, pp. 353–380). San Diego, CA: Academic Press.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303–306.

Fahle, M., & Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. *Current Biology, 6*, 292–297.

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

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, E. J., & Walk, R. D. (1956). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative and Physiological Psychology, 49*, 239–242.

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

Goldstone, R. L. (in press). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance.*

Goldstone, R. L., Schyns, P. G., & Medin, D. L. (1997). Learning to bridge between perception and cognition. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.), *Psychology of learning and motivation* (Vol. 36, pp. 1–17). San Diego, CA: Academic Press.

Goldstone, R. L., & Steyvers, M. (1999). *Attention to novel face dimensions*. Manuscript in preparation.

Goldstone, R. L., Steyvers, M., & Larimer, K. (1996). Categorical perception of novel dimensions. In G. W. Cottrell (Ed.), *Proceedings of the 18th annual conference of the Cognitive Science Society* (pp 243–248). Hillsdale, NJ: Lawrence Erlbaum Associates.

Grau, J. W., & Kemler-Nelson, D. G. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General, 117*, 347–370.

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

Harnad, S., Hanson, S. J., & Lubin, J. (1995). Learned categorical perception in neural nets: Implications for symbol grounding. In V. Honavar & L. Uhr (Eds.), *Symbolic processors and connectionist network models in artificial intelligence and cognitive modelling: Steps toward principled integration* (pp. 191–206). Boston: Academic Press.

Hillyard, H. C., & Kutas, M. (1983). Electrophysiology of cognitive processes. *Annual Review of Psychology, 34*, 33–61.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97*, 553–562.

Jakobson, R., Fant, G., & Halle, M. (1963). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America, 88*, 4966–4970.

Kohonen, T. (1995). *Self-organizing maps*. Berlin: Springer-Verlag.

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

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America, 63*, 905–917.

LaBerge, D. (1973). Attention and the measurement of perceptual learning. *Memory and Cognition, 1*, 268–276.

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.

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

McGraw, G., Rehling, J., & Goldstone, R. L. (1994). Letter perception: Toward a conceptual approach. In A. Ram & K. Eiselt (Eds.), *Proceedings of the 16th annual conference of the Cognitive Science Society* (pp. 613–618). Hillsdale, NJ: Lawrence Erlbaum Associates.

Miikkulainen, R., Bednar, J. A., Choe, Y., & Sirosh, J. (1997). Self-organization, plasticity, and low-level visual phenomena in a laterally connected map model of primary visual cortex. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.), *Psychology of learning and motivation* (Vol. 36, pp. 257–308). San Diego, CA: Academic Press.

Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on x-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 553–557.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*, 53–79.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9*, 441–474.

Palmer, S. E. (1978). Structural aspects of visual similarity. *Memory & Cognition, 6*, 91–97.

Peron, R. M., & Allen, G. L. (1988). Attempts to train novices for beer flavor discrimination: A matter of taste. *Journal of General Psychology, 115*, 403–418.

Perrett, D. I., & Oram, M. W. (1993). Neurophysiology of shape processing. *Image and Vision Computing, 11*, 317–333.

Perrett, D. L., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. D., & Jeeves, M. A. (1984). Neurones responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception. *Human Neurobiology, 3,* 197–208.

Pevtzow, R., & Goldstone, R. L. (1994). Categorization and the parsing of objects. In A. Ram & K. Eiselt (Eds.), *Proceedings of the 16th annual conference of the Cognitive Science Society* (pp. 717–722). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 297–314.

Repp, B. H., & Liberman, A. M. (1987). Phonetic category boundaries are flexible. In S. Harnad (Ed.), *Categorical perception* (pp. 89–112). Cambridge: Cambridge University Press.

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

Schank, R. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology, 3,* 552–631.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. (1998). Development of features in object concepts. *Behavioral and Brain Sciences, 21,* 1–54.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 31, pp. 305–354). San Diego, CA: Academic Press.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 681–696.

Shiffrin, R. M., & Lightfoot, N. (in press). Perceptual learning of alphanumeric-like characters. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.), *Psychology of learning and motivation* (Vol. 36, pp. 45–82). San Diego, CA: Academic Press.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing, II: Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190.

Smith, L. B. (1989). From global similarity to kinds of similarity: The construction of dimensions in development. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 146–178). Cambridge: Cambridge University Press.

Smith, L. B., Gasser, M., & Sandhofer, C. (in press). Learning to talk about the properties of objects: A network model of the development of dimensions. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.), *Psychology of learning and motivation* (Vol. 36, pp. 219–256). San Diego, CA: Academic Press.

Smith, P. L. (1990). Obtaining meaningful results from Fourier deconvolution of reaction time data. *Psychological Bulletin, 108,* 533–550.

Steyvers, M. (1999). Morphing techniques for generating and manipulating face images. *Behavioral Research Methods, Instrumentation, and Computers, 31,* 359–369.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Tipper, S. P. (1992). Selection for action: The role of inhibitory mechanisms. *Current Directions in Psychological Science, 1,* 105–109.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Weinberger, N. M. (1993). Learning-induced changes of auditory receptive fields. *Current Opinion in Neurobiology, 3,* 570–577.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49–63.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18,* 221–281.

9

The Proper Treatment of Symbols in a Connectionist Architecture

Keith J. Holyoak
John E. Hummel
University of California, Los Angeles

PHYSICAL SYMBOL SYSTEMS

A foundational principle of modern cognitive science is the physical symbol system hypothesis, which states simply that human cognition is the product of a physical symbol system (PSS). A symbol is a pattern that denotes something else; a symbol system is a set of symbols that can be composed into more complex structures by a set of relations. The term *physical* conveys that a symbol system can and must be realized in some physical way to create intelligence. The physical basis may be the circuits of an electronic computer, the neural substrate of a thinking biological organism, or in principle anything else that can implement a Turing machine-like computing device. Classical presentations of the PSS hypothesis include Newell and Simon (1976) and Newell (1980); more recent discussions include Newell (1990) and Vera and Simon (1993, 1994).

The PSS hypothesis, which implies that structured mental representations are central to human intelligence, was for some time uncontroversial, accepted by most cognitive scientists as an axiom of the field scarcely in need of either theoretical analysis or direct empirical support. In the mid-1980s, however, the hypothesis came under sharp attack from some proponents of connectionist models of cognition, particularly the advocates of models in the style of "parallel distributed processing," or PDP (Rumelhart, McClelland, & the PDP Research Group, 1986; more