Journal of Experimental Psychology:
Human Perception and Performance
1998, Vol. 24, No. 2, 441–462

0096-1523/98/$3.00

Figure–Ground Organization and Object Recognition Processes: An Interactive Account

Shaun P. Vecera
University of Utah

Randall C. O'Reilly
Carnegie Mellon University

Traditional bottom-up models of visual processing assume that figure–ground organization precedes object recognition. This assumption seems logically necessary: How can object recognition occur before a region is labeled as figure? However, some behavioral studies find that familiar regions are more likely to be labeled figure than less familiar regions, a problematic finding for bottom-up models. An interactive account is proposed in which figure–ground processes receive top-down input from object representations in a hierarchical system. A graded, interactive computational model is presented that accounts for behavioral results in which familiarity effects are found. The interactive model offers an alternative conception of visual processing to bottom-up models.

In a typical visual scene multiple objects partially occlude one another, which makes object recognition a computationally complex task. Traditional information-processing theories of visual perception have suggested that prior to object representation and recognition, an earlier stage of perceptual organization occurs to determine which features, locations, or surfaces most likely belong together (for examples, see Biederman, 1987; Kosslyn, 1987; Marr, 1982; Neisser, 1967; Palmer & Rock, 1994a). Such a hierarchical organization of processing seems to be logically required (Palmer & Rock, 1994b): For object recognition to occur, an object representation must receive inputs from features or regions of the visual field that correspond to the object that is to be recognized. Inputs from any other feature or region of the visual field are spurious and presumably make object recognition more difficult.

One particular aspect of perceptual organization that the visual system needs to determine is which regions in the visual field are figures and which regions are backgrounds. Only figural regions should be given as input to object representations. The study of figure–ground organization began slightly after the beginning of the century by the Gestalt psychologists, most notably Rubin. The work of these psychologists, as well as of more modern theorists, suggests that there are certain "rules" that the visual system uses to determine which regions are figure. For example, Rubin (1915/1958) reported that smaller regions tend to be perceived as figure. Also, studies of figural perception motivated by information-processing theory demonstrated that figural perception was influenced by factors such as symmetry and convexity (see Pomerantz & Kubovy, 1986, for a review). Such findings are consistent with the traditional theories of perception discussed above in which figure–ground organization is computed by using stimulus variables such as symmetry, area, and convexity. Following this figure–ground computation, the regions labeled as *figure* are then matched against object representations.

Although the hierarchical processing scheme is commonly advocated by most theories of visual processing, recent studies of figure–ground organization have challenged the traditional theory of visual processing. Peterson and her colleagues (Peterson, 1994; Peterson & Gibson, 1991, 1993, 1994a, 1994b; Peterson, Harvey, & Weidenbacher, 1991) have demonstrated that meaningful (or denotative) regions are more likely to be perceived as figure relative to less meaningful (or less denotative) regions. Similar findings were reported by Rubin (1915/1958), and Rock (1975) also briefly discussed this effect. For example, when viewing figure–ground stimuli that contain a more denotative, or familiar, region, research participants tend to report that the highly denotative (or meaningful) region is the figure. However, if the same display is rotated 180°, then the choice of figure is made primarily on the basis of stimulus factors such as symmetry or area (Peterson & Gibson, 1991). Similar results are found in experiments in which participants viewed figure–ground displays for a longer period of time (30 s) and reported reversals of figure and ground. Participants tend to report the denotative region as figure longer in upright displays, but when the displays are rotated 180°, the region that is favored by stimulus

Shaun P. Vecera, Department of Psychology, University of Utah; Randall C. O'Reilly, Department of Psychology, Carnegie Mellon University.

Randall C. O'Reilly is now at the Department of Psychology, University of Colorado at Boulder.

This research was supported in part by the Neural Processes in Cognition Program (National Science Foundation Award BIR 9014347) and by National Institute of Mental Health (NIMH) Training Grant 2T32 MH19102-06. This work was also supported by NIMH Grant MH47566.

We wish to thank Martha Farah, Ken Kotovsky, Jay McClelland, Mary Peterson, Dennis Proffitt, and Rich Zemel for discussion and comments on this work.

Correspondence concerning this article should be addressed to Shaun P. Vecera, Department of Psychology, 390 S. 1530 E. Room 502, University of Utah, Salt Lake City, Utah 84112. Electronic mail may be sent to Shaun.Vecera@m.cc.utah.edu.

factors is held as figure for longer (Peterson et al., 1991). This latter finding argues against an alternative interpretation that attributes familiarity effects to object or shape recognition rather than to an influence of familiarity on figure–ground organization itself. This is because the participants in this experimental paradigm are basing their behavior directly on their perceptions of figure–ground relations.

If one assumes that figure–ground organization must precede the activation of an object representation, then these results appear paradoxical. How can object representations influence figure–ground organization if the very goal of figure–ground organization is to provide input to object representations? Previous attempts to explain how past experience can influence perceptual organization focused on the notion of "preconscious oscillations" (Rock, 1975). This theory suggests that when someone views a figure–ground stimulus, one region is held as figure preconsciously. If that region (preconsciously) matches an object representation, then it is held as figure longer because of top-down input from the object representation. If the region does not match an object representation, then the assignment of figure is switched to the other region so that this second region is the preconscious figure. This account explains these experience effects within the framework of traditional theories of visual processing. Figure–ground organization still precedes object representation, and the processing between these two hierarchical stages is serial (i.e., a figure–ground organization is determined, and following this organization, the result is matched against object representations).

In contrast with the preconscious oscillation theory, Peterson and her colleagues (Peterson & Gibson, 1991, 1993, 1994a, 1994b; Peterson et al., 1991) have proposed an alternative account of these potentially paradoxical results: The observed orientation-dependent changes in figure–ground organization are due to object recognition processes that operate before figure–ground organization. That is, there are two types of cues that can determine figure–ground organization. The first type of cue is the gestalt, or stimulus cue, that emphasizes more general stimulus properties (e.g., symmetry, convexity, area). The second type of cue is an early orientation-dependent object representation process that is at work before figure–ground processes—a "prefigural" object representation that exists before figure–ground organization has been completed. In Peterson's (1994) words, "The second type of cue includes the outputs from an early object recognition process that operates before figure–ground relationships have been determined completely or even provisionally" (p. 110). This second process is thought to operate along both sides of the luminance contour in the figure–ground display. The result is that two sets of parts are determined, and these two sets of parts are matched against object representations that are stored in visual memory. If one of these sets matches an object representation, then that object representation provides input to the figure–ground processing stage, thus allowing object representations to influence figure–ground organization (see Peterson, 1994, for a review). This matching process occurs prior to any figural organization. Some computer vision theorists (e.g., Lowe, 1985) have also assumed that preliminary object identification can precede organization of the lower levels of representation.

Whereas Peterson's (Peterson & Gibson, 1991, 1993, 1994a, 1994b; Peterson et al., 1991) results clearly challenge the traditional models of visual perception, there is an alternative to the traditional theories of perception that predicts exactly these types of results. This alternative model is based on the principles of parallel distributed processing (PDP) models of information processing, particularly the ability of PDP models to exhibit interactive behavior (for reviews of PDP models see McClelland, 1993; Rumelhart, 1989; Rumelhart & McClelland, 1986). In this article we have presented a relatively simple PDP model of figure–ground perception that explains familiarity effects in figure–ground organization in terms of a hierarchically structured model in which the levels of processing *interact* with one another; no prefigural object processing is assumed. We propose that partial results from figure–ground processing can be sent to subsequent object representations. The object representations, in turn, can send activation back to the figure–ground units, providing top-down input before a stable figure–ground percept has been established. By this account, object representations can be viewed as another type of constraint on figure–ground organization, much as area, symmetry, and convexity constrain which region is perceived as figure. Unlike in Peterson's (1994) account, in our account it is not necessary for these constraints to be computed prior to figure–ground organization because processing is thought to be fully interactive.

Although Peterson and colleagues (Peterson, 1994; Peterson & Gibson, 1991, 1993, 1994a, 1994b; Peterson et al., 1991) considered such an interactive approach, they rejected it because they believe that interactive models can only facilitate lower level processing based on top-down inputs but not alter its outcome. This intuition is based in part on previous interactive models such as McClelland and Rumelhart's (1981) word superiority model; in this model, top-down support from word units to letter units allows the network to make faster discriminations between letters within words than between letters within nonwords. A stronger form of top-down influence is necessary to account for Peterson's results with an interactive model because the actual outcome of figure–ground organization is determined in part by top-down familiarity or denotivity influences. However, when McClelland and Rumelhart's model is presented with partial stimuli, the top-down influences are strong enough to fill in missing letters in a way that is consistent with known words. Although this is an example of top-down processing actually changing the outcome of lower level processing, it could be argued that merely "filling in" missing information is not the same as influencing the entire course of processing at the lower level, which is presumably what is observed in Peterson's experiments. Further, Peterson and Gibson (1993) showed that denotivity can actually compete with other bottom-up inputs instead of

merely resolving bottom-up ambiguity, which they argued further challenges an interactive account.

Although Peterson and colleagues have clearly presented an alternative to interactive models of visual perception (Peterson & Gibson, 1994a), it is not a foregone conclusion that the interactive-processing approach cannot sufficiently explain these results. In what follows, we have proposed an interactive model of figure–ground organization in an attempt to reconcile the logical requirements of hierarchical processing accepted by most theories of visual perception and Peterson's behavioral results. The purpose of our model is to simply demonstrate that an interactive model that incorporates hierarchical visual processing can account for familiarity (or denotivity) effects in figure–ground perception, reconciling the more traditional accounts of vision with Peterson's behavioral results. However, it should be noted that our model does not claim to simulate all of the relevant operations in visual processing that bear on our account of figure–ground organization, as we have simplified the model to capture the essential mechanisms that underlie the interactive processing account.

A PDP Approach

The principles we use to understand figure–ground organization in a PDP network are derived from the graded, random, adaptive, interactive, and nonlinear networks (GRAIN; McClelland, 1993) model and involve the following principles: (a) A unit's activation is a graded, sigmoidal function of the summed input to the unit; (b) activation is transmitted gradually in time; (c) processing is interactive based on between-module connections that are excitatory; (d) processing is competitive based on within-module connections that are inhibitory; and (e) activation across units is intrinsically variable.

The above principles provide a mechanistic account of figure–ground organization. Perhaps the most important principle for our account is *interactive information processing* (Principle c), in which processing at lower levels influences processing at higher levels, and vice versa. In multilayer networks, these influences are not constrained to immediately adjacent processing layers—a change in processing at one layer can be observed when processing is altered at a more distant layer. Thus, our approach to figure–ground organization relies on top-down projections from object representations to figure–ground processes to show effects of stimulus familiarity on figural organization. The interactive-activation account of the word superiority effect (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982) is probably one of the best known examples of interactive processing in which higher level information (word level) can impact on lower level processes (letter perception).

Graded processing (Principle a) has two important effects for our model. The first effect is that processing is not strictly sequential because partial products are propagated, or cascaded, throughout the network (McClelland, 1979). The second effect is that a single variable may only have a partial influence on other units, which allows multiple cues or constraints to simultaneously influence processing in the network. In relation to our simulations, the implications of graded processing for figure–ground organization are that a region can be considered partially figure during intermediate stages of processing, before the network has settled on a coherent interpretation of the image. This may be at odds with intuitive notions of figure–ground processing, in which there always appears to be a discrete figural region. However, it is important to note that whereas the processing is gradual, the network does indeed show discrete phase transitions, which might map directly onto these intuitions. This issue is revisited in more detail in the General Discussion section.

The final principle that is important for our account is *multiple-constraint satisfaction,* which emerges from the combination of the properties of graded and interactive processing (Principles a and c). As partial (graded) products of processing in different layers interact, the various constraints built into the weights, and the external inputs, jointly influence the activation state that results over cycles of activation updating or "settling." In our approach to figure–ground organization, different types of inputs can constrain the possible figure–ground organization. These inputs might be lower level stimulus cues, such as area or convexity, or they might be higher level inputs coming from object representations stored within the network. Each of these sources can influence (i.e., constrain) figure–ground processing without the need to postulate a prefigural object recognition process.

Our account of figure–ground organization is consistent with a number of other models of visual processing. Examples of these models include Cohen, Dunbar, and McClelland's (1990) work, in which they showed that attentional selection in the Stroop task can be simulated as a balance of constraints provided by the stimulus presented and by the task demands imposed on the participant by the experimenter. Another neural network model that relies on constraint satisfaction is the selective attention model (SLAM) of Phaf, Van der Heijden, and Hudson (1990), which accounted for participants' performance in several selective attention tasks, such as attentional filtering and the Stroop task. This model's processing is guided by constraints imposed by the stimulus and by the attributes to be attended. Finally, constraint satisfaction can also be observed in Marr and Poggio's (1976) account of binocular disparity; in this account of stereopsis, corresponding points between the two retinas are matched on the basis of constraints, such as continuity of surfaces. For example, in Marr and Poggio's model, two neighboring units that represented a patch of surface at the same depth plane were connected by an excitatory connection. This excitatory connection implements the constraint that these two units be mutually active, thus ensuring that the network maintained the continuity of some particular surface.

In what follows, we have presented a computational model of figure–ground organization. We assumed that processing can be hierarchical, with object recognition processes logically following figure–ground processes, as assumed by most theories of visual perception. No prefigural

shape recognition process is assumed, as was postulated by Peterson (e.g., Peterson, 1994). By our account, familiarity (or denotivity) effects are due to interactive processing among units in the figure–ground layer and units in the object representation layer. With such an interactive approach, the finding that one stage of processing can influence another stage of processing does not necessarily mean that the first process is either before or in parallel with the process it influences. Instead, higher level processes can interact with lower level processes. We thus offer our model as an "existence proof" of the interactive account: To the extent that our interactive account simulates the behavioral data, it needs to be considered a viable model of figure–ground perception. Indeed, our results show that our network can explain orientation effects, exposure-duration effects, and the combination of multiple cues to figure–ground organization.

The Model

Our network was based on a model developed and investigated by Sejnowski and colleagues (see Kienker, Sejnowski, Hinton, & Schumacher, 1986; Sejnowski & Hinton, 1987). We adopted this model as a framework for our simulations because it is capable of exhibiting some basic figure–ground organization, not because we feel that it is a model that simulates all aspects of visual processing. In essence, this model allows us to implement the more general principles of graded, interactive processing, which we feel are essential for explaining the range of figure–ground phenomena discussed above. We believe that our account, based on the general principles of PDP models, will hold given different assumptions about the details of the underlying processing.

The network, shown in Figure 1, has three processing levels. The first level processes the boundaries contained in the image; this level corresponds to simple image features (edges). This level is followed by a layer of units that represent which surfaces are figure. Finally, there are object representations that code for familiar shapes. Thus, the figure–ground units receive two types of information: bottom-up information from the edges present in the visual field and top-down information from internally stored object representations. Note that whereas the network can be divided into clear processing layers, this hierarchical organization is blurred by the interactive processing; as soon as information begins to propagate within the network, the information flow is bidirectional.

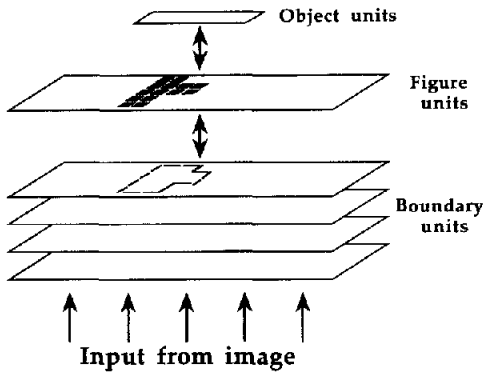

Figure 1. The architecture of the interactive model of figure–ground organization.

In the present model there were four types of edges, each corresponding to a simple visual contour and to the side of that contour on which the figural region was most likely to be located. For example, activation of a boundary-left unit signaled the presence of a vertical edge that had the figural region to its left. Similarly, activation of a boundary-right unit signaled the presence of a vertical edge that had the figural region to its right. Activation of either a boundary-above or a boundary-below unit signaled the presence of a horizontal edge that had the figural region above or below the edge, respectively. These four types of units are collectively referred to as the *boundary layer.* One could argue that such boundary units are psychologically implausible because they contain cues to figure–ground relations, and human observers can clearly perceive edges or boundaries without assigning figure–ground relations to either side of this boundary. However, the same is true of the boundary units in our model: The boundary units can represent simple edges without assigning figure–ground relations. This is because a single edge would activate boundary units corresponding to both figure–ground interpretations, which would result in *neither* side of the edge being labeled figure.

The figure units, when active, signaled the presence of the figural region. Thus, the network was presented with edge information, and from this information the network had to fill in the region that was most likely the figure by activating some set of figure units. The network solved this task by using both lower level constraints, provided by the edge units, and higher level constraints, provided by the object representations. Finally, it is important to note that the boundary information provided to the network was completely ambiguous from the view of the figure units; at any location the network did not know whether that region should be labeled figure or ground. Only the top-down information could bias the network to choose one particular figure–ground organization. Given this brief overview of the network, we now turn to the specific implementational details.

Connectivity

All of the connections described were symmetric. That is, for each connection described, there was also a reciprocal connection with the same weight value. This implementation is denoted $w_{ij} = w_{ji}$, which states that the weight from unit i to unit j equals the weight from unit j to unit i. Symmetric connections are significant because of Hopfield's (1982) convergence theorem, which states that a symmetric network settles into a stable pattern of activation (an energy minimum or a goodness maximum) given sufficient settling time.

The lowest levels of processing, the boundary units, consisted of an 11 × 16 array of units for the boundary-left

and the boundary-right layers and a 12 × 15 array of units for the boundary-above and the boundary-below layers. The next layer, the figure layer, consisted of a 12 × 16 array of units. Finally, the object representation level contained only two units, each corresponding to a different shape.

At each location in the visual field there were four different types of edge units, as described above, for an overall total of 712 edge units. In accordance with Kienker et al. (1986), opposite pairs of edge units were mutually inhibitory, as illustrated in Figure 2a. For example, a boundary-left unit at one location in the visual field inhibited the boundary-right unit at the same location because the figure cannot lie both to the left and the right of the edge; the figure must lie either to the left or the right. The boundary-above and boundary-below units were also mutually inhibitory for the same reason. (Exact values of all parameters, including the strengths of the connections, appear in Appendix A.)

Each of the figure units was connected by means of excitatory weights to its eight nearest neighbors (Figure 2b) to implement the constraint that the figural region tends to be connected and continuous. Between each pair of figure units

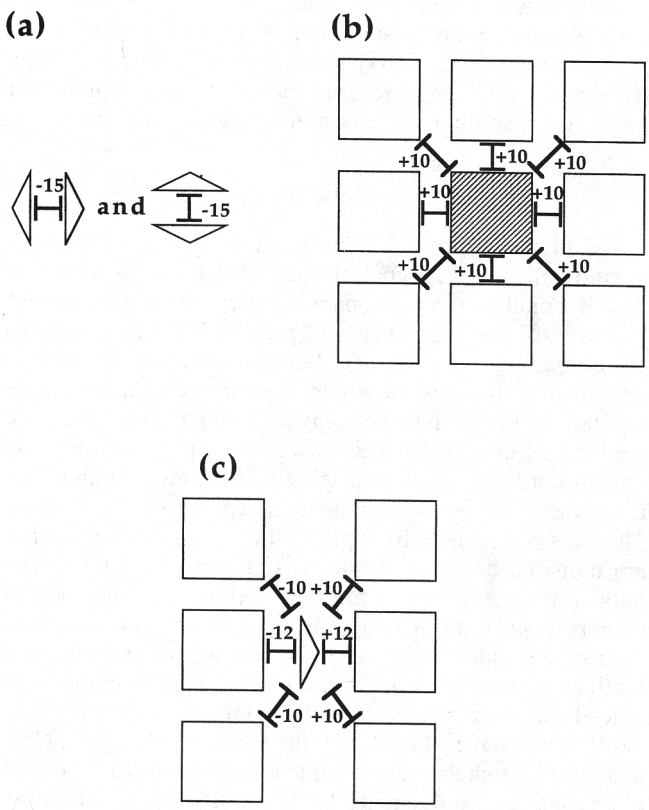

Figure 2. Patterns of connectivity in the network. (a) Opposing boundary units (i.e., left vs. right and above vs. below) inhibit one another. (b) A given figure unit has positive projections to and from its eight nearest neighbors. (c) Figure units (squares) receive both excitatory and inhibitory projections from boundary units (arrowheads). For example, the boundary-right unit shown will excite figure units to its right but inhibit figure units to its left. All connections are bidirectional.

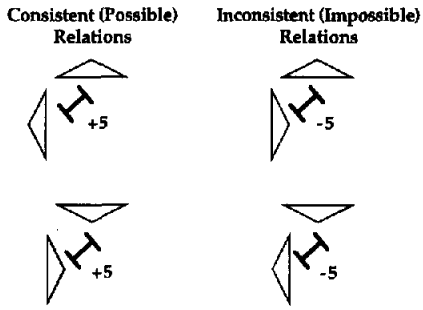

Figure 3. The patterns of connectivity among the boundary units. Some combinations of boundary units formed legal corners (consistent), whereas other combinations formed illegal corners (inconsistent). Units with a relationship consistent with a corner were joined by positive weights; units with a relationship inconsistent with a corner were joined by negative weights.

was a pair of edge units, oriented such that the boundary unit activated the figure units to one side and inhibited the figure units to the other side, as shown in Figure 2c. This pattern of connectivity between the boundary units and the figure units allowed the network to fill in the figural region to one side of the edge that was present in the visual field.

Because of the presence of corners in visual images, there was another pattern of connectivity among the boundary units, as shown in Figure 3. At a given location in the visual field certain combinations of vertical and horizontal boundary units are possible and are consistent with a corner, whereas other combinations of boundary units are impossible and do not correspond to any coherent visual input.[1] Possible combinations of vertical and horizontal boundary units (i.e., boundary units that corresponded to corners) were mutually excitatory, whereas impossible combinations of boundary units were mutually inhibitory. The bottom-up processing and figure–ground processing involve multiple-constraint satisfaction as the network attempts to determine which of the figure units are consistent with the boundary units and vice versa. This was the primary focus of the original Kienker et al. (1986) model. Also, although the pattern of connectivity among the boundary units and figure units was hand wired, it is possible for a network to develop similar patterns of connectivity through experience-based learning (Mozer, Zemel, Behrmann, & Williams, 1992).

The object units received input from the figure units. Each of the object units received excitatory weights from those figure units that corresponded to a particular shape; an example is given in Figure 4, which shows the two shapes known to the network used in Simulations 1 and 2. Thus, each object unit can be thought of as coding for a group of figure units that are simultaneously active. The object units only coded for a shape in a given orientation (i.e., they were

[1] This assumes a relatively clean image that does not contain multiple, overlapping objects. Processing such a complex scene would be difficult for the present network, and corners that are "inconsistent" for the present network might exist in a more complex scene.

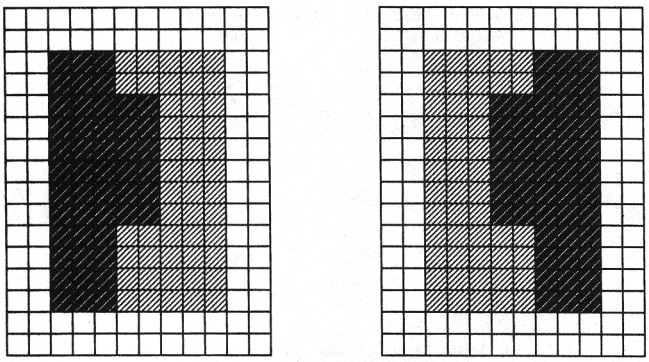

Figure 4. The object representations used in the model. The darker regions of each panel correspond to those figure units that project to, and receive projections from, the object units. One object unit represents a single object.

orientation dependent), which is consistent with findings from some behavioral studies (see Jolicoeur, 1985; Tarr & Pinker, 1989). However, we should note that the object representations were extremely simplified, as they do not incorporate the important properties of position, size, and other invariances known to be characteristic of human object recognition (Biederman & Cooper, 1991). This simplification is useful for understanding the basic behavior of our model. We revisit the issue of more realistic object representations in Simulation 3 and in the General Discussion. As noted earlier, our claims with this model are modest: We want only to demonstrate that an interactive model that incorporates a traditional hierarchical model of visual processing can simulate the effects of familiarity on figure–ground perception. As stated above, we do not claim that all aspects of our model are supported by the psychological and neurobiological data, nor does this have to be the case to demonstrate our point. The basic phenomena we observe in our model are present in systems that have different representations at several different levels of the system.

Updating Algorithm

Following standard algorithms (Hopfield, 1984; McClelland, 1993), the units in our model computed their net input from other units, and their activation was a sigmoidal function of the net input. All updating equations are given in Appendix A. The processing dynamics of our GRAIN model are briefly described here.

Using a network consisting of binary units (0, 1 outputs only) and symmetric connections (i.e., $w_{ij} = w_{ji}$), Hopfield (1982) used properties from statistical physics to mathematically prove that such a network converges to a fixed state. Hopfield (1984) later showed that this convergence behavior also holds for networks with continuous units and symmetric weights, as in the present simulations. In particular, for a continuous network there exists an "energy" function (also called a *Liapunov function*) that decreases monotonically as processing proceeds through time, thereby settling into a state of low energy. (The derivation of the energy function for the continuous model can be found in Hopfield, 1984, and is not presented here. See also Hertz, Krogh, & Palmer, 1991, and Movellan & McClelland, 1993.)

The global energy of the system can be thought of as the degree to which the current activations in the network satisfy the constraints given by both the input to the network and the patterns of connectivity among the units. Alternatively, the sign of the energy function can be reversed to define the "goodness" function (McClelland, 1993; Movellan & McClelland, 1993). A higher value of goodness suggests that the current pattern of activation is more consistent with, or fits well with, the constraints imposed by the input and the patterns of connectivity, whereas a lower goodness value suggests that the current pattern of activation is less consistent with, or fits poorly with, these constraints.

The use of noise in the activations is important because the network can reach a local maximum of the goodness function, which represents an incorrect or partial interpretation of the visual image presented to the network. Noise essentially allows the network to "jump" out of these local maxima, increasing the likelihood that the network will find the global maximum. As a result, the units in our network have random, Gaussian noise added to them, and we use a noise schedule in which the added noise is larger during the early stages of processing and is gradually reduced as processing proceeds (Kirkpatrick, Gelatt, & Vecchi, 1983). Kienker et al. (1986) discussed the importance of noise for multiple-constraint satisfaction in a network similar to ours.

Simulation 1: Network Performance

We first conducted a simulation to test the network's overall behavior. The goal of this simulation was to ensure that it could correctly separate figure from ground and "recognize" the shape that was presented (i.e., activate the corresponding object unit). Furthermore, we wanted to determine if the network would show the orientation effects reported by Peterson and colleagues (see Peterson, 1994, for a review of this work). The network was presented with an image of a figure–ground display with a single, ambiguous luminance contour separating two regions of the display. This was represented by input to the corresponding boundary units, an example of which is shown in Figure 5. The only difference between the two regions was that one of them received top-down input from an object representation. That is, one side of this otherwise ambiguous stimulus was familiar to the network. If the object representations can indeed influence figure–ground organization, then the network should have a bias to call the familiar side figure. Note that any bias that emerges would be entirely due to the object representations' influencing figure–ground organization *before the completion* of this processing, because either side of the ambiguous shape is equally likely to be called figure from the point of view of the lower level constraints on the network (i.e., the size or area constraints).

As a control condition, we rotated the shapes 180°. Because the object representations in the model are orientation dependent, rotating the stimulus would lessen the top-down bias for the shape because the familiar region in a

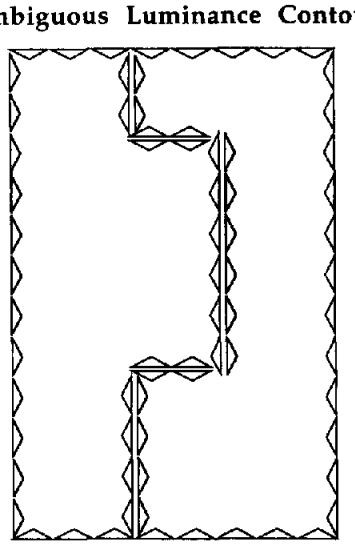

Figure 5. An example of the ambiguous input presented across the boundary units to simulate presenting the network with a figure–ground display. Note that the central contour is ambiguous—the network does not know which side of this contour should be called figure.

rotated display has a poorer fit to the object representations relative to the familiar region in an upright display. Thus, with rotated shapes there are few cues to determine which region should be figure; the network should, therefore, choose each region (i.e., the familiar region and the less familiar region) with approximately the same probability. A reliable difference between the upright and rotated conditions would replicate Peterson's (1994) behavioral results.

Method

The network was presented with one of four shapes: The familiar region could appear either to the left or right of the ambiguous central contour, and it could appear upright or rotated 180° from upright. The network cycled 400 times in which the activations of each unit were updated in accordance with Equation 3 in Appendix A. Each stimulus was presented by giving the appropriate boundary units additional net input of 60; this external input corresponds to the I_j term in Equation 1 of Appendix A.

Following the 400 time steps, the pattern of activation across the figure units was compared with all of the correct figure–ground solutions that the network could have settled on (i.e., those solutions that corresponded to a reasonable interpretation of the boundaries presented as input). To determine if the network settled into any local maxima, any discrepancies between the present solution and the correct figure were calculated. If there was any discrepancy between the number of units in the actual solution and in the correct solution, then the trial was excluded from the analyses reported below. This is a forced-choice procedure, but the network must choose exactly either the familiar region or the less familiar region. Note that this is a very conservative criterion, and as such it may tend to inflate the number of incorrect solutions.

Results

The network was tested with upright and rotated versions of the two familiar shapes. The four possible stimuli (upright, familiar region on left; upright, familiar region on right; rotated, familiar region on left; rotated, familiar region on right) were each presented on 100 different trials; however, the results for the upright and rotated shapes were collapsed across whether the meaningful region appeared on the left or on the right because there were no differences between the left and right positions. The resulting figure–ground solution was compared with the two possible solutions for a given stimulus. Again, if the solution did not match one of the two possible solutions perfectly, then the trial was excluded from the results reported. These imperfect solutions were removed prior to any further data analysis; thus, all of the results reported do not include all of the trials on which the network was tested.

For those trials in which a perfect solution was obtained, the probability that the network called the meaningful (or denotative) region as figure for the upright and rotated shapes appears in Table 1. As is evident from the results, when the stimulus was in the upright orientation, the network had a strong bias to call the familiar region figure. However, when the same stimulus was presented in the rotated condition, that bias was dramatically reduced. This difference was statistically reliable, as indicated by a z test on independent proportions ($z = 3.87$, $p < .001$; see Ferguson, 1981).

Trials in which the figure–ground solution did not perfectly match one of the two possible solutions for that stimulus were also analyzed. These incorrect solutions were compared between the upright and the rotated shapes. Of the 200 trials in which upright shapes were presented, the error rate was 34.5%; of the 200 trials in which rotated shapes were presented, the error rate was 30.5%. The difference between these two error rates was not statistically reliable ($z = 0.86$, $p > .30$). The same analyses were performed on the remaining simulations.

Finally, the number of incorrect solutions was also computed with a less conservative decision criterion to determine if the relatively large rates of incorrect solutions were due to our stringent criterion. The less conservative criterion involved excluding a figure–ground solution if it differed from one of the possible legal patterns by more than five units (instead of one unit). This criterion resulted in

Table 1
Results From Simulation 1: Figure–Ground Solutions and Stimulus Orientation

Summary statistic	Stimulus orientation	
	Upright	Rotated
Probability of choosing denotative region as figure	.801	.583
SD	0.035	0.042

Note. Standard deviations are based on the assumption that the probabilities came from a binomial distribution.

qualitatively similar results with only a moderate reduction in the number of incorrect solutions. In particular, for the rotated stimuli the denotative region was computed as the figural region 58.6% of the time; for the upright stimuli the denotative region was computed as the figural region 78.7% of the time. For the rotated shape trials the error rate was 30%; for the upright shape trials the error rate was 32%. Thus, a more lax decision criterion gives results similar to those of the strict criterion, suggesting that the large number of incorrect solutions is not entirely a function of the strict criterion.

Discussion

These results demonstrate that higher level object knowledge can indeed influence figure–ground organization, even when all of the lower level cues are identical between two regions. This supports our assertion that object representations can influence figure–ground organization in an interactive manner: Object representations can be partially activated by initial activation across the figure units, allowing the object units to send activation back to the lower processing levels. If there is enough of a match to a particular shape, then the corresponding object representation becomes more active and, as a result, sends more top-down activation to the figure–ground units. Although the error rates were moderate, they were similar for the upright and the rotated shapes.

Our results replicated those of Peterson and colleagues (e.g., Peterson & Gibson, 1991, 1994b). The mechanisms that give rise to our simulation results are consistent with interactive models of visual processing, suggesting that an interactive model can indeed explain how object representations can influence figure–ground processes.

Simulation 2: Exposure Duration

In Simulation 2 we investigated an aspect of the behavioral data that goes beyond the simple ability of object representations to affect figure–ground organization. Peterson and Gibson (1991, 1994a) reported that the meaningful regions in upright stimuli were more likely to be labeled figure than the same regions in rotated stimuli. However, this effect was determined by the exposure duration of the stimuli. That is, stimulus orientation interacted with exposure duration. For brief exposure durations (all durations less than 150 ms in Peterson & Gibson, 1991, and for the 14-ms exposure duration in Peterson & Gibson, 1994a), the difference in reporting the meaningful region as figure between upright and rotated shapes was much reduced, relative to longer exposure durations. Figure–ground processing can also occur without influences from object representations at short durations, but the top-down influence emerges as the exposure duration of the stimulus increases.

The effects of exposure duration were investigated by varying the number of processing cycles that the network was allowed to process the stimuli. The network was identical to that used in Simulation 1, but the stimuli were presented for 25, 50, 100, or 400 cycles.

Method

The procedure used in Simulation 2 was identical to that used previously, with the following exceptions. First, the stimuli were allowed to cycle for different amounts of time (25, 50, 100, or 400 cycles). Second, the strict decision criterion was not used because full figure–ground solutions would not have been present following a few processing cycles. Thus, use of a stringent criterion would have excluded most, if not all, of the trials. In Simulation 2 we instead had the network perform the equivalent of a forced choice on which side of the figure–ground stimulus appeared to be figure by scoring that side with more figure units active.

Results

The results of Simulation 2 appear in Figure 6. As is evident from the graph, stimulus orientation interacted with the number of processing cycles. Specifically, for the smallest number of processing cycles (25 cycles), there was no reliable difference between the upright and the rotated shapes ($z = -0.50$, $p > .60$). The differences for the other numbers of cycles (50, 100, or 400 cycles) were all highly reliable (for 50 cycles, $z = 2.30$, $p < .03$; for 100 cycles, $z = 2.10$, $p < .04$; and for 400 cycles, $z = 2.35$, $p < .02$).

Discussion

The results of Simulation 2 demonstrated that stimulus orientation interacted with the number of processing cycles, a finding which fits with the results presented by Peterson and Gibson (1991, 1994a). Observers who saw upright figure–ground stimuli at various exposure durations required a minimum exposure duration before reporting the denotative region as figure. The minimum exposure duration was different between two studies: In Peterson and Gibson's (1994a) article, exposure durations of 28 ms or greater were required in order to see denotivity effects, whereas in another article (Peterson & Gibson, 1991), exposure durations of 150 ms or greater were required. The results from Peterson and Gibson (1991) suggested that if the exposure duration was less than 150 ms, participants were likely to

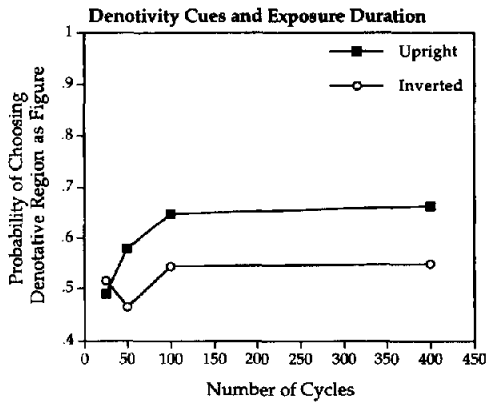

Figure 6. The results of Simulation 2. Stimulus orientation interacts with the number of processing cycles (or exposure duration).

use gestalt factors (specifically, symmetry) to determine which region appeared to be the figure, although in another study (Peterson & Gibson, 1994a) they found that symmetry cues were not used at the briefest exposure duration (14 ms). If the figure–ground stimulus was rotated 180° from the canonical, upright orientation, participants were likely to rely on gestalt factors to assign figure and ground, presumably because the inputs from object representations were diminished (see the "full figures" data in Table 2 of Peterson & Gibson, 1991).

This analysis is consistent with the operation of our network. Some minimum exposure duration must be met before object representations can be bootstrapped from the figure units to provide top-down input to the figure–ground computations. We should note that because of the intrinsic variability in the network, the object units might become activated very early in processing. However, this activation is not far enough above the noise level to reliably influence the figure units.

Simulation 3: Position Invariant Object Representations

We next turned to a shortcoming of our approach. In particular, the use of the simplified object representations in our model are problematic because of the well-known limitations of such simple template-like models (see Hummel & Biederman, 1992; Neisser, 1967; Pinker, 1984; Selfridge & Neisser, 1960, for reviews of the difficulties with template matching). Although some recent computer vision systems have effectively used a template-like procedure for object identification (e.g., Lowe, 1987; Poggio & Edelman, 1990; Ullman, 1989), some psychological evidence suggests that human object representation can be achieved over different transformations, such as position (e.g., Biederman & Cooper, 1991), a result inconsistent with template-matching models. Other evidence, however, is more consistent with template matching (Edelman & Bülthoff, 1992). Although we did not attempt to address all theoretical issues in both figure–ground perception and object recognition, we did want to ensure that the previous results could also arise in a system that used object representations that would more closely parallel the psychological data suggesting that object representations are not templates.

The critical property of the object representations for the success of Simulations 1 and 2 is that they were able to exert top-down influence *in a specific spatial location* corresponding to a familiar object. Thus, it seems that our account requires that object representations retain spatial information, which might appear to be fundamentally at odds with the psychological data just described. However, there is at least one way in which object representations could be both spatially invariant and yet still be able to exert a spatially specific top-down influence on earlier stages of processing. This could occur in a system that developed fully invariant representations over a number of intermediate processing stages, each of which extracts a greater degree of invariance from previous stages. Examples of this type of invariant object recognition have been proposed by both Mozer (1987, 1991) and Fukushima (1980), and this approach to translation invariance is depicted in Figure 7. This approach to invariant representations involves a convergence of information as one progresses upward in the hierarchy. For example, in Level 1 of Figure 7, there would be individual image features (e.g., oriented edge segments) in particular spatial locations (each of the circles in Level 1 of Figure 7 corresponds to a different spatial location). As one progresses to Level 2, the receptive fields of units in this layer are larger, thus allowing these units to start to code for more complex features, with less of a reliance on where that feature is located. This spatial collapsing continues, so that by the time information reaches Level 5, the features responded to by individual units are quite complex and, presumably, correspond to objects or components of objects. Furthermore, based on the convergence of information, the units in Level 5 are spatially invariant; they detect an object irrespective of the exact location the object occupied in Level 1. But, although the highest level representations in such a system have lost all information regarding spatial location, this information is still present in the intermediate representa-

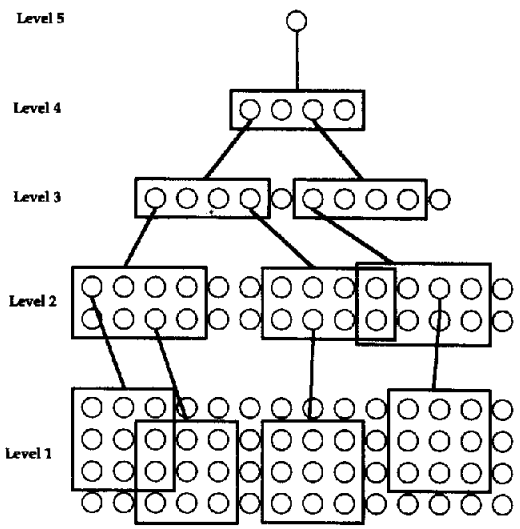

Figure 7. Illustration of how a spatially invariant object representation can interact with spatial positions occupied by objects. As one progresses from earlier levels to later levels in this network, receptive field size increases, resulting in spatial invariance in later levels (Level 5). However, the lower levels preserve spatial position, thereby allowing an invariant object representation to interact with locations. Also, although not depicted here, a network similar to this would code for more complex features the deeper one moved into the network (see Mozer, 1987, 1991). For example, Level 1 would only be involved in coding simple features such as an oriented edge in a specific location; later layers would code for more complex combinations of features, thereby resulting in object representations. We thank the International Association for The Study of Attention and Performance for permission to adapt this figure from "Early Parallel Processing in Reading: A Connectionist Approach" by M. C. Mozer, 1987, in M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* (p. 89), Hove, England: Erlbaum. Copyright 1987 by Erlbaum. Adapted with permission.

tions leading up to the object representations, and an interactive version of this type of network would indeed exhibit the ability for activity at higher, more invariant levels of processing to facilitate processing at lower, more spatially organized levels of processing.[2]

To demonstrate our model's performance with a simplified version of the multiple levels of invariant object processing (à la Fukushima, 1980, and Mozer, 1987; see Figure 7), we extended the model by including another layer of object-level representations. This layer contained four units and was positioned between the figure layer and the object layer. This intermediate object layer coded for familiar patterns in a particular retinal position (similar to the intermediate levels in Figure 7); the object layer itself received input from the different positions of a given object. Thus, the object layer had spatial invariance in that the two units coded for a shape irrespective of where that shape appeared (similar to Level 5 in Figure 7). As before, there were two familiar objects. Each of the four individual units in the intermediate object layer coded for one of the two familiar objects in one of two positions. For example, the first unit in this layer coded for Object 1 in Position 1, the second unit coded Object 2 in Position 1, and so on. Each of the two units in the object layer coded for one of the two objects irrespective of the location in which that object appeared, thereby making these object representations spatially invariant. The invariant object representations influenced processing by both receiving input from the appropriate intermediate object representations, thereby allowing objects to be recognized, and by sending top-down activation back to the intermediate object layer, thereby allowing a bias in processing familiar shapes relative to unfamiliar shapes.

Method

The network was tested as in Simulation 1. All parameters in the current network were identical to those used in the previous two simulations.

Results

The network was again tested with upright and rotated versions of the two familiar shapes. The four possible stimuli were each presented on 100 different trials, and the resulting figure–ground solution was compared with all possible solutions for a given stimulus. We used the criterion from Simulation 1: Any figure–ground organization that did not perfectly match one of the potentially correct solutions was excluded from the analyses.

The results appear in Table 2. As is evident from the results, when the stimulus was in the upright orientation, the network had a strong bias to call the familiar region figure. However, when the same stimulus was presented in the rotated condition, that bias reduced dramatically. This difference was statistically significant ($z = 4.72$, $p < .001$). These results parallel those from Simulation 1, suggesting that the higher level object representations can overcome the ambiguity of these stimuli. Thus, as before, object representations can directly influence figure–ground organization, even when all other stimulus factors are held constant between two regions.

Table 2
Results From Simulation 3: Position Invariant Object Representations

Summary statistic	Stimulus orientation	
	Upright	Rotated
Probability of choosing denotative region as figure	.919	.691
SD	0.023	0.044

Note. Standard deviations are based on the assumption that the probabilities came from a binomial distribution.

The number of incomplete solutions was also collected and averaged across the two different shape orientations. For the two upright shapes, the average percentage of incomplete solutions was 27%; for the two rotated shapes the average percentage of incomplete solutions was 44.5%. This difference was statistically significant ($z = 3.65$, $p < .001$), with reliably fewer incomplete solutions occurring when the shape was in the upright orientation.

Discussion

These results replicated our earlier results, suggesting that higher level object knowledge can indeed influence figure–ground organization. Furthermore, in the present simulation we used object representations that were more plausible than those used in Simulations 1 and 2. Note, however, that the intermediate object units used in the present simulation still have many of the difficulties involved with template models. Again, we are not attempting to solve all of the problems posed by object representation and recognition; rather, we are trying to understand the interactions that might exist between intermediate levels of perceptual organization and later stages of object representation. To this end, our model has succeeded in showing that an interactive approach can

[2] In addition, it is known that there is extensive interconnectivity between the ventral, object-based processing stream and the dorsal, spatial-based processing stream (see Desimone & Ungerleider, 1989; Goodale & Milner, 1992; Harries & Perrett, 1991; Ungerleider & Mishkin, 1982). Thus, these connections could be providing interactivity between spatial and object processing, which could lead to the ability of object representations to influence spatially localized processing, as is required in our model. Further, data from neglect patients, who suffer from lesions of the parietal lobe (the dorsal stream), show impairments of spatially based processing defined both by object-based as well as spatially based reference frames, indicating that object-based representations are apparently affecting the spatial-based processing taking place in the dorsal stream (see Behrmann & Moscovitch, 1994; Driver & Halligan, 1991; Farah, Wallace, & Vecera, 1993). Thus, the idea that invariant object-recognition processing can also exhibit spatially localized effects on processing in other areas (e.g., the figure–ground organization being modeled here) is sufficiently plausible to justify the simplified implementation of object representations used in our models.

explain the influence familiarity (or meaningfulness) has on figure–ground organization.

Simulation 4: *k*-WTA Algorithm

Although the results of the previous simulations are consistent with our claims, a significant point warrants further discussion. First, with both Simulations 1 and 3 the number of incomplete solutions is quite problematic. What caused the relatively large numbers of incomplete solutions? These improper solutions were most likely due to the unrestricted activation that could occur across the figure units, which permitted the figure units to settle into patterns that did not correspond to one of the possible solutions given the input. Such solutions correspond to the network settling into a local maxima (i.e., a nonoptimal solution). Other researchers have noted this difficulty with multiple-constraint satisfaction networks (Hinton & Lang, 1985; Mozer, Zemel, & Behrmann, 1992; Mozer, Zemel, Behrmann, & Williams, 1992). Specifically, an interactive, multiple-constraint satisfaction model such as ours must perform two searches (Mozer, Zemel, & Behrmann, 1992). The first search is for a good solution across the figure units; the other, simultaneous search is for a good solution across the object units. As Mozer, Zemel, and Behrmann (1992) noted, such multiple searches often lead to a large number of incorrect solutions.

One limitation of our model that encouraged these local maxima was that there were no inhibitory mechanisms to restrict the number of figure units that could be active at any one time. Thus, activation across the figure units would often begin to grow across the boundaries specified by the input to the network. Given that inhibition is a well-known aspect of neural processing, our failure to implement inhibition within the figure layer is a clear limitation of the model. However, we found it impossible to implement inhibition directly among the figure units because direct lateral inhibition is often unstable when multiple inputs are to remain active at the same time. To address this issue, we adopted a "*k*–winner take all" (*k*-WTA) algorithm to incorporate inhibitory processes across the figure units. This should significantly reduce the number of incomplete solutions because only a certain number of units (denoted by the parameter *k*) are allowed to be active on any particular layer. In Simulation 4 we attempted to replicate our previous results using this algorithm.

We used the activation function from Learning in Error-driven and Associative, Biologically Realistic Algorithms (LEABRA; see O'Reilly, 1995), which incorporates some additional, biologically motivated properties beyond the ones typically used in generic PDP models like the Boltzmann machine or back-propagation networks. For the present purposes, the relevant property is activity regulation, which is intended to capture the effects of inhibitory interneurons on pyramidal excitatory cells. The LEABRA model uses the *k*-WTA model of activity regulation, which stipulates that inhibition limits the maximum number of active units to some relatively fixed upper level, *k*. This form of activity regulation has been used to describe the effects of inhibitory circuitry in the hippocampus and other brain areas (Gibson, Robinson, & Bennett, 1991; McNaughton & Morris, 1987; O'Reilly & McClelland, 1994; Torioka, 1979; see Appendix B for the LEABRA activation equations).

Method

The procedure was identical to that used in Simulation 3, except that the network cycled for 200-time steps and used a slightly different noise schedule (see Appendix B for parameters). Note that *k*, the number of units active, was different for each layer. The value of *k* alone did not, however, bias the network to favor one region over the other region in the displays because the number of units was equal for the familiar region and the less familiar region (48 units in each). Any biases observed would be due to the top-down projections from the object units, not to the *k* parameter.

Results

The network was again tested with upright and rotated versions of the two familiar shapes, and the four possible stimuli were each presented on 100 different trials. We used the criterion from Simulations 1 and 3: Any figure–ground organization that did not perfectly match one of the potentially correct solutions was excluded from the analyses.

The results appear in Table 3. As is evident from the results, when the stimulus was in the upright orientation, the network again had a strong bias to call the familiar region figure; when the same stimulus was presented in the rotated condition, that bias reduced dramatically. This difference was statistically significant ($z = 3.07$, $p < .01$). These results parallel those from the previous simulations.

The number of incomplete solutions was also collected as in the earlier simulations. In contrast to the previous simulations, the average percentage of incomplete solutions was dramatically reduced: The error rate was 3.5% for the two upright shapes and 6% for the rotated shapes. Although there were fewer incomplete solutions when the shape was in the upright orientation, this difference failed to reach statistical significance ($z = 1.19$, $p > .20$).

Discussion

The results of Simulation 4 replicated our earlier results by using a *k*-WTA updating algorithm, but in the present

Table 3
Results From Simulation 4: Figure–Ground Solutions and Stimulus Orientation in k-WTA Network

Summary statistic	Stimulus orientation	
	Upright	Rotated
Probability of choosing denotative region as figure	.791	.650
SD	0.029	0.035

Note. Standard deviations are based on the assumption that the probabilities came from a binomial distribution. *k*-WTA = *k*–winner take all; *k* = parameter *k*.

simulation the number of incorrect or incomplete solutions has been significantly reduced. In addition, these results demonstrate that the interactive approach that we are advocating generalizes over different updating algorithms.

These results also suggest that a constraint satisfaction approach can be used without resulting in a large number of incorrect solutions (i.e., local maxima). Also, we were able to replicate the previous findings after reducing the number of processing cycles from 200 to 100, suggesting that the constraint satisfaction approach, when combined with the LEABRA *k*-WTA algorithm, can robustly simulate figure–ground results, even with a smaller time scale.

Simulation 5: Multiple Cues

Although the present series of simulations support an interactive approach to figure–ground perception, Peterson and Gibson (1993) have presented results that they believe argue against a "feedforward and feedback" (i.e., interactive) process. Peterson and Gibson (1993) presented participants with figure–ground displays that contained depth cues. These displays could either be black and white (B&W) stereograms or random dot (RD) stereograms. In each display type, there was a high-denotative (i.e., familiar) region and a low-denotative region. The depth cue could either cooperate or compete with the denotivity cue. For example, the high-denotative region could appear closer to the viewer (cooperation), or the low-denotative region could appear closer to the viewer (competition). When the denotative region was close to the viewer, this was said to be *cooperative* because regions that are closer to the viewer tend to be perceived as figure; thus the denotivity and depth cue both signaled that the same region (the denotative region) should be figure. Similarly, when the low-denotative region was closer to the viewer, this was said to be *competitive* because the distance of the low-denotative region suggested it should be figure while the denotivity of the more distant high-denotative region suggested it should be figure, thereby creating a competition between nearness and meaningfulness.

Using these types of displays, Peterson and Gibson (1993) found that figure–ground organization was different between B&W stereograms and RD stereograms. In B&W stereograms, both denotivity and depth cues constrained equally what participants would call figure. Specifically, in the competition condition, participants perceived the meaningful region to be figure about half of the time; the other half of the time they perceived the closer region to be figure, even though this region was low in denotivity. The cooperative B&W stereograms resulted in participants' perceiving the meaningful region, which was also closer to the participant, as figure most of the time. In contrast to these results, in the RD stereograms disparity alone and not denotivity determined which region would be figure. For example, in these stereograms, when the high-denotative region appeared closer to the participant, this region was called figure; however, when the low-denotative region was closer to the participant, this region was called figure, and denotivity did not appear to influence figure–ground organization, unlike performance in the B&W stereograms.

Peterson and Gibson (1993) argued that these results could not be due to interactive processes. In particular, if disparity cues acted as the lower level letter units in McClelland and Rumelhart's (1981) model of the word superiority effect, then in B&W stereograms higher level denotivity cues should "operate to facilitate correct fusion rather than to alter the outputs of fusion processes" (p. 423). In other words, they did not see how top-down cues (i.e., denotivity) could cause the lower level units to settle into a pattern that was actually *inconsistent* with the bottom-up input. For example, this would amount to McClelland and Rumelhart's network settling into the letter-level representation of *wave* in response to the nonword bottom-up input *mave*. That is, the top-down support for *wave* would actually have to override the bottom-up input of *mave* by changing all of the *m* features to *w* features in the first letter position. Thus, to account for their results, Peterson and Gibson (1993) would need to postulate that denotivity must in effect be a bottom-up-like cue (on the basis of the prefigural shape-recognition process) with respect to the figure–ground organization process, as shown in Figure 8a. Thus, by Peterson and Gibson's (1993) account, the denotivity cue can compete on equal footing with the other bottom-up cues such as disparity in determining what is viewed as figure. Denotivity is only a factor in the B&W stereograms because only certain types of contours (notably, luminance contours) allow prefigural shape-recognition processes to be conducted, whereas the RD stereograms, lacking luminance contours, have only disparity cues.

Although it is true that McClelland and Rumelhart's (1981) network did not exhibit the ability of top-down activation to override bottom-up inputs, there is no reason to believe that it is impossible in principle for interactive networks to have the kind of strong top-down influence necessary to account for Peterson and Gibson's (1993) results. However, it is not simply a matter of the strength of the top-down influence because the top-down units require input from the bottom-up inputs to become active in the first place. Thus, for top-down representations to compete with

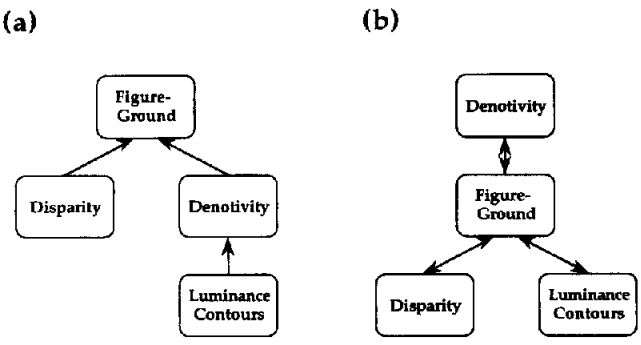

Figure 8. The two alternative models for combining denotivity and disparity cues. (a) Our interpretation of Peterson and Gibson's (1993) prefigural account, in which we assumed that no feedback occurs. (b) Our interactive account.

bottom-up ones, they require at least some degree of support from the bottom-up cues, otherwise they simply remain inactive and do not influence processing at all. This suggests that at least some degree of ambiguity in the bottom-up input is required to allow the top-down influences to affect processing.[3] The importance of ambiguity for the expression of top-down influences was emphasized in the original model of McClelland and Rumelhart (1981). They showed that when a stimulus word was partially occluded, the network would fill in the missing letter in a way that was consistent with a known word. When multiple words were consistent with the ambiguous input, the missing letter would be completed on the basis of a competition among the various word-level units.

We accounted for Peterson and Gibson's (1993) results not in terms of a competition among multiple top-down influences for resolving a partially ambiguous input but rather as a competition between top-down and bottom-up cues. We assumed that disparity cues operate in parallel with luminance contour cues, an assumption that is at least partially supported by the segregation of form and depth processing in the early cortical visual areas (see Livingstone & Hubel, 1987, 1988). Thus, the presence of the ambiguous bottom-up luminance contour enables the object representations to become partially activated and thereby to compete with the disparity cue (see Figure 8b). In the case in which the luminance contour is absent (i.e., in the RD stereograms), the object units never become partially activated in the first place, and processing is dominated by the bottom-up disparity cue.

The revised model used in Simulation 5 appears in Figure 9.[4] The model is identical to the previous models, except for the addition of a disparity layer. This layer was the same size as the figure layer (12 × 16 units) and served as an additional input to the figure units. This layer provided input to the figure units in parallel with input from the boundary units. The disparity units projected only to the figure units in a one-to-one manner. That is, a disparity unit at a given retinal location projected only to the figure unit in the same retinal location. All of these weights were positive (+5). Disparity cues were represented by activating a contiguous set of the disparity units, which corresponded to the single region that was closest to the network. For example, if a surface was supposed to appear closer to the network, then the disparity units that corresponded to that region would receive external input. The external input to these units was +36. All of the other disparity units would receive no external input. We took this representation across the disparity units to be the final product, or output, of a process that computed depth based on binocular cues (see Marr & Poggio, 1976, for an example of such a process).

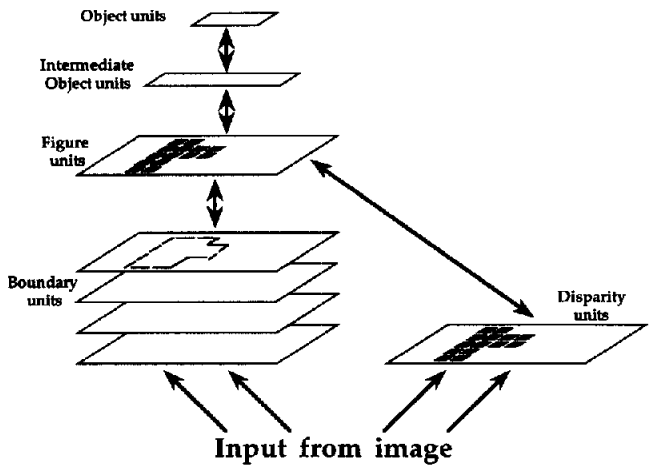

Figure 9. The network used in Simulation 5. This model included a layer of units that represented the disparity of a region.

Method

We simulated the effects of disparity cues with the network. The input was analogous to either a B&W stereogram or to an RD stereogram. In the B&W case, the input was presented to both the boundary units and to the disparity units. In the RD case, input was only presented to the disparity units; this was done because in an RD stereogram no luminance contour cues exist. Within each type of stereogram, the disparity cues could either cooperate or compete with denotivity. In the cooperative condition, the input across the disparity units was a meaningful shape (refer to Figure 4). In the competitive case, the input across the disparity units was the less meaningful shape (i.e., the other half of the stimulus in Figure 4).

The input in the four conditions was as follows. For the B&W, cooperative case, an ambiguous contour was presented across the boundary units, as with all of the previous simulations; along with this, the familiar, high-denotative shape was presented across the disparity units. In the B&W competitive case, the same ambiguous contour was presented across the boundary units; however, the less meaningful, low-denotative region was presented across the disparity units. For the RD, cooperative case, the high-denotative shape was presented across the disparity units. In the RD, competitive case, the low-denotative shape was presented across the disparity units. In both RD conditions there was no input presented along the boundary units.

Results

Incorrect solutions were judged as they were in the previous simulations, with one exception: No criterion was used in the RD stereogram conditions because these inputs typically did not precisely activate the figure units, so the solutions often did not perfectly correspond to a correct solution. This was due to the lack of any inhibitory connections between the disparity units and the figure units; this lack of inhibitory connections allowed activation across the figure units to spread across the shape's boundaries. The activation spread across the boundary due to the lateral connections among the figure units (see Figure 2b).

[3] Of course, there may be other ways of creating a top-down bias (e.g., experimental instructions, strategies). We are simply predicting that ambiguity is required in situations where this top-down influence has not been caused by other factors, as in the present simulations.

[4] The network used in Simulation 5 was the same network used in Simulation 4, but in the earlier simulation no input was presented across the disparity units.

Each of the four conditions (B&W cooperative and competitive; RD cooperative and competitive) was presented 100 times. The results appear in Figure 10. As is evident, the B&W stereograms resulted in a very different pattern of performance than did the RD stereograms. In particular, in the B&W stereograms the cooperative and competitive conditions were significantly different from one another ($z = 10.22$, $p < .001$). However, in the RD stereograms the cooperative and competitive conditions did not differ from one another ($z = 0.16$, $p > .80$). Finally, and most important, the difference between the two types of B&W stereograms was reliably different from the two types of RD stereograms ($z = 16.03$, $p < .001$), suggesting that stereogram type (B&W vs. RD) interacted with condition (cooperative vs. competitive).

Finally, the incorrect solutions for the B&W stereograms were infrequent. For the cooperative B&W condition the error rate was 2.5%; for the competitive B&W condition the error rate was 9%. This difference was statistically reliable ($z = 2.83$, $p < .01$), suggesting that fewer errors were made in the cooperative condition than in the competitive condition.

Discussion

The results of Simulation 5 replicated those presented by Peterson and Gibson (1993). Specifically, with B&W stereograms, the disparity cues and the denotivity cues both influenced figure–ground organization. This emerged as a difference between the cooperative and competitive B&W stereograms; the difference of the proportions was .47 (or 47%) in the model and approximately .30 (30%) in Peterson and Gibson's (1993) data, suggesting a good fit by our model. (Our approximations of the data from Peterson and Gibson's, 1993, article were averaged over three different amounts of disparity, and these data were the proportions of initial figure–ground reports.) In the RD stereograms, the results were dominated by the disparity cues alone because the bottom-up input was unambiguous (i.e., across the disparity units there was only one surface presented). The region that appeared closer (i.e., the disparity cue) was the region determined to be figure. The difference between the proportions in the cooperative and competitive stereograms was .005 (or .5%) in the model and approximately .06 (or 6%) in Peterson and Gibson's (1993) data, again suggesting a good fit by the model. (The data from Peterson and Gibson's, 1993, study are approximations based on Figure 7 in their article.)

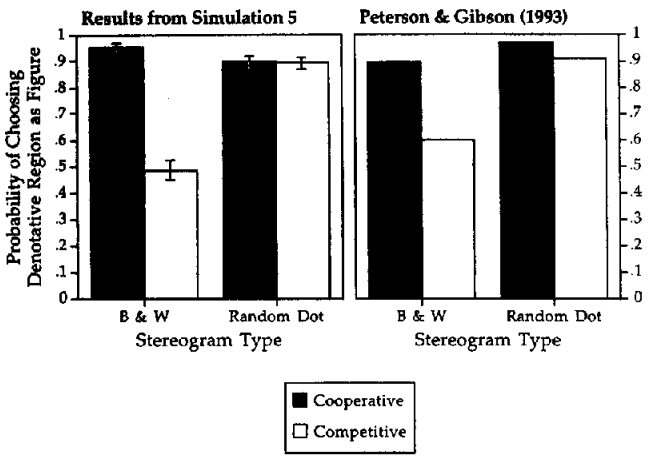

Figure 10. The results of Simulation 5. Stereogram type (black and white [B&W] vs. random dot) interacts with whether the two cues (denotivity and disparity) are cooperative or competitive. The model's data are presented with approximate values from Peterson and Gibson (1993).

Thus, in the present simulation we have also demonstrated that our interactive approach can account for the competition between multiple cues (i.e., denotivity and disparity cues), *even when these cues are at different levels of processing.* We postulated that the critical factor that determines when top-down cues are able to compete with bottom-up ones is the existence of some degree of ambiguity in the bottom-up input. In the present case, this ambiguity was present in the luminance contours, which could support either side of the contour as figure. This ambiguity allowed the denotative region of the display to become partially activated, thereby enabling the corresponding object unit to become active; this, in turn, provided top-down support for that region within the display. This result contradicts Peterson and Gibson's (1993) intuitions about what kinds of top-down influences can be found in interactive networks. However, we should acknowledge that it is not entirely obvious from McClelland and Rumelhart's (1981) interactive-processing model that results like those we obtained would occur in an interactive network. Indeed, we know of no other models of psychological results that show that top-down cues can actually override bottom-up cues. However, these kinds of strong top-down effects are not surprising at a purely computational level, as we would expect this kind of behavior from a wide range of interactive models under similar conditions.

One final point of discussion concerns the figure–ground organizations from the RD stereogram condition. Such a solution appears in Figure 11. The figural solution lacks the precisely defined edges that characterized figural solutions when luminance contours were the input. This "blurry" figure–ground solution is due to the absence of any inhibitory connections between the disparity units and the figure units. Note that these blurry solutions are not a failure of the network. Instead, this result suggests that human participants may have difficulty recognizing objects in RD stereograms, which is consistent with empirical reports (e.g., Peterson & Gibson, 1993, p. 421). Indeed, in viewing the RD stereogram stimuli used by Peterson and Gibson (1993, pp. 392–393), this seems informally to be the case. However, such a blurry solution is not the perception that one has of the disparity contour itself; instead, the disparity contour is quite sharp and highly distinct.[5] On the basis of our network, we suggest that the perception of the disparity

[5] Thanks to Steven Pinker for pointing this out.

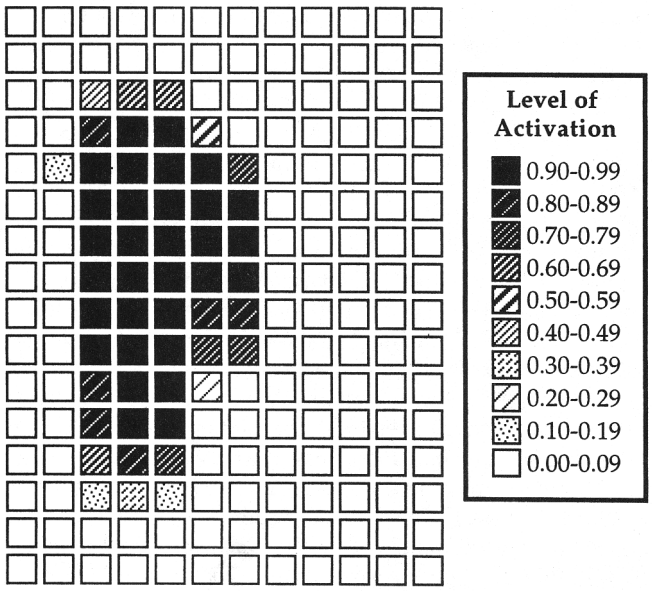

Figure 11. An example of a figure–ground solution that resulted after presenting a random dot stereogram as input. The squares correspond to the individual figure units, and the shading of the square indicates the activation of that unit (see figure legend). The absence of any luminance contours prevented the formation of a well-defined figural region, often preventing the network from recognizing the object. On this trial, however, the network did recognize the shape presented.

contour itself may be mediated directly by the disparity units, thereby resulting in a well-defined disparity edge.

General Discussion

In a set of five simulations we have demonstrated that a relatively simple PDP model can account for several phenomena in figure–ground perception. Not only was the network able to separate figure from ground correctly, but it also showed orientation-dependent effects and exposure-duration effects similar to those demonstrated by human participants. Furthermore, in Simulations 1–4 the network's performance was strongly influenced by the top-down information. This is because all of the lower level cues (e.g., area, symmetry, convexity) were identical between the two regions of the displays presented to the network; only the denotivity (familiarity) cue differed between the two regions of the stimulus. Finally, the network could exhibit competition between bottom-up and top-down cues when determining figure–ground relations, as shown in Simulation 5. Thus, we have demonstrated that top-down processing in an interactive model, can alter—not simply facilitate—the course of processing at lower levels, which contradicts Peterson's (e.g., Peterson & Gibson, 1994a) stipulation to the contrary. We now turn to a discussion of the implications of these results, as well as to additional issues raised by the present work.

Implications

A significant contribution of this work is that it accounts for the paradoxical data from studies of figure–ground organization without proposing a prefigural recognition process as proposed by Peterson and colleagues (e.g., Peterson, 1994). Our approach maintained a hierarchical scheme of information processing but adopted processing principles in accordance with PDP models of information processing. This allowed us to maintain the hierarchical organization of the more traditional models of visual perception (e.g., Biederman, 1987; Kosslyn, 1987; Marr, 1982; Neisser, 1967; Palmer & Rock, 1994a) without having to adhere to the sequential processing that has been associated with many of these models. We should note that our interactive account does not directly refute Peterson's account; what the model offers, instead, is a reconciliation between the traditional models, which often emphasize serial, feed-forward processing, and Peterson's data, which cannot be fully explained with the traditional models.

Although we have argued that our account is hierarchical, a cautionary note is in order. In massively interactive PDP networks, the intuitive notions of a level or of a hierarchy are blurred. Although the object units take input only from the figure units (and are therefore technically after the figure–ground processes), the parallel, interactive computations result in the object units operating in parallel with the figure units. We acknowledge that this approach to hierarchical processing blurs the distinction between our account and that offered by Peterson (1994) because figure–ground processing is occurring in parallel with object representation. However, our claim is that there is a difference between parallel, interactive processing and prefigural shape-recognition processes. The difference lies in how the object representations are activated, or "bootstrapped." In our model the object units take input from the figure units, but in Peterson's (1994) account object representations take input directly from luminance contours (e.g., the boundary units in our model).

Perhaps one of the most challenging findings for an interactive model of figure–ground processing like ours is the data from Peterson and Gibson's (1993) experiments that showed that denotivity and disparity cues competed with each other in influencing participants' figure–ground processing. Because denotivity is a top-down influence in our interactive model, accounting for this result requires that top-down influences be able to compete with bottom-up influences. That we were able to show how this can happen in our model and understand more generally the conditions under which effects of this nature can occur extends the range of phenomena that interactive models can account for. This result goes beyond simple intuitions about the nature of interactive processing and argues for the importance of simulation models for informing psychological theories. Further, we can predict on the basis of our model that the presence of some degree of ambiguity (even when other, unambiguous cues are also present) would be a necessary condition in order for top-down influences to be observed (assuming that the top-down influences have not been

activated by other sources, such as experimental instructions, strategic knowledge, etc.; see Footnote 3). We should also note, however, that bottom-up cues may often be sufficient; the main conclusion from our approach is that these cues alone do not fully characterize information processing.

Finally, there are other behavioral results that we have not directly simulated but that can be accounted for by our model. It is well established that symmetric regions are more likely to be perceived as figure than are asymmetric regions. Furthermore, these symmetry cues can combine with denotivity cues; Peterson and Gibson (1994a) showed that symmetric, high-denotative regions were more likely to be perceived as figure than were asymmetric, high-denotative regions. The same effect held for less denotative regions: Symmetric low-denotative regions were more likely to be seen as figure than were asymmetric, low-denotative regions. Thus, both symmetry and denotivity influenced figure–ground organization. To the extent that symmetry acts like the disparity cue used in Simulation 5, we can account for these results with our current model. This would likely be true even if the influence of symmetry was more of an emergent phenomenon, which might not be represented separately as disparity was in our network. Similar arguments could be made for other kinds of bottom-up cues, such as fixation (Peterson & Gibson, 1994b).

Additional Issues

Although our network accounts for the role of familiarity in figure–ground perception, there are additional issues that warrant discussion. Here we address five significant issues: (a) the locus of knowledge in the visual system, (b) the process of settling in the model, (c) the possible neural locus of figure–ground organization, (d) the relationship between figure–ground organization and image segmentation, and (e) the implementation of object representations in the current model.

In discussing our network, we have emphasized that the familiarity or denotivity cues arise from the higher level object representations. This suggests that knowledge about shapes occurs only in the weighted connections between the figure units and the object units and that familiarity effects in figure–ground organization are the result of interactions between these two layers of units. However, there is an alternative account for such familiarity effects also based on PDP models of visual processing.

Mozer, Zemel, Behrmann, and Williams (1992) have constructed a PDP model that learns to perform a type of perceptual organization; this network learns to segment two overlapping shapes apart from one another. This model, Multiple-object Adaptive Grouping of Image Components (MAGIC), consists of only a lower level, array-format representation in which image features (edges) are presented and a set of hidden units that takes its input from the feature representation. The network learns to group certain conjunctions of these image features. For example, the network might learn that a vertical edge and a horizontal edge in close retinal proximity form a T junction and that these two image features have a high probability of belonging to different shapes. Although MAGIC contains no object representations as our network does, MAGIC can still exhibit familiarity effects; MAGIC will segment two familiar shapes more quickly than it will segment two less familiar shapes (M. C. Mozer, personal communication, 1993).

When the results from MAGIC are contrasted with our simulations, does an explanation emerge regarding where familiarity effects arise in the human visual system? That is, could familiarity effects be explained as arising solely from lower level image statistics, as in MAGIC, or could familiarity effects be explained by object representations interacting with lower level representations? We argue that knowledge could exist in both lower level and higher level representations. Indeed, our network contains knowledge within both the boundary units and the figure units. This knowledge takes the form of weighted connections that specify how edges form corners and how edges constrain figural regions (see Figures 2 and 3). Thus, familiarity effects are most likely determined by knowledge at many different levels. Some tasks, such as those used by Peterson and colleagues (Peterson & Gibson, 1991, 1993, 1994a, 1994b; Peterson et al., 1991), may depend more heavily on knowledge at higher levels of representation, but other tasks may depend more heavily on knowledge of image statistics (e.g., vertex information or line termination information; see Waltz, 1975).

Another consideration regarding the locus of familiarity effects in visual processing, and the nature of visual processing more generally, is the duration of the settling process required by models such as the one we have used. Computationally, it is important that the buildup of activation is gradual so that constraints at different levels have a chance to influence the outcome of processing. There are two potential problems with this gradual settling, however. First, the overall processing time might be psychologically implausible—gradual settling in a network might take longer than is consistent with psychological data. Second, phenomenologically it seems that many aspects of visual perception are more discrete than graded (as in our model). Figure–ground organization seems to have this discrete character—it seems that one region is entirely the figure while another is entirely the ground. Regarding the first point, we have shown that by introducing further constraints on the activation states of units with a k-WTA activation function, settling times can be reduced from 400 cycles to 100 cycles. Regarding the second point, even though processing is fundamentally graded, the network can exhibit "phase transition" kinds of behavior, where processing undergoes a rapid transition into a figure–ground solution. To illustrate this latter point, in Figure 12 we present a plot of the goodness function as the GRAIN network settles into a correct figure–ground solution. (Recall that goodness is the degree to which the state of the network fits the constraints imposed by the weights. Higher values of goodness indicate the current state fits the constraints well.) As is evident, the network undergoes a rapid period of transition after some number of processing cycles. This transition corresponds to

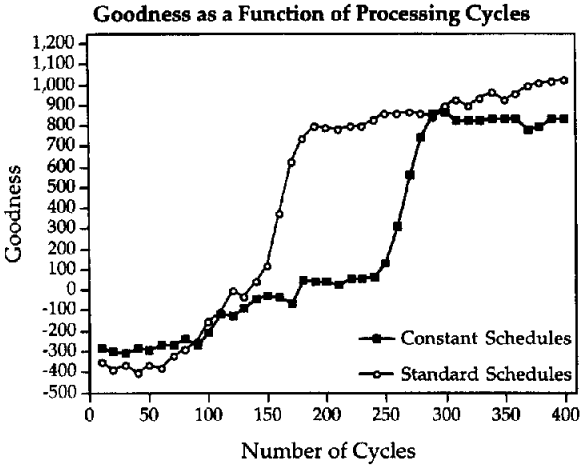

Figure 12. The goodness of the activation state of the network as it settles into a correct solution. The constant schedule uses constant noise and sharpening values; the standard schedule uses the schedules used in Simulation 1. Both schedules show a phase transition, where the network quickly settles into the correct, stable solution. For the constant schedules this occurs between Cycles 250 and 300; for the standard schedules this occurs between Cycles 150 and 200.

the development of a coherent and largely correct solution across the figure–ground layer. This phase transition behavior is not simply a byproduct of the noise and sharpening schedules; we see the same rapid transition behavior in a network using constant noise and sharpening levels (see Figure 12).

In our model, the figure units form a layer separate from both the boundary units and the object units. Whereas the separation of the figure units from the object units seems justified by most accounts of perception, is the separation of the figure units and the boundary units justified? Studies that have measured visual evoked potentials (VEPs) have revealed that early visual cortices may be responsible for some forms of perceptual segmentation (Bach & Meigen, 1991; Lamme, Van Dijk, & Spekreijse, 1992, 1993). However, VEPs may not allow a fine discrimination between adjacent cortical areas. More detailed single-cell recording studies have suggested that neurons in V1 may be modulated by figure–ground relations; V1 neurons gave larger responses to stimuli belonging to a figural region than to background regions (Lamme, 1995). This suggests that figure–ground relations may be computed early, at the same level at which edges are detected. However, other results have pointed to area V2 as being the neural locus of a "direction of figure" computation (see Sajda & Finkel, 1995; von der Heydt, Peterhans, & Baumgartner, 1984). These results suggest that figure–ground processes may occur at a different neural locus than does edge detection. Thus, the evidence concerning the separability of the boundary units and the figure units is somewhat inconclusive. Although we clearly separated these two layers in our model, we do not reject the possibility that figure–ground processes might emerge from edge-detection processes. Figure–ground organization may not be assignable to a distinct visual area.

Next, we should note that examining only the case of figure–ground organization may be somewhat limiting because in such behavioral experiments (and simulations), only a single stimulus is presented at any given time. What of complex visual scenes in which multiple objects are present? Presumably, before regions can be labeled figure or ground, different regions must be segmented apart from one another, or, alternatively, individual regions must be grouped together. This suggests that image segmentation or perceptual organization might precede figural judgments (see Finkel & Sajda, 1992, 1994; Sajda & Finkel, 1995, for a model that incorporates such a processing hierarchy). Could higher level object knowledge influence image segmentation? Some behavioral results suggest that image segmentation may indeed be an interactive process (see Vecera, 1993, for relevant behavioral results). The case of image segmentation is presumably quite similar to that of figure–ground organization in that object representations might constrain the processes by which local edge information is grouped. In the present network, such effects could be due to top-down information that is cascaded back from the object units through the figure units and finally to the boundary units. We should note, however, that image segmentation raises the difficult issue of binding of low-level features to shapes, which is obviously not addressed by our current model (cf. Hummel & Biederman, 1992).

Finally, the current implementation of object representations in our model is potentially problematic. Human vision is characterized by the ability to recognize objects across different spatial positions or different sizes (e.g., Biederman & Cooper, 1991). Could the present simulation results be specific to template-like matching systems (e.g., Simulations 1 and 2), or are the results more general? First, we showed the generality of our approach in Simulations 3–5 by allowing a more gradual buildup of spatially invariant object representations, in accord with other PDP models of invariant object recognition (e.g., Fukushima, 1980; Mozer, 1987, 1991). Second, the success of these PDP models suggests that our results should generalize to more realistic object representation schemes. As we discussed previously, Mozer's (1987, 1991) connectionist model architecture created spatially invariant object representations by gradually collapsing across space from one layer to the next, an idea originally proposed by Fukushima (1980). O'Reilly and Johnson (1994) showed that a self-organizing learning algorithm (Földiak, 1991) could develop a similar progression of increasingly invariant object representations. Finally, although more realistic object representations could be explored, our use of template-like representations does not detract from our main theoretical claim: Invariant object representations can influence figure–ground organization in a location-specific manner by means of bidirectionally connected intermediate representations that incrementally compute this invariance.

Toward a GRAIN Account of Visual Processing

The phenomenon of denotivity or familiarity effects in perception has been observed in many different tasks that tap

different aspects of visual processing. Perhaps the paradigmatic case is the word superiority effect (Reicher, 1969; Wheeler, 1970). Interestingly, the word superiority effect was explained in a manner similar to the prefigural account of figure–ground organization. Lawry and LaBerge (1981) proposed that word-level information was processed in parallel with or before letter information. This "prelexical" account suggested that word-level information might be bootstrapped from letter-feature information (i.e., edges and line segments), not from letter information itself. McClelland and Rumelhart's (1981) interactive account of the word superiority effect showed that this prelexical stage was unnecessary if processing was graded and bidirectional.

The situation in figure–ground organization parallels that of the word superiority effect. Peterson (1994) hypothesized a prefigural stage of processing, and we showed that this was an unnecessary assumption in an interactive model. However, whereas our interactive account explains the results from a number of behavioral experiments, the prefigural account also explains these results. Why should one favor our interactive model over the prefigural account? A strong argument in favor of our interactive model is based on parsimony, because our account retains the logical hierarchy of visual processing (Palmer & Rock, 1994b) and does not require the introduction of a new processing stage in addition to those commonly thought to exist.

Our interpretation of figure–ground processing is also consistent with the so-called Höffding function (Höffding, 1891; see also Neisser, 1967; Rock, 1962), which suggests that some sensory representation (or perceptual organization) must precede recognition. Sensory organization precedes recognition, but one does not need to assume that these two processes are sequentially ordered; instead, organization processes can be influenced by recognition processes. Thus, the Höffding step need not be viewed as a discrete, sequential step; instead, it can be thought of as a gradual, cascaded process.

Finally, a significant advantage to our interactive account is that we use a set of information-processing principles that are powerful enough to explain not only figure–ground organization but other effects in visual processing as well. For example, the GRAIN principles have been used to create working models in different visual domains (e.g., Stroop interference, binocular disparity, the word superiority effect), as well as across entirely different sensory domains (e.g., speech perception; see McClelland, 1991; McClelland & Elman, 1986). As a consequence, the GRAIN principles offer the possibility of providing the foundation of a more general information-processing framework, suggesting that interactive processing, in which information flows bidirectionally, might be a computational strategy used not only by the visual system but by other cognitive and perceptual systems as well.

References

Bach, M., & Meigen, T. (1992). Electrophysiological correlates of texture segregation in the human visual evoked potential. *Vision Research, 32,* 417–424.

Behrmann, M., & Moscovitch, M. (1994). Object-centered neglect in patients with unilateral neglect: Effects of left-right coordinates of objects. *Journal of Cognitive Neuroscience, 6,* 1–16.

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

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception, 20,* 585–593.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 2, pp. 267–299). New York: Elsevier.

Driver, J., & Halligan, P. W. (1991). Can visual neglect work in object-centered co-ordinates? An affirmative single-case study. *Cognitive Neuropsychology, 8,* 475–496.

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32,* 2385–2400.

Farah, M. J., Wallace, M. A., & Vecera, S. P. (1993). "What" and "where" in visual attention: Evidence from the neglect syndrome. In I. H. Robertson & J. C. Marshall (Eds.), *Unilateral neglect: Clinical and experimental studies* (pp. 123–137). Hove, UK: Erlbaum.

Ferguson, G. A. (1981). *Statistical analysis in psychology and education* (5th ed.). New York: McGraw-Hill.

Finkel, L. H., & Sajda, P. (1992). Object discrimination based on depth-from-occlusion. *Neural Computation, 4,* 901–921.

Finkel, L. H., & Sajda, P. (1994). Constructing visual perception. *American Scientist, 82,* 224–237.

Földiak, P. (1991). Learning invariance from transformation sequences. *Neural Computation, 3,* 194–200.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36,* 193–202.

Gibson, W. G., Robinson, J., & Bennett, M. R. (1991). Probabilistic secretion of quanta in the central nervous system: Granule cell synaptic control of pattern separation and activity regulation. *Philosophical Transactions of the Royal Society of London Series B, 332,* 199–220.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15,* 20–25.

Harries, M. H., & Perrett, D. I. (1991). Visual processing of faces in temporal cortex: Physiological evidence for a modular organization and possible anatomical correlates. *Journal of Cognitive Neuroscience, 3,* 9–24.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation.* New York: Addison-Wesley.

Hinton, G. E., & Lang, K. J. (1985). Shape recognition and illusory conjunctions. In *Proceedings of the Ninth Annual Joint Conference on Artificial Intelligence* (pp. 252–259). Los Altos, CA: Morgan-Kaufmann.

Höffding, H. (1891). *Outlines of psychology.* New York: MacMillan.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, 79,* 2554–2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences, 81,* 3088–3092.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480–517.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13,* 289–303.

Kienker, P. K., Sejnowski, T. J., Hinton, G. E., & Schumacher, L. E. (1986). Separating figure from ground with a parallel network. *Perception, 15,* 197–216.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220,* 671–680.

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review, 94,* 148–175.

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience, 15,* 1605–1615.

Lamme, V. A. F., Van Dijk, B. W., & Spekreijse, H. (1992). Texture segregation is processed by primary visual cortex in man and monkey. Evidence from VEP experiments. *Vision Research, 32,* 797–807.

Lamme, V. A. F., Van Dijk, B. W., & Spekreijse, H. (1993). Organization of texture segregation processing in primate visual cortex. *Visual Neuroscience, 10,* 781–790.

Lawry, J. A., & LaBerge, D. (1981). Letter and word code interactions elicited by normally displayed words. *Perception & Psychophysics, 30,* 71–82.

Livingstone, M., & Hubel, D. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience, 7,* 3416–3468.

Livingstone, M., & Hubel, D. (1988). Separation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240,* 740–749.

Lowe, D. G. (1985). *Perceptual organization and visual recognition.* Boston: Kluwer.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence, 31,* 355–395.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Marr, D., & Poggio, T. (1976). Co-operative computation of stereo disparity. *Science, 194,* 283–287.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology, 23,* 1–44.

McClelland, J. L. (1993). Toward a theory of information processing in graded, random, and interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV* (pp. 655–688). Cambridge, MA: MIT Press.

McClelland, J. L., & Elman, J. L. (1986). Interactive processes in speech perception: The TRACE model. *Cognitive Psychology, 18,* 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

McNaughton, B. L., & Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neurosciences, 10,* 408–415.

Movellan, J. R., & McClelland, J. L. (1993). Learning continuous probability distributions with symmetric diffusion networks. *Cognitive Science, 17,* 463–496.

Mozer, M. C. (1987). Early parallel processing in reading: A connectionist approach. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 83–104). Hove, UK: Erlbaum.

Mozer, M. C. (1991). *The perception of multiple objects: A connectionist approach.* Cambridge, MA: MIT Press.

Mozer, M. C., Zemel, R. S., & Behrmann, M. (1992). Discovering and using perceptual grouping principles in visual information processing. In *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society* (pp. 283–288). Hillsdale, NJ: Erlbaum.

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

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

O'Reilly, R. C. (1995, April). *Combined error-driven and associative learning as a model of neocortical learning.* Paper presented at the second annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.

O'Reilly, R. C., & Johnson, M. H. (1994). Object recognition and sensitive periods: A computational analysis of visual imprinting. *Neural Computation, 6,* 357–389.

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a tradeoff. *Hippocampus, 4,* 661–682.

Palmer, S. E., & Rock, I. (1994a). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review, 1,* 29–55.

Palmer, S. E., & Rock, I. (1994b). On the nature and order of organizational processing: A reply to Peterson. *Psychonomic Bulletin & Review, 1,* 515–519.

Peterson, M. A. (1994). Object recognition processes can and do operate before figure-ground organization. *Current Directions in Psychological Science, 3,* 105–111.

Peterson, M. A., & Gibson, B. S. (1991). The initial identification of figure-ground relationships: Contributions from shape recognition processes. *Bulletin of the Psychonomic Society, 29,* 199–202.

Peterson, M. A., & Gibson, B. S. (1993). Shape recognition inputs to figure-ground organization in three-dimensional displays. *Cognitive Psychology, 25,* 383–429.

Peterson, M. A., & Gibson, B. S. (1994a). Must figure-ground organization precede object recognition? An assumption in peril. *Psychological Science, 5,* 253–259.

Peterson, M. A., & Gibson, B. S. (1994b). Object recognition contributions to figure-ground organization: Operations on outlines and subjective contours. *Perception & Psychophysics, 56,* 551–564.

Peterson, M. A., Harvey, E. M., & Weidenbacher, H. (1991). Shape recognition contributions to figure-ground organization: Which routes count? *Journal of Experimental Psychology: Human Perception and Performance, 17,* 1075–1089.

Phaf, R. H., Van der Heijden, A. H. C., & Hudson, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22,* 273–341.

Pinker, S. (1984). Visual cognition: An introduction. In S. Pinker (Ed.), *Visual cognition* (pp. 1–63). Cambridge, MA: MIT Press.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343,* 263–266.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 36.1–36.46). New York: Wiley.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81,* 275–280.

Rock, I. (1962). A neglected aspect of the problem of recall: The Höffding function. In J. M. Scher (Ed.), Theories of the mind (pp. 645–659). New York: Free Press.

Rock, I. (1975). *An introduction to perception.* New York: MacMillan.

Rubin, E. (1958). Figure and ground. In D. C. Beardslee & M. Wertheimer (Eds.), *Readings in perception* (pp. 194–203). Princeton, NJ: Van Nostrand. (Original work published 1915)

Rumelhart, D. E. (1989). The architecture of mind: A connectionist approach. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 133–159). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations.* Cambridge, MA: MIT Press.

Sajda, P., & Finkel, L. H. (1995). Intermediate visual representations and the construction of surface perception. *Journal of Cognitive Neuroscience, 7,* 267–291.

Sejnowski, T. J., & Hinton, G. E. (1987). Separating figure from ground with a Boltzmann machine. In M. A. Arbib & A. R. Hanson (Eds.), *Vision, brain, and cooperative computation* (pp. 703–724). Cambridge, MA: MIT Press.

Selfridge, O. G., & Neisser, U. (1960). Pattern recognition by machine. *Scientific American, 203,* 60–68.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21,* 233–282.

Torioka, T. (1979). Pattern separability in a random neural net with inhibitory connections. *Biological Cybernetics, 34,* 53–62.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–585). Cambridge, MA: MIT Press.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32,* 193–254.

Vecera, S. P. (1993). Object knowledge influences visual image segmentation. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 1040–1045). Hillsdale, NJ: Erlbaum.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224,* 1260–1262.

Waltz, D. (1975). Understanding line drawings of scenes with shadows. In P. H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1,* 59–85.

Appendix A

GRAIN Equations and Parameters

Updating Algorithm

Following standard algorithms (Hopfield, 1984; McClelland, 1993), the units in our GRAIN model (Simulations 1–3) computed their net input from other units, and their activation was a sigmoidal function of the net input. Specifically, the net input to unit j was given by

$$\eta_j = \sum_i w_{ij} a_i + \theta_j + I_j, \tag{1}$$

where η_j is the net input to unit j, w_{ij} is the connection between unit j and unit i, a_i is the activation of unit i, θ_j is the bias (or threshold) of unit j, and I_j is the external input to the unit. This net input is passed through a sigmoidal activation function:

$$y_j = \frac{1}{1 + e^{-\gamma \eta_j}}, \tag{2}$$

where γ is gain (or sharpness of the sigmoid) and η_j is the net input to unit j as given in Equation 1. Finally, the activations of the units in the network were updated by moving the current activation toward the value of y_j. That is, the activation was changed gradually between the activation at time t and the new activation at time $t + 1$:

$$a_j(t + 1) = \epsilon(y_j - a_j(t)) + \nu(0, \sigma), \tag{3}$$

where $a_j(t + 1)$ is the activation of unit j at time $t + 1$, ϵ is the activation step size (or settling rate) parameter, $a_j(t)$ is the activation of unit j at time t, and $\nu(0, \sigma)$ is random, Gaussian noise with a mean of zero and a standard deviation of σ.

Parameters

The annealing schedule started the noise in the network at 10.0 and reduced the noise to 8.0 from Cycles 0–100; from Cycles 100–300 the noise was reduced from 8.0 to 4.0; and from Cycles 300–400 the noise was reduced from 4.0 to 1.0. The sharpening schedule, which changed the gain of the sigmoidal activation function, followed a progression similar to that of the annealing schedule. The sharpening schedule began at 1.0 and increased to 1.1 from Cycles 0–100; from Cycles 100–300 the sharpening increased from 1.1 to 1.5; and from Cycles 300–400 the sharpening was increased from 1.5 to 2.0. As the gain increases, the sharpness of the sigmoid also increases. This, in effect, makes the units more binary; that is, with a higher sharpening value the units' activations are more likely to be either 0.0 or 1.0 instead of intermediate values. The standard gain on the sigmoid was 0.50; this value was multiplied by the value of the sharpening schedule.

The units in the various layers each had different biases, or thresholds. The boundary units had biases of −45. The figure units had biases of −36 in Simulations 1 and 2 and biases of −38 in Simulation 3. The object units in Simulations 1 and 2 had biases of −250. The two object units in Simulation 3 had biases of −90, and the intermediate object units had biases of −250.

The external input to the boundary units was +60. The step size of the update function (ϵ in Equation 3) was 0.25.

Before cycling began on any given trial, the initial activations of the units were randomized with Gaussian noise. This noise had a mean value of 0.25 and a standard deviation of 0.25.

Finally, the values of the weights were as follows (see also Figures 2 and 3). The inhibitory weights among the opposing boundary units (e.g., between a boundary-left unit and a boundary-right unit) were −15. The boundary units projected to the figure units with excitatory weights of +12 and +10; the +12 weights corresponded to a boundary unit and a figure unit that were in the same retinal location; the +10 weights corresponded to the adjacent two units (see Figure 2c). The inhibitory weights among the boundary units and the figure units were similar to the excitatory connections: −12 and −10 for units that shared the same retinal position and adjacent units, respectively. Corners in the image were given by projections among the boundary units, as shown in Figure 3. These weights were +5 for relationships that were consistent with a corner and −5 for relationships that were inconsistent with a corner. Finally, the figure units projected to the object units with weights of +10. In Simulation 3 the intermediate object units projected to the object units with weights of +100. Note that all connections were reciprocal and that these reciprocal connections had the same weight values as the feed-forward connections.

Appendix B

LEABRA Equations and Parameters

Updating Algorithm

The particular implementation of the k-WTA algorithm used in LEABRA (Simulations 4 and 5) is based on the notion of a Relative Belief function (ReBel), which is derived in the context of Bayesian hypothesis testing. Units represent hypotheses that are evaluated by the extent to which they are supported by the current input pattern. The measure of relative belief is the product of two terms, one of which reflects the absolute level of support for a given hypothesis and the other of which measures its level of support relative to the other units in a layer. The absolute term is given by the likelihood function, $P(x_p|h_j)$, which is defined as a function of the goodness-of-fit between the input pattern x_p and the weights into unit h_j. We used a standard sigmoidal function of the net input as a measure of this fit, where η_j is the net input as defined previously in Equation 1:

$$P(x_p|h_j) = \frac{1}{1 + e^{-\gamma\eta_j}}. \tag{4}$$

Both the gain of this sigmoid γ and the offset or threshold θ_j are allowed to "float" to keep the resulting probability measure from becoming saturated at the tails of the sigmoid, which would reduce the sensitivity of the network to differences in the probability measure of the units. The offset θ_j is defined as the average net input of the units in the layer, and the gain γ is set so that the difference between the maximum net input and the average net input, when scaled by γ, is 5.0, which results in a sigmoidal activation value that is just at the saturation point.

The relative term is defined with respect to a null hypothesis, h_q, which reflects the level of support for other units in the layer. To get k-WTA behavior, h_q is defined as in between the probability of the kth and $(k + 1)$th most probable hypotheses in the layer, so that exactly k hypotheses will be more likely than the null hypothesis:

$$P(h_q) \equiv P(h_{k+1}) + q[P(h_k) - P(h_{k+1})]. \tag{5}$$

The parameter q determines where the null hypothesis lies between these two values. This is set to 0.25 in all simulations. To measure how probable a given unit is relative to this null hypothesis, the posterior probability of a given unit given the current input pattern is evaluated in the context of where the unit's hypothesis and the null hypothesis are considered to be mutually exclusive and exhaustive. This is denoted as $P_q(h_j|\, x_p)$. Thus, the ReBel function is:

$$\text{ReBel}(h_j, x_p, h_q) \equiv P(x_p|h_j)P_q(h_j|x_p). \tag{6}$$

Because h_j and h_q are considered to be mutually exclusive and exhaustive hypotheses in the relative term, Equation 6 can be manipulated using standard Bayesian techniques, resulting in the following:

$$\text{ReBel}(h_j, x_p, h_q) = P(x_p|h_j)\frac{1}{1 + \left(\dfrac{P(x_p\ h_j)P(h_j)}{P(x_p|h_q)P(h_q)}\right)^{-\gamma_r}}, \tag{7}$$

where the gain term in this equation, γ_r, is the gain of the relative probability term.[B1] This expression can be computed directly from the likelihood function $P(x_p|h_j)$ of the units in a layer.

Because the ReBel function has an upper limit set by the relative probability term in Equation 7, units that might have strong absolute support but are nevertheless below the null hypothesis have a low relative probability in this equation. To make the units more sensitive to the absolute levels of support, we defined the actual activations of units in the network to be a weighted combination of the ReBel function as defined in Equation 7 and their absolute probability as given by Equation 4. This has the effect of "softening" the strict k-WTA constraint of ReBel, and results in more robust settling performance:

$$\alpha_j \equiv \rho\ \text{ReBel}(h_j) + (1 - \rho)P(x_p|h_j) + \nu(0, \sigma), \tag{8}$$

where ρ is a parameter that determines the relative strength of the ReBel k-WTA constraint. A value of 0.8 was used in these simulations.

Settling in the network occurs by updating the prior probability term $P(h_j)$ with the current value of Equation 8. A certain number

[B1] This gain term arises because the Bayesian manipulation of the relative probability term ends up being equivalent to the application of a logistic function to the log odds ratio of the unit's posterior probability over that of the null hypothesis. The gain is thus the gain of this logistic function.

of cycles (50 cycles in the present simulations) are processed before updating begins in order to allow information to first pass throughout the network. Finally Gaussian random noise was added to the activations, as is shown by the $\nu(0, \sigma)$ term in Equation 8. The standard deviation of this noise was adapted according to an annealing schedule similar to that used in the GRAIN simulations.

Parameters

The LEABRA model in Simulations 4 and 5 had the following values for the parameters. The noise schedule started the noise in the network at 7.0 and reduced the noise to 4.0 from Cycles 0–50; from Cycles 50–150 the noise was reduced from 4.0 to 2.0; and from Cycles 150–200 the noise was reduced from 2.0 to 0.5.

The units in the various layers each had different offsets, or thresholds. The boundary units and figure units had offsets of +35; the intermediate object units in had offsets of +225; the two object units had offsets of +95.

The external input (stimulus gain) to the boundary units was +36. The prior delay, which was the number of cycles that progressed before units updated their activations, was 50.

The number of units active in each layer varied depending on the layer. The boundary-above and boundary-below layer were permitted to have 5 units active; the boundary-left and boundary-right layers had 12 units active; the figure layer and disparity layer were each allowed to have 48 units active; and the two object layers were each allowed to have 1 unit active.

Before cycling began on any given trial, the initial activations of the units were randomized with Gaussian noise. This noise had a mean value of 0.25 and a standard deviation of 0.1.

Finally, the values of the weights were as in the GRAIN model, with the following exceptions. First, there were no self connections on the two object layers; the *k*-WTA nature of the LEABRA algorithm incorporated these connections. Second, the connections from the figure units to the intermediate object units were +9. Third, the connections from the disparity units to the figure units were +5. As before, all connections were reciprocal, and these reciprocal connections had the same weight values as the feedforward connections.

Received November 29, 1995
Revision received December 10, 1996
Accepted January 22, 1997 ■