OBSERVATIONS

Graded Effects in Hierarchical Figure–Ground Organization: Reply to Peterson (1999)

Shaun P. Vecera
University of Iowa

Randall C. O'Reilly
University of Colorado, Boulder

An important issue in vision research concerns the order of visual processing. S. P. Vecera and R. C. O'Reilly (1998) presented an interactive, hierarchical model that placed figure–ground segregation prior to object recognition. M. A. Peterson (1999) critiqued this model, arguing that because it used ambiguous stimulus displays, figure–ground processing did not precede object processing. In the current article, the authors respond to Peterson's (1999) interpretation of ambiguity in the model and her interpretation of what it means for figure–ground processing to come before object recognition. The authors argue that complete stimulus ambiguity is not critical to the model and that figure–ground precedes object recognition *architecturally* in the model. The arguments are supported with additional simulation results and an experiment, demonstrating that top–down inputs can influence figure–ground organization in displays that contain stimulus cues.

According to many theories of visual perception, prior to object identification the visual system must solve the problem of determining which retinal locations belong together or bind with one another. Grouping locations within the retinal array allows the visual system to determine provisionally which locations are occupied by a single shape or object. To solve the problem of grouping or binding retinal locations from the same shape, the visual system relies on a number of organizational processes. These processes include the gestalt principles of organization (e.g., Wertheimer, 1923/1958) and figure–ground organization (e.g., Rubin, 1915/1958). However, not all theories of visual perception accept that organization precedes recognition. Specifically, Peterson (see Peterson, 1994, for a summary) proposed that the relationship between organization and recognition should be reversed from the traditional accounts to allow some recognition processes to occur prior to perceptual organization. The focus of our current exchange with Peterson (1999) is on figure–ground processes and how these processes relate to object-identification processes. Despite Peterson's (1999) criticisms of our interactive model (Vecera & O'Reilly, 1998; also see Vecera & Farah, 1997), we still contend that the interactive model provides the best computationally ex-

plicit account of how object representations can influence figure–ground segregation. In this reply, we support our position by clarifying our position, by discussing some definitional issues, and by presenting additional simulation results from our interactive model.

Figure–ground organization involves assigning one region in a visual display as the "figure," which lies in the foreground and eventually should be recognized, and another region as the "ground," which lies in the background and should not be recognized. As we have noted, most theoretical accounts of visual processing have placed figure–ground organization, as well as other modes of perceptual organization, prior to object identification in the visual processing stream (e.g., Biederman, 1987; Kosslyn, 1987; Marr, 1982; Neisser, 1967; Palmer & Rock, 1994a, 1994b; Vecera & Farah, 1997; Vecera & O'Reilly, 1998; Wallach, 1949). This placement is based on logical grounds: Presumably the retinal array must be organized before a perceiver can recognize the objects that appear in this array. However, despite this apparent logical requirement for figure–ground to precede object identification, some empirical results from Peterson and colleagues appear to be inconsistent with models that place figure–ground processes before recognition processes (see Peterson, 1994, for a summary; see Biederman, 1987; Kosslyn, 1987; Marr, 1982, for models that have difficulties explaining Peterson's data). Specifically, Peterson and colleagues reported that perceivers are more likely to see familiar regions as figure compared with less familiar regions, suggesting that object-level information is somehow influencing figure–ground processing (see Peterson, 1994, for a summary, and see Rock, 1975, for an earlier discussion of this finding). At issue are the architecture and processing dynamics of intermediate and high-level vision. The critical question, and the question at the heart of the present exchange, is *how* object-level information influences figure–ground organization.

Shaun P. Vecera, Department of Psychology, University of Iowa; Randall C. O'Reilly, Department of Psychology, University of Colorado, Boulder.

Correspondence concerning this article should be addressed to Shaun P. Vecera, Department of Psychology, 11 Seashore Hall E., University of Iowa, Iowa City, Iowa 52242-1407. Electronic mail can be sent to shaun-vecera@uiowa.edu.

Peterson's (1994; Peterson & Gibson, 1993, 1994) approach as to how object-level information influences figure–ground segregation was to reject the logical requirement that figure–ground precedes object identification. Under Peterson's approach, object representations initially receive input from luminance contours or edges, not from figure–ground processes (as would be required under other models; e.g., Biederman, 1987; Kosslyn, 1987; Marr, 1982). The input from these edge-based processes allows objects to be recognized before any figure–ground processing has occurred; thus, these object-recognition processes are termed "prefigural" recognition processes (Peterson, 1994). The outputs of the recognition process can then influence figure–ground organization, allowing more familiar, or high-denotative, regions to have a higher probability of being labeled as figure. We contend that Peterson's (1994, 1999) account is a feedforward model of visual processing in which (a) edges are extracted from the visual image, (b) prefigural object recognition occurs based on edge information, and (c) figure–ground organization is influenced by the outputs of the prefigural recognition processes. This position is corroborated by the presence of Pathway D in Figure 5 from Peterson's (1999) commentary (equivalent to the luminance contour pathway depicted in Figure 8a in Vecera & O'Reilly, 1998), which illustrates the prefigural recognition process.

Instead of abandoning the logical requirement of figure–ground preceding object identification, we developed an interactive account of figure–ground organization (Vecera & O'Reilly, 1998) in an attempt to reconcile the traditional hierarchical approaches with Peterson's empirical results (e.g., Peterson & Gibson, 1991, 1993, 1994; Peterson, Harvey, & Weidenbacher, 1991). In our model, top–down feedback from object representations partially guides figure–ground processes (Vecera & O'Reilly, 1998; also see Vecera & Farah, 1997), and figure–ground processes are influenced simultaneously by two forces, bottom–up stimulus cues contained in the image and top–down, feedback cues from visual object representations stored in visual memory. Constraint satisfaction processes that emerge from interactive models allow both sets of cues to partially guide the network in determining what region is figure. The top–down constraints allow the model to label as figure the familiar regions of figure–ground displays (Simulations 1–4 in Vecera & O'Reilly, 1998), and these top–down constraints can even override bottom–up, stimulus cues such as disparity (Simulation 5, Vecera & O'Reilly, 1998). In this model, object recognition does not need to occur fully prior to figure–ground segregation (i.e., there is no prefigural recognition based on edge or contour information).

In her commentary, Peterson (1999) critiqued our interactive approach on three fronts. First, she correctly pointed out that in our previous simulations we used only figure–ground displays that contained no low-level cues as to what is figure (e.g., cues such as size or symmetry; see Pomerantz & Kubovy, 1986, for a discussion of these cues). Peterson suggested that the interactive model requires this stimulus ambiguity in order for top–down inputs to influence figure–ground organization (see Vecera & O'Reilly, 1998, p. 453). She argued that this "ambiguity assumption" will result in the failure of the interactive model to explain the behavioral data in which *unambiguous* (i.e., biased) stimulus displays are used as stimuli. Second, Peterson argued that because of the ambiguous displays we used, our model does not and cannot perform figure–ground organization prior to object recognition.

Thus, the interactive model is not a hierarchical model that places figure–ground prior to object recognition, so our claims against prefigural object recognition are invalid. Third, Peterson acknowledged that we attempted to use unambiguous displays in Simulation 5 of our original article. However, she argued that because we used separate pathways in that simulation (one pathway for luminance cues and one for disparity cues), the pathway that provides the first input to figure–ground organization actually does not allow figure–ground processing to occur because of the ambiguous displays. Thus, the model does not demonstrate object familiarity cues overriding unambiguous stimulus input.

In this reply, we address each of these three points in turn and argue that they are not problematic for our interactive approach. First, we discuss our use of the ambiguous stimulus displays used in our original simulations. We clarify our meaning of "ambiguity" and the role of ambiguous stimulus displays in our original modeling efforts. Second, we discuss what it means for figure–ground organization to precede object-recognition processes. We challenge the sequential, discrete, noninteractive framework, embodied in traditional models of visual processing (e.g., Marr, 1982; Neisser, 1967; Wallach, 1949) for thinking about figure–ground processing and visual processing more generally. We argue that in an interactive, graded framework, the notion of which process comes first becomes an *architectural* issue, not a processing issue. Thus, it simply does not make sense to ask whether figure–ground *processing* precedes object processing or vice versa—both can occur simultaneously and in constant interaction. Third, we discuss the role of multiple cues and multiple pathways in our earlier model. We acknowledge that the use of multiple pathways complicates the interpretation of the model, but we disagree with Peterson's analysis of our model. On the basis of our discussion of ambiguity, we contend that our model can process unambiguous (biased) displays without needing separate pathways.

Following our reply to Peterson's three critiques, we corroborate our arguments with simulation results. We demonstrate that our interactive model exhibits top–down influences in *unambiguous* stimulus displays that are biased against top–down influences. These simulation results nullify any discussion of our model being unable to process unambiguous (biased) stimulus displays. These results also demonstrate that, contrary to Peterson's (1999) claims, not all stimulus cues to figure–ground organization need be carried by separate pathways in our model.

The Ambiguity Assumption

Peterson's (1999) Discussion of Ambiguity

In our original simulations, we used figure–ground displays that contained two regions. The two regions were identical to one another on all low-level, stimulus cues to figure–ground organization (i.e., they were identical on size, convexity, symmetry, etc.). Thus, the displays were ambiguous in terms of their bottom–up, stimulus information. However, one of the regions was familiar to the network (the high-denotative region), and the other region was unfamiliar to the network (the low-denotative region). We noted that *some degree* of stimulus ambiguity was required to support partially the top–down cues (i.e., the feedback from object representations).

Peterson (1999) discussed the potential consequences of our

model requiring ambiguous stimuli. In her comment she stated that "a fundamental assumption of [Vecera and O'Reilly's, 1998] ... model is that top-down influences ... can influence figure-ground segregation *only* when bottom-up influences ... are ambiguous" (p. 277, emphasis added). On Peterson's (1999) interpretation of our use of ambiguous displays, the problem that arises for our interactive model is that there are no bottom-up stimulus cues for the network to rely on in determining which region should be figure. That is, the network's figure layer cannot determine which of the two regions should be figure. As a consequence, neither region is called figure, and both regions of the figure-ground display simply are passed on to the object representations for processing (see Peterson, 1999, p. 279, for that analysis). The familiar region (i.e., the region that had an object representation in the model's object layers) in the display will match one of the object representations, which will then become activated and send top-down feedback to the figure layer, allowing the network to exhibit a bias to call the familiar, high-denotative region figure. Under this interpretation of our model, figure-ground processing does *not* occur prior to object-recognition processes because figure-ground organization is attempted but fails. Therefore, according to Peterson (1999), our model violates the "figure-ground-first" assumption that would be required to successfully argue against Peterson's (1994) prefigural account of figure-ground segregation in favor of a hierarchical account.

A consequence of the ambiguity assumption according to Peterson (1999) is that our interactive model would be unable to explain some of the behavioral data from Peterson and colleagues (e.g., Peterson & Gibson, 1993, 1994). Peterson (1999) noted that she has not always used ambiguous displays in her experiments. For example, some of the displays used contained binocular disparity cues in which one region in the figure-ground display appeared closer to the perceiver than the other region (Peterson & Gibson, 1993); these disparity cues could be either consistent with the familiar, high-denotative region, in which the high-denotative region appeared closer to the viewer, or inconsistent with the high-denotative region, in which the high-denotative region appeared further away from the viewer. Other studies (Peterson & Gibson, 1994) have contained symmetry cues that were either consistent with or inconsistent with the high-denotative regions.

The results from studies using those biased, unambiguous displays (e.g., Peterson & Gibson, 1993, 1994) are that the outputs from object-representation processes can still compete with the unambiguous stimulus input. For example, asymmetric, upright high-denotative regions are labeled as figure more often than asymmetric, rotated high-denotative regions (see Figure 3 in Peterson, 1999, for a sample display; also see Peterson & Gibson, 1994, for relevant results). Peterson's critique of our model suggests that our model would fail to simulate such results. The unambiguous display would cause the figure layer to label the symmetric, low-denotative region as figure. The object units would not receive any input from the asymmetric, high-denotative region, and the network would be unable to use top-down inputs to bias the network to prefer the high-denotative region as figure. Therefore, our model, if tested with unambiguous figure-ground displays, would not demonstrate any top-down influences.

Reply

We see an important interpretational difference between "ambiguity" in Peterson's (1999) commentary and our interactive view. This difference surrounds the definition of *ambiguity* in our model and how critical ambiguity is for the operation of our model. The section in our original article in which we explicitly discussed the role of ambiguity is the introduction to Simulation 5 (Vecera & O'Reilly, 1998, pp. 452–453). Unfortunately, we fear that our original discussion of ambiguity has been overstated as a critical factor in our model's operation. Although we note that ambiguity in the stimulus display is required in our model, we clearly state that "*some degree* of ambiguity in the bottom-up input is required to allow the top-down influences to affect processing" (Vecera & O'Reilly, 1998, p. 453, emphasis added). Our intent was to convey the fact that an *entirely* unambiguous input would bias the model so strongly that top-down inputs could not overcome the bottom-up biases (e.g., binocular disparity, in our simulation of Peterson & Gibson's, 1993, results). All "*ambiguous*" means in the context of our model is that there must be two regions in the figure-ground display; that is, there must be more than one figure-ground interpretation within the display. Furthermore, the two regions do not need to be identical to one another on stimulus-level cues to figure-ground organization.[1] Nor does one region need to be familiar in order for the display to be "ambiguous."

For purposes of clarification, we present three different figure-ground displays in Figure 1. The first of these displays (Figure 1, Panel A) is entirely unambiguous on our meaning of the term. Only a single region appears, and neither our network nor, we predict, a human observer could perceive the white region as figure. There is simply no stimulus-level evidence to support the percept of the white region as figure. Without additional cues to support the white region as figure, the strong bottom-up cues cannot be overcome. The display in Panel A of Figure 1 can be contrasted with the displays in Panels B and C of Figure 1. In both of these displays there are two regions, a black region and a gray region. Both of these regions are ambiguous in that the stimulus supports either region being held as figure. However, the two displays support figural interpretations differently; thus, the displays in Panels B and C of Figure 1 are not equally ambiguous. Specifically, the display in Panel C is less ambiguous than the display in Panel B because of the presence of area cues. All other cues being equal, the smaller, gray region should be more likely to be perceived as figure. That the larger, black region is present in Panel C means that this region in the display would help support this region as figure in those cases where the area bias was overcome. Again, we reiterate for clarification: All we mean "ambiguous" to imply is that there could be more than one figure-ground interpretation in a display and that some stimulus-level evidence exists for both of these interpretations.

---

[1] We also should note that in our original article, we discussed other ways in which top-down influences could exert their influence, such as through experimental instructions (e.g., "pay attention to the left side of the display") or participants' strategies (Vecera & O'Reilly, 1998, p. 453). Stimulus ambiguity is *not* the only mechanism whereby top-down feedback can influence earlier processing on our account of figure-ground organization.
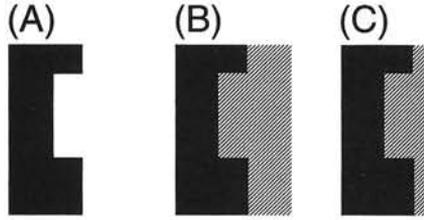
Figure 1. Elaboration of ambiguity in figure–ground displays. (A) An unambiguous (biased) display in which there is no image-level evidence for the white region to be figure. In such a display both our network and, we predict, human participants would have difficulty holding anything but the unambiguous black region as figure. (B) An ambiguous stimulus display that contains evidence for each of the two regions; that is, both regions are bounded and present in the image. Either region could be held as figure. (C) An ambiguous stimulus display in which stimulus-level cues (namely size) favors one region to be held as figure. Although this stimulus display is not entirely ambiguous (i.e., the two regions are not equal in size, convexity, symmetry, etc.), both regions are bounded and present in the image, thereby providing evidence for each region.

We contend that the results from Simulation 5 in our original article (Vecera & O'Reilly, 1998) demonstrate that top–down cues do, in fact, operate in the presence of unambiguous input. If completely ambiguous stimuli were required, as indicated by Peterson's (1999) discussion of the ambiguity assumption, then our model should not have allowed top–down object cues to compete with unambiguous, bottom–up disparity cues. However, Peterson's (1999) commentary indicates that she would disagree with our interpretation of Simulation 5. The point of contention that remains is the issue of figure–ground organization preceding object-recognition processes. In the interactive model the figure–ground layer occurs prior to object-representation layers; specifically, the figure layer is closer to the visual input than the object layers. But, Peterson challenged whether figure–ground actually occurs "first" in our model because figure–ground organization fails as a result of the ambiguous stimulus displays. We now turn to a discussion of what it means for figure–ground processes to come "first."

Figure–Ground-First

The Figure–Ground-First Critique

The figure–ground-first issue is important and relevant to the present exchange because this issue focuses on the interrelationships between figure–ground organization and object-recognition processes. Under most theoretical accounts of visual processing, figure–ground-first would simply mean that figure–ground organization precedes object recognition both hierarchically and temporally (see Biederman, 1987; Kosslyn, 1987; Marr, 1982; Neisser, 1967; Wallach, 1949, for examples). Peterson's (1999) commentary also is consistent with such an interpretation of figure–ground-first: For figure–ground organization to precede object recognition, the figure must be labeled on the "first pass" through the system (see Peterson, 1999, pp. 282–283). The figure–ground-first assumption can be contrasted with an "object-recognition-first" assumption, which characterizes Peterson's (1994) prefigural account. Under an object-recognition-first ac-

count, object-recognition processes would operate prior to figure–ground processes. At the heart of this issue lies the architecture of the visual system and the placement of these two visual processes. As we argue in the next section, there is another conception of figure–ground-first that applies to our model, which provides a more accurate description of the processing dynamics exhibited by the model.

Peterson (1999) stated that "figure–ground segregation entails: For two adjacent regions sharing a contour, the contour is assigned to one region (the figure), but not the other" (p. 283). Peterson questioned whether figure–ground organization actually precedes object representation in the interactive model. She noted that because our stimulus displays were ambiguous (i.e., contained no bottom–up, stimulus cues as to what is figure), the figure layer in our network could not possibly assign one region as figure on the first pass of activation through that layer (see Peterson, 1999, p. 283). That is, because the stimulus display was ambiguous, figure–ground organization was attempted but failed—neither region in the display could be labeled as figure. Because the figure layer could not organize this display, processing was deferred to the object layer (see Peterson, 1999, p. 279). Peterson concluded that because figure–ground organization fails as a result of ambiguous displays, there is no evidence to say that the figure layer is performing figure–ground organization (see Peterson, 1999, p. 283). Instead, this layer is simply passing the ambiguous information to the object layers. The object layers recognize the high-denotative region in the display (via a mechanism not specified by Peterson, 1999), thus causing the bias to label the high-denotative region as figure.

In Peterson's (1999) commentary, the critique of our ambiguous displays and the figure–ground-first critique are highly interwoven with one another. On the one hand, if figure–ground comes first in the interactive model, then unambiguous (biased) stimulus displays would abolish the top–down influence, contradicting Peterson's behavioral results (e.g., Peterson & Gibson, 1993, 1994). On the other hand, because figure–ground organization is prevented in the ambiguous stimuli, according to Peterson (1999) the network's figure layer is not performing figure–ground organization. Consequently, the interactive model is doomed either way: Either the model does not simulate the data from unambiguous displays or it is not a true hierarchical model that places figure–ground first.

Reply

We maintain our position that figure–ground organization precedes object-recognition processes in the interactive model. Our primary disagreement with both the traditional approaches to visual processing and Peterson's (1999) discussion of the figure–ground-first assumption centers on the nature of processing. We do not assume that figure–ground organization entails *fully* describing one region as figure and the other as ground. Further, we contend that such an "all-or-none" description of figure–ground processing might actually hinder the understanding of figure–ground mechanisms.

Peterson (1999) discussed figure–ground organization as an all-or-none process that obscures the processing dynamics in our model (see Peterson, 1999, p. 283, for an illustration of this all-or-none characterization). Although the *final output* of figure–ground processes is a fully defined figural region that "owns" the

central contour, we do not assume that figure–ground processing involves only the end state of the settling process. Instead, figure–ground segregation is a graded, cascaded process that begins at the initial activation of the figure units. When the figure layer receives ambiguous input from the boundary layers, the figure layer does not fail but rather uses the limited information at hand to begin the figure–ground settling process. Processing continues in both the figure and object layers simultaneously and interactively, with both the figure and object layers constraining one another. Figure–ground organization is a graded process, and any activation across any of the figure units constitutes figure–ground-organization processes. We would argue that figure–ground organization is not completed on a single time step as Peterson (1999) seems to assume. In our interactive model, the figure would not gain its "definite shape" on the first pass through the figure units, contrary to Peterson's (1999, p. 283) interpretation of figure–ground organization.

It is important to emphasize that in our model, figure–ground processing always takes many cycles of processing and is not, in principle, dependent on object-based input. For example, if the object layer were removed and a fully ambiguous input were presented, the network would still settle on one of the two figure–ground interpretations, with the choice being governed by intrinsic processing noise.

Also, we should be clear that the model does not need to try a particular figure–ground organization, send partial results to the object units, and then wait for top–down feedback based on this initial segregation of the display (see Peterson, 1999, pp. 279–280). It is simply not the case that the model must first attempt to fully fill one of the regions of the display as figure, prior to any top–down inputs. Top–down inputs from the object units are omnipresent because these units are partially activated by the random noise injected into all of the units' activations. Partial activation of the object units from this random noise will allow these units to exert some top–down influence on the figure units from the beginning of a trial (although this initial top–down influence may not be highly reliable). Thus, for example, in Simulation 5 from our earlier article (Vecera & O'Reilly, 1998), top–down inputs are directly competing with and influencing unambiguous (biased) stimulus input as the network settles.

In short, the question "what is meant by figure–ground-first?" is simply not a very useful question to ask within a graded, interactive framework. We regard this question as a hold-over from traditional, sequential models—one that causes more confusion than it resolves, especially in light of Peterson's behavioral data (e.g., Peterson & Gibson, 1991, 1993, 1994; Peterson et al., 1991). Furthermore, in trying to provide a definite answer to this question in the form of a "prefigural" object-recognition process, Peterson (1994, 1999) potentially perpetuated this confusion instead of embracing what we hear her behavioral results to be so clearly saying: Figure–ground processing and object recognition are intimately interrelated and should be viewed as simultaneously interacting, graded processes instead of sequentially ordered, discrete ones. Although Peterson endorsed simultaneous processing of figure–ground organization and object recognition (e.g., Peterson, 1994, pp. 110–111), she also indicated that interactive parallel distributed processing models are insufficient to explain her behavioral data (Peterson & Gibson, 1994, p. 258).

Having said that, there is still one very limited but nonetheless important sense in which the question "what is meant by figure–ground-first?" can be answered within a graded, interactive model. This is in the *architectural* sense, where the figure–ground processing layer is located first before object-recognition processes in the flow of information coming from the retina. Although this architectural distinction clearly has some influence on the overall nature of visual processing, it does not imply that figure–ground processing necessarily precedes object recognition temporally because interactivity implies that the architectural and temporal aspects of cognition can be dissociated.

Finally, we fully acknowledge that the behavioral predictions of our model and Peterson's (1994, 1999) model may be very similar. Nevertheless, we believe that our model provides a more parsimonious and coherent account of the behavioral phenomena, and one that is consistent with established principles of neural computation. We feel that Peterson's model (see Figure 5 of Peterson, 1999) leaves many questions of central importance unanswered. For example, how does the object-recognition system manage to process two very different kinds of inputs (e.g., edge-based inputs and figure–ground inputs), and how does it balance between them? How are the top–down inputs to the figure–ground system from the object-recognition system integrated with the bottom–up inputs? Our model provides answers to these and other related questions in terms of graded, interactive processing.

If Peterson (1999) had acknowledged that graded, interactive mechanisms provide perhaps the most plausible account of the kinds of processing that her model seems to require, then the sole difference between the two models would be in the existence of the direct pathway from the edge-based representation of the image to the object-recognition system in Peterson's model (Pathway D in Figure 5 from Peterson, 1999). If empirical evidence (biological or behavioral) unambiguously required the existence of such a pathway, we would have no hesitation in acknowledging its existence and incorporating it into our model. The main point of our model is that Peterson's (see Peterson, 1994, for a review) behavioral evidence to date does not constitute such unambiguous evidence. Thus, we see no reason at this point to abandon the simpler, hierarchically structured model that we have proposed.

The final issue that remains from Peterson's (1999) commentary is how multiple cues to figure–ground are combined with one another. The simulation that we presented to demonstrate that top–down familiarity cues could overcome unambiguous (biased) bottom–up, stimulus cues relied on a network that contained multiple processing pathways (Simulation 5, Vecera & O'Reilly, 1998). There are potential shortcomings in using a network that contains multiple pathways, and we now turn to the critique of this approach.

Multiple Cues in Figure–Ground Organization

Critique of Multiple Pathways

To demonstrate that top–down cues could override biased bottom–up input, we tested the interactive model with displays that contained disparity cues. Our goal was to simulate Peterson and Gibson's (1993) results in which participants could use denotivity as a cue to figure–ground organization in black and white (B & W) stereograms but not in random dot (RD) stereograms.

Peterson and Gibson (1993) interpreted this result as demonstrating that prefigural object-recognition processes operate on luminance contours (present in only B & W stereograms), not on disparity contours. The interactive model simulated the Peterson and Gibson (1993) data by adding another pathway to the model; this pathway contained disparity information, and the original boundary pathway contained the luminance contour information. The results of this simulation (Simulation 5) matched Peterson and Gibson's (1993) behavioral data (see Figure 10, Vecera & O'Reilly, 1998, for simulation results).

Despite the close match between our simulation data and her behavioral data, Peterson (1999) critiqued this simulation on two fronts. First, she questioned whether figure–ground organization is occurring prior to object-recognition processes (i.e., the figure–ground-first issue). Because we have just argued that graded figure–ground organization does come first (architecturally) in our model, we do not reiterate that argument here. Second, she argued that there is a conundrum for the interactive model concerning stimulus-level cues. On the one hand, if we were to place other stimulus cues (e.g., size, symmetry, convexity, etc.) in other pathways, as we did with disparity, then because the inputs from the boundary pathway remain ambiguous, object representations would be activated, allowing them to guide figure–ground organization before any figure–ground processing had occurred (and violating the figure–ground-first assumption of hierarchical models, thereby invalidating our model as a hierarchical model). On the other hand, if some stimulus cues were placed in the boundary pathway itself, then the input to the figure layer would be unambiguous and the model would have a bias to call one region figure (violating the ambiguity assumption). The object units would not be engaged, and no denotivity effects would be observed (preventing the model from simulating some empirical data; e.g., Peterson & Gibson, 1994).

Reply

The previous two sections have addressed Peterson's (1999) critiques concerning the combination of multiple cues in Simulation 5. We have argued that the interactive model can, indeed, overcome biased stimulus input that comes from unambiguous displays; relevant results appear in Simulation 5 of our original article (Vecera & O'Reilly, 1998). We have also argued here that if one assumes graded figure–ground processing, then figure–ground does come first in our model in terms of the model's architecture. Thus, the interactive model can, in principle, process multiple cues either across multiple pathways or in a single pathway (e.g., the boundary pathway). Whether multiple pathways or a single pathway are used should be constrained by other data; neurobiological data provided the constraints for separate disparity and boundary pathways in our earlier model.

Although we have provided arguments here against Peterson's (1999) critiques against the multiple pathways in our Simulation 5 (Vecera & O'Reilly, 1998) model, we acknowledge that some difficulties of interpretation remain because of the potential for different strengths of the cues carried by different pathways. Even though the processing dynamics in these pathways were identical (e.g., the activation function and its parameters), other factors, such as the weight differences and patterns of interconnectivity between the pathways, could have allowed the boundary pathway

to be favored over the disparity pathway. One possible consequence of this bias would be to give the boundary pathway, and the top–down inputs, an artificial advantage over the biased information carried by the disparity pathway. A much stronger test of the model would be to have all of the cues carried by a single pathway, such as the boundary pathway. We now turn to a simulation that tests the model exactly in this manner, where the stimulus cue of area (or size) is pitted against object familiarity, so that a figure–ground display contains both a large, high-denotative region and a small, low-denotative region.

Peterson's (1999) analysis of our model clearly would predict that the model could not overcome the size bias and refer to the high-denotative region as figure because of the unambiguous (biased) nature of the stimulus display (see Peterson, 1999, p. 282, for the basis of this prediction). Our foregoing analysis of our model would lead us to predict that top–down cues should be able to compete with lower level area cues. The results of this simulation will distinguish Peterson's (1999) interpretation of the interactive model from our own and test whether top–down cues can compete with stimulus cues in a model that contains only a single pathway.

Refuting the Critiques: Simulation Results

In this simulation, we use our original interactive model (Vecera & O'Reilly, 1998) of figure–ground organization to determine how different cues to figure–ground organization are combined. The general approach was identical to that in our earlier work (Vecera & O'Reilly, 1998), and the model was identical to our original model (with a few minor parameter differences, which we detail in the Appendix). For parsimony we assume that any cue that relies on luminance contour information, such as area, symmetry, or convexity cues, will be carried by the boundary pathway in the interactive model (although only one of these cues—size—is embodied in our network). Therefore, manipulating both denotivity cues and a bottom–up stimulus cue that relies on luminance contour information, such as area, would allow us to test whether our interactive model could overcome biased displays and demonstrate top–down influences. One cue to which the interactive model is sensitive is size (or area). In preliminary simulations that we conducted for our original simulations, we discovered that the model had a bias to choose smaller regions as figure compared with larger regions.

This size bias is the result of the interconnections among the figure units in which each figure unit was connected to its eight nearest neighbors. These interconnections implemented the constraint that figural regions should be smooth, continuous regions. The size bias is the result of a "gang effect" among the figure units: With smaller regions, there will be fewer intervening units between any pair of figure units, allowing these two figure units to more directly influence one another. In larger regions, there will be more intervening units, resulting in any two figure units having less of a direct influence on one another.

Predictions

In this simulation, we have pitted size cues against denotivity cues. Specifically, we tested the network on the critical display, in which the high-denotative region was larger than the low-denotative region. In such a display, size cues are competing with

denotivity cues, and information relevant to both cues is carried by a single pathway, the boundary pathway (the model's architecture was from Simulation 4 in Vecera & O'Reilly, 1998). Peterson's (1999) analysis of our model would predict that there should be a bias to choose smaller regions as figure, because the displays are not ambiguous. This is an entirely reasonable, and fortunately, testable prediction of the model. The stimulus cue of area would bias the network to choose the smaller, low-denotative region as figure, and the top–down inputs would never be engaged because figure–ground segregation would proceed based on size cues. Specifically, because top–down inputs from object representations would not be present according to Peterson (1999), there should be no influence of stimulus orientation on figure–ground organization: When the high-denotative region is in its canonical, upright orientation size cues should dominate just as much as when this region has been rotated 180° away from its canonical orientation. No orientation effects would be predicted because orientation effects are thought to tap orientation-dependent object representations (see Jolicoeur, 1985; Tarr, 1995; Tarr & Pinker, 1989). Thus, if object representations were not affecting processing, no effects of orientation should be observed.

By contrast, on the basis of our analysis of the interactive model and of figure–ground organization, we predict that top–down denotivity cues can influence figure–ground organization. Specifically, the network should choose the larger, high-denotative region as figure when this familiar region is in its canonical upright orientation compared with when it is rotated 180° from its canonical orientation. Our prediction is based on the simultaneous influences of both the top–down and bottom–up (stimulus level) cues; the top–down cues *do not* need to receive a provisional figure–ground solution before being engaged.

In our simulation, first we replicated our previous results by presenting the network with displays that contained no low-level cues as to what region was figure (i.e., the high-denotative and low-denotative regions were equal in size; the Same Size condition). We included a replication of our earlier results because of the minor parametric changes between the current network and our previous network (Vecera & O'Reilly, 1998). We then tested the same network with displays in which size cues opposed denotivity cues (the High-Denotative Region Large condition), as described earlier. The simulation methodology was similar to that used in our previous article (Vecera & O'Reilly, 1998), and the details appear in the Appendix of the current article.

Results

For the trials in which a coherent figure–ground solution was obtained, we computed the proportion of trials in which the network chose the familiar, high-denotative region as figure. Figure 2 presents these data for both the Same Size and High-Denotative Region Large conditions. In the Same Size condition, we replicated our previous simulation results. When the stimulus was in the upright orientation, the network had a strong bias to call the familiar region figure; this bias was reduced when the stimulus was rotated. The difference between the upright and rotated conditions was statistically reliable, as tested with a z test on independent proportions (Ferguson & Takane, 1989), $z = 3.75, p < .001$. This finding replicates the basic effects presented in our original article (Vecera & O'Reilly, 1998).
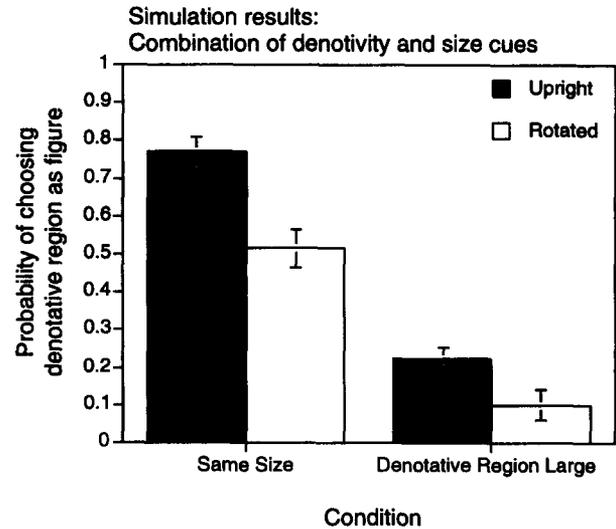

Simulation results:
Combination of denotivity and size cues

Figure 2. Results from the simulations. In the Same Size condition, top–down inputs result in the familiar region having a higher probability of being chosen as figure than the less familiar region, replicating Vecera and O'Reilly's (1998) results. In the High-Denotative Region Large condition, size cues favor the low-denotative (unfamiliar) region. However, top–down inputs still operate, allowing the model to choose the familiar region as figure with a higher probability than the less familiar region. (Note: Standard deviations are based on the assumption that the probabilities came from a binomial distribution; see Ferguson & Takane, 1989).

We performed the same comparison for the High-Denotative Region Large condition, in which the familiar region was larger than the less familiar region. If the top–down inputs from object representations are operating concurrently with the bottom–up, stimulus-level information, then the orientation of the stimulus should modulate what the network chooses as figure. Specifically, even when the high-denotative region is large, there should be a stronger bias to call the familiar region "figure" when the stimulus is in the upright orientation compared with when the stimulus is rotated. Inspection of Figure 2 indicates that, importantly, there is an effect of stimulus orientation: Although the familiar region is larger than the unfamiliar region, the network still has a stronger bias to call the familiar region "figure" when the display is upright than when it is rotated, $z = 2.36, p < .01$. Thus, familiarity cues can compete with this size bias in that the size bias is reduced when the top–down inputs are larger (upright condition) compared with when the top–down cues are smaller (rotated condition).

Discussion of Simulation Results

These simulation results demonstrate clearly that a top–down influence still exists even when unambiguous displays are used (the High-Denotative Region Large condition). If Peterson's (1999) analysis were correct, the initial bias should have been for the smaller, low-denotative region, and the object representations would not have been able to override this influence because the smaller region would have been filled as figure. There should have been no influence of stimulus orientation. Instead, the results are consistent with our analysis in which the top–down and bottom–up inputs interact with one another throughout processing,

allowing top–down information to actively influence figure–ground processing immediately after the onset of the display. The network does not need to fully specify one region as figure before top–down inputs can have their influence on figural processing, nor does the network need to perform "prefigural" object recognition on luminance contours to exert top–down influences on figure–ground segregation. Furthermore, the ambiguity required by the model is very minimal indeed—the network does not need a completely ambiguous stimulus but instead requires a stimulus that has *some* ambiguity (i.e., contains some evidence for each of the two regions, although this evidence could favor one of the two regions, as happens with size cues). A stimulus display that is unambiguously biased toward the less denotative figure–ground solution would place top–down inputs at a disadvantage because the stimulus array would contain no evidence for these top–down inputs.

The present results are consistent with Simulation 5 in our previous article (Vecera & O'Reilly, 1998), and these present results demonstrate that the ability of top–down inputs to compete with stimulus cues is not restricted to systems in which different cues are carried along different pathways (e.g., the boundary and disparity pathways in our earlier model). Our earlier finding that interactive models need not be merely facilitatory also is supported and extended by the present simulation; interactive models can actually *override* bottom–up cues, not just *facilitate* bottom–up cues.

Two additional points from this simulation warrant discussion. The first point for discussion is how do human participants perform when size cues are pitted against object familiarity cues? To our knowledge, the corresponding experiment has not been conducted with human observers. In most of her own research, Peterson has held stimulus-level cues, including size, constant between high-denotative and low-denotative regions (see Peterson & Gibson, 1991; Peterson et al., 1991, for examples). Peterson's recent work (Peterson & Gibson, 1994) pitted vertical-symmetry cues against object-familiarity cues (also see Peterson & Gibson, 1991; Peterson et al., 1991). Results from those studies demonstrated that object familiarity can compete with symmetry; in a display containing an asymmetric, high-denotative region and a symmetric, low-denotative region, the high-denotative region is labeled figure on 48% of the trials, indicating that object-familiarity cues can overcome symmetry cues (see Table 1, Peterson & Gibson, 1994). If symmetry cues dominated figure–ground organization, the low-denotative, symmetric region should have been chosen as figure overwhelmingly. Are similar results obtained for displays containing size and object-familiarity cues? In the following section, we present behavioral results from an experiment in which we investigated the combination of size and object-familiarity cues to test the predictions generated by our interactive model.

The second issue for discussion is the strong effect that size cues have in the network. Inspection of Figure 2 reveals that when a stimulus display contains a smaller region, that region is chosen as figure a majority of the time. For example, in the upright displays, the high-denotative region is chosen as figure approximately 23% of the time, indicating that the smaller, low-denotative region is chosen 77% of the time. Although the important and relevant comparison is between the upright and rotated displays, one may wonder if human observers show this degree of sensitivity to size cues or if the strong bias exhibited by the network is due to specific

parameters. Size is an important and powerful cue to figure–ground organization (see Pomerantz & Kubovy, 1986), and therefore we predict that human participants would likely show large effects of size similar to the model's performance. However, the model's sensitivity to size cues could be artificially large because of the weights among the figure units (see Figure 2 in Vecera & O'Reilly, 1998). Smaller weights among the figure units potentially could reduce the model's bias to call the smaller region figure, although we have not systematically explored this possibility. Another contributing factor to the strong size bias in the present simulation is the stimulus displays used. Subtle size manipulations were not possible to perform with the displays we used; increasing or decreasing the size of a region by 12 figure units was the smallest manipulation that we could perform because of the geometry of the displays and the object representations. Smaller manipulations (e.g., increasing or decreasing a region's size by 4 figure units) may reduce the strength of the size bias. The size bias exhibited by the model is relevant when comparing the model's performance to the performance of human participants, to which we now turn.

Behavioral Results

To determine if human observers could use familiarity or denotivity to overcome the stimulus-level cue of size as our model demonstrates, we conducted a brief experiment on figure–ground segregation. The participants were 14 University of Iowa undergraduates who participated for course credit. All participants had normal or corrected vision.

The stimuli consisted of figure–ground displays that depicted a high-denotative region—a profile of a human face—and a low-denotative region. The displays were constructed to place the denotivity cue in opposition to the gestalt figure–ground cue of size or area. Specifically, the denotative face was larger in size than the less denotative region. The denotative face region occupied 21,606 pixels, and the smaller less denotative region was formed by 8,814 pixels (all measurements were made with the NIH Image software package). The displays measured 4.13 cm wide by 9.17 cm tall, and participants viewed the displays from a distance of approximately 60 cm.

The two regions appeared in different colors, one red and one green. Half of the time the denotative region was red, and half of the time it was green. The orientation of the displays was manipulated also. Half of the displays presented the denotative region in its canonical, upright orientation, and half of the displays presented the denotative region rotated 180° away from its upright orientation.

Participants viewed displays for 100 ms, followed by a pattern mask, which was present for 200 ms. Following the pattern mask, participants indicated which region appeared as figure, either the red region or the green region. Following a response, there was an intertrial interval of 200 ms prior to the start of the next trial. We computed the percentage of time participants reported perceiving the denotative region as figure.

Based on the model's performance, we predicted that participants would perceive the upright face as figure more frequently than the rotated face despite the face region being larger than the less denotative nonface region. The results from our participants supported this prediction. In the upright displays, participants

perceived the denotative region as figure on 80.36% (SE = 6.51) of the trials, but in the rotated displays the denotative region was perceived as figure on only 64.29% (SE = 5.69) of the trials. The difference between the upright and rotated displays was statistically significant, $t(13)$ = 2.86, $p < .02$. The participants, like the model, overcame the size cue and reported the larger, denotative region as figure more frequently than the smaller, less denotative region. Our participants, however, showed less of a size bias than did the model. When shown large denotative regions, the model had a strong bias to label the smaller region as figure, reporting the larger, denotative region as figure only 16.34% of the time on average. By contrast, the human participants reported the larger, denotative region as figure 72.32% of the time on average. The source of this discrepancy likely is due to two sources: (a) the strong size bias in the model and (b) the use of a highly familiar shape (a face profile) with human observers. Detailed parametric manipulations of both the model's size bias and the human participants' familiarity bias will be required to provide a more detailed understanding of the combination of size and denotivity cues. As a first approximation, however, both the model and the human observers show that denotivity cues can override size cues, which is the most relevant result for the present exchange. Our behavioral results also are consistent with Peterson's published results comparing denotivity and stimulus-level cues such as binocular disparity and symmetry (e.g., Peterson & Gibson, 1993, 1994).

Summary and Conclusions

In this reply, we have argued that the interactive model does not require total ambiguity in the stimulus input in order to demonstrate top–down effects from object representations; that figure–ground organization should be viewed as a graded process; and that in the interactive model, top–down inputs combine with stimulus inputs concurrently, not sequentially. Most important, we supported our analysis of our model by providing additional simulation results. The results from our simulation show that top–down inputs can influence figure–ground organization even in the face of a competing stimulus cue, such as size. Behavioral results from human observers corroborated the behavior of our model.

The simulation results that we presented represent an important extension of the model. Although predictions about a model's behavior can be made based on computational principles, the predictions need to be empirically tested through simulation, especially when behavioral data is being simulated. The results from displays that contained size cues (the High-Denotative Region Large displays) demonstrate that the role of stimulus ambiguity is minimal in the model, consistent with the discussion in our previous article, in which we stated that "some degree of ambiguity" would be required in order to elicit feedback from object representations (Vecera & O'Reilly, 1998, p. 453). In that original article we did not establish the limits of the ambiguity required, so the current exchange with Peterson (1999) has been beneficial in not only clarifying our meaning of "ambiguity" but also in understanding better the role of ambiguous stimulus displays in our network.

One benefit of both the present exchange and our current simulation results is that they highlight the difficulties of describing macroscopic behavior, like figure–ground organization, with a model of the microstructure of behavior. For example, it is inaccurate to say that the model attempts figure–ground organization and fails (when the stimulus displays are ambiguous). It is also inaccurate to describe the model as holding simultaneously both regions as figure (Peterson's, 1999, "figure–figure first" discussion). The problem in describing the network's behavior with such terms is that these descriptions use a macrolevel phenomena (e.g., which region is figure?) to explain the workings of the microlevel phenomena. The figure at the macroscopic level is a relatively stable representation that is a statistical phenomena that emerges through time. During the early processing cycles, which could correspond to the first few milliseconds of processing in the visual system, there is no macroscopic figure, only a pattern of activation across the figure units. This early pattern involves small activation levels across the units, and these activation levels are strongly influenced by the intrinsic noise in the network (including the initial top–down inputs from the object units). As processing proceeds, the top–down and stimulus-level cues simultaneously influence the pattern of activation across the figure units, forcing this representation to be influenced less by intrinsic noise and more by object-familiarity cues and stimulus-level cues.

Exchanges such as this also are often beneficial to theoretical progress by clarifying issues and suggesting further research. Our own presentation concerning ambiguity has become clearer through this exchange, and we hope that now the relevant issues are understood better. Our understanding of Peterson's (1994) prefigural model is clearer also, which is both a blessing and a curse. Our improved understanding is a blessing, because we are able to distinguish the processing mechanisms of the two accounts. But, our improved understanding is also a curse, because it appears that it may be difficult to distinguish the models on empirical grounds. Converging evidence from other methodologies (e.g., neuroimaging data or data from neuropsychological patients) may help settle some of these issues, but the critical results are not at hand presently. Thus, at the present, why should one favor either of the two accounts? In this conclusion we offer two important reasons to favor the interactive model that we have proposed: parsimony and reliance on explicit computational mechanisms.

First, a comparison of the interactive model and Peterson's (Figure 5 from Peterson, 1999) framework suggests that the two approaches are very similar to one another. Pathway D in Peterson's (1999) Figure 5 corresponds to the prefigural recognition pathway. Our previous simulations (Vecera & O'Reilly, 1998), as well as the current simulation, indicate to us that this pathway is not required to explain the behavioral data. Thus it constitutes a free parameter in the system, and Occam's razor would favor the interactive account. Our previous model (Simulation 5, Vecera & O'Reilly, 1998) corresponds to Pathways A and C in Peterson's (1999) Figure 4, and the present model corresponds to her Pathway C.

Second, we have relied on explicit computational mechanisms to describe the operation of our model, demonstrating the sufficiency of our theory to account for the data. Although box-and-arrow models of cognition are important first steps in understanding mental processing, such models do not specify the processing dynamics in sufficient detail to make testable predictions, as they contain too many degrees of freedom. For example, what would happen to Peterson's (1999) system following lesions to one of the processing levels? Although one could ask the same question of

our model, our model has an advantage over Peterson's (1999) prefigural account in addressing this question: A working computational model, such as our interactive model, allows such questions to be addressed through *empirical* simulation. We would maintain that empirical simulation should always be favored to the "theoretical simulation" that box-and-arrow models require. Empirical simulation, such as that performed in this reply and in our original article (Vecera & O'Reilly, 1998), is not trivial. Often simulations do not work, and only through careful consideration of one's assumptions can such difficulties be overcome. Finally, our reliance on the explicit computational mechanisms of parallel distributed processing models makes our interactive model consistent with other models of visual processing without the need to postulate different processing mechanisms (e.g., Stroop interference and stereopsis; see Cohen, Dunbar, & McClelland, 1990; Marr & Poggio, 1976).

Of course, parsimony and explicit computational mechanisms should not provide the only assessment of our interactive model, or any model. Both empirical and simulation data will be needed to test the interactive model of figure–ground organization. Although it is unlikely that a single experiment will allow us to distinguish our interactive account from Peterson's (1994) prefigural account, the number of added pathways, processing mechanisms, and so forth will provide a telling sign as to which account provides the most elegant and the most explicit explanation of figure–ground organization.

References

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

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.

Ferguson, G. A., & Takane, Y. (1989). *Statistical analyses in psychology and education* (6th ed.). New York: McGraw-Hill.

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

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

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

Marr, D., & Poggio, T. (1976, October 15). Co-operative computation of stereo disparity. *Science, 194,* 283–287.

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

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

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

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. (1999). What's in a stage name? *Journal of Experimental Psychology: Human Perception and Performance, 25,* 276–286.

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. (1994). Must figure–ground organization precede object recognition? An assumption in peril. *Psychological Science, 5,* 253–259.

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.

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.

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

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

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review, 2,* 55–82.

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

Vecera, S. P., & Farah, M. J. (1997). Is visual image segmentation a bottom-up or an interactive process? *Perception & Psychophysics, 59,* 1280–1296.

Vecera, S. P., & O'Reilly, R. C. (1998). Figure–ground organization and object recognition processes: An interactive account. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 441–462.

Wallach, H. (1949). Some considerations concerning the relation between perception and cognition. *Journal of Personality, 18,* 6–13.

Wertheimer, M. (1923/1958). Principles of perceptual organization. In D. C. Beardslee & M. Wertheimer (Eds. and Trans.), *Readings in perception* (pp. 115–135). Princeton, NJ: Van Nostrand. (Original work published in 1923)
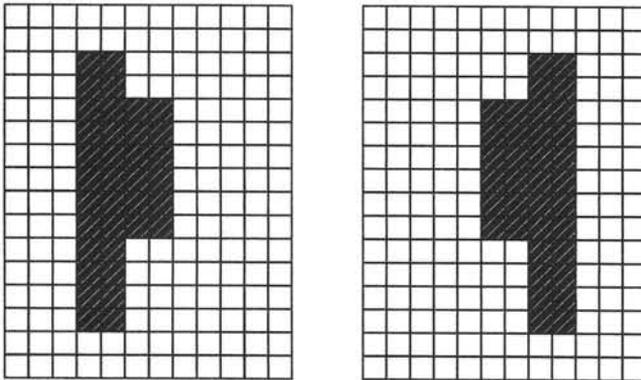
Appendix

Simulation Methods

Figure A1. The object representations used in the model. The black regions of each panel correspond to those figure units that project to, and receive projections from, the object units. There were two representations for each object to allow the model to exhibit spatial invariance, as discussed in Simulation 4 of Vecera and O'Reilly (1998).

The network used was similar to the Learning in Error-driven and Associative, Biologically Realistic Algorithms (LEABRA) network used in Simulations 4 and 5 in our earlier work (Vecera & O'Reilly, 1998). There were only minor differences between the current network and our original model. First, the strength of the weights between the figure layer and the intermediate object layer were changed (weights of +10 instead of +9), as were the strength of the interconnections among the figure units (+8 instead of +10). Second, the object representations used in the current model were smaller (i.e., had a fewer number of units) than the object representations in our earlier model, as shown in Figure A1. We relied on smaller object representations to manipulate the size of the two regions in the figure–ground display. All other weight values were identical to those reported in Vecera and O'Reilly (1998). Another parameter that changed for the present simulations was the offset (threshold) for the figure units. In the previous simulations, the figure units had offsets of +35, and the offsets used in the present simulations were +150. This larger offset was used because we increased the number of active units across the figure layer from 48 to 60 to accommodate the larger area in the figure–ground displays. Using a larger offset allows the network to produce reasonable figure–ground solutions when the smaller region is chosen as figure. Finally, the strength of the external input was slightly different between this model and the previous version; the external input to the boundary layers was +40 instead of +36. The remaining parameters were identical to those reported in Appendix B of Vecera and O'Reilly (1998).

The object representations and the stimulus displays were altered in this simulation to manipulate the size of the regions in the stimulus display. The object representations were smaller than those used in our previous simulations (see Figure 4, Vecera & O'Reilly, 1998, for the object representations). The smaller object representations are depicted in Figure A1. Each object representation received input from 36 figure units.

The stimulus displays used in this simulation are shown in Figure A2. The left panel of Figure A2 shows the displays from the Same Size condition, which is a replication of our earlier simulations (compare the left

panel of Figure A2 with Figure 4 in Vecera & O'Reilly, 1998). In the Same Size condition, the high-denotative and low-denotative regions had the same area. The right panel of Figure A2 shows the displays from the High-Denotative Region Large condition, in which the high-denotative region of the display, shown in black in the figure, is larger than the low-denotative region. The high-denotative region occupied 60 figure units, and the low-denotative region occupied 48 figure units.

In both the Same Size and the High-Denotative Region Large conditions, the high-denotative region appeared on the left side of the display on half of the trials and on the right side of the display in the remaining half of the trials. Also, half of the trials involved presenting the figure–ground display in its canonical, upright orientation, and the other half involved presenting the display rotated 180° from this upright orientation. Finally, stimuli in both conditions were presented by giving the appropriate boundary units additional net input. The network cycled 200 times in which the activations of each unit were updated according to Equation 8 of Appendix B in Vecera and O'Reilly (1998).

Following the settling process, the pattern of activation across the figure layer was compared with all of the possible correct figure–ground solutions. Any pattern of activation that reflected a local maxima (i.e., blends of the possible figure–ground solutions) was excluded from our analyses (see Vecera & O'Reilly, 1998). Only 9 of the 400 trials (2.3%) on which the network was tested resulted in an incorrect solution corresponding to a local maxima or "blend" (see Vecera & O'Reilly, 1998, for discussion of these incorrect solutions).
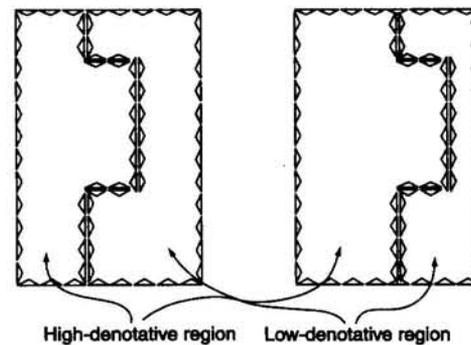

High-denotative region   Low-denotative region

Figure A2. The figure–ground displays used to test the model. The left panel shows the displays used in the Same Size condition, in which the high-denotative and low-denotative regions had the same size or area. These stimulus displays are ambiguous in terms of bottom–up, stimulus cues (e.g., area, symmetry, convexity, etc.). The right panel shows the displays from the High-Denotative Region Large condition. The familiar, high-denotative region is larger than the low-denotative region; thus, these displays are biased to favor the low-denotative region because of size cues.