

Is visual image segmentation a bottom-up or an interactive process?

SHAUN P. VECERA

University of Utah, Salt Lake City, Utah

and

MARTHA J. FARAH

University of Pennsylvania, Philadelphia, Pennsylvania

Visual image segmentation is the process by which the visual system groups features that are part of a single shape. Is image segmentation a bottom-up or an interactive process? In Experiments 1 and 2, we presented subjects with two overlapping shapes and asked them to determine whether two probed locations were on the same shape or on different shapes. The availability of top-down support was manipulated by presenting either upright or rotated letters. Subjects were fastest to respond when the shapes corresponded to familiar shapes—the upright letters. In Experiment 3, we used a variant of this segmentation task to rule out the possibility that subjects performed same/different judgments after segmentation and recognition of both letters. Finally, in Experiment 4, we ruled out the possibility that the advantage for upright letters was merely due to faster recognition of upright letters relative to rotated letters. The results suggested that the previous effects were not due to faster recognition of upright letters; stimulus familiarity influenced segmentation *per se*. The results are discussed in terms of an interactive model of visual image segmentation.

Imagine a rider atop a horse. There are clearly points of overlap between the two objects, such as where the rider's legs occlude the side of the horse. For the visual system to recognize these two distinct objects, each visual object representation stored in visual memory must receive input from the regions of the visual field that correspond to that object and only to that object. Visual image segmentation is the process by which the visual system groups or "binds" locations and features that are part of a single shape. Is image segmentation a purely bottom-up, stimulus-driven process or can stored knowledge of objects play a role in image segmentation?

Both answers seem equally plausible. Bottom-up image segmentation is consistent with Fodor's (1983) claims of information encapsulation in the peripheral or input sys-

tems. One could also argue a priori for bottom-up image segmentation because interactive image segmentation seems to pose a paradox: If the purpose of image segmentation is to group locations of an object in order to recognize that object, then how can object information be used to guide this process, since, presumably, the object hasn't been recognized until completion of the segmentation? Although Wertheimer (1923/1958) suggested that familiarity might influence perceptual organization, the work of the Gestalt psychologists, including Wertheimer, is typically viewed as an attempt to identify bottom-up heuristics for organizing the visual field. Heuristics for perceptual organization suggested by the Gestalt psychologists included proximity and similarity; others, such as area and symmetry, were suggested for figure-ground organization (see Rubin, 1915/1958). Note that "image segmentation" and "perceptual organization" can be viewed as opposite sides of the same coin. Grouping the features of an object is, in effect, the same as segmenting the features of different objects apart from one another. The result of each process is a representation in which the visual system knows which features belong together (and which ones do not belong together).

Research in computational vision has also focused on developing models of bottom-up segmentation processes. For example, in Marr's (1982) model of visual processing, the grouping of features represented in the raw primal sketch corresponds to the full primal sketch. However, this stage does not receive top-down input from the object-recognition stage, the 3-D model. Instead, these intermediate visual processes are completed before their

This research was conducted while the first author was at Carnegie Mellon University. The first author was supported in part by the Neural Processes in Cognition Program (NSF Award BIR 9014347), NIMH Training Grant 2T32 MH19102-06, and a Sigma Xi Grants-in-Aid of Research award; the second author was supported by ONR Grant N00014-91-J1546, NIMH Grant R01 MH48274, NIH Career Development Award K04-NS01405, and Grant 90-36 from the McDonnell-Pew Program in Cognitive Neuroscience. The authors wish to thank Marlene Behrmann, Muriel Boucart, Michael Corballis, Glyn Humphreys, Maureen Marron, Mike Mozer, Steve Palmer, Mary Peterson, Craig Stark, and an anonymous reviewer for comments on this work. Many thanks are extended to Jay McClelland for several discussions that led to Experiment 4 and also to Karen Klein and Kayt Sukel for assistance with the experiments and analyses. Correspondence concerning this article can be addressed to S. P. Vecera, Department of Psychology, University of Utah, Salt Lake City, UT 84112 (e-mail: shaun.vecera@m.cc.utah.edu).

results are passed on to the higher level visual processes (see also Marr, 1976). Ullman (1989) has proposed a model-based system for visual recognition in which object knowledge is used to guide the search for and the interpretation of features in the visual field. However, this system does not use this stored object knowledge to help guide any earlier image-segmentation processes. Other bottom-up heuristics implemented in computational vision systems include nonaccidental properties, properties such as parallelism and smooth continuation (see Lowe, 1985).

Recent connectionist models of image segmentation also have opted for bottom-up approaches to the problem. Mozer and his colleagues (Mozer, Zemel, & Behrmann, 1992; Mozer, Zemel, Behrmann, & Williams, 1992) trained a connectionist network to segment images consisting of two overlapping objects, say, two squares. The significant contribution of this work is that it exploits the ability of connectionist networks to learn, and the network discovers grouping principles instead of having to have the heuristics built in by the programmer. However, it is a bottom-up model in that it does not use higher level object representations to guide the segmentation process (however, see Grossberg, 1993, for a more interactive model of segmentation).

Although most work on image segmentation assumes purely bottom-up processes, the possibility that object knowledge influences image segmentation is far from being a straw-man argument. Much empirical research on higher levels of visual processing suggests that prior knowledge can partly guide processing. This suggests that interactive processing, in which top-down knowledge partly guides lower level processing, may be a general computational principle used by the visual system. Recent theoretical accounts of perceptual organization have also endorsed an interactive view (Palmer & Rock, 1994b).

A classic example of interactive processing in higher level visual processing is the word-superiority effect (Reicher, 1969; Wheeler, 1970), in which the perception of an individual letter is improved when it occurs in the context of a word as compared to when it appears either in a nonword or in isolation. McClelland and Rumelhart (1981) provided a mechanistic account of these results, suggesting that they could be due to partial activation of visual word representations interacting with intermediate letter representations.

Interactive visual processing has also been demonstrated by Prinzmetal and Millis-Wright (1984), who found that visual feature integration is influenced by cognitive and linguistic factors. In one experiment, subjects were presented with three letters that made either a word or a nonword. Each of the letters appeared in different colored ink, and one of the letters was either a G or an R. The subjects had to perform a conjunction search; they first reported whether the target letter was a G or an R and then named the letter's ink color. The subjects were more likely to perceive an illusory conjunction when the letters formed a word (e.g., "TAR"), than when they formed a nonword (e.g., "TVR"). Previous research has suggested

that illusory conjunctions are more likely to occur within a perceptual group than between perceptual groups (Prinzmetal, 1981); therefore, Prinzmetal and Millis-Wright's results suggested that letters that form a word are treated as a perceptual group, whereas letters that do not form a word are not treated as such a group. Thus, familiarity of a three-letter string can, presumably, influence the conjunction of features. Prinzmetal and Millis-Wright also found similar effects for the pronounceability of letter strings; more pronounceable strings resulted in more illusory conjunctions, suggesting that pronounceable strings were being treated as a perceptual group. These results argue that prior knowledge, in this case, knowledge of words and pronounceability, influences visual processes involved in conjoining shape and color information.

Interactive processing in high-level vision is also suggested by context effects in object naming (Palmer, 1975). Objects occurring in their correct context (e.g., a loaf of bread in a kitchen) were named more accurately than visually similar objects occurring in an improper context (e.g., a mailbox in a kitchen). Additionally, Biederman's (1981) findings that schema-level information influenced object recognition are consistent with interactive processing. He presented objects that violated relations used in constructing "well-formed" scenes; for example, a couch could be seen floating in the sky of a city scene (violation of support) or a fire hydrant could appear on top of a mailbox (violation of position). Subjects were less accurate in detecting objects that underwent a relation violation than they were in detecting objects that had not undergone such relation violations. Furthermore, as the number of violations increased, the less accurate subjects were at identifying the object.

There is also evidence that familiarity may influence more intermediate levels of visual analysis. Peterson and her colleagues have presented data that suggest that figure-ground organization may be influenced by object representations (see Peterson, 1994, for a review). Peterson and Gibson (1991) presented subjects with ambiguous displays and asked them to report which region appeared to be the figure. One region of the ambiguous display denoted a meaningful shape; the other region did not denote a meaningful shape. Using these types of stimuli, Peterson and Gibson (1991) showed that shape-recognition processes might contribute to figure-ground organization when the stimulus was in its canonical, upright orientation. This finding is inconsistent with most theories of visual processing that require figure-ground organization to occur prior to shape-recognition processes (e.g., Biederman, 1987; Marr, 1982; Neisser, 1967), and suggests that higher level knowledge of shapes might influence processing at an earlier stage of visual perception.

Many issues arise regarding the relationships between image segmentation and figure-ground organization. Figure-ground perception, as measured by Peterson and her colleagues (Peterson & Gibson, 1991, 1993, 1994a; Peterson, Harvey, & Weidenbacher, 1991), may be a later process than image segmentation, since features (i.e., edge contours) must first be organized before they can

be used to assist in the labeling of “figure” and “ground.” Indeed, segmentation precedes figure–ground organization in several computational vision models (e.g., Finkel & Sajda, 1992; Hummel & Biederman, 1992; Sajda & Finkel, 1995). For example, in Hummel and Biederman’s (1992) model, edges were segmented via temporal oscillations prior to more region-based processing (e.g., “blob” detection in their model; although this process was not equivalent to figure–ground organization, it did involve more region-based processing). Furthermore, Palmer and Rock’s (1994b) account of perceptual organization also indicates that edge-based processing occurs prior to region-based processing. Palmer and Rock (1994b) primarily focused on the case of a single edge with two regions; when multiple shapes were presented, as they were in the current experiments, not only would edge detection occur, but segmentation of those edges would also need to occur. In keeping with such computational and theoretical accounts, we assume that segmentation is used to group local contour elements, such as edges, in the visual field, and that after the edges have been grouped appropriately, figure–ground can be applied to the candidate regions that lie between various edges. Not all theorists may agree with such a scenario, however. The relationship between figure–ground organization and image segmentation will be revisited in the General Discussion.

In sum, previous findings suggest that both the bottom-up and the interactive models of image segmentation are plausible.¹ On the one hand, traditional approaches to vision within cognitive psychology and computational vision research have generally assumed purely bottom-up image segmentation, and there are machine vision systems using bottom-up image segmentation with at least modest success, as reviewed previously. On the other hand, there is abundant experimental evidence that suggests that interactive processing is found at many levels of the visual system. Surprisingly, there has been no direct attempt to address the issue of bottom-up versus interactive image segmentation using the methods of experimental psychology. In the present series of experiments, we asked which processing strategy the human visual system might actually use in performing image segmentation. In Experiment 1, we presented subjects with a simple segmentation task. Stimuli were two overlapping block shapes, and the subjects were asked to determine whether two probed locations fell on the same shape or on different shapes. To perform this task, the subjects needed to determine whether the probed locations were among the locations encompassed by a single shape or by multiple shapes.

To test between bottom-up and interactive image-segmentation models, we manipulated the availability of object information from internally stored object representations. Specifically, we manipulated the familiarity of the shape that was to be segmented. A bottom-up model would predict no effect for the familiarity of the shape. The image would be segmented by using properties of the stimulus itself, such as good continuation of a given line segment; the availability of memory representations of the

stimulus shapes should have no effect. However, an interactive model would predict an effect for the region’s familiarity. More familiar regions should be segmented faster than less familiar regions because segmentation would be partially guided by object knowledge stored in visual memory.

EXPERIMENT 1

In Experiment 1, we manipulated the familiarity of stimuli that were to be segmented by presenting upright letters, letters that were rotated by 180° (see Peterson, 1994, for a review of results using a similar rotation manipulation, and Rock, 1983, for further discussion of orientation and nonletter shapes), and nonletter shapes, derived from letters by moving one feature. This manipulation allowed us to vary the familiarity of the stimulus, with an upright letter being visually the most familiar given the many findings of orientation dependence in shape processing (see Gibson & Peterson, 1994; Jolicoeur, 1985; Tarr & Pinker, 1989). The letters and nonletter shapes may or may not differ from each other in their visual (or bottom-up) complexity. If, for example, the nonletter shapes are simpler than the letters or rotated letters, the nonletter shapes may be segmented faster. As a result, the critical test between bottom-up and interactive segmentation models is to compare performance between the upright- and rotated-letter conditions. This is because the low-level featural information (e.g., number of pixels, number of line segments) is similar in these two conditions. However, the two clearly differ in the degree of visual familiarity. Upright letters are seen much more frequently than letters that have been rotated 180°. As a result, if image segmentation is an interactive process, we would expect letters to be segmented faster than rotated letters. In contrast, if segmentation is a bottom-up process, there should be no effect of familiarity. Finally, comparisons with the nonletter shapes are potentially difficult to interpret, since the stimulus complexity (i.e., the “bottom-up” information) has not been controlled between the nonletters and letters. To the extent that the bottom-up information is similar between upright letters and nonletters, the interactive model predicts that the nonletter condition would be slower.

In the present experiment, subjects were presented with two overlapping transparent shapes, as shown in Figure 1, a manipulation inspired by previous experiments on perceptual grouping (see, e.g., Duncan, 1984; Rock & Gutman, 1981). Subjects were asked to determine if two Xs were on the same shape or on different shapes. Presumably, before the subjects can respond, they must segment the two objects apart from one another; if subjects hadn’t segmented the two shapes apart from one another, there would be no basis for a comparison of the Xs.

Method

Subjects. Sixteen Carnegie Mellon University staff and students served as subjects. All were paid for their participation. All were native English speakers and reported having normal or corrected-to-normal vision.

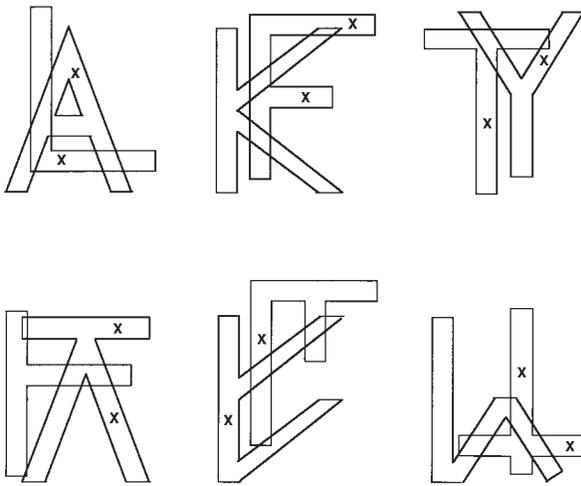


Figure 1. Examples of the stimuli used in Experiments 1 and 2. The Xs were rotated along with the shape so that they stayed in the same position with respect to the shape.

Stimuli. Six block letters were used as stimuli. The letters were A, F, L, T, and Y, and all were capital letters. All possible letter pairs were formed and superimposed on each other, similar to the examples shown in Figure 1, for a total of 15 stimuli. The subjects viewed the stimuli from a distance of approximately 60 cm. All of the individual shapes were 3.8 cm wide \times 5.0 cm tall, subtending 3.63° and 4.77° of visual angle, respectively. The nonletter shapes were created by altering the relationships among the features of the letters. Recall that the nonletter shapes are not perfect controls for the letters, since the bottom-up information they provide is different. That is, these nonletter shapes do not contain the same area as the letters, nor do they necessarily have the same visual complexity.

The overlapping letters were, on average, 4.32 cm wide and 5.17 cm tall, with a range of 3.9–4.8 cm wide and 4.1–5.6 cm tall. The overlapping nonletters were on average 4.95 cm wide and 5.85 cm tall, with a range of 4.2–6.3 cm wide and 5.4–6.9 cm tall. The shapes were overlapped so as to meet the following constraints: (1) two Xs could be fit onto the letters while maintaining the same spatial distance between the Xs when they appeared on the same shape and on different shapes, and (2) no feature of one shape perfectly overlapped a feature of the other shape (e.g., the vertical line in the letter F did not perfectly align with the vertical line in the letter L).

Two small Xs were located on the stimuli. Half of the time, the Xs were on a single shape (the “same” condition), and half of the time, one X appeared on each shape (the “different” condition). The subjects’ task was to respond as quickly and accurately as possible as to whether the Xs were on the same shape or on different shapes. The Xs were 0.3 cm wide \times 0.4 cm tall (0.29° \times 0.38° of visual angle) and appeared in a 12-point bold Helvetica font. The Xs were the same distance from each other in the “same” and “different” conditions, 2.8 cm. When the two Xs appeared on the same letter, one of them remained in the same position and the other was moved to create the stimuli for the “different” condition. For each individual display, which contained two overlapping shapes, there were two stimuli in the “same” condition and two in the “different” condition. For example, when A and L were superimposed, the Xs could appear either on the A or on the L (the two “same”-condition stimuli). When the two “different”-condition stimuli were created, one of the Xs on the A stayed in the same position and the other moved to the L; the other stimulus was created by leaving one of the Xs on the L and moving the other to the A.

The superimposed letters were presented in their normal, upright orientation in the “upright-letter” condition and in a 180° rotated orientation in the “rotated-letter” condition. These stimuli were identical except for the rotation. The nonletters were superimposed and presented in the orientation shown in Figure 1 in the “nonletter” condition.

Procedure. Stimuli were presented via a Macintosh Plus computer. Each subject received nine blocked presentations; the stimulus type was constant within a block. Prior to each block, the subjects were told the stimulus type of that block, but were told that they could ignore the stimulus type and that they should focus on determining whether the Xs were on the same shape or on different shapes. The subjects received three blocks of each stimulus type (upright letter, rotated letter, and nonletter). The order of block presentation was counterbalanced across subjects so that each block appeared equally often in each position.

Before beginning the experimental trials, the subjects received 36 practice trials. To avoid familiarization with the nonletters, the practice trials included intermixed upright- and rotated-letter shapes but no nonletter shapes. Following this practice, the subjects started on the nine blocks. There were 60 individual trials within each block, 30 with the Xs falling on the same shape and 30 with the Xs falling on different shapes. An individual trial began with a fixation pattern of five asterisks appearing on the screen in a plus-sign (+) pattern. The fixation pattern was 4.8 cm wide and 4.1 cm tall, subtending 4.58° and 3.91° of visual angle. The subjects initiated a trial by pressing the space bar with either the left or right hand. After the space bar was depressed, the fixation pattern remained on for 500 msec. The shapes and Xs were then simultaneously flashed for 200 msec. The screen was then blank while the subject responded. The subjects responded by pressing the “v” key for the “same” response (i.e., Xs on the same shape) and the “b” key for the “different” response (i.e., Xs on different shapes). Half of the subjects used the left hand for this response and half used the right; the hand that was not used for the response was used to initiate the trials by pressing the space bar.

Results

Only correct reaction times (RTs) were used in the analyses. Responses faster than 100 msec or slower than 2,000 msec were excluded; also, incorrect keypresses (e.g., a space-bar press) were excluded. Less than 2% of the total number of trials were excluded using these criteria. Subjects’ median RTs for each condition were analyzed with a repeated measures two-factor analysis of variance (ANOVA), with the factors being stimulus type (three levels, upright letters, rotated letters, and nonletter shapes) and X location (two levels, Xs on the same shape or on different shapes). Planned pairwise comparisons were conducted between the stimulus-type factor means.

The mean RTs and error data for the upright, rotated, and nonletter shapes appear in Table 1. As discussed above, an interactive image segmentation model would predict a main effect for stimulus type; this main effect was significant [$F(2,30) = 5.18, p < .02$]. The main effect for X location was also statistically reliable [$F(1,15) = 6.32, p < .03$]. The interaction was not significant at the .05 level, but tended toward significance [$F(2,30) = 2.81, p < .10$], a trend for which we have no interpretation.

Planned pairwise comparisons were performed between the stimulus-type factor means. The RTs to upright letters were significantly faster than RTs to nonletters [$t(15) = 9.22, p < .005$]. RTs to upright letters were also

Table 1
Mean Reaction Times (in Milliseconds)
and Percent Correct for Experiment 1

	Target Type					
	Nonletter		Rotated Letter		Upright Letter	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Same Shape						
RTs	547.97	25.50	528.91	25.57	510.47	22.23
% Corr.	86.9	1.7	86.9	1.8	88.8	1.7
Different Shapes						
RTs	563.50	24.58	571.75	29.57	545.44	38.36
% Corr.	92.3	1.2	93.3	1.1	93.8	1.3

significantly faster than RTs to rotated letters [$t(15) = 5.98$, $p < .03$]. There was no significant difference between RTs to nonletter stimuli and rotated letter stimuli [$t(15) < 1$].

The error data were also analyzed with a two-factor, repeated measures ANOVA. The pattern of results was similar to that observed in the RT data, suggesting that subjects did not sacrifice accuracy for speed. The main effect for stimulus type was significant [$F(2,30) = 3.99$, $p < .03$], as was the main effect for X location [$F(1,15) = 16.52$, $p < .001$]. The interaction was not significant [$F(2,30) < 1$]. Planned pairwise comparisons on the stimulus-type factor-level means revealed significant differences between the nonletters and upright letters [$t(15) = 7.56$, $p < .01$] and a marginal difference between the rotated letters and upright letters [$t(15) = 3.75$, $p = .062$].² There was no significant difference between the nonletter shapes and rotated letters [$t(15) < 1$].

Finally, because accuracy in this task was not near ceiling, we calculated measures of sensitivity (d') and bias (β) for each subject in this experiment in order to determine if the effects were due to changes in sensitivity, response bias, or both. The pattern of results was qualitatively similar to the RT and accuracy data. We tested the differences between d' and β between upright and rotated letters using a Wilcoxon nonparametric test because d' and β are often not normally distributed. Subjects were more sensitive in reporting whether the Xs were on the same shape or on different shapes when the shapes were upright letters than when they were rotated letters ($Z = 2.74$, $p < .007$). However, the biases for upright and rotated letters did not differ significantly from one another ($Z = 0.93$, $p > .30$). Thus, our results were not due to some response-bias difference between the upright and rotated letters.

Discussion

These results are consistent with predictions made by an interactive image-segmentation model. That is, objects in the visual field allow the visual field to be segmented more rapidly when the objects are familiar, in this case upright letters, than when they are the less familiar rotated letters or nonletter shapes. If image segmentation were a bottom-up process, no effect of stimulus familiarity would have been found.

Although these data argue for the use of object knowledge in image segmentation, two points warrant discussion. First, one might wonder why there is not a significant RT superiority of rotated letters over nonletters in this task. If image segmentation is interactive, as we argue, then shouldn't we expect an advantage for the rotated letters since they have some partial fit or match to internally stored letter representations that the nonletters do not have? As mentioned earlier, comparisons between the nonletter stimuli and the rotated letter stimuli are ambiguous because the two sets are not controlled for complexity. The nonletter shapes may be visually less complex than the set of letters we used in this experiment. Other possibilities are that the rotated letters are not able to activate stored letter representations at all in the time needed to segment the image, or that the nonletters are sufficiently similar to the letters that they do partially activate letter representations.

Second, while these data suggest that object knowledge can contribute to the segmentation process, it is possible that the blocked format of this experiment led subjects to strategically activate letter representations in the upright-letter blocks but not in the rotated-letter blocks. If image segmentation is interactive and if object knowledge is used automatically in segmentation, then changing to a mixed-block format should not alter the results. In Experiment 2, we tested the automaticity of the familiarity effects by presenting upright and rotated letters in a mixed-block format. The upright and rotated letters appeared within the same block of trials, which did not allow subjects to selectively activate memory codes of upright letters in order to use object knowledge to aid segmentation.

EXPERIMENT 2

The design of Experiment 2 was identical to that of Experiment 1, with two exceptions. First, the shapes appeared in mixed blocks. Second, because of the ambiguity in comparing the letter stimuli with the nonletter stimuli, upright and rotated letters were the only stimuli used. The predictions made by the interactive and bottom-up segmentation models are unchanged by these manipulations. Note, however, that if we failed to replicate Experiment 1 in the mixed-block format, this would suggest that subjects may have been selectively or strategically using object knowledge on only the upright-letter blocks in the first experiment. Although this finding would not completely rule out an interactive segmentation model, it would suggest that object knowledge could be used strategically only to facilitate image segmentation. However, a replication of Experiment 1 would suggest that use of object knowledge could be used "on-line" in a trial-by-trial fashion and not only by strategic use.

Method

Subjects. The subjects were 10 Carnegie Mellon University students who participated to fulfill a course requirement. All were na-

tive speakers of English, and all had normal or corrected-to-normal vision.

Stimuli. The stimuli were identical to those used in Experiment 1, except that the nonletter shapes were not used.

Procedure. The procedure was also identical to that used in Experiment 1, except that the upright and rotated letters were intermixed within the same block of trials. The subjects first received 60 practice trials, followed by 12 blocks with 60 trials each. Within each block, half of the stimuli were upright letters, half were rotated letters. Also, half of the time the Xs were on the same shape and half of the time they were on different shapes. Also, the subjects did not initiate the trials by pressing the space bar; instead, the fixation asterisks appeared for 1,000 msec prior to the presentation of the letters. The remainder of the procedures, including exposure durations and responses, were identical to those in Experiment 1.

Results

The analyses were the same as those used previously. The subjects' median RTs for each condition were analyzed with a repeated measures ANOVA, with the factors being stimulus type (two levels) and X location (two levels). Trimming due to incorrect keypresses or to short or long RTs excluded 1% of the total number of trials.

The mean RTs, percent correct, and standard errors for all conditions appear in Table 2. For the RTs, there was a main effect for stimulus type [$F(1,9) = 22.09, p < .002$], with RTs to upright shapes being faster than RTs to rotated shapes. There was also a main effect for X location [$F(1,9) = 12.60, p < .007$], with faster responses to Xs that were on the same shape than to Xs that were on different shapes. There was no interaction between these two factors [$F(1,9) < 1$]. The accuracy data showed the same pattern of results. There was a main effect for stimulus type [$F(1,9) = 9.97, p < .02$], with upright shapes receiving more accurate responses than rotated shapes. There was also a main effect for X location [$F(1,9) = 6.64, p < .03$], with Xs located on different shapes receiving more accurate responses than Xs located on the same shape. There was no interaction [$F(1,9) < 1$].

Finally, as in Experiment 1, we separated sensitivity and bias by performing a signal-detection analysis on the results. Again, we tested sensitivity (d') and bias (β) between the upright and rotated shapes, using a nonparametric Wilcoxon test. Sensitivity between upright and rotated letters was significantly different ($Z = 2.29, p < .03$), with subjects being more sensitive in judgments with the upright shapes. However, there was no differ-

ence in response bias between the upright and rotated shapes ($Z = 0.87, p > .30$), suggesting that both types of shapes evoked similar response biases.

We also compared the results from Experiment 1 (novel shapes excluded) with those from Experiment 2. There was a marginal effect for experiment [$F(1,24) = 3.38, p < .08$], with responses being slightly faster in Experiment 2 than in Experiment 1. There was a significant effect for target type [$F(1,24) = 11.11, p < .003$], as well as for X location [$F(1,24) = 20.54, p < .0001$]. None of the interactions were reliable (all $ps > .30$). The failure to find an interaction between experiment and target type suggests that the differences between upright and rotated letters were not statistically reliable between the two experiments. This suggests that the familiarity effect (i.e., the difference between RTs to upright letters versus RTs to rotated letters) was not reliably altered by switching from a pure block format to a mixed block format.

Discussion

These results replicate those of Experiment 1 and are consistent with an interactive model of image segmentation. The present results ruled out the possibility that subjects strategically activated letter representations only for the upright letters. Instead, the RT difference observed between the upright and rotated letters was not statistically influenced by presenting the stimuli in a mixed block in which subjects did not know the stimulus orientation beforehand.

While these experiments tentatively support an interactive image-segmentation model, there are two serious alternative explanations that need to be considered. The first alternative is as follows: The present results could be due to subjects' ability to scan between the Xs faster in a display containing upright letters than in a display containing rotated letters. Jolicoeur and his colleagues (e.g., Jolicoeur, Ullman, & Mackay, 1986) have shown that the RT to determine whether two dots lie on the same curve is a function of the distance of the two dots *along the curve*, not the Euclidean (i.e., straight-line) distance between the two dots. Thus, our results may be due to subjects' being able to scan faster through upright letters than through rotated letters.

To test this alternative we reanalyzed the results from Experiment 2. We first measured the distance between the two Xs when they were located on a single shape. (Note that the Xs were the same Euclidean distance from one another but that their distance along the shape varied slightly. We therefore had to measure the distance along the curve of the shape.) Then we determined the scanning rate for the upright displays and for the rotated displays for each subject. The scanning rate was calculated by performing a regression analysis on each subject's data, predicting (a) the RT for the upright shapes with the distance between the Xs (along the curve) and (b) the RT for the rotated shapes with the distance between the Xs (again, along the curve). If subjects are able to scan between the two types of displays differentially, then the scanning rate for the upright letters should differ reliably

Table 2
Mean Reaction Times (in Milliseconds)
and Percent Correct for Experiment 2

	Target Type			
	Rotated Letter		Upright Letter	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Same Shape				
RTs	464.0	16.30	445.50	18.08
% Corr.	88.7	2.7	92.3	1.9
Different Shapes				
RTs	505.35	23.0	490.55	21.03
% Corr.	93.1	1.7	94.7	1.3

from the scanning rate for the rotated letters. Our analysis, however, failed to find any reliable difference in the slopes [14.22 msec/cm for upright letters and 10.24 msec/cm for rotated letters; $t(9) < 1$].³

These results strongly suggest that there were no differences in scanning within an upright letter and a rotated letter in our stimulus set. Our stimuli were not designed to directly address this issue, however, so the range of distances between the Xs might not have been large enough to elicit scanning effects. While the larger issue of scanning or curve tracing is somewhat orthogonal to the question of bottom-up versus interactive segmentation, curve tracing may involve aspects of segmentation (or perceptual grouping). Therefore, curve tracing may exhibit familiarity effects similar to those observed in Experiments 1 and 2. Of course, it is beyond the scope of this paper to address this point, but the question of bottom-up versus interactive curve tracing would be a theoretically interesting and important issue for additional research.

Although the present results do not appear to be due to differential scanning effects within upright and rotated letters, the results could be due to subjects' making decisions about the probed locations after the recognition of the letters. This alternative suggests that decisions about the Xs (i.e., reporting "same" or "different") are influenced by recognition of the letters. Given that rotated letters will take longer to recognize than upright letters, longer responses to the Xs would result when the display contained rotated letters. Under this alternative account, letter familiarity does not influence the image segmentation process *per se*; instead, *recognition* finishes faster for the upright letters, permitting a more rapid response execution. Such a scenario could explain all of our results in a bottom-up manner: Bottom-up image segmentation occurs first and takes the same amount of time for upright and rotated letters. Following this segmentation, the letters are recognized, and the upright letters are recognized more quickly than the rotated letters. Finally, a response is executed, and this response is influenced by the RTs of the letters. Thus, faster RTs will result for upright letters, but this advantage is not due to faster segmentation.

EXPERIMENT 3

Could the previous results be due to the probed locations' being analyzed or responded to after full segmentation and recognition of the letters? If this account holds, then what happens if the display contains one familiar upright letter and one less familiar rotated letter? Since decisions about the probed locations might occur along with or after segmentation and recognition, there should be no difference between the occurrence of Xs on the upright letter and the occurrence of Xs on the rotated letter in displays containing two letters. This is because the alternative account proposes that both segmentation and recognition have been completed before any judgments concerning the Xs occur. On this account, both

letters would be recognized before a decision was executed; since the display contains one upright and one rotated letter, responses should be the same for trials in which the Xs appear on the upright letter and trials in which the Xs appear on the rotated letter. Furthermore, RTs to displays containing one upright and one rotated letter should fall midway between RTs to displays containing two upright letters and RTs to displays containing two rotated letters. However, if the results were due to simultaneous segmentation and judgment about the probed locations, then, in displays containing one upright and one rotated letter, we would expect faster responses to the Xs when they appeared on the upright letter than when they appeared on the rotated letter.

We addressed this alternative explanation in Experiment 3 by presenting subjects with new displays in addition to those used in the previous experiments. These new displays again contained two overlapping shapes, but one of these letters was an upright letter and the other was a rotated letter. The comparison of interest was when the Xs appeared on a single shape (i.e., "same" responses). On our interactive account, "same" responses should be faster when the Xs are both on the upright shape than when they are on the rotated shape. However, the alternative account would predict similar responses to these two conditions, because decisions concerning the probe occur during or following segmentation and recognition. Note that this comparison occurs between displays that are physically identical except for the location of the Xs.

Method

Subjects. The subjects were 10 Carnegie Mellon University students who either volunteered or participated to fulfill a course requirement. All were native English speakers and had normal or corrected-to-normal vision.

Stimuli. The stimuli were similar to those used in Experiments 1 and 2. The nonletter stimuli of Experiment 1 were not used; only letter stimuli were used. The "pure orientation" stimuli in which both letters were either upright or rotated were identical to those used in Experiments 1 and 2. The "mixed-orientation" stimuli contained one upright letter and one rotated letter, as shown in Figure 2. These stimuli were created by taking one of the letters in the original stimulus set and rotating it 180°. Following the creation of this first set of "mixed" stimuli, the entire set was rotated 180°. This had the effect of making the previously rotated letter upright, and vice versa. Thus, when the Xs were on an upright letter in one display, they were on a rotated letter in another display. As in the previous experiments, half of the time the Xs were on the same shape and half of the time they were on different shapes. Also, when rotations were performed, the Xs were rotated along with the letters. Although this placed the Xs in a different spatial location, it kept them in the same location with respect to the letter.

The mixed-orientation stimuli measured, on the average, 4.30 cm wide and 5.20 cm tall, with a range of 3.9 cm–4.8 cm wide and 4.9 cm–5.6 cm tall. All other aspects of the stimuli, including the distance between the Xs and the procedure used to place the Xs on the letters, were identical to those in Experiments 1 and 2.

Procedure. The procedure was identical to that used in Experiment 2. The stimuli appeared in a mixed-block format. The subjects first saw a fixation point composed of five asterisks in a plus-sign (+) pattern, and the fixation point was presented for 1,000 msec. Following presentation of the fixation point, the two letters were

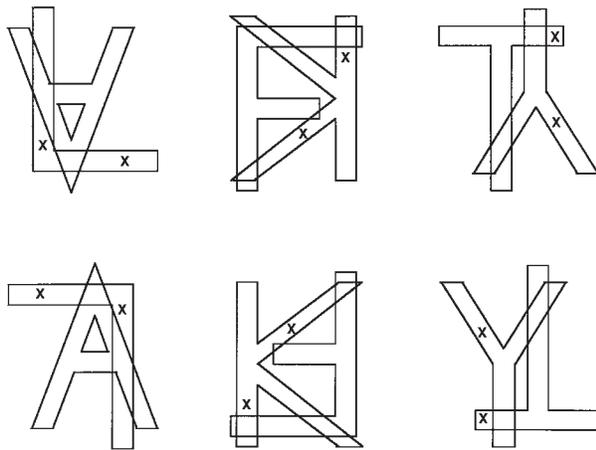


Figure 2. Examples of the “mixed-orientation” stimuli used in Experiment 3.

presented for 200 msec. The screen was then blank while the subject responded. The subjects again used the “v” and “b” keys to indicate “same” and “different” responses, respectively. Half of the subjects used the left hand for responding and half used the right hand. The subjects first received 60 practice trials, followed by eight blocks containing 60 trials each.

Results

The data were analyzed as before. The subjects’ median RTs for the conditions were analyzed with a two-way, within-subject ANOVA. The factors were stimulus type (rotated letters, upright letters, and mixed-orientation letters) and X locations (same shape vs. different shapes). The critical test was between the two types of “same” responses in the mixed-orientation stimuli; we tested this with a planned pairwise comparison.

The mean RTs and accuracy scores for all conditions appear in Table 3. For the RTs, there was a main effect for stimulus type [$F(2,18) = 8.01, p < .005$] as well as a main effect for X location [$F(1,9) = 17.95, p < .003$]. There was no significant interaction between these two factors [$F(2,18) = 2.628, p > .10$]. A planned comparison between the upright- and rotated-letter conditions was significant [$t(9) = 6.89, p < .02$], replicating the findings of Experiments 1 and 2.

Because of near-ceiling performance in the upright-letter condition, we could not estimate sensitivity and bias with the accuracy data. To determine if there was a speed-accuracy tradeoff in the data, we calculated a correlation between RT and accuracy. For the six conditions listed in Table 3, this correlation was not statistically significant [$r = 0.25, t(9) < 1$], suggesting that subjects did not sacrifice accuracy for speed. This result is consistent with the results of Experiments 1 and 2, in which the accuracy data paralleled the RT data.

Next, the test of the alternative hypothesis discussed above was performed by comparing RTs between the two types of “same” responses in the mixed-orientation stimuli. The mean RTs for these two orientations appear in

Figure 3. Clearly, subjects were faster to respond “same” when the Xs were both on an upright letter than when both Xs were on a rotated letter (487.50 vs. 518.20 msec, respectively). This difference was statistically reliable [$t(9) = 2.81, p < .03$].

Discussion

The results of Experiment 3 both replicate our earlier findings and rule out the possibility that subjects made decisions about the probed locations in parallel with or after segmentation and recognition of both letters. There was clearly a difference between “same” responses in the mixed-orientation stimuli: Subjects were faster when the Xs were on the upright letter than when the Xs were on the rotated letter (Figure 3). If the alternative account was correct, then there should have been no difference between these two conditions, because both letters would have been processed, irrespective of where the Xs were located.

Note that the RTs to trials in which the Xs were on the same shape (Table 3) reveals that the mixed-case condition did not fall midway between the upright- and rotated-letter conditions, in which both letters had the same orientation. While this may seem puzzling, it is important to recall that in Table 3 the mixed-case “same” condition contains RTs to Xs located on both upright and rotated letters. Thus, the relevant comparison is to break down the mixed-orientation “same” trials. When these results are compared with the upright-letter and rotated-letter conditions in Table 3, one sees that RTs are longer to mixed-case displays than to “pure-case” displays, in which the orientation is the same for both letters. The longer RTs to mixed-case displays may be due to the fact that these displays were confusing because of the mixed orientations. We analyzed all of the “same” trials with a two-factor ANOVA, with case type (mixed vs. pure) and orientation (upright vs. rotated) as factors. This analysis showed no main effect of case type [$F(1,9) = 3.99, p > .05$] but did show a highly reliable main effect of orientation [$F(1,9) = 25.55, p < .0009$], with RTs to Xs on upright letters being faster than RTs to Xs on rotated letters. Furthermore, there was no interaction [$F(1,9) < 1$], suggesting that the RT differences between the mixed-case and pure-case conditions were additive. Irrespective of the nonsignificant RT difference between the mixed-case

Table 3
Mean Reaction Times (in Milliseconds)
and Percent Correct for Experiment 3

	Target Type					
	Rotated Letter		Upright Letter		Mixed Orientation	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Same Shape						
RTs	509.60	28.30	474.55	27.69	507.25	28.89
% Corr.	91.2	2.5	97.0	1.1	91.2	2.1
Different Shapes						
RTs	575.85	41.59	569.90	45.29	598.55	44.08
% Corr.	96.2	1.1	96.1	0.8	94.3	1.1

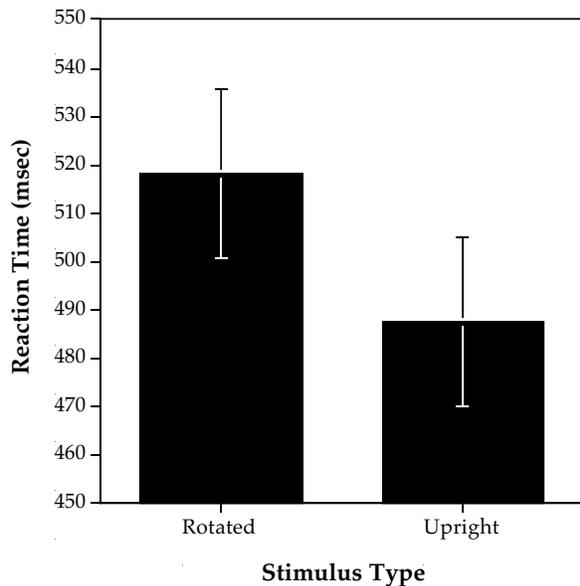


Figure 3. Mean reaction times to the “same” condition in the mixed-orientation stimuli (Experiment 3), broken down by whether both Xs were located on the upright letter or on the rotated letter. The data in Table 3 do not reveal this difference. Error bars are the within-subject confidence intervals (Loftus & Masson, 1994).

displays and the pure-case displays, the main result of importance is that in both types of display RTs were faster when the Xs appeared on the upright letter than when they appeared on the rotated letter.

While the results of Experiment 3 suggested that decisions about the Xs did not occur after letter recognition, these results (and the results of the previous two studies) could possibly be due to subjects’ localizing the Xs more rapidly when they were within an upright letter than when they were within a rotated letter. Two considerations suggest that this alternative is incorrect. First, in unpublished experiments, we found that subjects were just as fast to *detect* an X when it was located within an upright letter as when it was located within a rotated letter; thus, simple detection does not seem to be influenced by stimulus familiarity. Second, and more importantly, this alternative begs the question of how subjects were able to localize the Xs more rapidly in upright letters. One possibility is that some letter recognition occurred prior to X localization. But this is exactly the scenario that we argued against in Experiment 3: The subjects’ performance on the mixed-orientation displays suggested that recognition of both letters did *not* precede decisions about the Xs. Thus, differential rates of X localization between upright- and rotated-letter displays does not appear to fully account for the results we have presented.

The explanation tested in the present study assumed that segmentation and recognition occurred prior to judgments about the Xs. This account assumed that both of the shapes in the display were recognized before any decision about the Xs was made. Of course, this assumption may be incorrect. It might be that all of the pre-

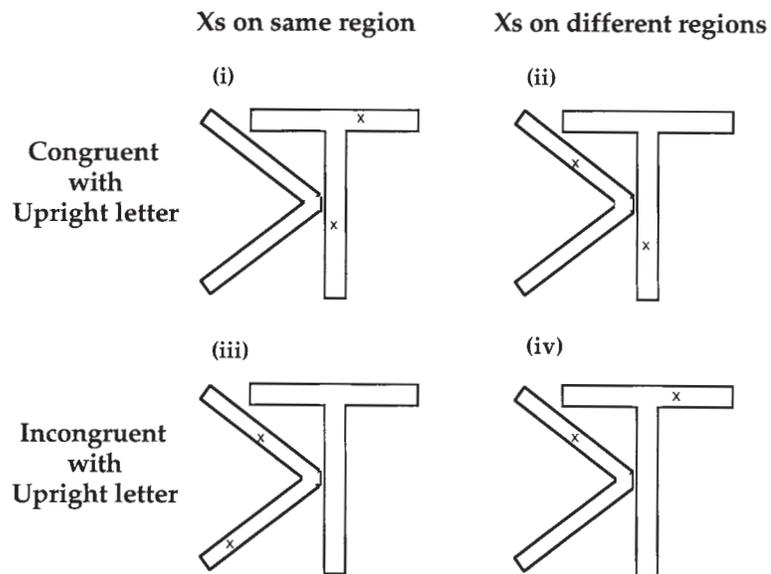
vious results could be accounted for by a bottom-up, feedforward system. The previously observed familiarity effects could be due to faster processing at the later level of object representation. Both shapes would not need to be recognized, as assumed in Experiment 3. Instead, in the “same” trials in Experiment 3, the Xs could be grouped with the shape in which they appeared, and this shape alone could be recognized. Since upright letters will be easier to recognize than rotated letters, one would expect to see RT differences like those observed in all three of the previous studies. To convincingly demonstrate that object recognition is the sole cause of these results, we needed to show that lower level grouping heuristics could actually be *overridden* by higher level object properties (e.g., familiarity). Specifically, could two regions that should not be grouped together (because they are unconnected) be incorrectly grouped if they appear to form a familiar shape?

EXPERIMENT 4

In the previous three experiments, all of the displays had both the lower level and the higher level information consistent with one another. That is, if the Xs were on a smooth, continuous, connected region, they were on the same shape, and any effects of stimulus familiarity could presumably be due to object recognition that occurred after the earlier segmentation processes. What would happen if the lower level cues were put in opposition with the higher level cues? Consider the case in which the Xs are on two physically separate regions. Lower level cues to segmentation would tend to segment these Xs as belonging to different shapes or regions on the basis of the grouping heuristics of connectedness and common region (see Palmer, 1992; Palmer & Rock, 1994b, for discussions of these heuristics). However, if these same two Xs were on physically distinct regions that were consistent with a familiar shape, would the Xs have a higher probability of being grouped together? We addressed this question in Experiment 4 by using stimuli such as those depicted in Figure 4.

The interactive and bottom-up accounts make differing predictions as to performance in a situation in which low-level grouping heuristics are put in opposition to higher level object properties. A bottom-up account, in which familiarity effects are due to faster recognition of upright letters relative to rotated letters, would predict that two unconnected regions would not be grouped with one another. Thus, in Figure 4b, the two unconnected regions that appear to form an emergent letter (the letter K) should not be grouped together. Furthermore, if one somehow modified the bottom-up account in order to allow physically separate regions to be grouped with one another, then the entirety of both separated regions should be grouped together; *there could be no partial grouping, because such partial grouping would imply that object information somehow modified the grouping, contradicting a bottom-up account.* For example, on such a modified bottom-up account, the *entire* upside-down T in Figure 4b

(a) Intact Letter Upright



(b) Emergent Letter Upright

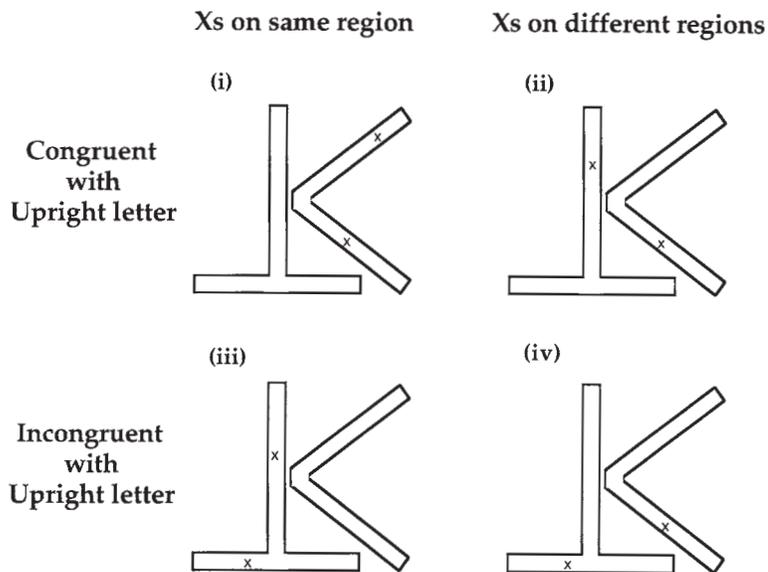


Figure 4. Examples of the stimuli used in Experiment 4. Note that in both the intact letter-upright condition and the emergent letter-upright condition, the Xs can be located so as to be consistent with being on an upright letter or inconsistent with being on an upright letter.

would be grouped with the rotated “V” shape that forms the emergent upright K. Such an account would predict that grouping in such displays would not depend upon where the Xs were located. That is, whether the Xs were congruent with being located on an upright letter (Fig-

ure 4b ii) or incongruent with being on an upright letter (Figure 4b iv), performance would be identical because the same letter (“K”) would be named in both cases.

By contrast, an interactive view would predict that object representations stored in visual memory could over-

ride, or influence, lower level grouping heuristics such as connectedness and common region (see Vecera & O'Reilly, in press, for an interactive model that makes such predictions). In Figure 4b, an interactive account would predict that two unconnected regions could be grouped together if they formed an emergent upright letter. If these unconnected regions did not form an emergent upright letter (as in Figure 4a), influence from object representations would be either weakened or nonexistent. Furthermore, the influence from object representations is not a global effect; instead, *regions that appear to form an upright letter would be preferentially grouped*. Thus, unlike the bottom-up alternative just discussed, only certain parts of the unconnected regions would be grouped. These parts would be those that formed an upright letter. In an interactive model, such a piecemeal segmentation could arise on the basis of various constraints in the image. If a parallel distributed processing (PDP) model began to settle on an interpretation consistent with an upright, but emergent, letter, this interpretation could cause other parts of the image to be segmented as part of a different shape or, conversely, could allow separate regions to be grouped together. If the evidence for an upright (but emergent) "K" were strong, as in the case of Figure 4b, minor differences, such as the lack of connectedness, might be overridden by the wealth of visual information consistent with a "K." Such a scenario is consistent with constraint satisfaction mechanisms in PDP systems (see Vecera & O'Reilly, in press, for a discussion of such processing in the context of figure-ground organization).

This "partial grouping" can be understood by comparing Figures 4b (ii) and 4b (iv). In Figure 4b (ii), the Xs both fall on parts that are congruent with being located on an upright letter; that is, the Xs are both on the letter "K," albeit an emergent "K." In Figure 4b (iv), while the Xs fall on physically separate regions, these regions are incongruent with being located on an upright letter, and these Xs should not be grouped together. Thus, any top-down contribution from object representations could be quite specific and isolated to only parts of the image. This stands in contrast to a bottom-up account, which would predict identical performance in the displays in Figures 4b (ii) and 4b (iv) because the (emergent) letters are identical in both displays. Thus, letter recognition should not differ, leading to similar performance in these conditions.

Before proceeding, we should note that Peterson and her colleagues (e.g., Peterson & Gibson, 1991, 1994a) have created similar situations with figure-ground stimuli. They reported results that demonstrated that object-recognition processes could override lower level cues (e.g., area or symmetry). Furthermore, Peterson (1994) proposes a model that does not appear to be either a bottom-up or an interactive model. As noted in the introduction, however, it is possible to construe this model as operating in a bottom-up manner, but additional work will be needed to address such issues. For now, we maintain the bottom-up versus interactive distinction and dis-

cuss other possible frameworks, including Peterson's, in the General Discussion.

In this experiment, the subjects were shown stimuli such as those in Figure 4 and were asked to respond to the lower level information; that is, subjects were asked to determine if the Xs were on the same region or on different regions. The subjects were explicitly told that the Xs were on the same region if they could trace from one X to the other without crossing any of the boundaries in the image. Responding to the lower level information means that if segmentation is interactive, subjects should be very inaccurate when the Xs fall on separate regions that are congruent with being located on a letter. Such a result would be observed as a statistical interaction between X location and the congruency of the Xs. This would demonstrate that higher level object representations could actually override lower level grouping heuristics.

Method

Subjects. The subjects were 18 Carnegie Mellon University undergraduates who participated in partial fulfillment of course requirements. All had normal or corrected-to-normal vision, and all were native speakers of English. Two subjects were excluded from the analyses due to either ceiling or floor effects (see Results section).

Stimuli. The stimuli consisted of three capital letters, K, L, and T. All possible pairs of these letters were arranged such that (1) one letter was upright and one was inverted, and (2) one letter was intact and the other was emergent. Examples appear in Figure 4. The resulting displays could appear in two orientations: (1) In the "intact-letter-upright" condition, the intact letter was in its canonical orientation (Figure 4a) and the emergent letter was rotated 180° from its canonical, upright orientation. This was a replication of the previous experiments. (2) In the "emergent-letter upright" condition, the emergent letter was in its canonical, upright position (Figure 4b) and the intact letter was rotated 180° from its canonical, upright orientation.

All of the displays were 4.4 cm tall (4.19° of visual angle). Because of the different widths of the individual letters, the displays also varied in width. The display containing the K and the T was 5.2 cm wide (4.95°); the display containing the K and the L was 6.4 cm wide (6.09°); and the display containing the L and the T was 4.0 cm wide (3.81°).

Two small Xs were placed inside the letters. The Xs were always the same spatial distance from one another, 2.5 cm, as measured by the straight line between the two Xs. The Xs appeared in a 10-point lowercase Helvetica font and measured approximately 0.2 cm wide \times 0.2 cm tall (0.19° \times 0.19°). The Xs were positioned so that they appeared either on the same region or on different regions. "Same" and "different" were defined according to whether or not one could trace from one X to the other and remain within a single contiguous region.

Construction of these stimuli resulted in three relevant factors. First, the emergent letter could appear upright or the intact letter could appear upright. Second, the Xs could appear on either the same region or on different regions. Third, the regions on which the Xs appeared could be either more familiar (i.e., an upright letter) or less familiar (i.e., an inverted letter). When the Xs appeared on a more familiar region, the Xs were said to be congruent with being located on an upright letter (irrespective of whether this letter was intact or emergent), whereas when the Xs appeared on a less familiar region, the Xs were said to be incongruent with being located on an upright letter.

Procedure. The stimuli were presented on a Macintosh Classic computer. The subjects sat approximately 60 cm from the comput-

er's monitor. Prior to the stimulus presentation, a fixation cross appeared for 800 msec. Following the fixation cross, a single stimulus was presented for 83 msec, followed by a pattern mask, which was presented for 200 msec. A brief exposure duration and a pattern mask were used to ensure that subjects were not at ceiling in their response accuracy. After the mask disappeared, the screen was blank and the subject responded. Following the response, the fixation cross reappeared, signaling the start of the next trial.

The subjects were asked to determine whether the Xs were on the same region or on different regions. The subjects were shown examples of the displays in order to ensure the clarity of these instructions. Responses were indicated via a standard keyboard; the "v" key was pressed if the Xs appeared to be on the same region, and the "b" key was pressed if the Xs appeared to be on different regions. Half of the subjects used the left hand to respond, and half used the right hand.

Each of the subjects received a total of 96 practice trials. Following this practice, each subject received four blocks of 96 trials each. The subjects were told that accuracy should be their primary concern, but they were also asked to respond in a timely manner. In accordance with these instructions, only accuracy was collected. RTs were not collected because pilot experiments had revealed high variability in the RT data; the accuracy data were much less variable both within and between the subjects. Note that this does not cause any difficulties in relating this study to the earlier experiments, because the intact-letter upright condition provides a replication of the previous studies.

Results

One subject was excluded from the analyses due to a ceiling effect (greater than 95% accuracy); another was excluded because of near-chance responding.⁵ The remaining 16 subjects' accuracy at determining whether the Xs were on the same shape or different shapes was analyzed with a repeated measures, three-factor ANOVA. The relevant factors were X location (same region vs. different regions), congruency (Xs congruent with an upright letter vs. incongruent with an upright letter), and orientation of the display (intact letter upright vs. emer-

gent letter upright). In addition, two main sets of findings are presented in order to investigate potential cooperation effects that familiarity may have on image segmentation as well as any competitive effects that familiarity may have on segmentation.

Omnibus ANOVA. The subjects' mean accuracy scores are shown in Figure 5. A three-way ANOVA revealed a significant main effect for the orientation of the display [$F(1,15) = 20.18, p < .0005$], with subjects being more accurate on the displays in which the intact letter was upright than on displays in which the emergent letter was upright. The main effect for X location (same vs. different) was not statistically reliable [$F(1,15) = 2.77, p > .10$]. The main effect for the congruency of the regions on which the Xs appeared (congruent with an upright letter vs. incongruent with an upright letter) was also not significant [$F(1,15) < 1$].

The two-way interaction between display orientation and X location was not significant [$F(1,15) = 2.63, p > .10$]. There were significant interactions between display orientation and congruency of the Xs locations [$F(1,15) = 5.71, p < .04$] and between X location and the congruency of the Xs locations [$F(1,15) = 31.94, p < .0001$]. Finally, and most importantly, these two-way interactions were subsumed by a three-way interaction that was statistically reliable [$F(1,15) = 8.22, p < .02$]. The three-way interaction suggests that accuracy in reporting the X location depends not only on the physical location of the Xs, but also on the congruency of the region(s) on which the Xs are located (i.e., whether the Xs are congruent with being located on an upright letter or not) and on whether the upright letter was intact or emergent. To better understand how these results rule out a simple recognition account of the previous studies, two additional planned analyses were performed.

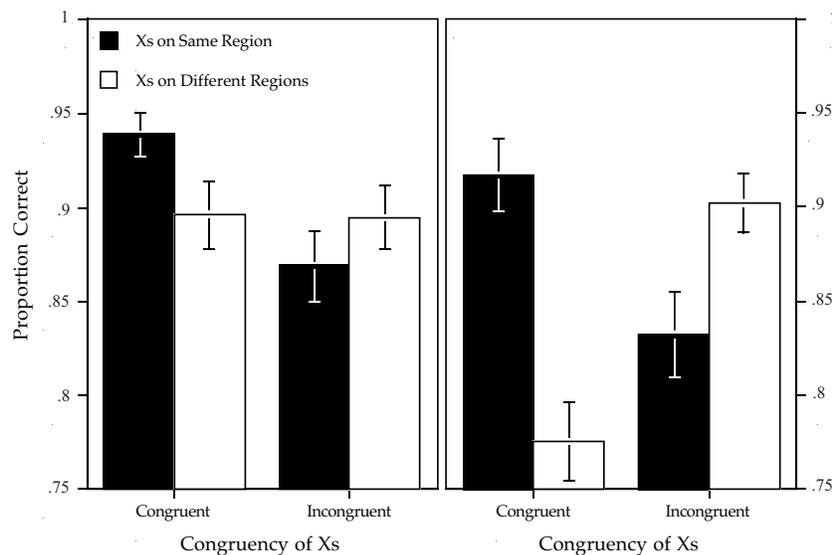


Figure 5. Results from Experiment 4. (a) The intact letter-upright condition. (b) The emergent letter-upright condition. See text for discussion.

Intact letters. We first analyzed the data from the intact-letter upright condition to ensure replication of Experiments 1 through 3. If higher level object information can cooperate with or facilitate image segmentation, then, when the stimulus configuration corresponds to a familiar shape in visual memory, top-down inputs from the object representation should partially influence segmentation. In particular, the top-down information should help increase the accuracy of the grouping. Consistent with this, when Xs were on the same region, the subjects were more accurate if this region was congruent with an upright letter than they were when this region was incongruent with an upright letter [$t(15) = 7.48, p < .02$].⁶ These data appear in Figure 5a (“same” conditions only). This result replicates the basic findings from Experiments 1 through 3, in which familiar shapes were segmented more reliably than less familiar shapes.

Emergent letters. The interactive model of segmentation predicts that when the stimulus information and the higher level object information conflict with one another the object information can override the stimulus information. Such a competitive, “overriding,” effect can be observed by analyzing the results from the displays in the emergent-letter upright condition (Figure 5b). Subjects’ accuracy with these displays was analyzed with a two-factor ANOVA, with factors being X location (same region vs. different regions) and the congruency of the region(s) on which the Xs were located (whether the Xs were located in places congruent with an upright letter or incongruent with an upright letter). There was a marginal main effect for X location [$F(1,15) = 3.20, p < .10$] and a nonsignificant main effect for congruency [$F(1,15) = 2.08, p > .15$]. However, there was a highly reliable interaction between these two variables [$F(1,15) = 29.33, p < .0001$]. This interaction suggests that the top-down information can actually *hinder* the lower level grouping when there are two physically distinct regions that, nonetheless, appear to be an upright letter; such separate regions tend to be (incorrectly) grouped together. When the Xs were on physically different regions, there was a highly significant difference between the congruent and incongruent conditions [$t(15) = 25.63, p < .0001$], with subjects being much less accurate when the Xs were congruent with being located on an upright letter. In this condition, the top-down information hinders the lower level grouping, because the grouping is not veridical; that is, the grouping was not faithful to the actual stimulus presented. This result stands in contrast to all of the previous results in which grouping, or segmentation, was always facilitated by familiarity. The results from the present experiment demonstrate that familiarity cannot only help, or facilitate, processing, but also hinder, or compete with, lower level cues to segmentation and grouping.

Discussion

There were two theoretically significant results from Experiment 4. First, consistent with Experiments 1

through 3, a cooperative effect was found between higher level and lower level information: In these studies, the subjects were more accurate in reporting Xs located on a single region if that region was congruent with an upright letter (i.e., if that region corresponded to an upright letter) than they were if that region was incongruent with an upright letter. Second, not only was a cooperative effect found, but a competitive effect was also found. When the Xs appeared on two physically distinct regions, the subjects were highly likely to report these Xs as belonging to the *same* region if these two regions formed an emergent upright letter. This likelihood of (erroneously) reporting that the Xs were on the same region was greatly reduced either (1) if the two regions were rotated 180° or, more importantly, (2) if the Xs were moved to locations that were incongruent with being located on an upright letter. Thus, depending on the placement of the Xs, higher level information about the letter could override the lower level cues of connectedness and common region.

While the cooperative results are consistent with both the bottom-up and interactive models of image segmentation, the competitive results are consistent with only the interactive approach to segmentation. If segmentation were a purely bottom-up process, then the Xs should not have been grouped together more when they appeared on regions congruent with a letter than when they appeared on regions incongruent with a letter. This is because the outputs of the lower levels of processing could not be altered by higher level representations. As noted previously, even if one attempted to modify the bottom-up account to allow grouping of unconnected regions, the entirety of the two separate regions should have been grouped together. No partial grouping could occur, because such partial grouping would imply that object information influenced the grouping, an implication that is fundamentally at odds with a bottom-up account. The results of Experiment 4 can be explained only by assuming that higher level object representations can influence the lower levels of representation—partial grouping of the two separated regions (Figure 4b) clearly occurs. Only familiarity cues from object representations stored in visual memory can provide for such partial grouping effects. This grouping via familiarity is lessened if (1) the emergent letter is rotated 180° from its canonical orientation (i.e., in the intact-letter upright condition) or (2) if the Xs are on two regions in the display that do not correspond to a familiar shape (contrast stimulus ii with stimulus iv in Figure 5b).

The finding that Xs on physically separate regions are strongly grouped when they fall on regions congruent with an upright letter and less strongly grouped when they fall on regions incongruent with an upright letter is incompatible with a bottom-up account of segmentation. However, the lower level segmentation stage could contain knowledge about objects that might allow only certain parts of two separated regions to be grouped together. We will discuss such a modified bottom-up account in the General Discussion.

GENERAL DISCUSSION

In a series of four experiments, we have demonstrated that image segmentation can be influenced by the familiarity of the shape being segmented. We also demonstrated that these results were not due to subjects' delaying their response to the image segmentation task until they had recognized the shapes being segmented. Finally, we have also demonstrated that lower level grouping heuristics, such as connectedness and common region, could be overridden by familiarity cues and that partial grouping could occur; that is, parts or subregions of two physically distinct regions could nevertheless be grouped, provided those two regions appeared to form an upright letter. As outlined in the introduction, these results are consistent with the hypothesis that image segmentation is an interactive process.

While we have attempted to argue for an interactive approach to image segmentation, there are several other issues that warrant discussion. In what follows we will discuss our interactive account and an alternative model that has been proposed for figure-ground organization. We will also discuss the generality of our results as well as the locus of knowledge in the visual system.

The Interactive Account

On the basis of our knowledge-based effects in segmentation, we have attempted to argue for an interactive model of image segmentation. When viewed from the perspective of a traditional, hierarchical/sequential model of visual processing, knowledge effects in segmentation seem paradoxical. If we assume that the goal of sensory segmentation is to provide the correct inputs to object representations, then how can object representations influence segmentation prior to receiving inputs from the earlier segmentation processes?

This apparent paradox disappears if one adopts the concepts of interactive, cascaded information processing (see McClelland, 1979; McClelland & Rumelhart, 1981; also see Palmer & Rock, 1994a, for an application to perceptual organization). In cascaded systems, in which preliminary results of partial processing at one stage are available to the next stage, feedback from a later stage can guide processing in an earlier stage (McClelland, 1979). Thus, this paradox can be resolved by an appeal to mechanisms used by connectionist or parallel distributed processing (PDP) networks (Feldman & Ballard, 1982; McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986). Specifically, while the image is being segmented, partial activation is being sent to higher, object-level representations. These partial results could be due to bottom-up processes, such as Gestalt grouping heuristics. However, as matches are made with higher level object representations, as presumably happens more strongly with upright letters, these representations, in turn, send activation back to the image representation. This top-down activation will reinforce groupings that correspond to fa-

miliar objects and thus allow the segmentation process to finish faster than it would if there was no such activation. Furthermore, such top-down influences may be able to actually alter, not just facilitate, processing at lower levels (Vecera & O'Reilly, in press). This overall scheme is similar to that suggested by the interactive activation model of word recognition (McClelland & Rumelhart, 1981).

This framework leads to image segmentation as being conceptualized as a graded and dynamic process instead of as an "all-or-none" process (see, e.g., Höffding, 1891). Specifically, the visual field may not be segmented in a single time step; rather, perceptual organization might reach a stable solution (e.g., some certain segmentation) only after a certain amount of time. During this time, partial results of the segmentation process are sent to higher level object representations, as discussed above. Again, if matches are made, then top-down activation can partly guide the segmentation process. Of course, this framework can also account for the segmentation of images containing completely novel objects. In this case, the process proceeds either (1) without support from object representations or (2) with partial support if there is some degree of similarity between the novel object and stored object representations. This explains subjects' ability to segment the nonletter shapes used in Experiment 1.

An Alternative Model?

While interactive mechanisms have become increasingly popular in cognitive science, there are alternative models that can describe how object-recognition processes could impact upon earlier processes. Indeed, one such alternative model has been proposed to explain behavioral results from figure-ground organization.

In numerous studies, Peterson and her colleagues (Peterson & Gibson, 1991, 1993, 1994a, 1994b; Peterson et al., 1991; see Peterson, 1994, for an overview) found that object or shape-recognition processes contributed to figure-ground organization. They concluded that figure-ground organization was not necessarily a precursor to object (or shape) recognition, as would be assumed by more traditional, hierarchical models of visual processing (e.g., Biederman, 1987; Kosslyn, 1987; Marr, 1982; Neisser, 1967). Rather, object recognition could operate in parallel with, or prior to, figure-ground organization, as seems to be suggested by a "prefigural" account (e.g., Peterson, 1994). They proposed that shape-recognition processes could operate along both edges of a luminance contour. Thus, in this scheme, object-recognition processes operate before figure-ground organization.

Could such a "prefigural" account apply to the present results? To adopt this type of account to our results, we would also have to hypothesize that image segmentation operated in parallel with or after object recognition. However, object-recognition processes could, in principle, operate prior to segmentation, instead taking inputs from contour information that presumably exists prior to

segmentation (see Peterson, 1994, for a summary of such a model).

Although a “prefigural” model fully explains the figure-ground results, others have argued against such a model. For example, Palmer and Rock (1994a) have argued against Peterson’s “prefigural” model on logical grounds. One of the goals of figure-ground organization is to provide appropriate inputs to object representations; thus, figure-ground organization must precede object-recognition processes. To account for the behavioral results, Palmer and Rock (1994a) proposed cascaded, interactive information flow between figure-ground organization and object-recognition processes.

Furthermore, Vecera and O’Reilly (in press) used a graded, interactive PDP network to demonstrate that such a network could simulate many of Peterson’s behavioral results. In this interactive approach to figure-ground organization, object-recognition processes operated *after* figure-ground organization and took their inputs from the earlier figure-ground processes. Yet the interactive model still exhibited an influence of object knowledge over figure-ground organization, replicating the behavioral results (see Peterson, 1994, for an overview of these results). The “prefigural” and interactive accounts might be difficult to distinguish from one another empirically. However, the interactive approach is particularly powerful because one does not need to reject the hierarchical notions of information processing that were implemented in the more traditional vision models; the only revision required of the traditional models is to allow for graded, interactive information processing.

While the simulations of figure-ground processing discussed above argue for an interactive approach, some may perceive discrepancies between the Vecera and O’Reilly (in press) model and the present results. In their model, Vecera and O’Reilly noted that ambiguity in the input image was particularly important in engaging top-down influences from object representations. Is this a flaw of the model (and the interactive approach), given that the displays used here may not have been ambiguous? We think not, for two reasons. First, the displays used in Experiment 4 were ambiguous in the sense that they could be organized in multiple ways (i.e., use only bottom-up knowledge or use both bottom-up and top-down knowledge). This is consistent with the model, because, when standard figure-ground stimuli were presented, the inherent ambiguities (i.e., the fact that either of two regions could be figure) were sufficient to invoke top-down influences. The same type of ambiguities are present in the displays used here; the probed locations could be on the same shape or on different shapes. Second, and more importantly, the importance of ambiguity in the model became evident only when *unambiguous* input was presented. If an unambiguous, yet unfamiliar, shape is presented (as in the case of a random-dot stereogram; see Peterson & Gibson, 1993), top-down influences are *unable* to overcome powerful bottom-up cues. *Some degree* of ambiguity is required to overcome bottom-up cues

in such situations (Vecera & O’Reilly, in press). The main conclusion from these simulation studies is that while bottom-up cues are sometimes sufficient for processing, these cues do not act alone; top-down cues, on the basis of familiarity, also appear to influence perceptual organization.

Finally, while our interactive account may seem fundamentally at odds with the “prefigural” account articulated by Peterson (1994), there may be some commonalities. For example, in an interactive model, particularly one with cascaded information processing, object-recognition processes and perceptual organization processes would overlap temporally. This would, in effect, allow object recognition and perceptual organization to occur in parallel, which could be consistent with Peterson’s “prefigural” model. Of course, the prefigural and interactive accounts also differ in some respects, and such differences should surely be the focus of future studies and computational investigations.

Generality of the Results

A second issue that deserves discussion concerns the generality of the present results. We have used a rather limited stimulus environment; our shapes were all transparent letters, which may have biased the results to favor interactive processing instead of bottom-up processing. It might be that such impoverished stimuli take longer to process because of the lack of several real-world cues, such as occlusion, texture, or color. As a result, this hypothesized increase in processing time may have provided enough time for feedback from object representations to guide segmentation.

While our stimuli do restrict the extent to which we can argue for interactive segmentation as the norm, there are considerations that suggest our results would generalize to displays with additional cues. First, the result that figure-ground organization may also be an interactive process (see discussion previously) indicates that interactive processing may generalize to other displays and tasks. Second, we have unpublished results from other displays and tasks that are also consistent with an interactive account. We have replicated the basic effects reported here with a simple “inside-outside” judgment; if subjects are asked to determine whether an X is inside or outside a shape, they are faster if this shape is an upright letter than if it is a rotated letter. The overall RTs in this task were, on average, 465 msec, which is faster than most of the RTs in the present tasks. Thus, evidence can be found for interactive segmentation in tasks that can be performed more quickly than the present series of tasks. Furthermore, we have also collected data with displays similar to those used in Experiments 1 through 3 except that they contained occlusion cues—one letter occluded the other. While these displays are still quite simple, they do contain additional cues, yet subjects still show faster segmentation times to the upright letters than they do to the rotated letters. Finally, the numerous findings of interactive processing in other visual tasks, such as the word superiority effect, suggest that interactive pro-

cessing may be a more general information-processing strategy. As such, we would expect our results to generalize to other, more complex displays.

Knowledge in the Visual System

Finally, the interactive account that we have attempted to develop implies that object or shape knowledge resides at a relatively late stage of information processing (i.e., at the level of object representations stored in visual memory). However, we should note that this does not preclude the possibility that other forms of visual knowledge might exist at other, earlier stages of visual processing. Indeed, an approach based on the mechanisms of PDP networks should also allow for knowledge throughout the visual system.

Interactive image segmentation is one form of *knowledge-based* image segmentation. There are several mechanisms by which knowledge can influence image segmentation. As just mentioned, knowledge could be stored only at the level of object representations. However, knowledge could also be stored at the same level of representation as the image. Local conjunctions of features could, in principle, be learned and stored in the connections among the feature representations in the image. This is essentially how Mozer et al.'s (Mozer, Zemel, & Behrmann, 1992; Mozer, Zemel, Behrmann, & Williams, 1992) model of image segmentation codes statistical information contained in images.

These simulation results indicate that knowledge may be coded at all levels of visual processing. Knowledge of image statistics, such as vertex information or line-termination information (e.g., Waltz, 1975), may be stored in weighted connections at earlier levels of processing. Knowledge of more global shape characteristics, such as the overall shape of an object, might be coded at later levels of visual processing. Such a view of visual knowledge suggests that there may be no "information encapsulation," but, instead, that our perception may be the result of statistical regularities that have shaped our visual systems at all levels.

REFERENCES

- BIEDERMAN, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213-263). Hillsdale, NJ: Erlbaum.
- BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.
- DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, **113**, 501-517.
- FELDMAN, J. A., & BALLARD, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, **6**, 205-254.
- FINKEL, L. H., & SAJDA, P. (1992). Object discrimination based on depth-from-occlusion. *Neural Computation*, **4**, 901-921.
- FISHER, R. A. (1948). Combining independent tests of significance. *American Statistician*, **2**, 30.
- FODOR, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- GIBSON, B. S., & PETERSON, M. A. (1994). Does orientation-independent object recognition precede orientation-dependent recognition? Evidence from a cuing paradigm. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 299-316.
- GROSSBERG, S. (1993). A solution of the figure-ground problem for biological vision. *Neural Networks*, **6**, 463-483.
- HÖFFDING, H. (1891). *Outlines of psychology*. New York: Macmillan.
- 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.
- JOLICOEUR, P., ULLMAN, S., & MACKAY, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition*, **14**, 129-140.
- KOSSLYN, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, **94**, 148-175.
- LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, **1**, 476-490.
- LOWE, D. G. (1985). *Perceptual organization and visual recognition*. Boston: Kluwer.
- MARR, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London*, **275**, 483-524.
- MARR, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- 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., & 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.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models*. 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 Conference 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.
- PALMER, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, **3**, 519-526.
- PALMER, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology*, **24**, 436-447.
- PALMER, S. [E.], & 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., & 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 & Performance*, **17**, 1075-1089.
- PRINZMETAL, W. (1981). Principles of feature integration in visual perception. *Perception & Psychophysics*, **30**, 330-340.
- PRINZMETAL, W., & MILLIS-WRIGHT, M. (1984). Cognitive and linguistic factors affect visual feature integration. *Cognitive Psychology*, **16**, 305-340.

- REICHER, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, **81**, 274-280.
- ROCK, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
- ROCK, I., & GUTMAN, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 273-285.
- RUBIN, E. (1958). Figure and ground. Translated in D. C. Beardslee & M. Wertheimer (Eds.), *Readings in perception* (pp. 194-203). Princeton, NJ: Van Nostrand. (Original work published 1915)
- RUMELHART, D. E., & MCCLELLAND, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 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.
- TARR, M. J., & PINKER, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, **21**, 233-282.
- ULLMAN, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, **32**, 193-254.
- VECERA, S. P., & O'REILLY, R. C. (in press). Figure-ground organization and object recognition processes: An interactive account. *Journal of Experimental Psychology: Human Perception & Performance*.
- WALTZ, D. (1975). Understanding line drawings of scenes with shadows. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 19-91). New York: McGraw-Hill.
- WERTHEIMER, M. (1958). Principles of perceptual organization. Translated in D. C. Beardslee & M. Wertheimer (Eds.), *Readings in perception* (pp. 115-135). Princeton, NJ: Van Nostrand. (Original work published 1923)
- WHEELER, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, **1**, 59-85.
- WINER, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.
- WOLF, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage Publications.

NOTES

1. There may also exist other classes of models that do not fall neatly within the "bottom-up" and "interactive" alternatives. For example, Peterson and her colleagues (see Peterson, 1994) have proposed a model of figure-ground perception in which prefigural object-recognition processes operate before figure-ground organization. Such a model may, however, be interpreted within a bottom-up framework (see Vecera &

O'Reilly, in press), and we reserve discussion of this specific model for the General Discussion.

2. Because of this marginal effect, we conducted a meta-analysis on the comparisons between upright and rotated letters in Experiments 1 and 2. Experiment 3 was not used because of ceiling effects on the accuracy data. Accuracy was higher with the upright letters than with rotated letters in both experiments, and this effect was significant across both studies when tested with a Winer combined test (see Winer, 1971; Wolf, 1986)— $Z_c = 3.34$, $p < .001$. This analysis was also confirmed with a Fisher combined test (see Fisher, 1948; Wolf, 1986)— $\chi^2(4) = 15.98$, $p < .01$.

3. This same finding was confirmed in a regression of all of the "same" responses for the upright and rotated letters across all subjects. If the scanning rates within upright and rotated letters were different, then there should be a reliable interaction between letter orientation and the distance between the Xs. Results from a stepwise regression demonstrated that the RT data were best fit by the letter orientation and the distance factors, but the interaction term did not significantly increase the amount of variance captured by the regression model. Including the letter orientation and distance factors accounted for 13.7% of the variance ($R^2 = .137$), while adding the interaction to this model accounted for no additional variance ($R^2 = .137$).

4. There were three exceptions to this. These three cases could not be transformed this easily because the resulting overlapping pattern did not leave enough open space in which to place the Xs on both letters while maintaining the same distance between the Xs in the "same" and "different" conditions. These three stimuli were the overlapping A and Y, F and L, and K and F, and these displays were created by rotating one of the letters and then repositioning that letter so as to facilitate placement of the Xs.

5. These subjects' data were excluded on the basis of overall accuracy; to prevent experimenter bias from influencing rejection of these subjects, the specific pattern of results was not examined. A recent reexamination of these data, however, showed that these 2 subjects showed patterns that were qualitatively similar to those of the remaining 16 subjects. The primary differences were in baseline accuracy.

6. This effect held for the emergent letter upright conditions: For "same" trials, subjects were more accurate when the Xs were on regions that were congruent with an upright letter relative to when the Xs were on regions that were incongruent with an upright letter [$t(15) = 11.29$, $p < .005$]. This same effect was highly significant when the data from the intact letter upright and emergent letter upright conditions were pooled [$t(15) = 18.58$, $p < .0007$].

(Manuscript received November 3, 1994;
revision accepted for publication December 18, 1996.)