The initial identification of figure-ground relationships: Contributions from shape recognition processes

MARY A. PETERSON and BRADLEY S. GIBSON
University of Arizona, Tucson, Arizona

We tested whether shape recognition computations contribute to the initial determination of figure and ground by asking observers to report about which region of briefly exposed masked stimuli appeared to be figure. The region on one side of a figure-ground contour always denoted a meaningful shape. In *full versions* of the stimuli, Gestalt variables favored the other, less denotative, region; in *half versions* of the stimuli, Gestalt variables were absent. If shape recognition inputs do not contribute to figure-ground computations, the less denotative region should be seen as figure in the full stimuli, and both regions should be seen as figure equally often in the half stimuli. Results show otherwise, suggesting that shape recognition routines do contribute to the initial determination of figure and ground, provided they have been completed, which seems to require approximately 150 msec for canonically oriented stimuli.

Perceptual theories must take a position on the question of when, in the stream of visual processing, pattern identification occurs. Most recent theorists have chosen to assume that figure-ground organization precedes shape recognition (e.g., Biederman, 1987; Marr, 1982), reasoning that one cannot recognize a shape until one has perceived it.

The process of perceiving a shape includes the determination of figure-ground relationships. Indeed, the ability to recognize a shape is coupled to figure-ground organization in that the region to which the figure-ground contour is assigned has shape and meaning, whereas the other region is shapeless and simply appears to continue behind the figural region, as the Rubin vase/faces stimulus demonstrates.

The Gestalt theorists proposed that the assignment of a figure-ground contour is governed by a number of variables, all of which can be measured on the physical stimulus. Using displays devoid of meaningful shapes, they showed a tendency for figure-ground contours to be assigned to regions that were symmetric, enclosed, more convex, or smaller in area, thereby demonstrating that figure-ground relationships *can* be determined without shape recognition input. However, because the meaningfulness of the shape denoted by the regions competing for the figure-ground contour was not manipulated, those experiments do not show that figure-ground computations *cannot* weigh shape recognition inputs.

Opinions about whether or not figure-ground organization must precede shape recognition hinge on solutions to the *Hoffding step*, the step at which a stimulus contacts the appropriate representation for recognition among the myriad of stored memory representations. In some views, the Hoffding step is simplified somewhat by the prior identification of figure-ground relationships (or, in Marr's terminology, of the "2½-D sketch"), so that attention can then be directed to identifying the figural region (Neisser, 1967). Accordingly, within current theoretical frameworks, the Hoffding step is accomplished by first identifying the figure, then partitioning the contour into parts delimited by concave regions (as defined from inside the figure), and mapping the resultant parts onto the best-fitting combination of representational components in shape memory (e.g., Biederman, 1987; Hoffman & Richards, 1985; Marr, 1982).

In our view, however, it is not mandatory that the identification of the figural region necessarily precede the operation of all shape recognition processes. Questioning the same assumption, Peterson, Harvey, and Weidenbacher (in press) proposed that contours might be partitioned from both sides simultaneously before figure-ground relationships are determined. This proposal is at least logically possible, given the substantial evidence that luminance contours can be identified preattentively (e.g., see Marr, 1982). If figure-ground contours can be partitioned from both sides prior to the determination of figure-ground relationships, then the outcomes of the resultant shape recognition routines might serve as input to figure-ground computations.

Peterson et al. (in press) tested this hypothesis using the figure-ground stimuli depicted in Figures 1D and 1E.[1] These stimuli were biased by Gestalt variables to favor the center-as-figure interpretation, yet were drawn so that
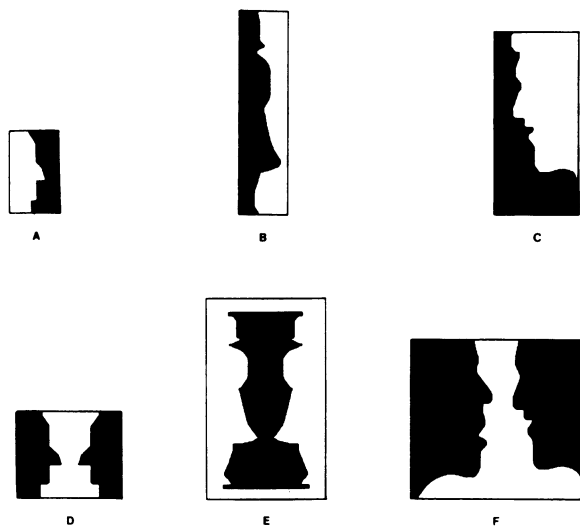
Figure 1. Half stimuli are depicted in A, B, and C; full stimuli are depicted in D, E, and F.

the surrounding region denoted a meaningful shape, whereas the center did not. (Hereafter, the center and the surround will be referred to as the *less denotative* region and the *more denotative* region, respectively.) Peterson et al. (in press) proposed that shape recognition processes would contribute to figure–ground computations only when they were completed within some time period critical for the determination (or redetermination) of figure–ground relationships. They tested this proposal by manipulating the orientation of the stimuli. Jolicoeur (1985) and others have shown that incrementally disorienting stimuli from their canonical uprights incrementally increases the time required for shape identification. However, there is no reason to expect that disorienting figure–ground stimuli will affect the computation of the Gestalt variables. Hence, the stimulus bias toward the center-as-figure interpretation should not be changed by disorientation.

Peterson et al. (in press) found that when the stimuli were upright rather than inverted, the more denotative surround was both *maintained* as figure longer and *obtained* as figure more often by reversal out of the less denotative center-as-figure interpretation. These results indicated that figure–ground reversal computations can receive input from at least one shape recognition route, provided that the shape recognition computations are completed rapidly enough.

THE EXPERIMENT

We tested whether or not input from shape recognition routines contributes to the *initial* determination of figure–ground relationships by examining figure–ground responses to briefly exposed upright and inverted stimuli. We compared performance with two types of stimuli: Half of our stimuli were *full stimuli*, depicting a center and a surrounding region (see Figures 1D, 1E, and 1F). For

these full stimuli, the Gestalt variables of relative area, convexity, and/or symmetry, and/or enclosure favored the center-as-figure interpretation and denotivity favored the surround-as-figure interpretation. The other half of our stimuli were *half stimuli*, in which a vertical contour separated two regions that were minimally differentiated by Gestalt variables,[2] but that differed in their denotivity; specifically, one region was more denotative than the other (see Figures 1A, 1B, and 1C).

Different groups of observers viewed the stimuli for 50, 100, 150, 200, 300, or 400 msec, and reported which of the two regions appeared to be figure. Performance with half stimuli, in which no Gestalt variables distinguish between the two regions, was compared to performance with full stimuli, in which Gestalt input favors the center as figure and shape recognition input favors the surround as figure.

We expected that, at some exposure duration, the more denotative region of the upright half stimuli would be chosen as figure more often than would the less denotative region, indicating that shape recognition inputs can contribute to figure–ground organization when Gestalt variables are absent and providing some measure of the time necessary to complete the relevant shape recognition computations.

For the full upright stimuli, we expected very different outcomes depending on whether initial figure–ground organization is determined by Gestalt variables alone or by Gestalt variables in concert with shape recognition inputs. Specifically, if Gestalt variables operate first to identify the figural regions and shape recognition routines are then initiated to identify only the figural region, then observers should initially favor seeing the center as figure. Moreover, this bias toward center as figure should last at least long enough for the center shape to be recognized and for a reversal out of the center-as-figure organization to occur. Time courses for these processes are difficult to assess, but there is some evidence that reversals do not generally occur within the first 500 msec of viewing (Ohrbach, Zucker, & Olson, 1966).

Alternatively, if both Gestalt variables and shape recognition inputs operate jointly to determine figure–ground relationships in the full upright stimuli, then observers' responses should reflect that interaction. Specifically, we should obtain approximately equal reports of the center and surround as figure when the strengths of the combined inputs favoring the center and the surround are approximately equal, and we should obtain more surround-as-figure reports when the combined inputs favoring the surround are stronger than those favoring the center, and vice versa.

Given the findings of Peterson et al. (in press), we expected to obtain fewer reports that the more denotative region appeared to be figure in the inverted versions of both half and full stimuli.

Method

Stimuli and Apparatus. The stimuli were the drawings shown in Figure 1. There were two versions of each of the full and half stimuli, with the black/white color relationships reversed.

Table 1 lists the denotivity of each region of each stimulus, as measured in a study in which observers listed all the objects or shapes resembled by the region. (For details, see Peterson, Rose, Gibson, & Vezey, 1991.) The two regions competing for the figure–ground contour differ in denotivity, and there is some variation in the degree of denotivity even among the more denotative regions. Table 1 also lists measures of the Gestalt variables of relative area and convexity, and indicates whether or not each region is symmetric and/or enclosed.

The stimuli were presented on a Princeton Ultrasync monitor at a distance of 105 cm from the subjects. The figures subtended visual angles ranging from 2.1° to 4.7° in height and 1.2° to 4.0° in width. Each stimulus had a 0.65° gray border.

Procedure. Each subject served in 1 of 12 conditions (2 conditions of stimulus orientation [upright or inverted] × 6 exposure duration conditions). Within each condition, each stimulus was exposed twice, once in each black/white combination. Order of stimulus presentation was counterbalanced. No stimulus was repeated until after all stimuli had been viewed once, and no more than three stimuli with the same color denotative region followed consecutively.

Each stimulus was preceded by a fixation point and followed immediately by a 100-msec pattern mask. A chinrest was used. The subject pressed one of two keys to indicate which region of the figure–ground stimulus appeared to be figure, with response mapping counterbalanced across subjects. The 12 experimental trials were preceded by a discussion of figure–ground organization and by three practice trials with the Rubin vase/faces stimulus.

Results and Discussion

For each stimulus, we computed the probability of a surround-as-figure response.[3] We averaged these proba-

Table 1
**Strength of Gestalt and Shape Recognition Measures
for Each Region of Each Figure–Ground Stimulus**

	Half Stimuli					
	Figure 1A		Figure 1B		Figure 1C	
	center	surround	center	surround	center	surround
Denotivity of Regions						
	0.05	0.61	0.20	0.55	0.05	0.70
Gestalt Variables						
Area	0.50	0.50	0.48	0.52	0.49	0.51
Cvx	0.63	0.37	0.65	0.35	0.70	0.30

	Full Stimuli					
	Figure 1D		Figure 1E		Figure 1F	
	center	surround	center	surround	center	surround
Denotivity of Regions						
	0.22	0.61	0.28	0.58	0.28	0.92
Gestalt Variables						
Area	0.41	0.59	0.40	0.60	0.28	0.72
Cvx	0.63	0.37	0.65	0.35	0.69	0.31
Symm	1.00	0.00	1.00	0.00	0.00	0.00
Encl	0.00	0.00	1.00	0.00	0.00	0.00

Note—Area = relative area. Cvx = relative convexity of the two regions, measured by extending a plumb line to both the top and the bottom of the frame from the deepest concavity into each region, and taking the area outside that plumb line as a gauge of how much that area extends into the other area. Symm = symmetry. Encl = enclosure; presence of these two variables is indicated by 1.00, absence by 0.00.

Table 2
**Proportion of Observers Reporting the Surround as Figure
for Each Stimulus at Each Exposure Duration**

	Exposure Duration (msec)					
	50	100	150	200	300	400
Half Figures						
Upright						
Figure 1A	.58	.54	.92*	.79*	.96*	.96*
Figure 1B	.54	.62	.96*	.75*	.71*	.75*
Figure 1C	.62	.79*	.83*	.79*	.88*	.88*
Inverted						
Figure 1A	.50	.62	.67	.79*	.75*	.54
Figure 1B	.42	.62	.50	.54	.62	.50
Figure 1C	.38	.42	.58	.75*	.67	.67
Full Figures						
Upright						
Figure 1D	.21†	.29†	.38	.58	.71*	.67
Figure 1E	.25†	.29†	.38	.33	.42	.54
Figure 1F	.54	.67	.79*	.79*	.83*	.83*
Inverted						
Figure 1D	.42	.17†	.50	.21†	.42	.38
Figure 1E	.33	.12†	.29†	.08†	.29†	.38
Figure 1F	.58	.33	.46	.42	.67	.71*

*Above-chance surround-as-figure reports.   † Below-chance surround-as-figure reports, which correspond to above-chance center-as-figure reports [$t(11) >$ or < 1.79].

bilities across subjects and tested whether they differed from chance. Results are shown in Table 2.

Half figures. In the 150-msec condition, and in all longer exposure conditions, we obtained above-chance reports that the more denotative region was figure for all the upright half stimuli. Chance performance was obtained in shorter exposure conditions, except the 100-msec condition in which the observers reported seeing the more denotative region of Figure 1C as figure more often than chance.

Performance with inverted half stimuli differed from performance with upright half stimuli. In most exposure conditions, the observers performed at chance. A few above-chance reports that the more denotative regions appeared to be figure were not reliable in longer exposure conditions.

These results show that the outputs of shape recognition routines can influence figure–ground organization when they are completed rapidly enough (as they were in the upright condition, but not in the inverted condition), at least when they do not compete with Gestalt variables. Moreover, the different time courses over which shape recognition inputs were available for Figure 1C versus Figures 1A and 1B suggest that shape recognition computations may be completed faster for regions of greater denotivity.

Full figures. Figure–ground reports about upright full stimuli varied with figure type, which in itself suggests some interaction between Gestalt variables and shape recognition variables. For Figure 1F, the observers never favored the center-as-figure interpretation (i.e., their surround-as-figure reports never fell below chance). In the 150-msec condition (and in all longer exposure conditions), the observers' surround-as-figure reports for

Figure 1F were above chance, suggesting that a denotivity value of .92 for the surround was sufficient to overpower the Gestalt variables of greater convexity and smaller relative area. The results were different for Figures 1D and 1E, for which below-chance surround-as-figure reports were obtained in both the 50-msec and the 100-msec conditions.

Must these below-chance choices of surround as figure for upright versions of Figures 1D and 1E imply that Gestalt variables first determine which region is to be seen as figure and that shape recognition routines are only then initiated for those figural regions? We think not, for the following reasons.

First, below-chance surround-as-figure reports were not obtained consistently across figures in the briefest exposure conditions.

Second, if Gestalt variables were determining figure-ground organization at the briefest exposures, then performance with inverted stimuli should have been the same as performance with upright stimuli. That was not the case in the 50-msec condition, although it was the case in the 100-msec condition. Performance with both upright and the inverted versions of Figures 1D and 1E in the 100-msec exposure condition may reflect the early operation of symmetry as a Gestalt organizing principle, since symmetry was the only Gestalt variable reliably distinguishing Figure 1F from Figures 1D and 1E. Alternatively, since symmetry can be detected rapidly (Garner, 1974), the observers required to make a two-alternative forced-choice response following a 100-msec masked presentation may have responded on the basis of symmetry differences because that was the only detectible difference between the two regions at that point in processing. But the mere fact that visual routines identifying symmetric regions were completed does not imply that figure-ground relationships were determined.

Third, if Gestalt variables were first defining the center as figure, then the center should have been maintained as figure at least long enough for shape recognition routines to be conducted and for some process to cause a reversal into the surround-as-figure organization. Consequently, one might expect to obtain below-chance surround-as-figure reports in longer exposure conditions as well. However, results show that the observers perceived the center and the surround as figure approximately equally often when viewing upright full versions of Figures 1D and 1E for 150 msec or longer, with some above-chance surround-as-figure reports. Remember that the presence of inputs from shape recognition routines can be signaled by chance performance, because Gestalt and shape recognition inputs are expected to interact. The results obtained with half stimuli had indicated that shape recognition inputs should be operative in the 150-msec exposure condition. Our findings suggest that, for the full versions of upright Figures 1D and 1E exposed for 150 msec or more, the combined inputs favoring the center as figure may have been as strong as the combined inputs favoring the surround as figure. Thus, these results are consistent with the interpretation that, immediately upon exposure to a stimulus, both routines detecting symmetry and shape recognition computations are initiated for the regions on both sides of a luminance contour.

Fourth, in the full figures, fixation was always located on the center region. Harvey, Peterson, and Gibson (1990) have recently found that fixated regions are likely to be figural regions. Hence, choices of the center as figure may reflect the operation of a fixation variable rather than, or in addition to, the Gestalt variables.

Below-chance surround-as-figure reports were obtained for inverted full versions of Figures 1D and 1E exposed for as long as 300 msec, as expected on the basis of previous work with stimuli disoriented from their canonical upright.

CONCLUSION

The results of this experiment are consistent with Peterson et al.'s (in press) proposal that luminance contours can be partitioned from both sides simultaneously prior to the determination of figure-ground relationships. Our evidence, obtained using brief masked presentations of figure-ground stimuli, indicates that initial figure-ground computations weigh shape recognition inputs, as well as inputs from routines assessing other variables such as symmetry, relative area, and so forth. Further research must identify the rules for combining these different inputs.

REFERENCES

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

GARNER, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.

HARVEY, E. M., PETERSON, M. A., & GIBSON, B. S. (1990, April). *Mechanisms of intention in figure-ground perception*. Poster presented at the Rocky Mountain Psychological Association Meeting, Tucson, AZ.

HOFFMAN, D. D., & RICHARDS, W. A. (1985). Parts of shape recognition. In S. Pinker (Ed.), *Visual cognition* (pp. 65-96). Cambridge, MA: MIT Press.

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

MARR, D. (1982). *Vision*. New York: Freeman.

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

OHRBACH, J., ZUCKER, E., & OLSON, R. (1966). Reversibility of the Necker cube: VII. Reversal rate as a function of figure-on and figure-off durations. *Perceptual & Motor Skills*, **22**, 615-618.

PETERSON, M. A., HARVEY, E. M., & WEIDENBACHER, H. J. (in press). Shape recognition contributions to figure-ground reversals. *Journal of Experimental Psychology: Human Perception & Performance*.

PETERSON, M. A., ROSE, P. M., GIBSON, B. S., & VEZEY, E. (1991). *A new corpus of figure-ground stimuli*. Manuscript in preparation.

NOTES

1. Peterson, Harvey, and Weidenbacher (in press) used a slightly different version of Figure 1D.

2. Where the two regions were differentiated by Gestalt variables, those variables favored the less denotative region as figure.

3. Let "surround" = *high denotative*, for the half figures. Below-chance surround-as-figure reports indicate above-chance center-as-figure reports.