The reference frame of figure–ground assignment

SHAUN P. VECERA
University of Iowa, Iowa City, Iowa

Figure–ground assignment involves determining which visual regions are foreground figures and which are backgrounds. Although figure–ground processes provide important inputs to high-level vision, little is known about the reference frame in which the figure's features and parts are defined. Computational approaches have suggested a retinally based, viewer-centered reference frame for figure–ground assignment, but figural assignment could also be computed on the basis of environmental regularities in an environmental reference frame. The present research used a newly discovered cue, lower region, to examine the reference frame of figure–ground assignment. Possible reference frames were misaligned by changing the orientation of viewers by having them tilt their heads (Experiments 1 and 2) or turn them upside down (Experiment 3). The results of these experiments indicated that figure–ground perception followed the orientation of the viewer, suggesting a viewer-centered reference frame for figure–ground assignment.

Figure–ground assignment involves organizing a visual scene into occluding foreground figures and occluded backgrounds. Assigning regions as either "figure" or "ground" is important because before an object can be recognized, attended to, or acted upon, that object must be isolated from the nonobject "grounds" that lie between objects. Thus, figure–ground assignment forms a foundation for, and provides the inputs to, high-level visual processes.

Gestalt psychologists were the first to recognize the importance of figure–ground assignment and to demonstrate several cues for determining figural regions. Smaller, symmetric, horizontally or vertically oriented, and convex regions are more likely to be perceived as "figure" than as "ground" (Palmer, 1999; also see Pomerantz & Kubovy, 1986; Rock, 1975, 1995; Rubin, 1915/1958). Several figure–ground cues have been added to the Gestalt psychologists' original list of cues: Spatial frequency (Klymenko & Weisstein, 1986) and temporal structure (Lee & Blake, 1999) influence figure–ground assignment; regions depicting familiar objects are perceived as figures (Peterson, 1994, 1999), as are regions to which exogenous spatial attention is summoned (Vecera, Flevaris, & Filapek, 2004); and regions in the lower portion of a stimulus array are perceived as "figures" more frequently than are regions elsewhere (Vecera, Vogel, & Woodman, 2002).

A general issue in visual cognition is the reference frame used to represent the shape and other spatial properties of an object. A *reference frame* is a coordinate system used to define the locations of a shape's features or parts (Pinker, 1984; Tarr, 1994). As in Cartesian coordinates, a reference frame establishes a set of parameters (e.g., an origin, axes, metric directions, etc.) from which features and parts can be defined. The visual system has several potential reference frames for representing shape. For example, an object's parts could be defined in an object-centered reference frame in which the parts are represented with respect to an origin on the object itself. Many have claimed that such an object-centered reference frame is the most advantageous frame for object recognition, because the object's representation will remain stable across changes in position, size, and orientation. Another reference frame available for visual representation is a viewer-centered frame, in which a shape is represented with respect to its visible surfaces and parts. A viewer-centered reference frame thus defines a shape with respect to what an observer can see; these frames can be defined relative to the observer's retina, head, or body. Shapes can also be represented in an environment-centered reference frame, in which the shape is defined with regard to an environmental origin, such as gravitational upright or the location and orientation of walls in a room.

Visual reference frames have been studied extensively in the domain of visual object recognition (see Biederman, 1987; Biederman & Gerhardstein, 1993, 1995; Rock, 1973; Rock & DiVita, 1987; Tarr, 1995; Tarr & Bülthoff, 1995, 1998; Tarr & Pinker, 1989), but the reference frame of figure–ground assignment has been studied little. Object-centered reference frames are ruled out by orientation-dependent familiarity effects in figure–ground assignment: Familiar regions are perceived as figures only when they appear in their canonical, upright orientation (Peterson, 1994, 1999). Viewer- and environment-centered reference frames have not been studied, perhaps because of methodological difficulties. Studies of viewer- and environment-centered visual reference frames need

to disentangle the different frames by changing the orientation of either the stimulus or the viewer, and it is difficult to disentangle these frames in figure–ground displays. Consider a figure–ground display that contains a convex region and a concave region that share a central contour. The convex region is preferred as the figure. The region will remain convex, however, if the stimulus or the viewer is rotated, making it impossible to determine whether the convex figure is represented relative to the image on the viewer's retina or to an origin on the convex region itself (an object-centered reference frame).[1] This inability to distinguish different reference frames holds for most gestalt figure–ground cues, but lower region, symmetry, familiarity, and orientation are possible exceptions to this rule.

In the present study, the lower-region cue was used to disentangle viewer- and environment-centered reference frames of figure–ground assignment. Participants are more likely to perceive a region in the lower portion of a figure–ground display as the "figure" than to so perceive regions higher in the display (Figure 1A). This lower-region preference can be used to investigate the reference frame of figure–ground assignment, because rotating the viewer can disentangle different reference frames. For an upright viewer, the lower figural region in a display could be defined based on its position with respect to the viewer (viewer-centered representation), its position in the display (object-centered representation), or its position in the environment (i.e., the lower region is always closer to the floor; environment-centered representation). However, if the viewer observes this stimulus with a 90° head tilt, then the retinal image rotates, misaligning the viewer- and object- and environment-centered representations. Specifically, because the retinal image is tilted approxi-

mately 90°, the lower region will then be represented on the left or right of the retina (depending on the direction of head tilt), allowing this region to be perceived to the left or right of the display. If figure–ground assignment occurs in a viewer-centered reference frame, a head tilt should abolish the lower-region preference, because the retinal representation has rotated. If figure–ground assignment occurs in either an object- or an environment-centered reference frame, the lower-region preference should persist across the head rotation.

Several accounts of figure–ground assignment have proposed that figure–ground processes occur relatively early in the visual processing hierarchy in a retinotopic, viewer-centered reference frame in which a shape is represented relative to its retinal position and orientation[2] (see Grossberg, 1997; Kosslyn, 1987, 1994; Marr, 1982; Sejnowski & Hinton, 1987; Vecera & O'Reilly, 1998, 2000). The use of such a reference frame for figure–ground assignment is not a foregone conclusion, however. Ecological factors could favor the use of an environmental reference frame. In the case of the lower-region preference, environmentally and gravitationally lower regions are perceived as physically closer to viewers when these regions arise from opaque objects, raising the possibility that figural lower regions may be defined in terms of environmental coordinates.

The reference frame of figure–ground assignment was tested in three experiments in which participants performed either an explicit figure–ground task (Experiment 1) or a visual short-term memory matching task (Experiments 2 and 3). In Experiments 1 and 2, participants viewed displays with head either upright or tilted 90° to the left or right. In Experiment 3, participants viewed displays with head either upright or upside down.
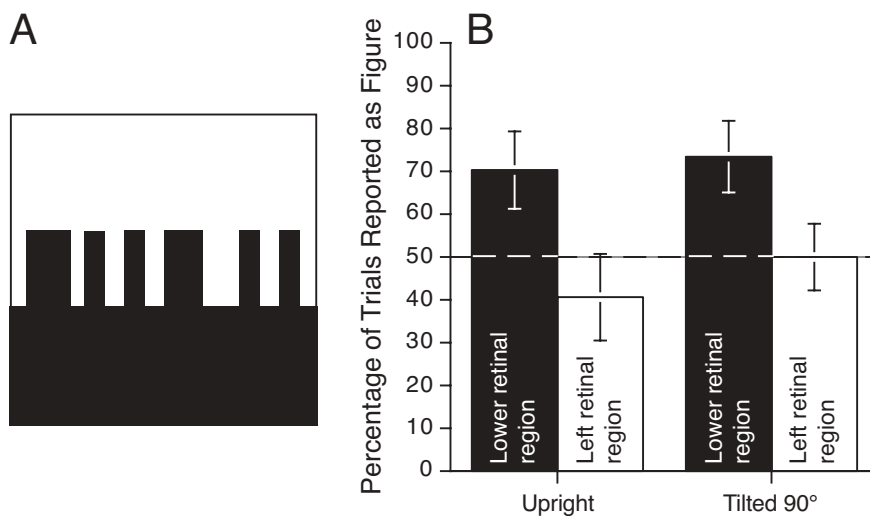

Figure 1. (A) An example of the lower-region preference. Most observers perceive the lower (black) region as the foreground figure. (B) Results of Experiment 1, in which the lower-region preference follows head tilt, suggesting that figure–ground assignment is based on a retinotopic reference frame. The error bars are 95% confidence intervals for comparisons with chance (50%).

In all experiments, participants viewed both displays with environmentally defined upper and lower regions and others with environmentally defined left and right regions. If figure–ground assignment occurs in the retinotopic, viewer-centered representation, then in Experiments 1 and 2, the head tilt should cause left/right displays to exhibit a retinal lower-region preference, and in Experiment 3, upside-down viewing should cause the environmentally defined *upper* region to exhibit a retinal lower-region preference. If figure–ground assignment occurs in an environment-centered reference frame, then the same figural assignment will be produced in the head-upright and head-tilted/upside-down conditions.

EXPERIMENT 1

Method

Participants. Sixteen University of Iowa undergraduates with normal or corrected-to-normal vision volunteered for course credit.

Stimuli. The figure–ground stimuli were similar to those in Vecera et al.'s (2002) Experiment 8. Stimuli were figure–ground displays in which two regions, one red and one green, shared a central, irregular contour (Figure 1A). The irregular shape of the shared contour was created by dividing the contour into 16 smaller regions that were equal in area. Then, each of these 16 regions was randomly assigned to either the red or the green region, with the constraint that 8 of the convexities be assigned to each region to ensure equal convexity and area on both sides of the shared contour. Four random contours were generated using this procedure. Four versions of each contour were then generated, corresponding to four variations of red/green color and orientation of the display (red on left or right; red upper or lower), resulting in 16 displays. Each display measured 8.3º square. The red/green color values were those used in previous studies (Vecera et al., 2002).

Procedure. Participants were instructed to report the color of the region that appeared to be the foreground figure. Prior to testing, each participant was shown Rubin's (1915/1958) face–vase figure to illustrate the principle of figure–ground assignment. Participants were told that either the faces or the vase, but not both, could be perceived as lying in the foreground and would appear to be closer than the other region. Participants were asked to try to perceive as "figure" both the faces and the vase in alternation.

As a between-subjects manipulation, participants viewed the figure–ground displays with head either upright or tilted 90º to the left or right; in the tilted-head condition, half of the participants were assigned to tilt to the left and half to the right. Participants viewed 16 randomly presented figure–ground displays, with equal numbers of left/right and upper/lower displays and of red/green combinations. Each trial began with a 500-msec fixation cross, followed by the figure–ground display, which was visible until participants responded. There was then a 100-msec interval before the start of the next trial. Participants responded with which color (red or green) they perceived as the figure. Responses were made using a response box with a red key (on the left of the box) and a green key (on the right of the box). The response box rested on the table in front of the participant, and the buttons were aligned with the environmentally defined left and right sides of the monitor. Finally, the ceiling and walls of the room and the computer table were fully visible during all experiments, providing participants with ample cues to the environmental layout.

Results and Discussion

Figure 1B depicts the percentage that retinally defined lower or left regions were reported as "figure." The results demonstrate that figure–ground assignment followed the retinal, not the environmental, orientation of the stimuli. When viewed with a tilted head, the environmentally defined left or right region, whichever appeared in the lower viewer-centered position, produced a lower-region preference, and the environmentally defined lower-region preference was abolished when these upper and lower regions were viewed with a tilted head.

These observations were confirmed by comparing the lower- and left-region preferences with chance (50%). Lower regions are perceived as "figure" above chance, but left regions are not (Vecera et al., 2002). These findings were replicated when participants viewed displays with upright heads; the lower region was perceived as "figure" above chance (70.3% of trials) [$t(7) = 2.9$, $p < .03$], but the left region was not (40.6% of trials) [$t(7) = 1.4$, $p > .20$]. Most importantly, when these displays were viewed with a tilted head, a lower-region preference emerged for the *viewer-defined* lower region. The viewer-defined lower region was perceived as "figure "above chance (73.4%) [$t(7) = 3.1$, $p < .02$], whereas the viewer-defined left region was perceived as "figure" at chance (50%). Figure–ground assignment followed from the viewer- rather than the environment-centered orientation of the stimulus, indicating that figure–ground assignment occurred in a viewer-centered reference frame. However, there are potential problems with explicit figure–ground reports, as reported, for example, by Driver and Baylis (1996) or Vecera et al. (2002). Experiments 2 and 3 used an indirect measure of figure–ground assignment to overcome the possible limitations of explicit report.

EXPERIMENT 2

Participants in Experiments 2 and 3 performed a visual short-term memory task that provides an implicit measure of figure–ground assignment. This task, developed by Driver and Baylis (1996), requires participants to determine which of two test shapes was present in a figure–ground display (see Figure 2A). Participants are faster to match test shapes that correspond to a figure than those that correspond to a ground (Driver & Baylis, 1996), even when figure–ground assignment is not explicitly mentioned. If figure–ground assignment occurs in viewer-centered coordinates, then faster memory-matching times would be expected for viewer-defined lower regions, regardless of head orientation. Specifically, participants should be faster to match viewer-defined lower regions than upper regions, but no systematic preference should emerge between viewer-defined left and right regions.

Method

Participants. Sixteen University of Iowa undergraduates with normal or corrected-to-normal vision volunteered for course credit.

Stimuli. Stimuli were similar to those for Experiment 1, except that 12 random contours were generated to increase the stimulus set. For each contour, a mirror image was created to further increase the number of stimuli. There were 96 displays, with orientation (left/right or upper/lower) and color appearing equally in the four different positions (upper, lower, left, and right).
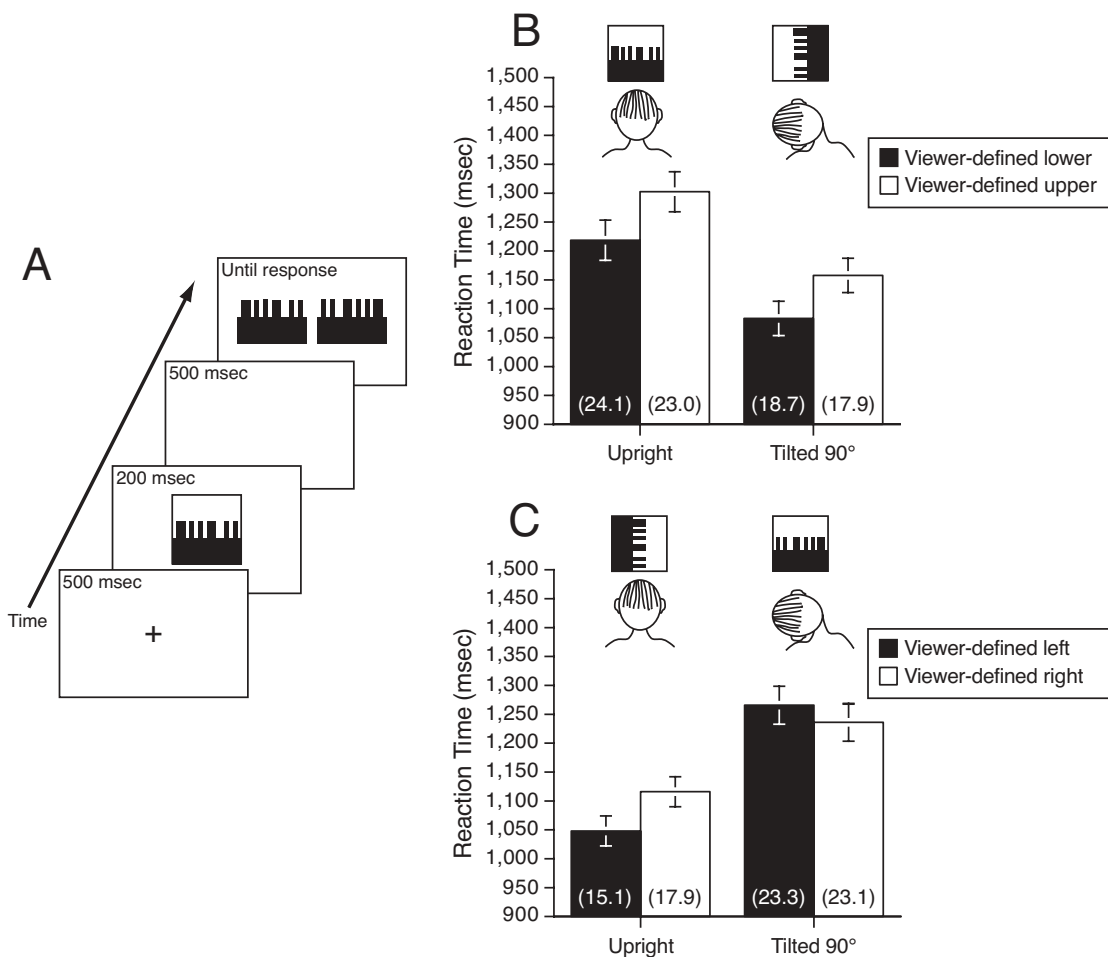
Figure 2. (A) Order and timing of events in the visual short-term memory matching task. (B) Retinally defined lower regions are matched faster than retinally defined upper regions. (C) Retinally defined left and right regions do not show any systematic figural preferences. Both when heads are upright and when they are tilted, participants match regions that are lower in retinal coordinates faster. The numbers inside the bars are the average error percentages per condition, and the error bars are 95% confidence intervals on the upper versus lower and left versus right pairwise comparisons.

Procedure. Participants viewed a figure–ground display and were asked to remember the two regions for a test that occurred 500 msec later (Figure 2A). The regions in the test display were mirror images and appeared on the environmentally defined left and right sides of the monitor (Figure 2A). Participants were instructed to respond quickly and accurately but were not informed about figure–ground assignment. Each participant received 384 trials with head upright and 384 with head tilted. Half of the participants performed the head-upright condition first, half performed the head-tilted condition first. In the head-tilted condition, half of the subjects were randomly assigned to tilt left and half to the right.

For every stimulus, each of the two regions was probed equally, and the correct response appeared equally on the left and right sides of the monitor. In the head-tilted condition, the test shapes appeared above and below fixation in viewer-centered coordinates; however, participants responded based on the environmental position of the shape (i.e., region on the left or right side of the monitor). Participants reported the test shape that had appeared in the figure–ground display using a button box that was oriented as in Experiment 1. Participants received 96 practice trials that were not analyzed and

were allowed a break every 96 trials during the test. The display types (left/right and upper/lower) were intermixed.

Results and Discussion

Only correct responses with reaction times (RTs) below 2,500 msec were analyzed; this trimming excluded less than 8% of the data. Median RTs were computed for all conditions and analyzed with a repeated measures analysis of variance (ANOVA) using head orientation and viewer-centered region tested (lower, upper, left, and right) as factors. There were no speed–accuracy trade-offs. The RT results, shown in Figure 2, indicate that the lower-region preference followed head orientation: When participants viewed a left/right display with tilted heads, the retinally defined lower region appeared as "figure." When participants viewed an upper/lower display with tilted heads, neither region had a figural advantage, be-

cause the regions were to the left and right of each other retinally.

The ANOVA revealed no main effect of head orientation; responses were not significantly different in the head-upright (1,171.3 msec) and head-tilted (1,185.7 msec) conditions [$F(1,15) < 1$]. There was a main effect of region tested. Responses to the four tested regions differed from one another [$F(3,45) = 5.3, p < .005$], with fastest responses to viewer-defined lower regions (1,151 msec) and slowest responses to viewer-defined upper regions (1,230.1 msec). Planned comparisons revealed that the lower-region preference followed head orientation, consistent with a viewer-centered coordinate system. In the head-upright condition, lower regions were matched faster than upper regions (1,218.6 msec vs. 1,302.5 msec) [$t(15) = 2.6, p < .03$], replicating the lower-region preference. Most importantly, in the head-tilted condition, *viewer-defined* lower regions were matched faster than upper regions (1,083.3 msec vs. 1,157.6 msec) [$t(15) = 2.7, p < .02$].

Finally, the omnibus ANOVA revealed a significant interaction between head orientation and region tested [$F(3,45) = 36.3, p < .005$]. As is evident in Figure 2, the interaction indicates an environment-based influence on RTs: Participants were faster to respond to environmentally defined left/right displays than to environmentally defined upper/lower displays. The environmental effect does not change the lower-region preference, which was based on viewer-centered coordinates.

The environment-based influence might involve the shape-matching component of this task, not the edge-assignment component: If environment-based left–right reversals of the test shapes (Figure 2A) are more perceptually similar than environment-based top–bottom reversals (as would be encountered in left/right displays), an orientation × region tested interaction would result. This explanation predicts that the same orientation × region interaction would result even when figure–ground requirements were minimal (e.g., if the prime display was a single region, not a figure–ground display). As discussed later, an environmental reference frame for the shape-discrimination component of this task would be consistent with other studies that have investigated reference frames (Rock, 1973, 1983). We are currently investigating the possibility that multiple reference frames are used in performing this shape-matching task.

Regardless of the observed environmental contribution, Experiment 2 suggests that figure–ground assignment occurs in a viewer-centered reference frame. When participants tilted their heads, the figure–ground assignment based on the lower-region cue followed head orientation. If an environmentally defined left/right display is viewed with a left-tilted head, the right region appears "lower" to the viewer, and the shared contour is assigned to this region. These results are consistent with theories that propose that figure–ground assignment is computed in a retino- or spatiotopic viewer-centered reference frame (e.g., Kosslyn, 1987, 1994).

One question arising from Experiments 1 and 2 concerns the strength of the viewer-centered influence on figure–ground assignment. Specifically, could the viewer-centered reference frame override the environment-centered reference frame if the two were opposed? This issue was addressed in Experiment 3 by asking participants to view figure–ground displays with head either upright or upside down. The latter viewing was the result of participants bending over and looking at the display through their legs, which rotated the retinal image 180°. If figure–ground assignment again follows head orientation, then the lower-region preference should *reverse* when participants view upper/lower displays: An environmentally defined upper region, now a viewer-defined lower region when viewed upside down, should appear as "figure." Such a finding would indicate that a viewer-centered reference frame could override an environment-centered frame.

EXPERIMENT 3

Method

Participants. Sixteen University of Iowa undergraduates with normal or corrected-to-normal vision volunteered for course credit.

Stimuli and Procedure. The stimuli and procedure were identical to those in Experiment 2, with one exception: Participants viewed the displays either sitting on the floor with their heads upright or standing, bent over, and looking at the computer monitor through their legs with their heads upside down.

Results and Discussion

The results were analyzed as in Experiment 2. Trimming eliminated less than 7% of the data, and there were no speed–accuracy tradeoffs. The results, shown in Figure 3, replicated the results from Experiment 2 by showing that lower-region figural assignment followed head orientation. When participants viewed an upper/lower display while upside down, the viewer-defined lower region (the environmentally defined upper region) appeared as "figure."

The ANOVA revealed no main effect of head orientation; responses were not significantly different in the head-upright (1,185.7 msec) and upside-down (1,095.2 msec) conditions [$F(1,15) = 1.6$, n.s.]. There was a main effect of region tested. Responses to the four tested regions differed from one another [$F(3,45) = 5.7, p < .005$], with fastest responses to viewer-defined left regions (1,074 msec) and slowest responses to viewer-defined upper regions (1,242.6 msec). Planned comparisons revealed that the lower-region preference followed head orientation, consistent with a viewer-centered coordinate system. In the head-upright condition, lower regions were matched faster than upper regions (1,222.3 msec vs. 1,301.7 msec) [$t(15) = 2.2, p < .05$], again replicating the lower-region preference. Most importantly, in the upside-down condition, viewer-defined lower regions were matched faster than upper regions (1,047.8 msec vs. 1,183.4 msec) [$t(15) = 2.5, p < .03$].

Finally, there was a marginal interaction between head orientation and region tested [$F(3,45) = 2.4, p < .10$].
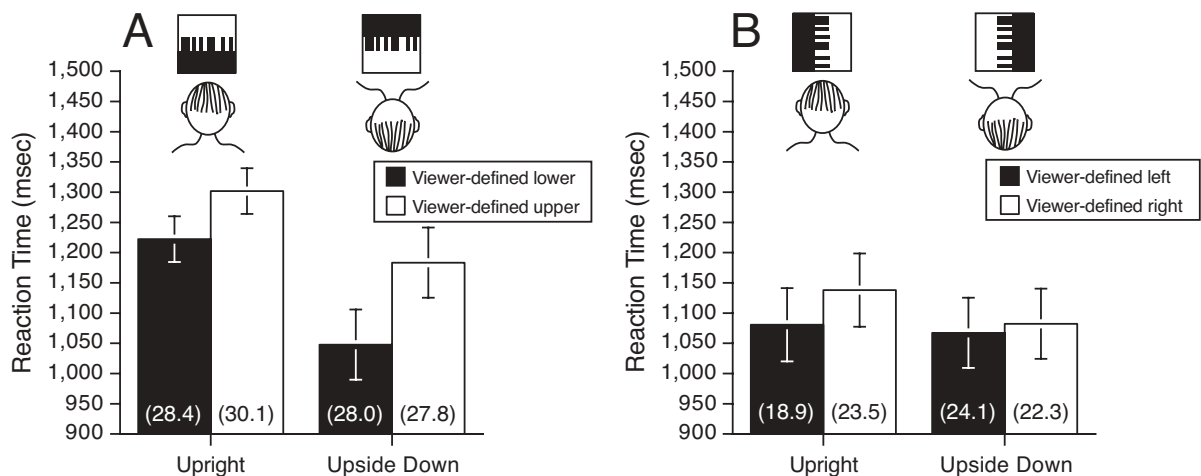
Figure 3. Results from Experiment 3 (upright and upside-down viewing). The lower-region figural advantage rotates with the head orientation. (A) When participants view a figure–ground display while upside down, the retinally defined lower region (the environmentally defined upper region) is matched faster than the retinally defined upper region (the environmentally defined lower region). (B) Retinally defined left and right regions do not show any systematic figural preferences. The numbers inside the bars are the average error percentages per condition, and the error bars are 95% confidence intervals on the upper versus lower and left versus right pairwise comparisons.

This interaction suggests an environment-based influence on RTs, which likely arises from the shape-discrimination rather than the figure–ground (or edge-assignment) component of the task. As in Experiment 2, this environmental contribution did not affect the lower-region preference, which was based on viewer-centered coordinates.

These results demonstrate that figure–ground assignment occurs in a viewer-centered reference frame, even when the viewer-centered reference frame is directly opposed to an environmental reference frame.

GENERAL DISCUSSION

Most accounts of figure–ground assignment have hypothesized that this visual process operates in a viewer-centered (possibly retinotopic) representation. Unfortunately, because of the nature of most gestalt figure–ground cues, there was no empirical support for the use of such a reference frame in figure–ground assignment. The present experiments used a new figure–ground cue, lower region, to demonstrate that figural assignment does indeed use a viewer-centered reference frame. Changing the retinal image, either by tilting the head or viewing a display while upside down, affects figural assignment. The region that falls in the viewer-centered lower position is perceived as "figure."

One potential concern with the present findings is that they may only hold for the lower-region cue. However, there is no evidence to suggest that the lower-region cue is different from other figure–ground cues. In previous work (Vecera et al., 2002), we demonstrated that the lower-region cue influences the assignment of the contour shared by two regions, just as the symmetry and area cues do (Driver & Baylis, 1996). Lower regions also produce

more stable figures that undergo fewer figure–ground reversals. Converging evidence from other figure–ground cues (e.g., orientation and symmetry) could ensure the generality of the viewer-centered reference frame.

The present results differ from those of other studies that have used head-tilt manipulations to investigate reference frames. Rock has shown that perceiving simple shapes is not dependent upon the retinal image of those shapes (Rock, 1973, 1983). For example, when viewing a square with a 45° head tilt, a viewer will continue to perceive a square, not the diamond that appears in the viewer-centered retinal image. However, when a shape is rotated in the frontal plane, the perception of that shape can be altered because of the assignment of directions (e.g., "top," "bottom," etc.) to regions of the shape. This assignment can be based on gravitational (i.e., environmental) or other visual information (see Rock, 1983). Although the retinal and environmental representations are misaligned when the head is tilted, in simple shape perception head orientation is taken into account and used to derive the object- or environment-centered description of the shape. However, in figure–ground assignment this is not the case. Instead, the retinal representation of the stimulus determines the figure–ground solution.

Although the cause of the differences between Rock's findings and the present results are not obvious, one straightforward possible cause would be the difference in the tasks that participants performed. In Rock's research, participants performed a long-term shape memory task that might depend on object identification (e.g., reporting whether a previously viewed shape was "old" or "new" when viewed with a tilted head). Such a task might benefit from the use of an object- or environment-

centered reference frame. In the present experiments, participants performed a short-term matching task based on edge assignment. Edge assignment can occur in an early, retinotopically mapped representation (Sejnowski & Hinton, 1987; Vecera & O'Reilly, 1998, 2000; Zhou, Friedman, & von der Heydt, 2000), possibly minimizing the need to represent the geometry of the figural region in an object-centered reference frame. However, the visual system uses several reference frames, and there may be situations in which edge and figure–ground assignment require a non–viewer-centered reference frame.

The present results support the use of a viewer-centered representation in figure–ground assignment, consistent with both computational views of figure–ground processes (e.g., Kosslyn, 1987; Sejnowski & Hinton, 1987; Vecera & O'Reilly, 1998, 2000) and neurophysiological findings that place these processes in early visual cortex (Lamme, 1995; Zhou et al., 2000). Figure–ground processes appear to operate within viewer-centered, intermediate-level vision, allowing them to provide inputs to high-level visual processes such as object recognition, selective attention, and visuomotor action.

REFERENCES

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

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 1162-1182.

Biederman, I., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995). *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1506-1514.

Driver, J., & Baylis, G. C. (1996). Edge-assignment and figure–ground segmentation in short-term visual matching. *Cognitive Psychology*, **31**, 248-306.

Grossberg, S. (1997). Cortical dynamics of three-dimensional figure–ground perception of two-dimensional pictures. *Psychological Review*, **104**, 618-658.

Klymenko, V., & Weisstein, N. (1986). Spatial frequency differences can determine figure–ground organization. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 324-330.

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

Kosslyn, S. M. (1994). *Image and mind*. Cambridge, MA: MIT Press.

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

Lee, S.-H., & Blake, R. (1999). Visual form created solely from temporal structure. *Science*, **284**, 1165-1168.

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

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

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

Peterson, M. A. (1999). What's in a stage name? Comment on Vecera and O'Reilly (1998). *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 276-286.

Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, **18**, 1-63.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 1-46). New York: Wiley.

Rock, I. (1973). *Orientation and form*. New York: Academic Press.

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

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.

Rock, I. (1995). *Perception*. New York: Scientific American Books.

Rock, I., & DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, **19**, 280-293.

Rossi, A. F., Desimone, R., & Ungerleider, L. G. (2001). Contextual modulation in primary visual cortex of macaques. *Journal of Neuroscience*, **21**, 1698-1709.

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

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

Tarr, M. J. (1994). Visual representation. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 503-512). San Diego: Academic Press.

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

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1494-1505.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey, and machine. *Cognition*, **67**, 1-20.

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

Vecera, S. P., Flevaris, A. V., & Filapek, J. C. (2004). Exogenous spatial attention influences figure–ground assignment. *Psychological Science*, **15**, 20-26.

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

Vecera, S. P., & O'Reilly, R. C. (2000). Graded effects in hierarchical figure–ground organization: A reply to Peterson (1999). *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1221-1231.

Vecera, S. P., Vogel, E. K., & Woodman, G. F. (2002). Lower region: A new cue for figure–ground assignment. *Journal of Experimental Psychology: General*, **131**, 194-205.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, **20**, 6594-6611.

NOTES

1. One might argue that neurophysiology could determine the reference frame of figure–ground assignment. Some neurophysiological studies have suggested that figure–ground assignment occurs in the primary visual cortex, which is retinotopically mapped (Lamme, 1995), suggesting a viewer-centered reference frame. However, other studies have challenged this conclusion and suggested that figure–ground assignment may occur in the extrastriate visual cortex (Rossi, Desimone, & Ungerleider, 2001) and use any number of reference frames.

2. There are several coordinate systems for viewer-centered representations, including retinal position, head position, and body position. Theories of figure–ground assignment have primarily discussed the retinal position of figures; in this study, the term *viewer-centered* is likewise used to mean a representation that represents a stimulus relative to its retinal position and orientation.