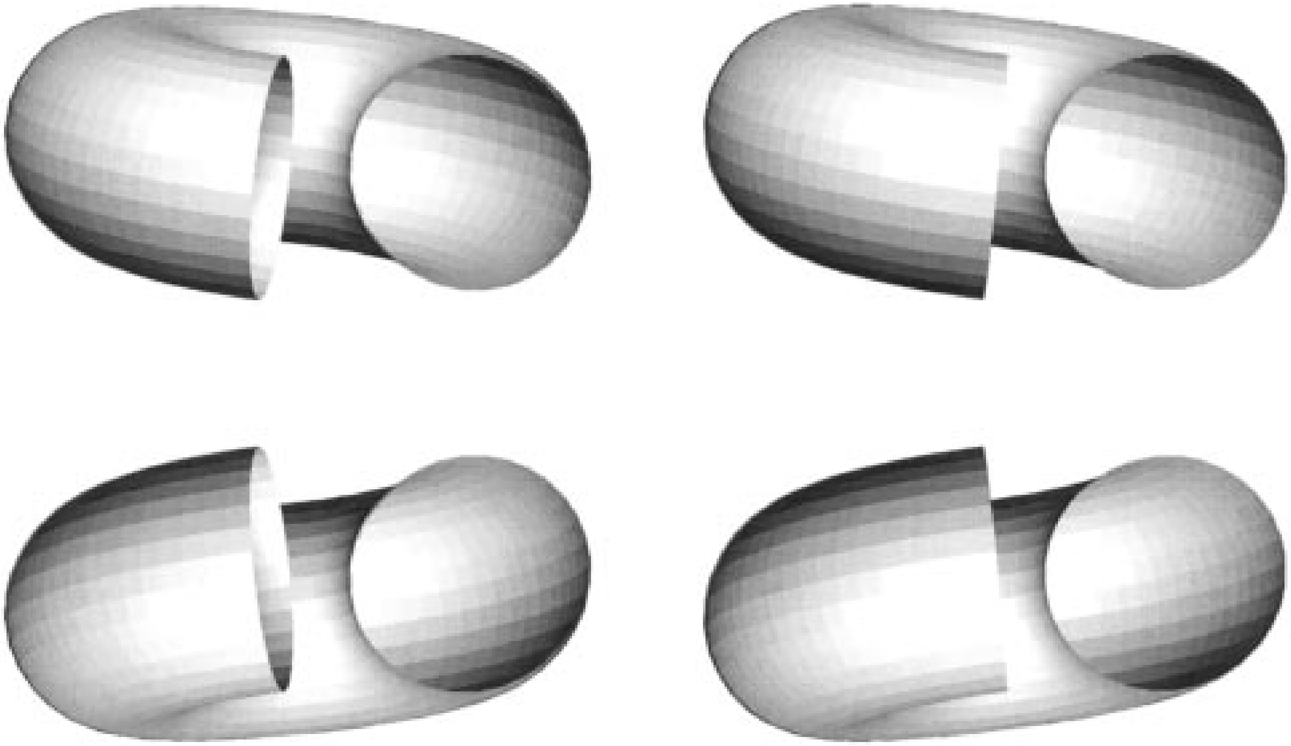

Figure 1. In both pairs of figures, the shape is easily determined from the shading cues, although the lower figure incorrectly appears to have a ball at one end. This is due to the expectation that all light sources are above. The pairs of figures may be fused by using a stereoviewer or free-fusing to provide unambiguous depth cues. In the upper pair, where the stereo and shape-from-shading cues agree, the right end of the tube is correctly seen as an opening. In the lower pair of figures, the disparity cues and the shape-from-shading cues are in opposition. Even with the presumably unambiguous disparity cues, the ball remains at the end of the tube. The shape-from-shading cues are dominant, suggesting that it is processed prior to disparity. Studies in which one type of shape cue is pitted against another are often used in psychophysical studies, with the actual physiological measurements lagging behind.

the representation of shape. The best-understood recognition process is the motion pathway that passes through striate cortex, to the middle temporal motion area (MT/V5). In MT/V5 there is a representation of the velocity of the image for different parts of the visual field (Albright, Desimone, and Gross 1984). This motion representation is further developed in the medial superior temporal area (MST) in which neurons are found that respond to environmental optic flow for spatial vision (Tanaka et al. 1986; Duffy and Wurtz 1991). Beyond MST, the motion signal passes to the parietal cortex (area 7a; Siegel and Read 1997) in which optic flow signals are further processed and combined with eye position information. Both 7a and MST project to the anterior polysensory temporal area (STPa) which has neurons that represent both flow and apparently 3-D shape (Bruce, Desimone, and Gross 1981; Anderson and Siegel submitted).

Running roughly in parallel to the processing of visual motion is the analysis of disparity cues. At each step from striate to MT/V5 to MST to 7a, neurons are found that are tuned to disparity (Poggio and Poggio 1984; Roy, Komatsu, and Wurtz 1992; Gnadt and Mays 1995). Little is known as to how binocular cues are used for shape representation.

A more temporal cortical stream represents shape using luminance and color cues. Neurons have been described that represent all sorts of luminance cues, such as orientation (Hubel and Wiesel 1977) and borders (von der Heydt, Peterhans, and Baumgartner 1984). Geometrical figures (Tanaka 1993), as well as shapes as complex as faces (Gross 1973), may be represented by temporal cortical neurons. Color analysis surely is used in object identification, although little formal work has been done. Surprisingly, the dependence of these neurons upon parameters of motion (Perrrett et al. 1985) and disparity are as yet little explored. Such studies are crucial, as the psychophysical ability to describe shape (a putative temporal stream analysand) does not deteriorate when motion or disparity (a putative dorsal stream analysand) is the underlying representation.

In summary, the visual perception of 3-D structure utilizes motion, disparity, and luminance. Psychophysical studies have defined the limits of our ability, while computational studies have developed a formal framework to describe the perceptual process as well as to test hypotheses. Anatomical and physiological results have provided essential cues from functional systems.

See also HIGH-LEVEL VISION; MID-LEVEL VISION; MOTION, PERCEPTION OF; TOP-DOWN PROCESSING IN VISION

—Ralph M. Siegel