

What visual perception tells us about mind and brain

Shinsuke Shimojo*[†], Michael Paradiso[‡], and Ichiro Fujita[§]

*Division of Biology/Computation and Neural Systems, California Institute of Technology, Pasadena, CA 91105; [‡]Department of Neuroscience, Brown University, Providence, RI 02192; and [§]Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan

Recent studies of visual perception have begun to reveal the connection between neuronal activity in the brain and conscious visual experience. Transcranial magnetic stimulation of the human occipital lobe disrupts the normal perception of objects in ways suggesting that important aspects of visual perception are based on activity in early visual cortical areas. Recordings made with microelectrodes in animals suggest that the perception of the lightness and depth of visual surfaces develops through computations performed across multiple brain areas. Activity in earlier areas is more tightly correlated with the physical properties of objects whereas neurons in later areas respond in a manner more similar to visual perception.

Neuroscience research over the past 40 years has revealed that there are roughly 30 different visual areas in the primate brain, and that within these areas there are parallel streams of processing and distinct modules (1, 2). But how is neuronal activity in the different areas related to our conscious visual perception? How can our unitary visual experience be based on neural activity spread across distinct streams of processing in multiple brain areas? The answers to these questions have profound implications for our understanding of the relationship between mind and brain. Whereas earlier pioneering work focused on the delineation of visual areas in the brain and the neurons' basic response properties, recent research attempts to expose the roles different areas play in perception and the extent to which there are hierarchies of visual computations.

Conscious visual experience is thought to be based on activity in visual areas of cerebral cortex, which receive input from the retina. Early cortical structures are organized topographically with regard to the visual world. This topography can be exploited to investigate the role of different visual areas in perception. For example, neuronal activity in visual cortex can be locally blocked by transcranial magnetic stimulation (TMS) and the effect on visual perception in the corresponding portion of the visual field can be assessed. Kamitani and Shimojo (3) briefly (40–80 ms) presented a large grid pattern to human observers, and after a delay of 80–170 ms, a single pulse of TMS was given to the occipital lobe. The TMS caused the observers to perceive a disk-shaped patch of homogeneous color in the visual field on the opposite side from the side of the brain given TMS (TMS-induced scotoma). When the visual stimulus was a grating composed of parallel lines rather than a rectilinear grid, the scotoma was distorted and appeared to be an ellipse with its short axis along the contours. This contour-dependent distortion appeared to reflect long-range interactions between neurons selectively responsive to similar orientations (4). Interestingly, the color perceived inside the scotoma was consistent with that of the background, which was presented after, not before, the grid or grating. Thus there appears to be filling-in backward in time to compensate for the local information blocked by the TMS. This is just one example from a large body of evidence suggesting that neural activity in early visual cortex is necessary for conscious experience of perception, and that neuronal connections and interactions at these levels are reflected in the content of perception.

Perception is actually much more complex than a simple topographical representation of the visual world. Its primary goal is to recover the features of external objects—a process termed unconscious inference by von Helmholtz (5, 6). What we see is actually more than what is imaged on the retina. For

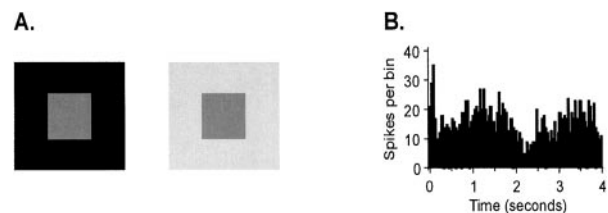


Fig. 1. (A) Lightness induction. The small gray squares are identical but the one surrounded by black appears lighter than the square surrounded by white. (B) The response of a V1 neuron to a lightness induction stimulus. The receptive field of the neuron was centered on a uniform gray square. The luminance of the surrounding area was sinusoidally modulated. The cell's response was synchronized to the surround modulation and correlated with the perceived lightness of the central patch, even though nothing changed within the receptive field. [Reproduced with permission from ref. 14 (Copyright 2001, National Academy of Sciences).]

example, we perceive a three-dimensional world full of objects despite the fact that there is a simple two-dimensional image on each retina. In general, a particular retinal image may correspond to more than one object. For example, a circular patch of light on the retina could result from viewing a cylinder on end or a round ball from any perspective. Thus perception is inevitably an ambiguity-solving process. The perceptual system generally reaches the most plausible global interpretation of the retinal input by integrating local cues, as will be illustrated in the case of lightness perception next.

Black-and-white photographs make it clear that lightness alone conveys a great deal of information. The perception of lightness is far from a “pixel-by-pixel” representation of the light level on the retina. It is actually strongly influenced by context. Thus a gray piece of paper appears darker if it is surrounded by white than black (Fig. 1A). Although this deviation of lightness perception from physical reality might appear to be a case of a perceptual error, the spatial interactions underlying it may have an important perceptual purpose. We perceive surface lightness to be constant across surprisingly large changes in ambient illumination, a phenomenon called lightness constancy. In this example, as in other cases of perceptual constancy, the lighting

This paper is a summary of a session presented at the third annual Japanese–American Frontiers of Science symposium, held September 22–24, 2000, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: TMS, transcranial magnetic stimulation; IT, inferior temporal cortex.

[†]To whom reprint requests should be addressed. E-mail: sshimo@cns.caltech.edu.

and viewing conditions affect the retinal image of objects, and extensive spatial integration and normalization are performed to recover the constant attributes of the objects themselves.

At what point in the visual pathway from retina to the many cortical visual areas does the neural activity correlate with what we perceive? Do neurons in the retina, primary visual cortex (V1), and higher-level cortical areas contribute to perception equally? Or instead, does perception have a specific locus in the brain? To tackle these questions, Paradiso and coworkers (7, 8) assess the computations neurons perform in different visual areas and the extent to which neural responses correlate with either the physical or perceptual attributes of objects. They found that responses of neurons in the retina and visual thalamus depend on light level but they do not correlate with perceived lightness. These neurons appear to primarily encode information about the location of contours in the visual scene. Only in V1 were cells found that had responses correlated with perceived lightness (Fig. 1B). They also found that the average response of neurons in V1 is lightness constant. Thus the response of the neurons is relatively immune to changes in overall illumination—a property without which lightness would be of little behavioral value. These findings suggest that lightness information is first explicitly represented in visual cortex and that responses correlated with visual perception build in stages across multiple visual areas. The results combined with findings from other labs suggest that early visual processing focuses on the extraction of object contours, secondary processing stages are involved with the computation of lightness and later processing assigns color to objects.

As mentioned previously, the visual system has the difficult task of understanding a complex three-dimensional world from two-dimensional images on each retina. Images of objects at a distance other than at the fixation plane are projected to different relative positions on the two retinas. The relative position difference, called binocular disparity, provides an important cue for the brain's computation of distance. However, there is much more to distance perception than the interpretation of binocular disparity. Consider a retinal image of a cross with crossed disparities (disparities that lead to perception of objects closer than the plane of fixation) added to the ends of the horizontal arms. Because of the disparities, the vertical edges of the horizontal arms can be unambiguously determined as being closer to the observer, whereas the depth of the horizontal edges remains ambiguous because there is no fixed disparity between the two retinal images. Two different three-dimensional objects are equally consistent with the retinal image: a horizontal bar in front of a vertical bar and a cross with horizontal arms bent forward. However, humans and monkeys almost always perceive the former (9, 10). The brain selects one interpretation among the possible surface structures.

1. Ungerleider, L. G. & Mishkin, M. (1982) in *Analysis of Visual Behavior*, eds. Ingle, D. J. G., Goodale, M. A. & Mansfield, R. J. W. (MIT Press, Cambridge, MA), pp. 549–586.
2. Felleman, D. J. & Van Essen, D. C. (1991) *Cereb. Cortex* **1**, 1–47.
3. Kamitani, Y. & Shimojo, S. (1999) *Nat. Neurosci.* **2**, 767–771.
4. Gilbert, C. D., Das, A., Ito, M., Kapadia, M. & Westheimer, G. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 615–622.
5. von Helmholtz, H. (1909/1910) *Handbuch der Physiologische Optik* (Leipzig, Hamburg, Germany).
6. Southall, J. P. C., ed. (1962) *Helmholtz' Treatise on Physiological Optics* (Dover, New York).

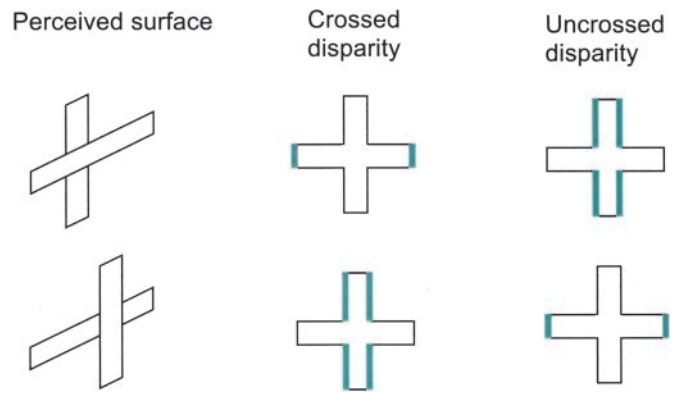


Fig. 2. (A) The relationship between disparity type and location and surface depth order perceived. Responses of IT neurons to these four stimuli were tested to determine whether their activity correlates with the perceived surface structure or with the type of disparity.

The inferior temporal cortex (IT) represents the final stage of the visual pathway crucial for object recognition. Neurons in IT respond to shape, color, or texture. Recent studies show that many IT neurons also convey information on disparity (11) and disparity gradients (12). These findings lead to a new view that IT is involved in some aspects of depth perception. Indeed, the activity of some IT neurons encodes information on the relative depth order of surfaces rather than the local absolute disparity cues of the stimulus. For example, a population of IT neurons responds more strongly to a horizontal bar in front of a vertical bar than to a vertical bar in front of a horizontal bar, regardless of whether crossed or uncrossed disparities are added (Fig. 2). Other cells prefer different surface structures. This behavior of IT neurons is in contrast to that of disparity-selective V1 neurons that respond to local absolute disparity (13). Thus, the pathway from V1 to IT transforms information about binocular disparity that is based on the optics of the eye into a perceptually relevant representation of information about surface structure.

The studies of lightness perception and depth perception lead to a similar conclusion about the relationship between brain activity and conscious visual perception. Rather than being based on neural activity in one special area, visual perception involves progressive computations spread across multiple brain areas. Both early areas, as in the TMS study, and later areas, as in the study of area IT, are involved in perception. The visual system masterfully recovers information about the objects in our environment based partly on processes of integration and normalization and partly on hard-wired probabilities of what objects are most likely to result from particular retinal images.

7. Macevoy, S., Kim, W. & Paradiso, M. A. (1998) *Nat. Neurosci.* **1**, 616–620.
8. Rossi, A. F. & Paradiso, M. A. (1999) *J. Neurosci.* **19**, 6145–6156.
9. Nakayama, K. & Shimojo, S. (1992) *Science* **257**, 1357–1363.
10. Uka, T., Tanaka, H., Kato, M. & Fujita, I. (1999) *Vision Res.* **39**, 2399–2410.
11. Uka, T., Tanaka, H., Yoshiyama, K., Kato, M. & Fujita, I. (2000) *J. Neurophysiol.* **84**, 120–132.
12. Janssen, P., Vogels, R. & Orban, G. A. (2000) *Neuron* **27**, 385–397.
13. Cumming, B. G. & Parker, A. J. (1999) *J. Neurosci.* **19**, 5602–5618.
14. MacEvoy, S. P. & Paradiso, M. A. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 8827–8831. (First Published July 10, 2001; 10.1073/pnas.161280398)