Chapter in press for a Psychology Reader:
Foundation for the Advancement of
Behavioral and Brain Sciences

Human Vision as a Reality Engine

Donald D. Hoffman
Department of Cognitive Science
University of California
Irvine, CA 92697

Introduction

Vision feels easy. We simply open our eyes and look. Without apparent effort we see a three-dimensional world packed with objects, colors, textures and motions. The apparent ease of vision makes it natural for us to suppose that vision is a simple process, no more complex than taking pictures with a camera.

But research in cognitive neuroscience reveals that roughly half of the brain's cortex, perhaps 50 billion neurons and 10 trillion synapses, are engaged when we simply open our eyes and look (e.g., Chalupa & Werner, 2003; Gazzaniga, 2004). Why should half of our most sophisticated computational power be engaged in vision? This is overkill if vision is, like a camera, just taking pictures. A digital camera needs nowhere near that much circuitry to take pictures.

So why engage all this computing power for vision? Research in psychophysics and computational vision provides a remarkable answer. Your visual system is a reality engine: It creates all the depths, shapes, objects, colors, textures and motions that you see. The term "reality engine" here is borrowed from the field of virtual reality, where it refers to the powerful computers and sophisticated software used to generate the impression of immersion in a virtual world. In the case of human vision, enormous computing power is needed to create the visual worlds in which we find ourselves constantly immersed. And that's why roughly half of your most sophisticated computing power is recruited for the job.

Admittedly, it doesn't feel like we construct the visual world. It feels like the visual world is always there, and we simply take pictures of it. But the reason it feels this way is that our reality engine is so fast and effective that it creates what we see as fast as the eye can move. It is the very power and efficiency of our reality engine that makes us unaware that it even exists. But we can find telltale signs of its operation if we know where to look.

For example, check out the two boxes in Figure 1. Are their gray tops the same shape, or different? They certainly look different. The one on the left looks long and narrow, the one on the right short and fat. But they are identical. You can verify this with a ruler, or by using tracing paper to copy the top of one box, and then place it on the top of the other. You'll see it fits perfectly. So why does it look to us like the tops have different shapes? The reality engine of vision is at work, creating impressions of shape and depth, using rules that normally work well. But, in this figure, the normal operation of the reality engine constructs a visual reality that contradicts what we measure with a ruler.

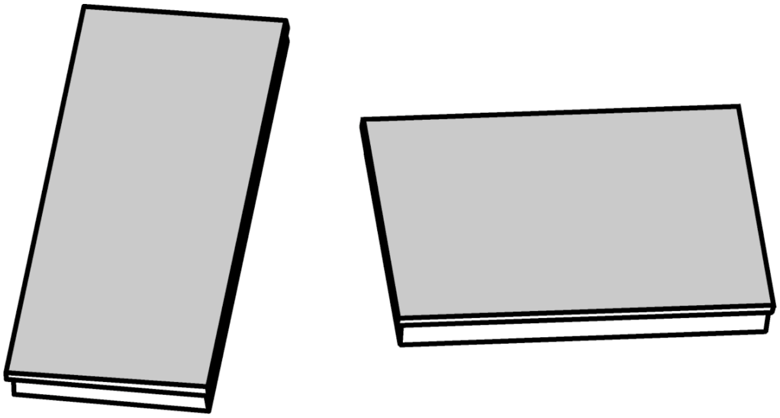

Figure 1. Identical box tops. The gray top of the box on the left has the same length and width as the gray top of the box on the right. Yet they look different. This unmasks an operation of our visual reality engine, and shows that it can construct realities that contradict what we measure with a ruler. This figure was inspired by a similar demonstration with tables created by Shepard (1990), which first appeared in Kubovy and Pomerantz (1981).

Another illustration of your reality engine at work is shown in Figure 2. When you view this figure, you see a ghostly cube floating in front of black disks. The white lines of the cube are sharp, and brighter than the background. But if you cover two adjacent black disks, you will see that no line extends between them. If you then uncover those two disks, you will again see a line between them. That line is entirely the creation of your reality engine. You can think of the reality engine as trying to create the most probable interpretation of the image. It notices the precise alignments of the white cutouts on adjacent black disks. Your reality engine decides that it is improbable that this alignment is accidental. It decides instead that this alignment is probably due to white lines that occlude the disks. So it hallucinates these lines, making them a bit brighter than the background .

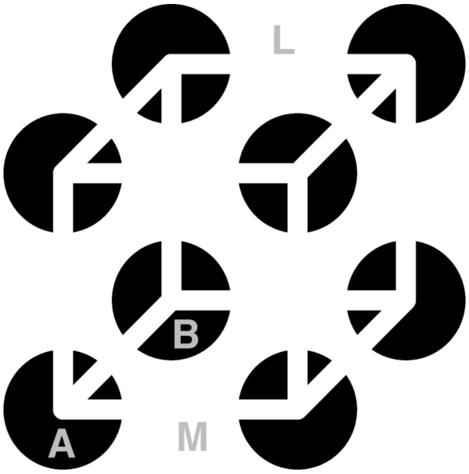

Figure 2. The subjective Necker cube. The white edges of the cube are entirely constructed by your visual reality engine. So also are the cubes that you see. The first subjective Necker cube figure was created by Bradly and Petry (1977).

Once your reality engine has constructed the illusory lines between the black disks, it then notices that these lines have a regular pattern, consistent with being the projection of a cube. It decides that this is probably no accident. It decides instead that a cube probably occludes the black disks. So it hallucinates the cube, and makes it float in front of the disks.

If you watch the cube for a while, you might see it suddenly flip. Sometimes you see a cube with corner *A* in front and sometimes a cube with corner *B* in front. Call these *cube A* and *cube B*. Your reality engine decides that cube A and cube B are both likely interpretations of this image, and so it first constructs one, then the other, repeatedly in an infinite loop.

But there are more cubes here. Think of the black disks as holes in a white sheet of paper. You look through these holes, and behind them see a white cube. Again, you sometimes see a cube with corner *A* in front and sometimes with corner *B* in front. So here are two more cubes that your reality engine constructs. Call them *cube A′* and *cube B′*.

Now comes a strange question. Which of the four cubes is there when you don’t look? Cube A, or A′? Cube B, or B′?

The answer must be that no cube is there when you don’t look. After all, each cube you see looks three-dimensional, but the page on which Figure 2 appears is flat. Therefore a three-dimensional cube is only there when you look, and only exists so long as your reality engine creates it (Hoffman, 2000). When you don’t look, the best description of what you’ll see next is a “superposition,” i.e., a description of the probability of seeing each cube. If, for instance, all four cubes are equally likely to be seen, then each has a probability of one fourth. When you next look, this superposition of probabilities "collapses," and you see one of the four cubes.

Notice the lines labeled *L* and *M*. Sometimes *L* is on the front of a cube, sometimes it is on the back. Similarly for *M*. But, whenever you see *L* in front you see *M* in back and vice versa: Your construction of the cubes "entangles" *L* and *M* to always be opposite.

One last observation. Notice that when you see the cubes in front of the disks, the edges have a ghostly quality, but when you see the cubes behind, the edges are no longer ghostly, but instead have a paradoxical quality of being invisible and yet substantial or solid. This quality of ghostliness versus solidity is also the creation of your reality engine.

Evolution and User Interfaces

Our reality engine constructs everything we see. So the question naturally arises, What is the relation between our visual constructions and the world that we don't construct, the world that exists whether or not we exist?

This question has a clear answer in evolutionary theory: Our sensory systems are shaped by natural selection to allow *homo sapiens* to survive long enough to reproduce within its niche. *Homo sapiens*, like, say, the cockroach, is a species whose sensory systems guide adaptive behavior within a particular niche. We don't expect the sensory systems of a cockroach, or a maggot, or a nematode, to give detailed insight into the true nature of reality. Instead we expect them to give simple signals, suited for a particular niche, that help avoid predators and find mates. Similarly for *homo sapiens* (Pinker, 1999).

So evolutionary theory leads us to think of our sensory systems as constituting a species-specific user interface. A user interface, like the windows interface on your laptop, is useful because it does *not* resemble what it represents. For instance, a file icon on your computer screen might be red, rectangular and in the upper left corner of the screen. But of course this doesn't mean that the file itself in the computer is red, rectangular, or in the upper left corner of the computer. The icon is there to guide our behavior, not to resemble the file. The icon *hides* the complexity of the computer, all its diodes and resistors and megabytes of software. That's what makes the user interface useful. And that's what makes the sensory systems of *homo sapiens* useful. Our sensory systems don't resemble the complex reality that they represent, but instead present a greatly simplified set of icons that let us behave adaptively in our niche (Hoffman, 2000).

Medical Implications

If our visual system is a species-specific reality engine, then damage to the visual system should impair our ability to construct visual realities. There are many fascinating examples of this.

For instance, there is an area of cerebral cortex in the so-called “lateral fusiform gyrus” whose activity is correlated with the perception of faces. If this “Fusiform Face Area,” or FFA, is damaged, say by a stroke, then, in many cases, the patient can be normal in every respect except one: The patient can no longer recognize faces by looking at them. Even the faces of family members or friends, or the patient’s own face seen in a mirror, are no longer recognizable (Bruce & Young, 1998; Sacks, 1998). However the patient can still recognize people by the sound of their voice, indicating that it is the specifically visual ability to recognize faces that is impaired. Patients with FFA damage can often recognize facial expressions, even though they cannot recognize facial identity: A different area of cerebral cortex in the posterior superior temporal sulcus (PSTS) has activity correlated with the perception of facial expression. If a patient has damage to the PSTS but not to the FFA, then in many cases they can recognize facial identity but not facial expression. This leads to serious social difficulties, since we rely on our ability to quickly read facial expressions in our daily social interactions.

Damage to other visual areas of cerebral cortex is correlated with other visual deficits (Behrmann, 2001). Bilateral damage to visual area V8 is correlated with “achromotopsia,” an inability to see color. The patient describes the world as appearing only in shades of gray. In some cases the achromatopsic patient can no longer dream in color or imagine colors. Similarly, damage to visual area V5 is correlated with an inability to see motion. Damage to visual area V1 is correlated with “scotomas,” regions of the visual field in which the patient is blind.

Because of these correlations, visual deficits can be invaluable in diagnosing neurological disorders. For instance, Alzheimer’s Dementia (AD) frequently damages visual areas of cerebral cortex. Some undiagnosed AD patients first complain that they cannot see. Ophthalmological examination reveals that their eyes are fine, but further examination reveals that regions of visual cortex are probably damaged with the plaques and tangles of AD. Research is in progress to use visual deficits, and concomitant changes in brain activity as measured by electroencephalograms (EEG), to diagnose AD before clinical symptoms appear, so that medical interventions can be introduced as early as possible (Sneddon et al., 2005).

Technological Implications

If the visual system is a sophisticated reality engine that constructs our visual worlds, then we should be able to study its operation, reverse engineer it, and build computer systems that mimic it. To “reverse engineer” a complex system means to study its structure and function to try to understand how it works, often with the goal of trying to build a new system that has the same capabilities. To reverse engineer human vision means to study it through psychophysical, neurophysiological and brain imaging experiments to learn the secrets of its operation.

Indeed this is proving possible. We now have computer vision systems, inspired by insights from biological visual systems, that take images streaming in from video cameras and construct depth, shapes, colors, motions and objects (Shapiro & Stockman, 2001). These systems are sufficiently advanced that they can autonomously drive cars on highways, using only computer vision to "see" the road.

Currently, the big open problem in computer vision is object recognition. It has proved remarkably difficult to recognize three-dimensional objects from arbitrary viewpoints, under arbitrary lighting, and with other objects obscuring them. The reason is that a single object can lead to dramatically different images as it is rotated, occluded, or presented under different lighting. A computer vision system must recognize that these widely different images are in fact images of precisely the same object. The technical difficulties involved make us admire the speed and accuracy with which human vision solves this problem. Reverse engineering visual object recognition, both in humans and other primates, has led to the surprising discovery that biological visual systems can recognize a three-dimensional object by using neural representations of just a few of its two-dimensional views. This has lead vision scientists to discover theorems and algorithms for recognizing three-dimensional objects given only a few two-dimensional views. In this way, reverse engineering biological vision has provided key insights for solving a difficult technological challenge.

Computer vision systems are already in use in industry, the military, and homeland security. Once the problem of object recognition has been sufficiently resolved, we can expect computer vision systems to serve as prosthetic devices for the visually impaired, to be used widely in industry for flexible automated assembly, and, inevitably, to be used for smart bombs and other military applications.

Philosophical Implications

What is the relation between our perceptions and the world that exists independent of us? This question has puzzled philosophers at least since Plato, and remains puzzling to this day. Advances in understanding human visual perception and the human visual system have helped to sharpen this question and to provide rigorous experiments that constrain our ideas. The question can now be posed as follows: What is the relationship between brain activity and conscious experience? Vision researchers are now studying many "neural correlates of consciousness," or NCCs. By getting detailed information about conscious visual experiences and their neural correlates, researchers hope to fashion a scientifically rigorous theory (Blackmore, 2003; Koch, 2004). We have already mentioned the correlations between neural activity in V8 and conscious color experiences, neural activity in V5 and conscious motion experiences, and neural activity in the FFA and conscious recognition of faces. These and other similar correlations are now being studied in great detail using brain-imaging techniques such as EEG, functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and magnetoencephalography (MEG).

This is a fascinating problem, in part because we do not yet have a single scientific theory that can explain how neural activity could cause, or be, conscious experiences. This might be surprising. But to get a feel for the difficulty of this problem, just consider what neural activity is. It is primarily various ions, such as sodium and potassium ions, passing through holes in neural membranes; it is various neurotransmitters, such as serotonin and dopamine, diffusing across synaptic clefts and binding to receptors. How can ions running through holes or neurotransmitters diffusing through clefts cause, or be, my conscious experience of the smell of garlic or the sound of a flute? Why shouldn't such neural activity, whether of single neurons or of interacting systems of neurons, go on without any conscious experiences at all? These are difficult and open questions that engage many vision researchers today.

SETI Implications

Are we alone in the universe, or might there be other intelligent beings on other planets? The Search for Extraterrestrial Intelligence (SETI) program has, for decades, been looking for evidence of extraterrestrial intelligence. Suppose that SETI succeeds. How will we communicate with the extraterrestrials? Should we use Greek? English? French? Given that humans have enough trouble understanding foreign human languages, it seems unreasonable to assume that extraterrestrials would do any better. Perhaps, then, we can use the same strategy that many traffic signs use: Instead of presenting the information in a specific language, present it by means of images. Such images are universally understood by humans, despite their linguistic differences. Perhaps images will also be universally understood by sufficiently intelligent extraterrestrials. With this hope in mind, the Pioneer 10 and Voyager 1 and 2 spacecraft were sent into space with many images of life on earth (Drake & Sobel, 1992).

But if vision is a species-specific user interface, shaped by the selection pressures of a particular ecological niche, it is unlikely that images will have the same interpretation for a civilization near the star Alpha Centauri as they have on earth. It is more likely that our interfaces will differ, and that considerable effort will be required to translate between them (Hoffman, 2007). The difficulty we have communicating with, say, dolphins, suggests how difficult such translation can be. If SETI does make contact, vision researchers may be at the forefront of efforts to establish reliable means of interstellar communication.

Conclusions

Vision is full of surprises. Its apparent ease hides its astonishing complexity and computational power. As we better understand vision we better understand ourselves, our evolutionary history, and the nature of our conscious experiences. We learn to create automatic vision systems for driving, manufacturing, homeland security, and the visually impaired. Perhaps, too, we will learn to succeed at interstellar communication.

Suggested Further Reading

Hoffman, D.D. (2000). *Visual intelligence: How we create what we see*. New York: W. W. Norton.

Livingstone, M.S. (2002). *Vision and art: The biology of seeing*. New York: Henry Abrams.

Pinker, S. (1999). *How the mind works*. New York: W. W. Norton.

References

Behrmann, M. (2001). *Handbook of neuropsychology, Second edition: Disorders of visual behavior*. Amsterdam: Elsevier Science.

Blackmore, S. (2003). *Consciousness: An introduction*. Oxford: Oxford University Press.

Bradley, D.R. & Petry, H.M. (1977). Organizational determinants of subjective contour: The subjective Necker cube. *American Journal of Psychology*, *90*, 253–262.

Bruce, V. & Young, A. (1998). *In the eye of the beholder: The science of face perception*. Oxford: Oxford University Press.

Chalupa, L.M. & Werner, J.S. (2003). *The visual neurosciences*. Cambridge, MA: MIT Press.

Drake, F. & Sobel, D. (1992). *Is anyone out there? The scientific search for extraterrestrial intelligence*. New York: Delacorte Press.

Gazzaniga, M. S. (2004). *The cognitive neurosciences III*. Cambridge, MA: MIT Press.

Hoffman, D. D. (2007). Images as interstellar messages. In D. Vakoch (Ed) *Between worlds: The art and science of interstellar message composition*. Cambridge, MA: MIT Press.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Englewood, Colorado: Roberts and Company.

Kubovy, M. & Pomerantz, J. (1981). *Perceptual organization.* Hillsdale, NJ: Erlbaum.

Sacks, O. (1998). *The man who mistook his wife for a hat*. New York: Touchstone.

Shapiro, L. G. & Stockman, G. C. (2001). *Computer vision*. New York: Prentice Hall.

Shepard, R.N. (1990). *Mind sights*. San Francisco: Freeman.

Sneddon, R., Shankle, W. R., Hara, J., Rodriquez, A., Hoffman, D.D., & Saha, U. EEG detection of early Alzheimer's disease using psychophysical tasks. *Clinical EEG Neuroscience*, 2005, 36, 141–150.