

Human Vision as a Reality Engine

Donald D. Hoffman

University of California, Irvine

- ***Please tell us about your current position and research interests.***

I am a professor of cognitive science at the University of California, Irvine. I also have appointments in the Departments of Philosophy and Computer Science. This might sound like an odd combination, but there is a simple reason: I want to understand how human vision works, hence the cognitive science. I am also interested in how our subjective perceptions are related to objective reality, hence the philosophy. Finally, I am interested in creating models of perception and reality that are precise enough to build working robotic-vision systems, hence the computer science.

- ***What got you interested in studying vision?***

My first interest has been to understand conscious experiences, such as the experience of pain, the taste of mint, or the blue color of a clear sky. Do our conscious experiences arise from brain activities? If so, precisely how? If not, what is the relationship between conscious experiences and brain activities? Vision provides a concrete and fascinating arena to study these questions.

- ***What has been the real-world impact of this work?***

It is natural for us to wonder what kind of creatures we are and what, if anything, makes us special in the universe. For me, the most exciting possible impact of the work on conscious visual experiences is that it might illuminate these perennial questions and help us better understand ourselves. As an added benefit, however, vision research can lead to practical applications such as automatic vision systems for driving, manufacturing, homeland security, medical image processing, and the visually impaired.

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. This apparent ease leads naturally to the assumption that vision is a simple process, no more complex than taking pictures with a camera. But research in cognitive neuroscience reveals that roughly one third of the brain's neocortex—billions of neurons and trillions of synapses—are engaged when we simply open our

eyes and look. Why should much 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.

Why does vision engage all this computing power? 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 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. That's why roughly half of your most sophisticated computing power is recruited for the job.

Admittedly, we don't feel as if we're constructing the visual world; it seems to always be there with us. We feel this way because our reality engine is so fast and effective that it creates what we see as fast as the eye can move. Thus, the very power and efficiency of the reality engine blocks our awareness of its existence. We can, however, 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 are they different? They certainly look different. The one on the left looks long and narrow, the one on the right short and fat. Nevertheless, they are identical. You can verify this with a ruler, or by using tracing paper to copy the top of one box, and then placing it on the top of the other. You'll see that 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 can measure.

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.

To understand what's happening, 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 and decides that this alignment is probably not accidental but is due instead

to white lines that occlude the disks. So it hallucinates these lines, making them a bit brighter than the background.

Once your reality engine has constructed the illusory lines between the black disks, it notices that these lines have a regular pattern, consistent with being the projection of a cube. It decides that this is probably no accident, concluding 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 (it might take a little time to see it). 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; its existence depends on your reality engine (Hoffman, 2000).

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 relationship 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 those systems to give simple signals, suited for a particular ecological niche, that help avoid predators and find mates. Similarly for *Homo sapiens* (Pinker, 1999).

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 this doesn't mean that the file itself is red, rectangular, or in the upper left corner of the screen. The icon is there to guide behavior, not to resemble the file. The icon *hides* the complexity of the computer, all of 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; Hoffman and Singh, 2012).

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. The neurological literature provides many fascinating case histories of people struggling to live with such conditions.

For instance, activity in an area of cerebral cortex called the lateral fusiform gyrus is correlated with the perception of faces. If the fusiform face area, or FFA, is damaged, say, by a stroke, the patient may be normal in every respect except one: He or she may be unable to recognize faces by looking at them—even the faces of family members, old friends, or the patient's own face seen in a mirror (Bruce & Young, 2012). Patients can still recognize people by voice, indicating that the specifically visual ability to recognize faces is the impaired function. Many people with FFA damage can still recognize facial expressions. Activity in a different area of cerebral cortex in the posterior superior temporal sulcus (PSTS) is correlated with the perception of facial expression. A patient with damage to the PSTS but not the FFA can usually recognize facial identity but not facial expression. This condition leads to serious social difficulties because we are so dependent on the 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 achromatopsia, an

inability to see color, in which the world appears only in shades of gray. Some people with achromatopsia can no longer dream in color or imagine colors. Similarly, damage to visual area V5 is correlated with an inability to see motion, while area V1 damage is correlated with scotomas, regions of the visual field in which the patient is blind.

Because damage to specific areas of the visual cortex and particular behavioral deficits are so closely correlated, visual deficits can be invaluable in diagnosing neurological disorders. For instance, some people with undiagnosed Alzheimer disease (AD) 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 (EEGs) 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 advanced enough to 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. These technical difficulties make neuroscientists admire the speed and accuracy with which human vision solves this problem.

Efforts to reverse engineer 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 insight has led to the discovery of 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, homeland security, and medical image processing. 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 the basis for 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. Detailed information about conscious visual experiences and their neural correlates may enable neuroscientists to fashion a scientifically rigorous theory (Blackmore, 2012). I 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, positron emission tomography, and magnetoencephalography.

This problem is fascinating in part because we do not yet have a single scientific theory that can explain how neural activity could cause, or be, conscious experiences. If you find this surprising, 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 interacting systems of neurons, go on without any conscious experiences at all? These difficult and open questions currently engage many vision researchers and neuroscientists (Koch 2012).

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 extra-terrestrials? Should we use Greek? English? French? Given that humans have enough trouble understanding foreign human languages, extraterrestrials aren't likely to 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. All humans, despite their linguistic differences, understand such images; perhaps they 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. Instead, our interfaces will probably differ, and considerable effort will be required to translate between them. The difficulty we have communicating with, say, dolphins, suggests how complicated 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.

Conclusion

Vision is full of surprises. Its apparent ease hides its astonishing complexity and computational power. As we better understand vision, we will better understand ourselves, our evolutionary history, and the nature of our conscious experiences. We will create automatic vision systems for driving, manufacturing, homeland security, medical image processing, 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.

Hoffman, D.D. (2009). [The interface theory of perception](#). In *Object Categorization: Computer and Human Vision Perspectives*, S. Dickinson, M. Tarr, A. Leonardis and B. Schiele (Eds). Cambridge University Press, pages 148-265.

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

References

Behrmann, M. (2001). Disorders of visual behavior. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (2nd ed.). Amsterdam: Elsevier Science.

Blackmore, S. (2012). *Consciousness: An introduction, 2nd Edition*. 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. (2012). *Face perception*. New York: Psychology Press.

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

Hoffman, D.D. & Singh, M. (2012). Computational evolutionary perception. *Perception*, 41, 1073-1091.

Koch, C. (2012). *Consciousness: Confession of a Romantic Reductionist*. Cambridge, MA: MIT Press.

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

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

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

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

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 Roger Shepard (1990), which first appeared in Kubovy and Pomerantz (1981).

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. Drake Bradley and Heywood Petry (1977) created the first subjective Necker cube figure.