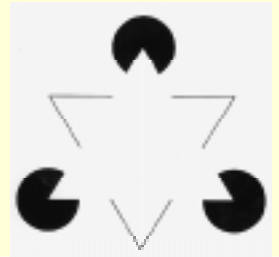


Our brains have to make hypotheses about what we think is out there in the real world. We hope that these hypotheses are right, because a mistake could be fatal.



How We See

by Richard A. Andersen



Driving down a street (as here in La Cañada Flintridge, not far from Caltech), generates an optical flow of motion signals. Surrounding objects seem to radiate out from the focus point, expanding toward the edges of the field of vision. Processing of these complex signals occurs in the higher levels of the brain.

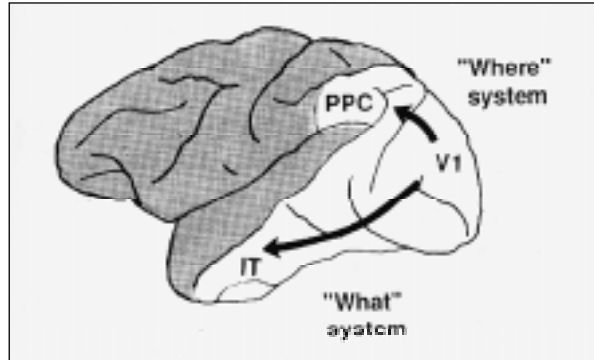
When we look around us, seeing is so effortless that we think we naturally perceive what is actually out there in the world. But, in fact, the brain works very hard at reconstructing its own reality—what we refer to as neural representation. In the well-known Kanizsa triangle shown above, you can see illusory contours that are created by the occlusions, the lines, and the little Pac-Man figures. These “contours,” and the perceived variations in brightness lie entirely within your brain and do not exist in the real physical world. Because the brain is often faced with an ambiguous, ill-defined environment, it’s very useful to be able to reconstruct such lines. Our brains have to make hypotheses about what we think is out there in the real world. We hope that these hypotheses are right, because a mistake could be fatal. In terms of evolutionary pressure, the brain has evolved over time to create its own reality that meshes with the world in such a way as to enable the organism to survive.

Neurobiologists believe that at least a third of our approximately one hundred thousand genes are exclusively involved in brain function. With that limited number of genes, we can’t completely specify all the complex connections and structure in the brain, so during some periods of development, the brain has to look to the outside world for assistance in forming its structures. At a very early age, for example, the brain becomes plastic for vision; during this critical period information from the two eyes, which compete with each other, is used to actually set up the appropriate neural machinery for depth perception.

The understanding that the visual system actually constructs images of reality has led to an exciting revolution not only in neuroscience but also in the field of philosophy. A new school of philosophy called neurophilosophy has reconsidered what the nature of reality and the nature of knowledge are, based on what we now know about how the brain works and about the changes that occur in our neural networks over the course of development.

Visual information required to construct this representation of the world comes in through the eyes and is projected on the retina; then the optic nerve sends this information to the thalamus, which passes it up to the primary visual cortex (called V1), where simple aspects of the visual scene are first analyzed. Then information is projected out to cortical areas around the primary visual cortex, and they process the visual image more elaborately; here is where the more complicated cognitive functions take place. The information travels along two processing streams—one to the upper part of the brain and the other to the lower part. In 1982, two neuropsychologists from the NIMH, Mort Mishkin and Leslie Ungerleider, proposed that the pathway to the upper part of the brain was the “where” pathway, which tells us the location of an object. They labeled the lower route the “what” pathway, because it seems to handle information about the object itself. Patients with

Right: This schematic view of a monkey brain shows the "what" and "where" pathways. Both pathways begin at the back of the brain in the primary visual cortex (V1); the "what" pathway proceeds through the visual cortex in the lower part of the brain to the inferotemporal cortex (IT), and the "where" pathway leads to the posterior parietal cortex (PPC) in the upper part.



injuries, or lesions, to the upper pathway can identify objects and the differences between objects, but can't tell where they are. With lesions to the lower area a person can tell where things are but can't identify them. Lesions in this area can cause an interesting syndrome called prosopagnosia, in which people can't identify faces, including their own. This object-based pathway is also important for the perception of color.

A typical lesion in the upper, or "where," pathway might leave a patient unable to pour a liquid into a glass. He can see the glass and he knows it's a glass, but he can't figure out where the glass is with respect to his body. Another one of the deficits from damage to this pathway is the inability to attend to the area of space opposite to the hemisphere that was damaged.

Monkeys have visual functions similar to ours. They see color the way we do; they see motion and depth; they perceive objects; they make eye movements in the same ways that we do. So they make ideal animal models for studying the human brain, because we can do experiments with monkeys that we obviously can't do with people. We have several rhesus monkeys who participate in experiments for a period of years. Recently we have been successful in placing them in zoos for their retirement. A common technique for studying the visual system introduces very fine (about the diameter of a human hair) wire electrodes into a monkey's cerebral cortex. We park these electrodes near nerve cells. During the experiments the monkeys are awake and performing different tasks that they've been trained to do, such as moving their eyes toward a stimulus, reaching toward a target, or pressing a button for a juice reward. In this way the monkeys "tell" us what they see. As they do their tasks, the electrodes record the activity of the nerve cells. Then we can correlate the activity of specific cells with the behaviors or perceptual experiences the animals have.

The illustration at the bottom of the opposite



Above: A microelectrode is placed near a neuron to record its electrical activity (photo by Fritz Goro). A recording is shown at right, on the opposite page; the changes in activity are due to the appearance and disappearance of stimuli.



Above: The Rose Bowl prank of 1961 is a good example of population coding, one of the brain's strategies for processing visual information. A single perception is stored across many units, but it only makes sense when all the units are combined.

To understand the brain and how it processes visual images, we not only have to know what each single element is saying, but also what the whole ensemble of activity is saying.

page shows the type of signal that we record on one of these electrodes. Time is plotted along the x-axis, while the y-axis displays the membrane potential, or electrical activity, coming from one of these nerve cells. When we shine a light or present a stimulus to the animal, a cell that is involved in the perception of that stimulus begins to fire action potentials—pulses that are the communication method for nerve cells. These signals will then be transferred via synapses to other nerve cells to which this nerve cell projects. This synaptic transmission is how messages get sent through the cerebral cortex, and by tapping into this system with our electrodes, we can determine the locations of very specific types of visual processing that the brain uses to reconstruct reality.

The brain uses five basic strategies in its visual processing: population coding, functional localization, parallel processing, hierarchical processing, and association. A single neuron in the brain looks at only a small piece of the world. This fragmentation actually starts in the retina, which has the image of the whole visual field on it, yet a single cell receives its input from only a tiny part of that image. So we have to realize that each time we record a signal from one of these nerve cells, we're seeing only a small part of the entire visual message. This brings us to a concept known as population coding—the idea that a whole perception is stored across many, many units. Our brains are a bit like TV sets; we can think of neurons as corresponding to the pixels on the screen. Of course, a normal TV screen measures about 600

by 400 pixels, while the brain contains about a hundred billion cells. Each one of these cells can change its activity over a certain range to store a small bit of the “picture.” A simple example of population coding can be seen in the great Rose Bowl prank of 1961, where each University of Washington fan knew only that he or she was holding up a white or a dark card and, fortunately for the Caltech students who pulled off the prank, no one person could see the whole message. When they all flipped their cards in unison, they inadvertently spelled out CALTECH. To understand the brain and how it processes visual images, we not only have to know what each single element is saying, but also what the whole ensemble of activity is saying together.

A second important feature of how the brain works is known as functional localization. This concept refers to the fact that different parts of the cortex are specialists in particular visual processes. At the turn of the century, a German neuroanatomist, Korbinian Brodmann, divided the human brain into about 50 different areas simply by looking at sections of it under a microscope and noticing the differences in nerve-cell structure or packing density in different cortical regions. With the advent of microelectrode recording techniques, neurophysiologists in the 1970s began dividing the brain up into areas based on different functional activities as well. Often these functional areas corresponded to Brodmann's anatomical ones; for example, V1 was his area 17. But others, like Brodmann's area 19, turned out to contain many different cortical areas delineated by functional differences. It's also important, in dividing up the cortex, to notice that one area might connect to some areas and not to others, so that different cortical areas have specific connectivities between them. About 35 different cortical areas have been identified as being involved with vision in monkeys, and there are probably even more in our own brains.



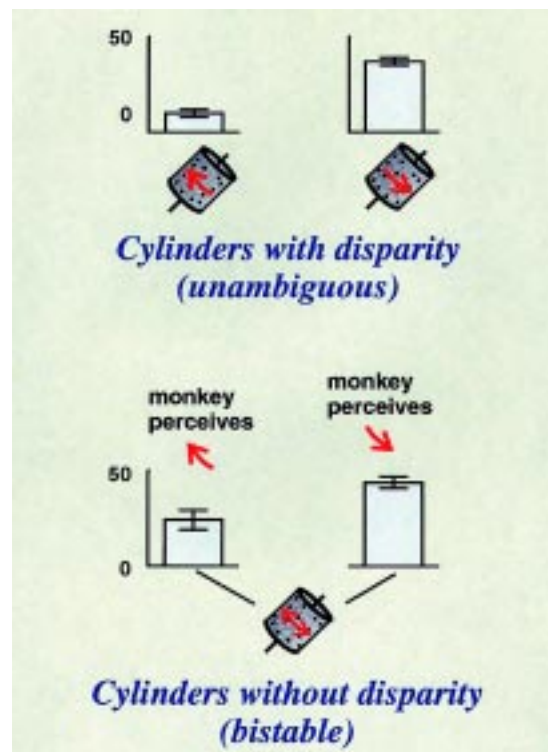
dots, however, we use high-speed, computer animation to generate these 3-D structure-from-motion stimuli. When we project such an image onto a flat computer monitor screen, we lose the depth information that we would normally get from looking at the cylinder with two eyes, but, amazingly, due to the motion signals, we can still perceive a revolving hollow cylinder. This computer simulation demonstrates that the brain is able to use motion signals to reconstruct three-dimensional depth. It is most interesting, however, that, since there's no depth information contained in the projected stimulus, the direction in which the cylinder appears to be rotating is ambiguous. Sometimes you may see it rotating clockwise, other times counterclockwise. And it appears to shift directions; we refer to this spontaneous shifting as a bistable percept. An example of another bistable percept is illustrated at right: the well-known Necker cube illusion. Some people will see the upper square as being in front, and others will see the bottom square in front. If you look at it for awhile, you'll see it flip spontaneously. (Sometimes it helps if you concentrate on one point and then on another to see the flipping.)



The diagram at right shows the activity of an MT neuron to rotating cylinders. The upper two bar graphs indicate the amount of activity (in action potentials per second) when unambiguous cylinders containing depth cues rotated counterclockwise (left) and clockwise (right). This MT cell preferred counterclockwise rotation. The lower bar graphs illustrate activity from this same MT neuron, but in both cases with the same bistable cylinder that lacks depth cues. When the monkey perceived the cylinder rotating in the clockwise direction, the cell was less active (left) than when he perceived the identical stimulus rotating counterclockwise (right).

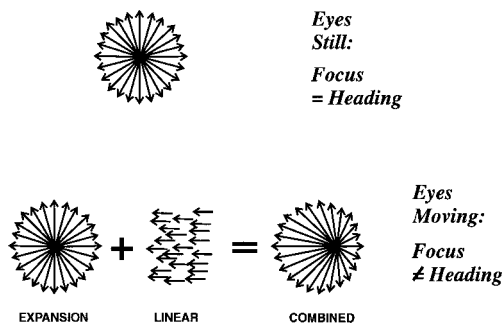
Postdoc David Bradley, grad student Grace Chang, and I trained monkeys to tell us with eye movements which direction they saw the cylinders rotating; we then recorded from their MT neurons. In some trials we added in stereoscopic depth cues in the computer display using an anaglyph technique similar to that used in the old 3-D movies of the 1950s. We found that when the monkey looks at a rotating cylinder with depth cues, the

cylinders are unambiguous, and certain cells will prefer certain directions of rotation. For example, when the cylinder is rotating counterclockwise, it will generate a lot of activity in a given cell. But when it's rotating in the opposite direction, the same cell is much less active. Because of the stereoscopic depth cues added to the dots, the cell is sensitive to the three-dimensional structure of the cylinder. In the bistable state, however, in which the cylinder is projected on a two-dimensional surface and there is no depth information, the monkey still tells us the direction he thinks the cylinder is rotating. Sometimes he says it's rotating one way, sometimes the other. When he thinks it's rotating counterclockwise, the nerve cell reliably reports this by the activity it generates corresponding to its perception. This result indicates that we've tapped into the area of the cortex that is analyzing this depth from motion, and we can actually see in the nerve-cell activity what the monkey is *perceiving*. And even though the information on his retina remains the same, the cells respond differently, indicating that the changes in perception—of which way the cylinder is turning—are occurring in this part of the brain.

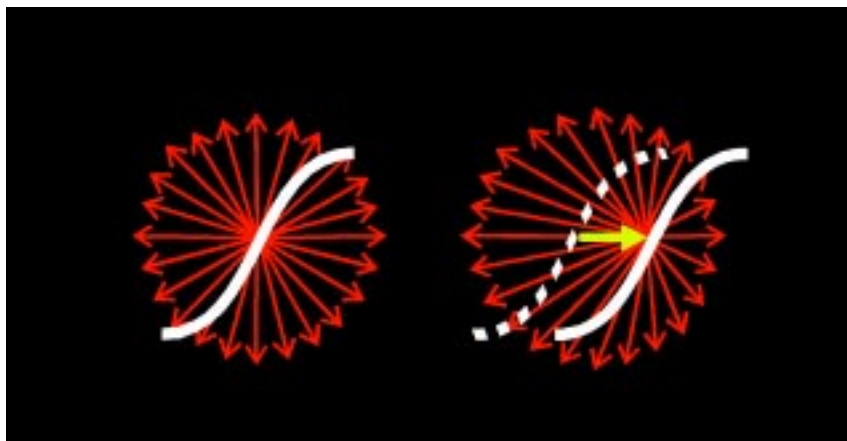


If we continue upward along the motion pathway's hierarchical organization we come to a tiny area called MST (medial superior temporal area), which is about half the size of the nail on your pinkie finger. Humans and monkeys both have an MST; it's specialized for helping us to navigate through the world using motion information. While you're driving along a highway or walking along a street, you generate motion signals. These signals are called optical flow. At the point—or focus—toward which you're headed, there's very little motion, but around this focus point motion appears to radiate out, speeding out toward the edges of the visual field like an expanding circle. We call this spot the focus of expansion; it corresponds to the direction in which you're heading, and it gives you useful information about

The upper stimulus shows that when the eyes are still, the focus of expansion corresponds to the direction of heading. The lower part of the diagram indicates that when the same stimulus is generated by the same self-motion direction, but the eyes are also tracking to the right, generating linear motion to the left, the two motions combine on the retina to form a pseudo-focus displaced in the direction of the eye movement. This pseudo-focus does not correspond to the direction of heading, which is still straight ahead.



where you're going in the world. Cells in MST are tuned to these sorts of expanding stimuli generated by motion and also to the location of the focus. Now, a problem occurs when you're moving through the world in one direction but you begin to track something with your eyes—say a freeway sign—that may be off to the side. Moving your eyes introduces a motion of your visual field in the opposite direction. For example, if you hold a finger in front of you and follow it with your eyes as you move it to the right, you'll notice that everything behind it moves to the left. With a rightward eye movement, you've introduced a leftward motion onto the eye. If you're also moving at the same time, this retinal motion gets combined with the expansion signal, shifting the focus toward the direction in which the eye is moving. If our brains were, in fact, using only this new focus to guide us through the world, when we looked at a sign on the freeway we'd run into it,



An MST neuron tuning curve superimposed on an expansion pattern shows that the cell responds at half its maximum response when the focus is straight ahead (left). When the eyes are moving to the right, the focus shifts to the right, but so does the tuning curve of the MST neuron (right).

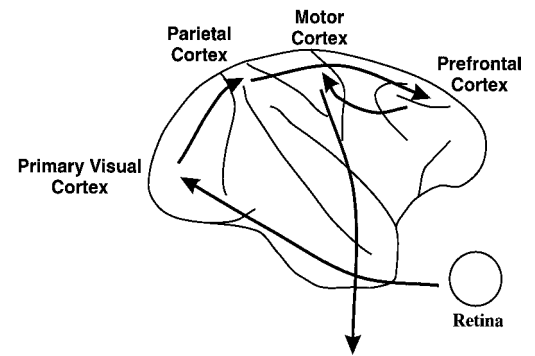
because that would be the point where the image is now stabilized, with everything else radiating out from it.

But we know we don't do that. To find out what's going on in the brain during this process, we (David Bradley, Marsha Maxwell, and Krishna Shenoy from my lab; Marty Banks, a professor at UC Berkeley; and I) have recorded from nerve cells in MST. The tuning curve of such a cell (which describes the frequency of the electrical signal coming from a cell) for an expanding stimulus is shown at lower left. If the expansion point is straight ahead, this cell is firing at about half activity; if the expansion point is over to the right, the cell is very active, and if it's to the left, the cell's not active at all. If we then have the monkey move its eyes so that it shifts the eye's focus in the direction of the eye movement (the equivalent of looking at the freeway sign), we find that the nerve cells shift their tuning curves to compensate for the eye movement. The cell continues to fire at half activity, indicating that the monkey knows it hasn't changed its heading. What we think is happening is that the areas in the front part of the brain that are sending out signals to move the eyes are also sending signals back into the perceptual areas saying: "The eye is moving; shift your receptive fields to compensate for it so that you still perceive locations in the world as being the same." This mechanism is called efference copy or corollary discharge, and it explains why, when we move our eyes around and shift the images on our eyes, the world still appears stable. We are using information about what we're doing with our eyes to stabilize the visual world. Thus we can see that there is a hierarchy from V1, which measures motion, to MT, which extracts the 3-D structure of surfaces in motion, to MST, which helps us navigate through the world.

The final processing strategy that I'll discuss is association. The bouncing red ball has now been divided up so that it's processed along three different streams—motion, color, and shape. But since we view the world as a unitary entity, at some point we need to begin bringing this information back together again into one picture. This binding of features back together occurs at the highest levels of the visual cortex, in the visual association areas.

A few years ago, our lab described an area called LIP (lateral intraparietal area), which is important for perceiving visual space and is located in the upper "where" processing stream. LIP is also important for making eye movements by gathering information from the visual cortex and sending it to the front part of the brain to move the eyes. However, we not only move our eyes to locate visual stimuli, but also to identify auditory stimuli. We know that our brains can perceive a sound location as easily as a visual location, but auditory information is collected in a very different way. It is assembled from auditory cues

The schematic diagram of the brain at right illustrates the pathway of visual information that leads to visually guided movements. Visual information first passes through the primary visual cortex and proceeds through the posterior parietal cortex to frontal lobe structures. The primary visual cortex is responsible for sensation, and the motor cortex for sending out commands to make movements. Evidence shows that the early neural correlates of plans to make movements appear in this pathway as early as the posterior parietal cortex.



arriving at the two ears, while visual information is imaged on the retinas in the eyes. The brain has to combine these two very different types of signals to come up with a single, unified spatial representation. To this point, we had tested LIP neurons only with visual signals. We were, however, interested in how this high-level processing area might combine or “associate” features of external stimuli to locate them in space. So we developed an auditory localization task.

It turned out that when Brigitte Stricanne, Pietro Mazzone, and I recorded from nerve cells in the LIP area (which is a part of the posterior parietal cortex), we could also map tuning curves or receptive fields for auditory stimuli. We had the monkey sit in a room with his head facing straight ahead, keeping his head always in the same position. He did, however, have to move his eyes to look at three different locations in the room. We played tones sequentially from speakers in different locations in order to map the cell’s preferred location in space. When the animals looked in the three different directions, the preferred auditory location actually shifted in space

But since we view the world as a unitary entity, at some point we need to begin bringing this information back together again into one picture.

by the same amount as the shift in gaze direction. In other words, the selectivity of the cell to the sound moves with the eye. This finding shows that the auditory signals have been mapped onto the same coordinate frame as the visual signals, which also move with the eyes. We say that both the auditory and visual signals are in an eye-centered reference frame. Auditory and visual information have been brought together and associated in LIP to form a single common perceptual representation of the world.

In the last couple of years we have begun to investigate how sensory signals lead to decisions and plans for action. Working in such a high-order area as the posterior parietal cortex, with so many fascinating neural activities, we have wondered if *intentions* might be hatched here. Since the posterior parietal cortex lies between sensory areas and motor areas and acts as an interface between them, it seemed a likely candidate for the location of the neural correlates of intention. In experiments published in March in *Nature*, Larry Snyder, Aaron Batista, and I trained our monkeys

to do one of two tasks when directed by a signal. On a green signal light they were to reach in the dark for the remembered location of a briefly flashed target; a red signal light told them to make an eye movement (saccade) to the target instead. They had to memorize the target’s location over a delay of one to one and a half seconds before they acted. We measured the activity of specific neurons during this delay and discovered that the neurons fired not only to a specific location in the visual field but also according to whether the monkey was planning to look at or reach for the target. Moreover, the cells selective for eye movements were confined to area LIP, the saccade area, and the reach-selective cells were confined to a reach area abutting LIP. This anatomical segregation shows that a motor plan, guided by the visual perception, originates here in the culmination of the “where” pathway, and that the intended response, rather than the visual information, may be the determining factor in organizing how neural computations are made within the area. This may be the place where our thoughts begin to turn into actions, and where our spatial perception is mapped not only by what our senses tell us but also by how we plan to use that information. □

This article was adapted from a talk given by Richard Andersen at Seminar Day in May 1996. Andersen, the James G. Boswell Professor of Neuroscience, came to Caltech in 1993 from MIT, where he had been a member of the faculty since 1987. Before that sojourn on the other coast, he had earned his BS at UC Davis (1973) and his PhD (1979) at UC San Francisco; he was a postdoc at the Johns Hopkins School of Medicine but returned to California as a faculty member at the Salk Institute and UC San Diego. Andersen’s research is supported by the National Eye Institute (part of NIH), the Office of Naval Research, the Sloan Foundation, and the Human Frontiers Scientific Program.

PICTURE CREDITS:
24, 30 – David Bradley;
26, 28-31 – Richard Andersen; 26 – Fritz Goro; 27 – Bruce Whitehead