VISION AND BRAIN

HOW WE PERCEIVE THE WORLD

JAMES V. STONE

Vision and Brain

Vision and Brain

How We Perceive the World

James V. Stone

The MIT Press
Cambridge, Massachusetts
London, England

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please e-mail special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Stone Sans and Stone Serif by Toppan Best-set Premedia Limited, Hong Kong. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Stone, James V., Dr.
Vision and brain : how we perceive the world / James V. Stone.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-51773-7 (pbk. : alk. paper)
1. Visual perception. 2. Vision. I. Title.
BF241.S77 2012
152.14–dc23
2011052297

10 9 8 7 6 5 4 3 2 1

To Verity, Chris, Dorian, and Bob

Some may fear that such a materialistic outlook, which regards the brain as a kind of super machine, will take the magic out of life and deprive us of all spiritual values. This is about the same as fearing that a knowledge of human anatomy will prevent us from admiring the human form. Art students and medical students know that the opposite is true. The problem is with the words: if machine implies something with rivets and ratchets and gears, that does sound unromantic. But by machine I mean any object that does tasks in a way that is consonant with the laws of physics, an object that we can ultimately understand in the same way we understand a printing press. I believe the brain is such an object.

—David Hubel, *Eye, Brain, and Vision* (1988)

Philosophy is written in this grand book—the universe—which stands continuously open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and interpret the characters in which it is written. It is written in the language of mathematics . . . without which it is humanly impossible to understand a single word of it; without these one is wandering about in a dark labyrinth.

—Galileo Galilei (1564–1642), *The Assayer*

Contents

Preface

This book tells the story of how we see. It is a story with only two characters, the physical world and the brain. The physical world is a cunning, deceitful character, full of lies, or worse, half-truths. It is not to be trusted at any time, nor at any cost. The brain is a flawed detective with a loaded die for a compass, working on lousy pay with fuzzy data, and a strict, sometimes literal, deadline. But over eons of evolutionary time, the brain has always had one crucial advantage: it knows that the physical world has to play by certain rules, rules that are ultimately derived from the laws of physics. Armed with this singular insight, the brain tests and retests, millisecond by millisecond, multiple competing hypotheses about what in the world might have produced the evidence of its own eyes, ruthlessly casting aside red herrings and fallguys one by one, by one, until there is only a single suspect who does not have a rock-solid alibi: and that is the one chosen by the brain, that is what we *see*.

Scientific Revolutionaries

If you really want to understand how science works then don't read this book. Read books written by scientists who were revolutionaries, scientists who simply stepped over the cutting edge of conventional science to create whole new fields of research. For physics, read Albert Einstein's *Relativity: The Special and the General Theory* (1920) or Richard Feynman's lectures (Feynman, Leighton, & Sands 1964); for evolution, read Charles Darwin's *On the Origin of Species* (1859); and for information theory, read Claude Shannon and Warren Weaver's *The Mathematical Theory of Communication* (1949). These are the scientists who defined whole new areas of investigation, scientists who created a new research field, and made it shine. Their books are beacons of insight and wisdom. More importantly, these books demonstrate how scientific skepticism can be used to ask the right questions, and how not to be fooled by the physical world, or by yourself (as Richard Feynman said, “The first principle is that you must not fool yourself, and you are the easiest person to fool”).

Vision and Brain: How We Perceive the World is intended as a companion to John Frisby’s and my more comprehensive book, *Seeing: The Computational Approach to Biological Vision* (2010). A small amount of the material in *Seeing* is inevitably duplicated here (after all, there’s only one human visual system to describe).

The title of this book is a little ambiguous. The question implicit in the title is: “How does the brain see the world? At one level, the answer might be a description of how individual neurons react in response to particular retinal images. But this would be like asking how a video camera works and being told the changing voltages of every component inside the camera. In one sense, this is a complete answer, but it is essentially useless in terms of understanding how a video camera works. At a different level, the answer could consist of a description of which different brain regions seem to be activated by different visual features, such as color, motion, or faces. But this would be like being told which of the various electronic modules in the camera was responsible for each different function without being told how each module works. Admittedly, these two extreme scenarios are caricatures, but they do give an impression of the different approaches to understanding vision. Now supposing we assume that the brain, in contrast to a video camera, behaves according to a set of well-defined principles, principles that dominate its processing at every level within the system. One consequence of making this assumption is that it forces us to reformulate questions about vision and the brain in an increasingly rigorous and mathematically succinct manner. Armed with these questions, the answers obtained take the form of rigorous and transparent mathematical models. Even though the details of such models would normally demand some mathematical expertise, the principles on which they are based are, in essence, simple, and even obvious. *Vision and Brain* is intended to provide an account of these principles, the principles that seem to underpin how the brain sees the world, but without the distracting mathematical details that such rigor usually entails. Accordingly, this book provides an account of the principles on which vision and brain function may be based, but it requires that readers have only a high school level of mathematical competence. In writing what I hope is an accessible survey of modern findings on vision, I’ve been inspired by Marr’s computational framework (1982), the efficient coding hypothesis (Atick & Redlich 1992; Attneave 1954; Barlow 1961; Laughlin 1981), Bayesian inference (Bayes 1763), and information theory (Shannon & Weaver 1949).

We begin with an overview of vision. This is followed by an examination of the particular nature of the problems faced by the brain in attempting to solve the problems of vision, but portrayed in the context of general principles that may allow it to do so. The all-too-thin line between accessibility and the larger (usually complicated) truth is trodden with some care, and when the occasional wobble leaves the reader on the wrong side of this line, reassurance is provided in the form of frequent informal summaries. This book is based on the philosophy that every key idea should be accompanied by an explanatory figure, and these figures should also help to keep the reader on the right side of that narrow line. And if the text appears overly simplistic in places, this is because there are no depths to which I will not sink (in terms of simplicity, diagrams, and analogies) in order to explain a complex idea to willing but inexpert readers. So, if you want to understand the scientific basis of human vision then read on.

Acknowledgments

Thanks to my wife and colleague, Nikki Hunkin, for all sorts of help and advice and to our children, Sebastian and Teleri, for helping to ensure that at least some of the ideas expressed here possess the simplicity demanded by the crystal-clear logic of a child's mind. Thanks to David Buckley for his enthusiasm for vision, to Pasha Parpia for long conversations on the nature of brain science, to John DePledge for making me think hard about Satosi Watanabe's ugly duckling theorem (which is why it was removed), to Steve Snow and also Royson Sellman for pointing out the limits of stereopsis and parallax, and to Chris Williams for comments on Bayes's theorem.

Several colleagues and friends have generously given up their time to read a draft of this book. I wish to express my gratitude in particular to David Buckley, Stephen Eglen, John Frisby, Nikki Hunkin, Stephen Isard, Pasha Parpia, and Tom Stafford. I hope that psychologists Buckley, Frisby, Hunkin, and Stafford find the mathematical ideas in this book clear, and I apologize to mathematician Isard, physicist Parpia, and computational biologist Eglen for bastardizing some beautiful theorems to force them through the word colander that an introductory textbook demands (and to SE for both).

Thanks to Bob Prior, Susan Buckley, Katherine Almeida and many others at MIT Press for advice in completing the various tasks involved in writing a book.

Finally, thanks to the University of Sheffield for standing by the noble tradition that a university has a duty to share the knowledge of its academics not only with students, but with the widest possible audience.

The Party Trick

It was a Christmas party in 2010, and I sat munching snacks. He just came up to me, shook my hand, and said hello. Then he took two rubber bands out of his pocket; he looped one over my thumb and forefinger and one over his, except that, together, our bands formed a one-link chain. And I realized that he and now I were part of the entertainment. Then he did something extraordinary. He moved his hand back, so that his link pulled my link toward him, and his rubber band simply *passed through* mine. Apparently, the link had been broken, except that both rubber bands were intact. He looked at me expectantly as the applause broke out. "Very impressive," I said, but my voice was a flat monotone. "'Wow' is the usual response," he said. "I teach vision," I said.

It took me a little time to work out why I said that. But, once I had, I wish I'd said this:

> I teach vision, about how the brain manages to see the world, despite objects that hide among their own shadows, the lousy lens of the eye, the back-to-front photoreceptors, the cussed nature of rainbowed light, the blood vessels that run amok over the retina and shield photoreceptors from that light, the squelching of 126 million photoreceptors' information into the one million cables of the optic nerve, the transformation from a fuzzy photograph-like retinal image to a series of Morse code-like blips that race into the brain's recesses, where the retina's image is corrupted by the stark unreliability of neuronal components that are the shrapnel of the Cambrian explosion from 543 million years ago. We know a large number of facts, a virtual archipelago of clues, about how this machine works, but we do not understand why it works as well as it does. That it works at all seems little short of miraculous, and the more facts we gather about each part, the less we seem to understand about the whole, and the more miraculous it seems. So when you seem to pull your rubber band through mine, I know it's a trick, but my incomprehension is immeasurably trivial compared to the ancient arc of steely ignorance that shields us from the knowledge of how we manage to see any rubber band. That's the mystery. Not how my visual system can be tricked so that it sees something that could not possibly exist, but how any visual system, annealed in the unknowing cauldron of evolution, *sees anything*.

1 Vision: An Overview

And here we wander in illusions.
—William Shakespeare, *Macbeth*

Impressions

Photons smash into the eye's darkness, deforming light-sensitive molecules, causing a cascade of chemical fallout, ending only when a stream of spikes escapes through the backdoor of the eye. These are coded messages from the dazzling world. Unscrambled, they command pictures in the head, with spectral order, and a particular beauty. They burst un-bidden, from eyes that cannot themselves see, and shower the brain with information, until, in that dark place, these messages let there be light.

All that you see, all that you have seen, and everything you will ever see is delivered to your brain as a stream of digital pulses whizzing along the fragile threads of salty, fluid-filled cables that are your nerve fibers. These nerve fibers, or *neurons*, are the only connection between you and the physical world, and the pulses they deliver are the only messages you can ever receive about that world. Once the enormous implications of this fact are appreciated, it no longer seems remarkable that we can experience illusions and even hallucinations. Instead, what does seem remarkable is that anyone ever sees anything at all, or at least anything that actually exists.

How We See: The Brain as a Detective

Seeing is very easy—and very hard. Look around you: it is easy to see the world. Now try to explain how you see that world, and you will begin to find out how hard seeing really is. But you are in good company because the truth is that nobody knows exactly how we see, even though there is an enormous body of knowledge about optics, the eye, and the brain. This book is a brief account of that knowledge.

The brain is constantly doing its best to find out what in the world is responsible for the image on the retina. In essence, this is the central problem of vision. It's as if the brain is a detective at the scene of a murder, with the retinal image as the body. But this is a case that would defeat even the great Sherlock Holmes. There are many clues, but they're scattered over a large area. The suspects are many, but devious, and all of them have both motive and opportunity. But the worst of it is that all the suspects who don't have rock-solid alibis appear to be equally and utterly guilty. In the face of so much evidence, the brain, drawing on some 3.8 billion years of inherited evolutionary experience, does what it does best: it makes an intelligent guess. Its trick is to arrive at a guess as good as it can possibly be while taking into account every quantum of evidence and every iota of past experience—a guess that is *optimal*.

Illusions: How the Brain Fails?

We can find clues to how the brain works in the ways it fails to work when confronted with particular optical illusions. First, the brain has to find ways to see beyond the image in order to see what is probably in the world, by compensating for information *missing* from the retinal image, as shown in figures 1.1 and 1.2. Second, the brain has to see what is in the world by sometimes disregarding information that is *present* in the retinal image. For example, in order to see the patches on the cube in figure 1.3 (plate 1) as different shades of gray, the brain has to disregard the fact that they are the same shade of gray on the page. Let's briefly consider several illusions in terms of how the brain fails to work.

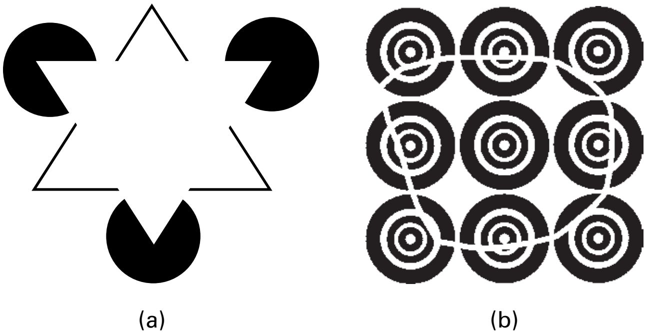


Figure 1.1
(a) Kanizsa triangles show that some lines are more apparent than real. (b) Variant on the Kanizsa illusion. The white curve is apparent, but not real, within and between the nine sets of concentric disks. Created by author.

(a) (b)

Figure 1.2
Disappearing lines of the World Wildlife Fund panda logo between 1961 (a) and 2006 (b). Lines around the head and back that are real in (a) are merely apparent in (b). Reproduced with permission.

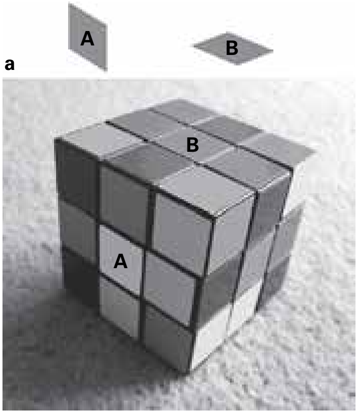


Figure 1.3 (plate 1)
Edited image of a Rubik's cube. The patch labeled A on the side of the cube and the red patch labeled B on the top of the cube share the same color on the printed page, as can be seen when they're shown isolated above the photograph. This illusion was created simply by copying the color of B to A using a graphics application. Based on an illusion by Ted Adelson. From Frisby & Stone 2010.

Illusory Lines: Triangles and Pandas

For the brain to see what's in the world rather than simply what's in a retinal image, it sometimes has to "fill in the gaps" in the image. Although doing this works in general, it can also result in optical illusions, such as in the Kanizsa triangles shown in figure 1.1a. At first glance, it is not obvious that this is an illusion at all: the picture simply seems to portray two overlapping triangles. On closer examination, however, it's apparent that one triangle has edges only within the cut-out regions of the black disks. The illusion is also supported by the gaps in the black lines, which suggest that something (i.e., another triangle) is covering up parts of the black lines. Together, these fool the brain into guessing that there are two overlapping triangles, even though there clearly are not. Surprisingly, the Kanizsa triangles do not disappear on ce the basis of the illusion is known. It is as if the visual system uses a self-contained logic of its own, which is immune to the knowledge that the triangles are illusory. Artists are well aware of the brain's ability to "fill in the gaps," as shown in figure 1.2 by the disappearing lines in the World Wildlife Fund logo between 1961 and 2006.

Recognizing Objects: Cubes, Rings, and Pianos

For the brain to see the three-dimensional structure of what is in the world, it has to ignore all but one of the possible 3-D structures that could have given rise to the retinal image. This is demonstrated in figure 1.4, where two different 3-D interpretations can be perceived. If you look at the star in the center of the figure, at one moment the rings appear to be viewed from above, but at the next moment they seem to be viewed from below. A more extreme example is shown in figure 1.5, where the shape in the retinal image (shown at the bottom) could have been formed by any one of the three 3-D objects shown above it. This, in turn, implies that any one of the 3-D objects shown would have generated *exactly the same* retinal image. How does the brain know which one?

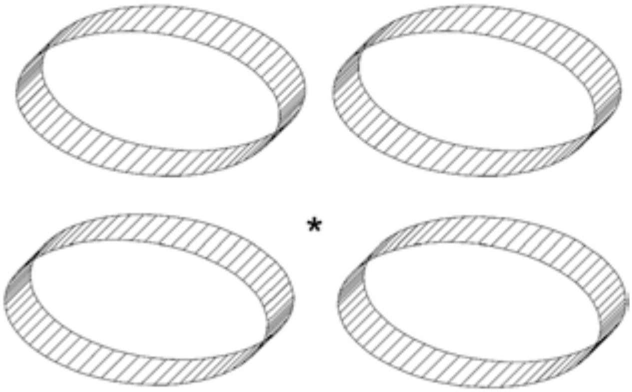


Figure 1.4

Look at the star in the center of the figure. At one moment, the rings appear to be viewed from above, but at the next moment they seem to viewed from below. Created by author.

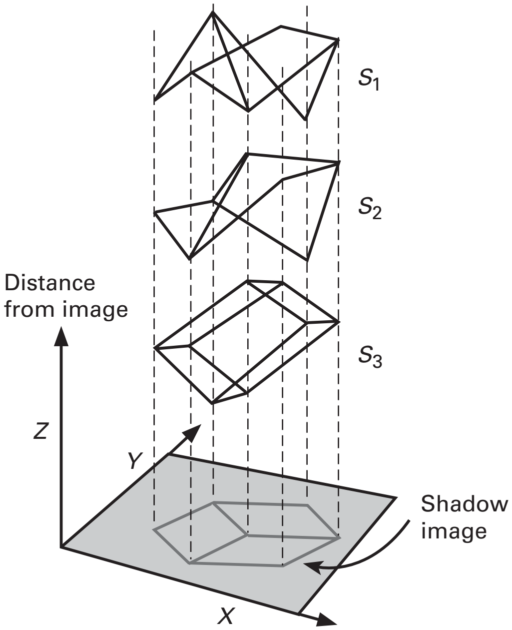


Figure 1.5
The fact that every retinal image could have been produced by many possible scenes is a major problem for the brain. Each of three different wire-frame objects yields exactly the same image (bottom). Adapted with permission from Kersten & Yuille 2003.

Fortunately, the brain can rely on its previous experience to exclude 3-D shapes only rarely encountered in the past. Because cubes are more common than all the other 3-D shapes that could have generated this retinal image, the brain is able to choose the cube as the single most probable 3-D interpretation of the inherently ambiguous shape on the retina.

This general rule of thumb, or *heuristic*, of excluding rare, and therefore improbable, 3-D interpretations is smarter than it might first appear. This is because the 3-D shapes shown in figure 1.5 are three examples from an infinite number of 3-D shapes, each of which would have generated the same the retinal image. Thus, by choosing a cube as the most probable 3-D shape, the brain effectively excludes an infinite number of other possible 3-D interpretations.

The ambiguity of retinal images has been used to good effect by artists such as Shigeo Fukuda, whose jumbled piano, shown in figure 1.6, is a compelling example. Here the piano seen in the background is actually the

Figure 1.6
Piano or pile of wood? It all depends on your point of view. By Shigeo Fukuda.

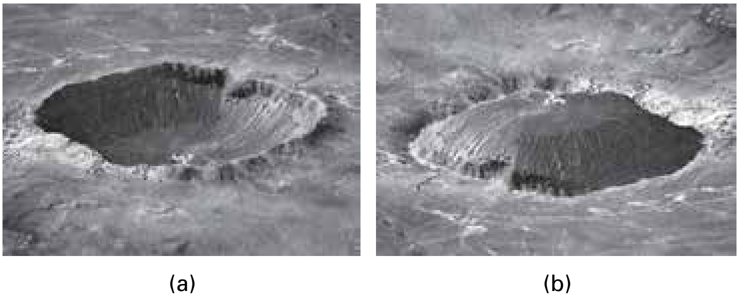
reflection in a mirror of the jumble of wood shown in the foreground. In other words, if you were to stand where the mirror is then the jumble of wood would appear to be a piano. As with the example of a wire-frame cube, the image of a jumble of wood has an infinite number of possible 3-D interpretations (e.g., a 3-D jumble of wood). Moreover, all but one of these interpretations have probably never been seen before and are therefore rejected by the brain. In contrast, the single 3-D structure that is consistent with the retinal image is, of course, a piano.

Although these examples include 3-D objects deliberately designed to generate the required images, the general point remains: experience is required to disambiguate images. More importantly, using previous experience to interpret ambiguous retinal images can be made mathematically precise by making use of the *Bayesian framework*, a topic to which we will return in chapter 6.

Perceiving Three-Dimensional Shape: Shading, Craters, and Faces

The contours apparent in figure 1.7 (plate 2) are almost entirely based on the subtle changes in shading across the image, which is an important visual cue regarding the three-dimensional structure of a surface. Crucially, shading can be used to find a 3-D structure only if the viewer knows where the light source is in relation to the viewed scene. This becomes obvious if we trick the visual system into assuming the light is coming from the wrong direction. Thus both pictures in figure 1.8 depict the Barringer meteor crater in northern Arizona, except that one is upside down. But,

Figure 1.7 (plate 2)
Silk sheet. The three-dimensional shape perceived here is almost entirely due to subtle changes in shading.



(a) (b)

Figure 1.8
Are these pictures of a hill or a crater? It all depends on your point of view. Barringer Meteor Crater, reproduced with permission from the United States Geological Survey.

in both cases, the brain assumes that the light comes from *above*. This is true for figure 1.8a, which is therefore perceived as a crater. However, for figure 1.8b, this assumption forces the brain to perceive the crater as a hill (because a hill is all it could be if the light came from above). As with the example of wire-frame objects in figure 1.5, the brain is forced to interpret an image that is fundamentally ambiguous by adding its own bias. For figure 1.8b, this bias takes the form of an assumption about where the light source is likely to be, and a reasonable guess seems to be "from above."

Whether this "light-from-above" assumption is innate or based on previous experience is not known for sure. One of the first scientists to study this effect, David Brewster (1826), noted that children did not always interpret such pictures as if light came from above. More recently, it has been found that, as children grow older, they have an increasing tendency to perceive pictures like those in figure 1.8 as adults do (i.e., figure 1.8a as a crater and figure 1.8b as a hill; Thomas, Nardini, & Mareschal 2010; Stone & Pascalis 2010).

However, it is not hard to show that the assumption that light comes from above is more of a broad hint than an assumption that is rigidly applied to all images. When confronted by the image of a face defined almost entirely by changes in shading, as in figure 1.9, the brain appears willing to discard the light-from-above assumption, in favor of a more compelling assumption (more compelling in the context of faces, that is).

Figure 1.9 shows four frames from the movie of a rotating hollow mask, where the light comes from above. For the first three frames (a–c), the assumption that light comes from above leads to the perception of a convex (sticking-out) face. However, the fourth frame (d) shows the *concave* (hollow) inside of the mask, but it still appears to be *convex*.

When compared to the number of times normal (convex) faces are encountered, hollow faces are exceedingly rare. Consequently, the brain excludes the possibility that this could be a hollow face. But this choice

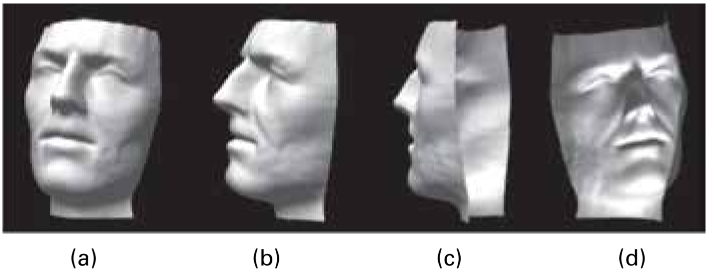


Figure 1.9
Face mask seen from four different positions with light coming from above. (a–c) Views from the front and sides create perceptions of a normal convex face. (d) View from the rear of the mask creates perception not of a concave face, as it "should" do, but again of a convex face. To achieve this incorrect perception, the brain is forced to assume that the light is coming from below. Courtesy Hiroshi Yoshida.

poses a dilemma. For the face in frame d to be seen as convex, the brain must assume that the light is coming from *below*. In other words, the image in figure 1.9d can be perceived either as a convex face with light from below or as a concave face with light from above. Self-evidently, the brain chooses the first interpretation because that's what we *see*. Indeed, the bias for seeing faces as convex is so strong that it is almost impossible for most adults to see frame d as a hollow face with light coming from above.

In short, it seems that the assumption that faces are convex effectively vetoes the assumption that light comes from above in the presence of images of faces. This, in turn, suggests that there may be a hierarchy of assumptions, with more important assumptions over-riding lesser assumptions in the interpretation of particular types of images. As we will see in chapter 6, these types of assumptions can be formalized within the Bayesian framework.

Shades of Gray and Grays of Shade

A piece of coal in bright sunshine reflects many times more light than a piece of white paper does in the relative darkness of a moonlit night. So why does the paper still appear as white and the coal as black? A clue to the answer can be gleaned from pictures such as the one in figure 1.10. Even though the upper surface is perceived as if it is darker than the lower surface, both surfaces have the same *albedo* (shade of gray) on the printed page.

The reason for this illusion is that the three-dimensional structure of the scene implies that the lower surface is in shadow, and that the upper surface is well lit. If this were the case for a 3-D object then the only way for the upper and lower surfaces to have the same gray level on the page (and therefore in the retinal image) would be if the upper surface were darker than the lower surface *on the object*. So this illusion has a rational basis in terms of the 3-D scene depicted. Again, the brain manages to see beyond the image on the retina to the underlying lightness of surfaces in the world. Moreover, it is precisely this ability to see past the retinal image that gives rise to the illusion depicted here.

Color and Shade

Look again at figure 1.7 (plate 2). The silk material does not change color, even though the changing three-dimensional contours of the silk sheet ensure that the image is darker or lighter in different places. The brain has managed to correctly interpret these image changes as changes in surface orientation rather than changes in the color of the material itself. Here the

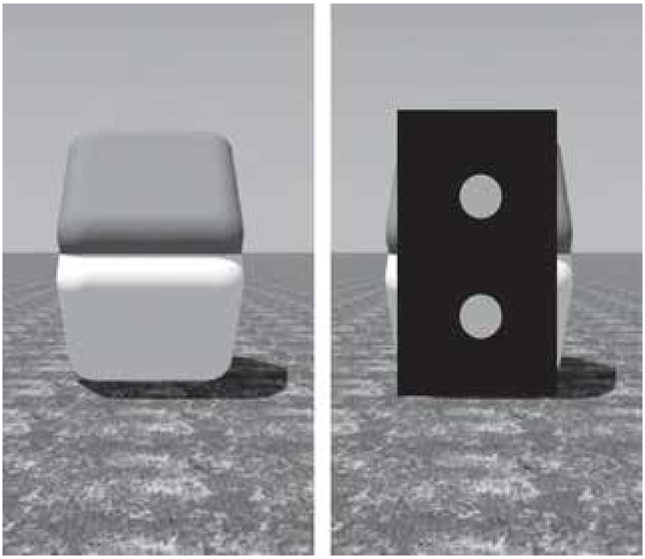


Figure 1.10
Surfaces in shadow are perceived as being lighter. (a) Both the upper and lower surfaces have the same albedo (shade of gray) on the page, which can be seen in (b), where a black card with holes has been placed in front of the object. The illusion of unequal lightness on the page seen in (a) is due to the fact that the upper surface is in the light whereas the lower surface is in shadow on the three-dimensional object. Reproduced with permission.

(less than obvious) illusion is that we see a constant color across the silk sheet, despite the changing shade of red across the image.

The converse of this is shown in figure 1.3 (plate 1). Here two patches of color on different sides of a Rubik's cube yield identical patches of color on the printed page, which are nevertheless perceived as different colors on the three-dimensional object. As we have already seen with figure 1.10, this can be explained by the shadow that falls over the patch on the side of the cube. Thus, the brain seems to reason, if both patches have the same color in the image and one of them is in shadow on the object then the shadowed patch must be lighter on the object. Notice that, to perceive this illusion, the brain must first work out the 3-D structure of the scene portrayed; otherwise, the two patches are perceived to be the same color (can be seen by viewing the patches in isolation, as in the top part of figure 1.3b).

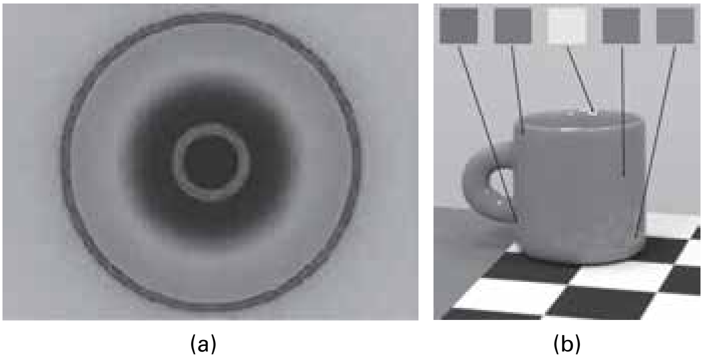

(a) (b)

Figure 1.11 (plate 3)
Color illusions. A single color at two different locations on the page gives rise to the perception of two different colors in (a), whereas many different colors on the page give rise to the perception of a single color in the world in (b). The square patches at the top of the figure show the colors on the page from the locations indicated by the arrows. Reproduced with permission from (a) MacLeod (2003) and (b) Brainard & Maloney (2011).

Thus, in figure 1.7 (plate 2), the brain effectively ignores *changes* across the image in order to perceive the same color across the three-dimensional surface of the silk sheet. Conversely, in figure 1.3 (plate 1), the brain effectively ignores the *similarity* of colors of two patches in the image in order to perceive two different colors on the object's surface. A similar example of this can be seen in figure 1.11b (plate 3). In both cases, the brain takes account of the three-dimensional structure of the scene in order to interpret the colors depicted in the image.

Brains, Vision, and Bird Flight

Because human vision depends on neurons, it is tempting to think that all of the examples presented here can be explained in terms of the particular neuronal machinery of the brain. However, one of the revolutionaries of vision science, David Marr (1982, p. 27) famously used the analogy of flight to argue against studying *only* neurons in order to understand vision: "Trying to understand vision by studying only neurons is like trying to understand bird flight by studying only feathers: it just cannot be done."

Many of the early pioneers of human flight mistakenly believed that human flight should imitate bird flight in almost every detail. Accordingly, they built machines covered in feathers or machines with wings that flapped. In so doing, they mistook the particular trappings of how birds fly for the general principles of flight. For Marr, such principles emerge from a *computational analysis*, an analysis that considers not only how birds fly, but how flight can be achieved by any animal or machine, with or without feathers, with or without flapping. In other words, a computational analysis considers the *bare necessities* of attaining flight and the *principles* that *underpin* flight, without being constrained to copy the materials (feathers) and methods (flapping) used in any of the particular biological examples that exist in the natural world (e.g., birds).

In retrospect, it seems obvious that the Wright brothers' first powered flight in 1903 succeeded precisely because they took great pains to ensure that they understood the general principles of controlled flight (figure 1.12). More importantly, they understood that flight does not necessarily require feathers or flapping.

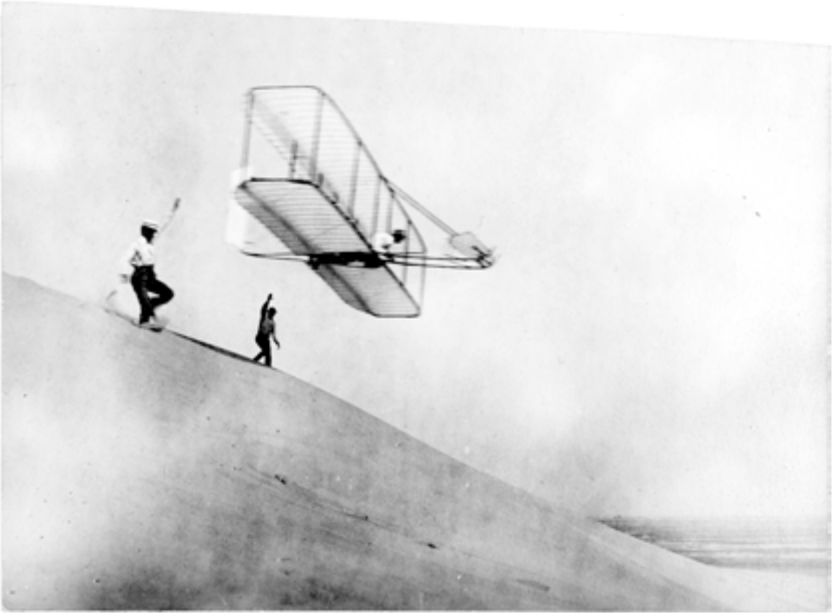


Figure 1.12
"Trying to understand vision by studying only neurons is like trying to understand bird flight by studying only feathers: it just cannot be done" (Marr 1982, p. 27). The Wright brothers were the first to successfully fly in 1903 because they understood the computational nature of the problem of flight, whereas many inventors insisted on mimicking the superficial aspects of flight (e.g., by building machines with flapping, feather-covered wings). This is a picture of the Wright brothers' 1902 glider.

Similarly, Marr argued that a proper understanding of the problem of vision can only be achieved by a full understanding of the computational principles on which vision is based. A computational analysis of vision considers the bare necessities of seeing and the principles of seeing, without being constrained to use the same materials (neurons) and methods (chemical messengers) used by the particular biological examples that exist in the natural world (brains). The abstract nature of Marr's computational framework is entirely compatible with recent developments in computational neuroscience, which suggest that the efficient coding hypothesis and information theory have much to offer the science of seeing; these ideas will be explored in subsequent chapters.

Marr's choice of the term *computational framework* is unfortunate, suggesting as it does that vision must function like a computer. Perhaps a more accessible term would be *informational framework* because Marr was keen to emphasize the nature of the information being processed without necessarily referring to the particular nature of the machinery (e.g., neurons or silicon chips) that happened to process that information. So, for Marr, it would matter little whether the device that is seeing was built from neurons, silicon, or wood, provided it sees and, more important, sees in the same way we do. But, you might ask, how could a device built from silicon see as we do?

We can explore this question by taking an extreme example from the realms of the philosophy of artificial intelligence. Quite simply, if one of the neurons in your brain were to be replaced with an electronic device that was, in every other respect, identical to that neuron, do you think you or the rest of your brain would notice? No. Now, suppose this was repeated, one by one, for all of the neurons in your brain. You would be, to all intents and purposes, the same person. Indeed, you would argue that you *are* the same person now as you were last year, even though many individual proteins in your brain have been replaced with (probably) identical ones. The point is that, as far as the *function* of each individual neuron is concerned, it does not matter what it is made of. Similarly, when it comes to considering how the brain performs the many tasks implicit in vision, it does not matter that it happens to be made of neurons. It does matter if we want to find out how a neuron helps solve the problems of vision, but it does not matter if we want to find out how any system, whether it is made of neurons or silicon, solves the problems of vision.

The above quotation from Marr was his protest against a purely reductionist approach, which would involve searching for the essence of vision only in the neuronal machinery of the brain. Thus, even though Marr did take account of findings from neurophysiology, he was also aware of the pitfalls of relying too heavily on such findings in order to understand vision. In essence, he proposed that, in formally describing how the brain solves the many problems implicit in vision, we should take account of neurophysiological, anatomical, and psychophysical data.

Because the following detailed discussion is fairly abstract, you may wish to return to it after first reading the remainder of the book. Marr's computational framework embodies three distinct, though not entirely independent, *levels of analysis*:

1 *Computational theory* What is the nature of the problem to be solved, what is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?
2 *Representation and algorithm* How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm (method) for transforming input to output?
3 *Hardware implementation* How can the representation and algorithm be realized physically?

At the level of *computational theory*, the nature of the visual task to be solved should be formulated as precisely as possible. For example, this might entail finding the shape of three-dimensional objects from shading information. An *algorithm* is another word for a precise method. At this level, the amount of shading at each point in an image might be represented as a single number, and the local orientation of each corresponding point on the surface might be represented as a small plane (which requires two numbers to represent its orientation). The algorithm used to transform image gray levels into surface orientation could make use of the constraint that the local orientation of most objects varies smoothly from place to place on the object (although we would need more details to make such an algorithm work in practice). Finally, the *hardware implementation* could be executed using neurons, or a computer, for example.

Although each analytical level places constraints on the level below it, it does not *determine* its exact form. For example, the problem of obtaining shape from shading (see figure 1.7, plate 2) at the computational level specifies the nature of the problem for the algorithmic level, but this could be solved by a number of different algorithms at that level. Provided an algorithm solves the problem specified at the computational level, it doesn't matter which algorithm is used. In turn, each algorithm could be implemented at the hardware level by using either neurons or computers.

Again, provided a hardware implementation executes the specified algorithm, it doesn't matter which implementation is used. There is thus a cascade of influence, from the computational to the hardware level.

In reality, these levels are rarely independent of one another, and influences travel up and down between the levels. Moreover, it's difficult to find even one algorithm that solves a given computational problem and even one hardware implementation that executes the specified algorithm. In its defense, however, the computational framework forces the essential elements of any theory of vision to be specified with transparency and precision. This means that the strengths and flaws of a theory are easy to spot, which ensures that improvements to the theory are relatively easy to make. The computational framework is essentially both a recipe for how to study human vision, but it is also a recipe for how to study the human brain.

Unfortunately, Marr died (aged 35) before he could bring his work to fruition. But he left a rich legacy of ideas and a precise framework for how these ideas could be developed, ideas which have inspired subsequent generations of vision scientists.

Conclusion

The examples presented in this chapter are not only intriguing, they are also informative. They demonstrate that, to see the world as it is, the brain has to *interpret* the retinal image, which is full of lies and omissions. The brain's task is to see beyond these, to ignore the lies and fill in the omissions. Fortunately, the lies and omissions don't change from day to day but are law-like and systematic, and this systematicity allows the brain to correctly interpret all but the most ambiguous retinal images. Using Marr's framework to consider the problem of vision from a computational perspective, especially in the context of the efficient coding hypothesis and information theory, can inform us not only *how* the brain sees the world, but *why* the brain sees the world as it does.