

TheNational

The eye of the beholder

Last Updated: January 09, 2009 8:30AM UAE / January 9, 2009 4:30AM GMT

Jeremy Freeman looks at a new book about vision, a charming and erudite account that loses sight of the most important player: the brain.

A Natural History of Seeing: The Art and Science of Vision

Simon Ings
WW Norton and Co
Dh85

In the 1950s, computer scientists made the bold proclamation that building machines capable of vision would take about a decade. There was an innocent hope: take a computer, put two cameras on it, have it do some maths with the camera input, and voila — the robot should be able to see. But it's been half a century, and machines still perform miserably at visual tasks that humans accomplish effortlessly. The annual

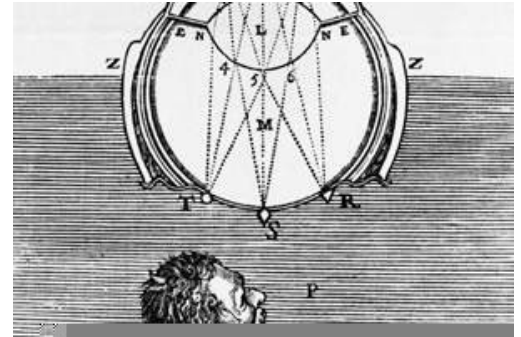
Robocup tournament – a football competition for state-of-the-art robots – is a great example. Watching the robots play, it's hard not to see a pack of confused five-year olds, chasing the ball one minute and getting distracted by the sidelines the next. Vision is a large part of the problem. The robots need to see the ball and estimate its motion in real time. This is so hard that the competitors have agreed to use a bright orange ball; the neon colour contrasts with the grass, which gives the robots a more simple visual cue. The judges would like to use an official black and white ball, but that's far too difficult for the robots to recognise. Maybe in another 10 years they can make the switch.

Why is it so hard to make machines that see? It's easy to put cameras on a robot. But it's incredibly difficult to decide what to do with the input. Our eyes are indeed a little like cameras in our head. And light bouncing off objects in the world is indeed focused through our eyes onto the retina, much as light is focused through a camera lens onto film. But that's where the analogy ends. Cameras merely reproduce images, while a functional visual system needs to infer properties of the image: How many objects are present? What are they? What is their texture? Are they moving, and if so, how fast?

The German physiologist Herman von Helmholtz best appreciated that perception is really a set of inferences. He realised that, rather than merely receiving sensory input like a camera, we continually estimate the sources of our sensory input. Our brain forms a "best guess" as to what's out there in the world by combining our current sensory input with our prior expectations. The vision theorist Horace Barlow came up with a perfect illustration: Try to recall the last time you tripped while trying to dodge a shadow. Having trouble? A shadow across a pavement can produce the same retinal impression of an edge as does a doorstep. So why don't we avoid stepping into shadows? The visual brain must incorporate knowledge of the world. Put elegantly by Barlow: "What can these signals for lines, edges, textures, movements, disparities and colours mean without any background knowledge of the world? Are they not like single letters without a language?"

In his recent book *A Natural History of Seeing*, the science journalist Simon Ings prefers to focus the story of vision on the eyes. He is direct about his choice: "this is a book about the nature of the eye." Lest the reader fail to appreciate the grandeur of his task, Ings explains, "This has not been done before... no one has put the eye at the centre of a sprawling and epic story." In what follows, Ings does tell a sprawling story, and does put the eye at the centre of it. What Ings fails to convey is that the eye – remarkable device that it is – plays only a supporting role when it comes to seeing.

Over several chapters, Ings discusses the evolution, physiology and chemistry of the visual system's front end – the eye, the photoreceptors and the retina – in tremendous depth, combining a delightful narrative pace with a textbook's worth of detail. Unlike many practicing scientists, who dumb down their work when writing for a popular audience, Ings indulges the nitty gritty of biology. He is also especially good at characterising pioneering researchers without resorting to the stock, egghead-on-a-mission clichés of popular science writing ("Dr. So-and-So has a receding hairline, glasses, and a penchant for thinking faster than he speaks"). From Ings we learn, for example, that the physicist Ernst Mach "caught the experimental bug" after he tried "stuffing



Sight unseen: Detail from a drawing of the eye by Descartes from his *Dioptrics*, published in 1637. Courtesy Science Photo Library

putty under his eyelids to stop his eyes” from moving. Through an ahistorical mix of such facts and personalities, Ings’ narrative manages to successfully convey “what vision feels like as a subject of study and wonder:” multifaceted, all consuming, and most of all, perplexing.

However, the most perplexing and rewarding avenues of vision research concern not the eye, but the brain – the seat of perceptual inference. And Ings makes only brief forays into the more brain-based, inferential aspects of visual perception. Compared to his thorough descriptions of the eye, these diversions feel more like pseudo-philosophy – weighty rhetorical pondering without experimental fact or rigorous theory. Ings’s treatment of depth-perception is a perfect example. How do we infer depth in three dimensions when the images projected onto our retina are only two-dimensional? We get one of our cues from something called stereopsis. It works like this: our two eyes are separated by a few centimetres, so our two retinas receive slightly shifted copies of the visual scene; “when we fuse our left-eye and right-eye views of the world,” Ings writes, “the tiny discrepancies between them give us a sense of depth.” But how do we – and more specifically, how do our brains – actually accomplish this complex inference?

Ings gives a neuro-mumbo-jumbo response: “When both views are laid over each other” in the brain, “a special class of neurons spots the inconsistencies: a spark in the left eye doesn’t quite overlap with a spark in the right eye.” While this is not exactly wrong, it manages to make the process seem both more magical and more well understood than it is. It gives the reader the – admittedly appealing – image of a little box of neurons looking at the input from the two eyes and spitting out a depth, maybe even in centimetres. If only it were that easy! Certain neurons are indeed sensitive to discrepancies between the inputs of the two eyes, but the devil is in the details. For instance, when you look at a single object, the discrepancy between the input from the left eye and the input from the right eye constitutes what scientists call the absolute disparity of that object. When you look at two different objects – say, a bear and a tree, one standing in front of the other – each object presents its own absolute disparity, and the difference between those disparities is their relative disparity. It turns out that some neurons are only sensitive to absolute disparity, but others, recently discovered, are also sensitive to relative disparity. Fine depth judgements depend on calculations from both. On top of all that, our expectations also influence depth estimates: the known structure of a familiar object can override these disparity-based signals, especially when the signals are inconsistent with what we expect. Clearly, depth estimation is not as simple as a bunch of neurons “spotting” the difference between the two eyes. Ings makes it sound like magic; the real answer is more complex – and more interesting.

Vision theory has made impressive (if not entirely conclusive) strides in the modern era, beginning with the pioneering work of Helmholtz and Barlow. But Ings’s narrative founders on ancient, first-order theories of seeing. He highlights a problem that perplexed the early Greek philosophers of vision: “Visual impressions do not arrive muddled in a sensory soup... We see objects, not splodges, or waves, or impressions.” Why is that? It’s a great question, and substantial contemporary research and theory is devoted to it. But for some reason Ings starts with the early Greeks and then never makes it past Descartes. Early Greek theorists debated whether things out in the world emit “objectness” particles onto our eyes, or whether our eyes project a “visual ray” onto objects in order to see them. Both theories are, of course, hopelessly wrong, and Ings corrects them, describing successive attempts to explain how light reflects off of objects and is focused by the eye. He concludes the account with one of Descartes’ drawings of the visual system. The drawing details the geometry of how light is projected onto the retina, and at the bottom of the drawing is an observer, “studying the back of the eye,” presumably making sense of the visual input – a personification of the mind. For Ings, this *deus ex machina* reminds us of the “dreadful conundrum” faced by all vision theorists: We know the visual system is not just a camera. We know it must infer or represent the world instead of merely reproducing it. But how does it work? Ings leaves the theoretical discussion at that.

Admittedly, the “dreadful conundrum” is still with us. But considerable effort has been devoted to resolving it. For example, scientists have found that several populations of visual neurons seem to respond selectively to particular objects and shapes, such as faces, and research groups in both neuroscience and computer science have developed models of how the visual system might assemble representations of objects from their constituent parts. In these models, neurons near the front end of visual processing detect simple visual features, like lines and edges, and successively advanced levels of processing integrate or combine these features into textures, contours, shapes and, ultimately, objects. These theories do not resolve the Greek philosophers’ confusion about where “objectness” comes from, but they point in the right direction.

The key lingering conundrum facing these theories is that scientists know it is impossible for the brain to represent every possible visual stimulus we might encounter – every object, every texture, every face. As Barlow realised, the visual system needs compact, efficient representations. Luckily, the visual world is highly structured and full of redundancies. Modern computers exploit such regularities when they compress digital photographs into tiny file sizes. According to Barlow’s theory of efficient coding, the brain uses similar tricks to represent visual information efficiently. The brain never represents the same thing twice, and if something in the world doesn’t change, the brain doesn’t bother re-representing it. The challenge is identifying which regularities the brain exploits.

Ings stumbles on the theory of efficient coding during his discussion of retinal physiology. He explains that neural coding in the retina is primarily sensitive to image contrast; whenever we perceive an image, many of the “large expanses of light and dark... are effectively evened out to a medium grey while areas of high contrast are massively exaggerated.” This approach is consistent with efficient coding. “By reporting only the lines of contrast,” he writes, “the retina avoids having to prepare endless, uninteresting, and massively redundant reports about plain surfaces.” There is an exciting theoretical idea lurking here, one that extends well beyond the retina – but Ings doesn’t take us there.

Ings does a phenomenal job explaining how the eye works, but his book is purportedly about vision. I worry that readers will walk away thinking that the eye is the primary subject of vision science, and that the brain is still the magic interpreter – its function left to the musings of philosophers and its mechanisms too complicated for rigorous theory and experiment. Our intuitions about the mind and brain already tend towards dualism; like Descartes, we imagine a little man sitting behind our eye, looking at the input and making sense of what comes in. No one actually believes this, of course. But it’s a tempting illusion when faced with the truth: that billions of neurons in the brain infer properties of the visual world, and that the functioning of these neurons is wholly responsible for subjective perceptual experience. Pioneers like Helmholtz and Barlow realised this truth, and argued that we should be able to use a set of theoretical principles – such as expectation-based inference and efficient coding – to understand how the eyes and the brain perceive. However, we will only succeed in that goal if we stop thinking of the brain as a magic box sitting behind the eyes.

Jeremy Freeman is a doctoral student in neural science at New York University. His work has recently appeared in the Journal of Vision.

Have your say

Please [log in](#) to post a comment