First Expedition: Vision

We began the book with a question: how can we know number? The more general question is how can we know anything? The philosophy of knowledge is epistemology, and though we’re here for science not philosophy, a small break will help set out the issues.

Written work on epistemology traces at least as far back as Plato and Aristotle, about 300 BCE. Plato compared us to people living in a cave (Figure 114), chained so that we can only see shadows of the world, projected in front of us. Plato is examining the gulf: on the one side, senses, which tell us about the world, and on the other, understandings of the world. For Plato, all that the senses can tell us about reality is plausible myths: mere stories. To get to the core of reality, an individual needs to use reason. Mathematics is an essential part of that reason:

...the Platonic classification of existence [has] two orders. The higher is the realm of unchanging and eternal being ...[containing] the objects of rational understanding ...namely, arguments of mathematics and dialectic which yield a securely grounded apprehension of truth and reality. The lower realm contains ‘that which is always becoming’ – passing into existence, changing, and perishing, but never having real being. This is the world of things perceived by our senses. ...sense can only state a fact ...the reason why can only be apprehended by the higher faculty of understanding.


We’ll be after something different: not the chasm between senses and reason, but that between reality and the senses. Even taking Plato’s ‘plausible myths’ into account, Kant (see p35) suggested we can make progress. It is as though, while we see only shadows, we can still work out the patterns of those shadows. Kant made a distinction between the bounded and the limitless, and he noted in particular that science has the character of the limitless: while we cannot know everything, we can make continual progress. We’ll see how that applies, starting with visual senses. with vision: what can vision tell us about the world? Figure 115 shows how a flower looks to us, in sunlight; then to a bee, in ultraviolet light. It’s conjectured that flowers evolved markings to direct bees to nectar; the payoff for the plant is pollination.
Yes, our vision is limited – yet, we use UV lights to do ‘bee.’ Again: Athanasius Kircher used a microscope to examine the blood of plague victims; he noted ‘little animals’ which he believed caused the disease; see Figure 116. Having a microscope to extend his vision allowed Kircher to guess at the germ theory of disease.

Before Kircher, though, there were parts of reality we couldn’t experience. Will there be whole chunks of science and medicine we can’t do, because we don’t have the right ‘microscopes’? In cosmology, some theories posit universes parallel but unconnected to ours – is there a way to detect them? (see p138)

We’ll begin with experiments on vision – from the 1950’s. The article *What the Frog’s Eyes Tells the Frog’s Brain* (p138) discusses experiments presenting different kinds of visual stimuli to frogs, then recording which stimuli cause a particular cell in the eye to respond. The article posits these stimuli determine what a frog sees (compare p133). On this hypothesis, what the frog can see is: differences in contrast (possibly representing an insect standing out from the background), convexity (possibly representing the shape of prey), a moving edge and its direction (possibly a moving insect) and dimming of light (possibly indicating a predator in back). Figure 117 gives an artists interpretation of what a pond might look like to a frog: it’s nothing like what we would see. Is our vision also hiding much of the world?

Johannes Kepler in the 1600’s based his theories of planetary motion on the observations of Tycho Brahe. He knew the atmosphere distorted light, and his computations took that into account. He wondered whether the eye also distorted observations. Kepler believed that the lens of the eye focuses light on the lining at the back of the eye, opposite the lens: retina, (figure 118). To Kepler, this may have been by analogy to the artists tool for drawing in perspective, the camera obscura (Figure 119). We now know the retina contains cells responsible for converting light to electrical charges, which, interpreted by the brain (and the retina itself!) constitute vision. How does this work?
Using mammalian retinas, Nobel laureate Santiago Ramon y Cajal (Figure 120) used a microscope and developed innovative cell staining techniques to elucidate the cellular structure of the retina, diagrammed in Figure 121. He wrote:

[... ] the retina is a genuine neural center, a sort of peripheral cerebral segment whose thinness, transparency and other qualities render it particularly favorable to histological analysis. In fact, though its cells and fibers are essentially similar to those of other centers, they are arranged in a more regular fashion, different types of cells being distributed in distinctly different layers.

Cajal, La retine des vertebres, 1892.

Cajal is saying the retina is a protrusion of the brain. He identified ordered layers of cells (Figure 121), suggesting that the eye doesn’t just gather light: the layers he discovered process the light, sending the results on to the brain. Find the limitations in the processing and we may find the limitations of vision.

To help think about processing in the retina, we’ll compare it to something simpler: a modern high-megapixel digital camera. Perhaps camera vision will help explain human vision. These kinds of analogies are behind many kinds of attempts to understand the brain: even today, people compare our brains to computers. And indeed, a modern camera is very much a computer. However: the eye-brain system is not like a camera; vision is much more like an odd kind of movie.

In reality (and this is very obvious) human vision is video, not photography. Even when staring at a photograph, the brain is taking multiple “snapshots” as it moves the center of focus over the picture, stacking and assembling them into the final image we perceive. Look at a photograph for a few minutes and you’ll realize that subconsciously your eye has drifted over the picture, getting an overview of the image, focusing in on details here and there and, after a few seconds, realizing some things about it that weren’t obvious at first glance.


What vision does, then, is a complex interaction between eye and brain; scenes aren’t recognized all at once; the brain will move the eye to detect other parts of the visual field, and then will put all those together, somehow, to construct perception.

However, the individual ’snapshots’ could be thought of as analogous to a camera picture, and we’ll continue with the comparison.
Both the camera and the eye have a lens to focus light on a region that can recognize the incoming light and translate the color and intensity of the light into an electrical response; we’ll call the response a signal. This is analogy, and by itself it can distort how we think about vision; see p139.

In modern cameras, the region that takes light and changes it into an electrical signal is called a CMOS sensor (CMOS refers to both the design and the materials of a chip; these kind of chips are resistant to noise and consume little power). The chip has a rectangular array of photodiodes shown in Figure 122. When light hits the photodiode, it generates an electron. Photodiodes store electrons; the number of electrons stored is proportional to brightness of the incoming light.

Photodiodes only detect brightness, so colored lenses overlay each photodiode, and the color has be be generated by the camera processor. This makes a mosaic; Figure 123 shows a mosaic from a simple sensor, with only $180 \times 80$ photodiodes. Figure 124 shows what we’d see from a slightly better sensor. It looks blurred because the mosaic limits resolution, that is, how much detail we can see.

The retina, on the other hand, uses cells to recognize light: rods and cones. The rods do black and white vision; cones detect red, green and blue colors. Light energy causes a molecule to decompose; in the rod, the molecule is rhodopsin. Even one photon of light causes rhodopsin to decompose within picoseconds, triggering a chain of reactions that result in a change of the charge across the cell membrane.

Back to the camera. After collecting the electrons that will be used for the picture, the electrons are led out of the sensor, and converted to a voltage. If you think of millions of electrons, you can have millions of possible voltages – far too much information for the camera. At this point, voltages are assigned to one of a small numbers of different levels. The number of levels is determined by the number of bits used in the camera circuitry; a twelve-bit sensor can handle $2^{12} = 4096$ levels. The assignment of numbers to a limited collection of levels is called quantization; see p143. We met up with quantization in Sections 8 and 9, when we discussed how computers represent numbers, and how scientific data is recorded.

If we think of a photodiode as a small box for collecting electrons, we can imagine a very bright light could overfill the box. In this case, the electrons stored in the photodiode overflow into nearby diodes. Figure ?? shows overflow; the effect is called camera bloom.

In contrast, retinal cells respond logarithmically to light: even one photon can cause a response, but brighter lights cause smaller re-
sponses. This allows us to see a wider range of dark/light than a camera; correspondingly, the output from cameras has to be logarithmically adjusted to match our eyes; the adjustment is called gamma correction.

But the retina too can also overload – at the molecular level. Rhodopsin decomposes quickly in response to light, but takes longer to rebuild. This causes what’s called the ‘theater effect’: on a very bright day, light saturates the rhodopsin; if you leave sunlight to enter a dark room (like a movie theater), the delay in rebuilding rhodopsin means there’s not enough available, so there’s a short time when you can’t see very well.

Size is another limitation: neither we nor the camera can see molecules or even viruses; Figure 125 suggested why: the size and number of the photodiodes limit the amount of detail. Our 20 megapixel camera has $5384 \times 3752$ photodiodes, each is $1.12 \mu m$ by $1.12 \mu m$ ($\mu m$ is a micrometer, $10^{-6}$ meters). Each photodiode gives rise to a $1.12 \mu m$ pixel or picture element, and these are equally distributed across the picture frame. In the high-rez picture (Figure 124), light changes very quickly as we go from one tiny pixel to the next; we call these quick changes high-frequency. In contrast, the low-rez sensor (Figure 125) has large pixels, or big blocks where the light is constant. This is called low-frequency information.

The smallest object we, or our camera, can see is also an issue of how we process high-frequency information. Each pixel in Figure 125 comes from one photodiode; we can think of it as a little box to hold light. The camera pushes incoming light into these boxes; information processing theory calls this sampling. Sampling is at the foundation of all scientific/medical data gathering, as we saw in Section 9. For now, the question is what happens when the camera or eye tries to push high frequency information into low frequency boxes.

We’ll give a simple analogy; for the precise details, see Chapter 1 Section 9.3. When you try to push light into boxes and run out of boxes, the light has to be placed in boxes that have already been used. You can see an example of this in Figures 126 and 127: the left side of Figure 127 shows the bricks in a castle wall, as they should look. The right side shows the effect of putting the information into boxes already used. The effect is called aliasing; it distorts the original picture, and can add the appearance of patterns that were never in the original. Cameras avoid aliasing by adding a layer of material over the sensor, blurring the picture slightly. This leaves low frequency information alone, but limits high frequency information: tech people use the term ‘low-pass’ filter (a filter, e.g. a coffee filter, allows small things
to get through but blocks big things. It’s a ‘small thing pass’ filter).
Since the high frequency information contains the small changes, the
filter reduces the resolution of pictures.

Unlike a CMOS sensor, rods and cones not equally distributed. The
retina has about 120 million rods and some 7 million cones. It seems
to avoid aliasing. Research suggests this is because the cells recognizing
light are not arranged in a regular pattern like a CMOS sensor;
they randomly deviate from that pattern (Figure 128). See John I Yel-
loft Jr, Spectral Analysis of Spatial Sampling by Photoreceptors: Topological

There are also limitations in how quickly cameras and eyes can re-
spond. For a camera to take a picture, all the accumulated electron
charges in the photodiodes of the sensor are dumped, using the tran-
sistor circuitry shown in Figures 130 and 129. The dump is parallel,
and goes to the camera’s central processor, which is limited in speed.

A modern camera can take 22 pictures per second; faster than that,
one picture blurs into another. Engineers design delays to stop taking
too many pictures too quickly. Very fast changes are thus invisible to
the camera.

We have the same issue: experiments show we can’t see a difference
between two pictures if they appear for less than 16ms to 13ms. So the eye could process about 60 frames per second, if it acted like a
camera.

To begin with,

...at 20/20 vision, the human eye is able to resolve the equivalent of a
52 megapixel camera (assuming a 60° angle of view). However, such
calculations are misleading. Only our central vision is 20/20, so we
never actually resolve that much detail in a single glance. Away from the
center, our visual ability decreases dramatically, such that by just 20° off-
center our eyes resolve only one-tenth as much detail. At the periphery,
we only detect large-scale contrast and minimal color... a single glance
by our eyes is therefore only capable of perceiving detail comparable to
a 5-15 megapixel camera (depending on one’s eyesight). However, our
mind doesn’t actually remember images pixel by pixel; it instead records
memorable textures, color and contrast on an image by image basis.
In order to assemble a detailed mental image, our eyes therefore focus on
several regions of interest in rapid succession. This effectively paints our
perception. The end result is a mental image whose detail has effectively
been prioritized based on interest.

Cameras vs. The Human Eye, https://www.cambridgeincolour.com/
tutorials/cameras-vs-human-eye.htm
For now, let’s do some comparative numbers.

i) Camera: $2 \times 10^7$ pixels, at 12 bits, and 22 pictures per second; we get about $1.8 \times 10^{12}$, or two trillion bits per second.

ii) Eye: 127 million, or $1.27 \times 10^8$ receptors. At 16 bits (too low, but for comparison ...), and 60 images per second, we get about $1.2 \times 10^{11}$, or 120 billion bits per second.

In the camera, these bits either go directly to memory, or to a processor; for example, one which uses jpeg compression. Figure ?? shows that the information from rods and cones go through several layers of cells. The round cells at the top of the figure are *ganglion* cells; the ganglia are the final stop before the signal is transferred to the areas in the brain responsible for vision.

In the layers before the ganglia, the light signal is represented by changes in the charge of cellular membranes; this can be transmitted very quickly. The ganglia, however, transmit the signal as a series of pulses. A fast (mylenated) nerve fiber can transmit about 2500 pulses a second. But this is misleading, for several reasons.

First, even resting cells put out spikes; the issue is not how many spikes there are, but how much information the spikes give. Second, not all ganglion cells are the same; the experiments on the frog eye discussed on p125 were done on ganglion cells. Further experiments identified many different kinds of ganglion cells, with different rates of spiking; see the survey article *Functional Architecture of the Mammalian Retina*, Heinz Wässel and Brian B. Boycott, Physiological Reviews Vol. 71, No. 2, April 1991.

The computation that balances information and firing rates was performed by Kristin Koch *et. al* in *How Much the Eye Tells the Brain*, Current Biology 16, 1428 -1434, July 25, 2006. For the million ganglia of the retina, the amount of information that can be transmitted is about 875,000 bits per second. As the retina is putting out 120 billion bytes per second, there’s an enormous mismatch. Some of the information from the eye has to be eliminated before it reaches the brain.

The layers of cells between the rods and cones and the ganglion cells (the middle layer in Figure 131) seem to be like the jpg compression software in a camera: they reduce, or compress, what comes from the rods and cones:

...the world that you see is not the world that exists – it has been heavily retouched by your retina. The modified image uses less computational power than the raw form because, before being sent to the brain, it is packaged into more than 30 representations that emphasize specific fea-
tures of the visual scene. The content of these messages is partially understood.


Researchers are now deciphering those partially understood messages, though much work remains. The first layer of cells that respond to changes in rod/cone membrane charge is the layer of horizontal cells. These cells connect a rod or cone to those nearby. One message they can carry is, roughly ‘Wow, it sure is bright. Let’s tone things down.’ More accurately,

The horizontal cell […] measures the average level of illumination falling upon a region of the retinal surface. It then subtracts a proportionate value from the output of the photoreceptors. This serves to hold the signal input to the inner retinal circuitry within its operating range, an extremely useful function in a natural world where any scene may contain individual objects with brightness that varies across several orders of magnitude. The signal representing the brightest objects would otherwise dazzle the retina at those locations, just as a bright object in a dim room saturates a camera’s film or chip. . . .

The Neuronal Organization of the Retina, Richard H. Masland, Neuron 76, October 18, 2012

Another function that we understand is, roughly, ‘Hey! I’m all lit up. If you guys would quiet down, everyone could see me.’ Masland again:

[…] objects neighboring a bright object have their signal reduced […] in the extreme, the area just outside a white object on a black field is made to be blacker than black. This creates edge enhancement […] .

Edge enhancement allows us to pick out objects from a background. In this way, the eye can construct the appearance of individual objects (for us, an important step in counting individual objects!).

The next layer consists of bipolar cells, then amacrine cells which connect directly to the ganglia. There are back-and-forward connections among these cells, and their interaction is complex:

In the inner retina, roughly 42 types of mostly inhibitory amacrine cell modulate bipolar cell output. Although some amacrine cell circuits have been studied in depth, we still understand little about the general principles by which amacrine cell circuits help to decompose the visual scene into the parallel channels carried by the bipolar cell.

Katrin Franki et. al., Inhibition decorrelates visual feature representations in the inner retina. , Nature 542, February 23, 2017
Again, researchers use the language of signal processing: the different kinds of information generated from retinal ‘circuits’ are referred to as ‘channels’. The action of these channels ‘decorrelates’ features. A correlation between two signals determines how much they have in common: for example, Are they simultaneous? Do they respond to the same stimulus? TV channels, for example, are decorrelated: football doesn’t blend in and out of cooking shows. The idea of uncorrelated channels carrying different kinds of information goes back to the original experiments on amphibian vision. The experimenter provides a stimulus, say a moving dot projected on a screen, and records what shapes or sizes or speeds or directions cause a cell to respond. Different stimuli cause different cells to respond; the conclusion of this work is (roughly) that one stimulus is ‘represented’ by one group of neurons in the visual cortex; see Figure ?? Later work showed that this stimulus/response specificity begins in the retina: as Masland remarked above, compression begins in the retina.

For a discussion of ‘channels’, see p139.

But: Franki et. al. point out, "a deeper understanding of the functional diversity of bipolar cells and its origin is lacking." In consequence, experimenters don’t know what stimuli activate which channels. Masland again: "The challenge is how to choose test stimuli, and how to interpret the bipolar-cell responses to them. It is unlikely that naturalistic responses can be achieved using spots of light, striped patterns... ."

Franki et. al. chose a very complex visual stimulus: they used a light flickering at different rates (frequency) and changing intensity (amplitude); see Figure 133. They then used statistical techniques to detect clustering of responses. We don’t yet know what these clusters ‘mean’, so don’t know what the eye is telling the brain. The channels are statistically decorrelated, at least in the retina, but what this means is up in the air. We know the eye is withholding some kinds of information, because vision presents too much information for the system to handle. But we don’t know what; see p143.

This kind of issue arose in signal processing when pictures, video and music became digital. A single CD recording could hold only about 12 or so songs (using the ‘RedBook’ recording standard, Figure 134). To get more, engineers developed compression, like jpeg, mpeg, mp3. These lose some of the original; what we can afford is the subject called compressed sensing.
Charles Darwin puts in a word here. His brother Erasmus reminded him of Plato’s allegory that we understand ideas like numbers because, between an endless series of deaths and rebirths, our soul has directly experienced abstract concepts. Darwin wrote in his notebooks that instead of a preexisting soul, we have pre-existing monkey ancestors. That is, Darwin contended perception evolved to help organisms survive and reproduce; in evolutionary terms, the way we see the world reflects what was useful as our species evolved. The flowers and the bees seem to confirm this: what we don’t need, we can’t see.

Is it true?

...it has long been assumed that neurons are adapted, at evolutionary, developmental, and behavioral timescales, to the signals to which they are exposed.
Because not all signals are equally likely, it is natural to assume that perceptual systems should be able to best process those signals that occur most frequently. Thus, it is the statistical properties of the environment that are relevant for sensory processing. Such concepts are fundamental in engineering disciplines: Source coding, estimation, and decision theories all rely heavily on a statistical ‘prior’ model of the environment.

The experiments on frog vision relied on an artificial stimulus – a moving black dot, for example. A cell in the brain firing in response was considered to recognize dots moving in that direction. A model of the frog’s environment, but not a statistical one. The chirp of Franki (p132) isn’t even an attempt to represent the environment. Moreover, one cell may take part in recognizing many other motions and will stimulate other neurons throughout the brain. As Peiran Gao, Eric Trautmann et. al. note,

...how can we record on the order of hundreds of neurons in regions deep within the brain, far from the sensory and motor peripheries, like mammalian hippocampus, or pre - frontal, parietal, or motor cortices, and obtain scientifically interpretable results that relate neural activity to behavior and cognition? ...we could be completely misleading ourselves: perhaps we should not trust scientific conclusions drawn from statistical analyses of so few neurons, as such conclusions might become qualitatively different as we record more.
The issue of which neurons react is the coding problem; as it could be as many as $10^9$, the issue is complicated. One way to reduce the complexity is this: if a large numbers of neuron respond in a similar way (are correlated) then deciphering the coding is simpler. This is called a low-dimension response; it is the type of response that investigators found in the frog.

If the neurons all react in a unique way, the code is said to be high-dimensional; this is the dimensionality problem. Very low-dimension codes can’t distinguish different visual scenes, like a child to whom every moving vehicle is ‘car’. High dimension codes are highly sensitive to differences in a scene, as though the absence of a single forest leaf could identify a tree as a freeway. This is a typical difficulty for artificial intelligences, Figure 136.

If we assume that visual systems are highly adapted to the environment, we’d believe in something like low-dimensionality. The scientific problem is in providing natural images to an animal, and recording simultaneously from thousands of neurons. An experiment recording from over 10,000 neurons was recently reported in High-dimensional geometry of population responses in visual cortex by Carsen Stringer, Marius Pachitariu et. al., Nature Vol 571, 18 July 2019.

Here we recorded the simultaneous activity of approximately 10,000 neurons in the mouse visual cortex, in response to thousands of natural images. We found that stimulus responses were neither uncorrelated (efficient coding) nor low-dimensional. Instead, responses occupied a multidimensional space . . . . These findings suggest that the population responses are constrained by efficiency, to make best use of limited numbers of neurons, and smoothness, which enables similar images to evoke similar responses.

There’s another issue. After the experiments on ‘one direction of motion=one neuron’ in the frog, it was understood that neurons recognizing direction grouped together physically:

The modular organization of nervous systems is a widely documented principle of design for both vertebrate and invertebrate brains of which the columnar organization of the neocortex is an example. The classical cytoarchitectural areas of the neocortex are composed of smaller units, local neural circuits repeated iteratively within each area. Modules may vary in cell type and number, in internal and external connectivity, and in mode of neuronal processing between different large entities; within any single large entity they have a basic similarity of internal design and operation. Modules are most commonly grouped into entities by sets
of dominating external connections. [...]. The set of all modules composing such an entity may be fractionated into different modular subsets by different extrinsic connections. Linkages between them and subsets in other large entities form distributed systems. The neighborhood relations between connected subsets of modules in different entities result in nested distributed systems that serve distributed functions.


It’s an attractive picture: small groups of neurons clump together to process certain features of the visual environment; the brain has many different groups, each recognizing a single feature: the recognition of features is ‘distributed’. In cognitive science, there are extensions of this idea: the functions of the brain are carried out by modules, each of which performs a particular cognitive function. See Jerry Fodor, Modularity of Mind: An Essay on Faculty Psychology, MIT Press 1983, as well as Marvin Minsky, The Society of Mind, Simon & Schuster 1988.

But this picture leads to a question: if the visual field is distributed, where and how is the information re-assembled? This is called the binding problem. One way to say it is this: you’re playing tennis, and a ball is coming towards you. Part of the brain recognizes the motion, another part the spin of the ball, another part the color, another the shape.

Why do all of these stick together? Why doesn’t the color peel off from the shape? There’s some recent research that hints at what may be happening – and, once again, it has to do with dimensionality. The experiment records neuron behaviors in the visual cortex of the rat. Figure 137 shows the idea: the eyes detect objects; the visual response is sent through nerve pathways to a specific area of the brain, where the modules responsible for recognition sit, the visual cortex. An experiment testing this was run by Carsen Stringer, Marius Pachitariu et al. ([Spontaneous behaviors drive multidimensional, brainwide activity, Science 364, 255 19 April 2019]). They recorded from roughly 10,000 sites in the visual cortex of a rat moving about in a completely dark room. There was tremendous activity, of high dimension – but the activity must have certainly been noise. Was it?
In part two of the experiment, infrared cameras recorded the rats as they ran, groomed themselves, moved their heads or ears, etc. As it happened, these non-visual behaviours correlated with what had been assumed to be ‘noise’.

Recording more than 10,000 neurons in mouse visual cortex, we observed that spontaneous activity reliably encoded a high-dimensional latent state, which was partially related to the mouse’s ongoing behavior and was represented not just in visual cortex but also across the forebrain. Sensory inputs did not interrupt this ongoing signal but added onto it a representation of external stimuli in orthogonal dimensions. Thus, visual cortical population activity, despite its apparently noisy structure, reliably encodes an orthogonal fusion of sensory and multidimensional behavioral information.

Not only was the visual cortex aware of non-visual information, areas outside of it were aware as well. It seems that visual and non-visual activities are linked from the very beginning, rather than being processed separately. How this linking works, and what it means, will need further research.

Regrettably, this makes traditional kinds of experimentation – ‘change only this, record only that’ – much more difficult; one has to analyze system-wide changes across many dimensions. An interesting question is whether our minds are capable of understanding this complexity. Some see artificial intelligence as a solution, but this also has problems: see pp 134 and 144.
Notes for First Expedition: Vision

If we think back to the Babylonian Systems A and B, p15, these mathematical models are a bit like Plato’s ‘myths’. These ancient astronomers didn’t expect their equations to capture a complete reality; they considered it a useful approximation, and hoped it will explain or lead to new truths—though these again may be only approximations.

p124 The difference between our view of a flower and that of a bee shows our eyes aren’t like those of other organisms; Figure 138 shows a more extreme example, a sea urchin. The red areas are the ‘eyes’—more accurately, cells that respond to light. The urchin crawls about on the sea bed, detecting light without lens or eyeball, as though the whole organism were an all-directional eye.

Color vision is always trickier. Newton showed that light passing through a prism (Figure 139) could be broken up into an entire range of colors. He used the term spectrum for this range of colors (from the Latin ‘to look’, cognate to ‘spectacles’, ‘inspect’, etc). But this is a vast range of colors; most animal eyes can detect only a few of them. For humans, the main colors are Red, Green and Blue: RGB. The human brain then does a bit of mathematical juggling: it takes the relative amounts of R, G and B, and constructs new colors, as an artist would mix paints; see Figure 140.

The mantis shrimp (Figure 141) has receptors for twelve distinct colors. As an analogy, in Figure 139 the prism breaks light into its different colors, and the black dot is a photoreceptor for each color. These 12 receptors form a representation of the colors present; called a spectral analysis. "Why use 12 color channels when three or four are sufficient for fine color discrimination? Behavioral wavelength discrimination tests ... revealed a surprisingly poor performance, ruling out color vision that makes use of the conventional ... coding system. Instead, our experiments suggest that stomatopods use a previously unknown color vision system based on temporal signaling combined with scanning eye movements, enabling a type of color recognition rather than discrimination." Hanne H. Thoen et. al., A Different Form of Color Vision in Mantis Shrimp Science vol 343 24 January 2014; see also Michael F. Land and Dan-Eric Nilsson, Animal Eyes, Oxford University Press; 2nd edition 2012. Although the mantis vision system is not fully understood, these authors suggest that the shrimp scans across its receptors one at a time, rather than merging them as the human brain does. This requires less co-operation amongst the receptors; it also doesn’t require a layer of cells to integrate the responses of the different receptors. This makes it likely that the shrimp can
respond more quickly to prey.

These exotic kinds of visions are studied as *Visual Ecology*; species take basic structures like light sensors (photoreceptors), and then those structures become adapted towards detecting the parts of its environment assisting reproductive fitness. Since species have different needs, it’s no surprise our unaided eyes can’t detect the reality of other organisms; see the article in Science News, Vol. 190, No. 2, July 23, 2016, p 35. For a detailed study, see Thomas W. Cronin *et. al.*, *Visual Ecology*, Princeton University Press 2014.

Many cultures believe in an invisible world of ancestors, of helping and harmful spirits. The Japanese festival of *obon* is a contemporary survival of such beliefs. During the festival, families visit the gravesite of their ancestors and symbolically carry the spirits back home, where they are offered food.

We dismiss the belief in invisible worlds as superstition, but some modern cosmological/quantum theories theorize many universes. The *many worlds* interpretation of quantum mechanics posits that every possible outcome of an experiment actually occurs in some world; new universes are constantly being birthed (see for example Max Tegmark, *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*, Vintage, 2015). The inflationary theory of the origin of this universe (see Alan Guth’s *The Inflationary Universe*, Basic Books, 1988); suggests many universes sprouted at the same time as ours, with different laws of physics. See also Yasunori Nomura, *The Quantum Multiverse*, Scientific American June 2017.


For an introduction to how the brain processes visual information, see the book *From Neuron to Brain*, by John G. Nicholls and A. Robert Martin, p??

The frog work is from the late 1960’s; we now know a great deal more about vision.

*In the vertebrate visual system, all output of the retina is carried by retinal ganglion cells. Each type encodes distinct visual features in parallel for transmission to the brain. How many such output channels exist and what each encodes are areas of intense debate . . . we show that the mouse retina harbours substantially more than 30 functional output channels.*

Baden et. al. note that some of the visual features (or ‘output channels’) extracted for presentation to the brain are: local motion, direction of motion, and illumination. Most of the channels, however, are kinds of information we don’t yet understand, and perhaps can’t conceptualize; see the note on p144, below.

Similarly,

*The retina actually performs a significant amount of preprocessing right inside the eye and then sends a series of partial representations to the brain for interpretation. We came to this surprising conclusion after investigating the retinas of rabbits, which are remarkably similar to those in humans.*


The retina sends a series of images to the brain; since the images change over time, the authors call these ‘movies’ (Figure 143). They identify twelve different kinds of movies that the retina generates and sends on: some show the edges of a scene, some show brightness, or reflectance, and, as in the mouse, some show information we have no name for. The brain then integrates these movies into what we call vision.

p127 In using the term *signal* here we’re using the language of modern information processing. Much of it was developed in Bell Labs, a research unit of the AT&T corporation. The research was concerned with very general properties of electrical *signals*, how those could *encode* voice, or pictures, or ... *information*, and *transmit* it over wires, undersea cables, satellites, microwave systems . . . *channels*. Initially developed for telegraph and telephone systems, it was extended to computers. This language is currently used to describe both computers and nervous systems; we’ll use terms like *signal*, *information*, *channel* and *signal processing*. We also will be careful: nerves and wires, brains and computers, are not the same. *Signal processing* sounds very scientific but is just an analogy. Nerves are much more complicated than wires. Here’s more detailed discussion about the use of the word ‘channel’ in talking about the retina and ganglion cells. Those uninterested can skip directly to p144.

p132. The word ‘channel comes from the Latin *canalis*, ‘pipe’. Information-theory channels take an input, and with some probability produce an output (see Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc, 1991). In a telegraph: you key the input ‘dot-dash’, the channel is miles of wire, and, 90% of the time, ‘dot-dash’ is the output. You could think of the body’s system for regulating blood pressure as a channel.
Figure 144 shows how this works. A baroreceptor on a vein or artery senses pressure; there’s your input. The receptor translates this to an electrical signal, which is sent to the brain. The brain controls pressure by signals to the heart through sympathetic or parasympathetic nerves. These release hormones which either slow (norepinephrine) or increase (epinephrine) the heartbeat; those hormones are the output of the channel.

‘Channel’ as ‘pipe’ may be too simple an idea for the above. Information theory had in mind something more like a headphone cable. An mp3 player emits varying voltages representing sounds; the cable transmits these as electromagnetic waves, and the headphone changes this to sound. In this example, it’s very clear what input and output mean.

Clear, but inaccurate. The easiest way to see the problem is by thinking about a prism (Figure 145). The low frequency red light travels more slowly through glass than the high-frequency blue light, so the two frequencies are bent slightly differently by the glass, and the prism separates them out.

Earphone cables act like prisms: bass and treble sounds travel at different speeds, and arrive at the ear at different times. In severe cases, this produces a mushy, blurred sound. People who are serious about music (the author) spend tons of money (not the author) on cables that minimize blurring. Even then, you’re advised to ‘burn-in’ your cables, by playing a hundred or so hours of the music you like. Apparently, the cables can adapt (or de-adapt) themselves to your music.

If you think of a channel as a pipe, the idea ‘water in one end, water out the other’ is appealing and simple. Take something more complicated, even just voltages traveling through wires, and the simplicity is lost.

In fact, water travels more slowly nearer the pipe, and more quickly though the center. And, as most water contains dissolved solids, those can wind up encrusting the inside of the pipe, slowing things further. So even the simplest kind of pipe adapts to its inputs.

We can think about ‘channels’ made of neurons, and use information theory, as with ganglion cells in the retina. Abstractions like this can introduce ideas and questions from other sciences: how much information can these channels transmit? How is it encoded? How does this system deal with noise? With partial information?
However, ‘nerves-as-channels’ are, instead, actually complex systems: a single neuron can behave more like a computer than as a wire. We’ll start with the stereotype of a brain cell, shown in Figure 146. There’s the purple cell, with its nucleus in black, and little finger-ish extensions called dendrites, and little yellow blobs called glial cells. Let’s start with the fingers: dendrites. In our simplified picture, dendrites are extensions from the main cell; they can interact with other neurons. Old theories viewed nerve cells as wired to each other through dendrites. The dendrites would pick up electrical impulses from other neurons, and the sum of all the impulses would cause the entire neuron to emit an electrical impulse – if the sum was large enough. Dendrites were to be the wires connecting a circuit element (the neuron) to other circuit elements.

In this picture, a single neuron is a ‘circuit element’ – while the neuron has internal components (DNA, mitochondria, ATP molecules, etc.), that structure was seen as irrelevant to the input/output nature of the neuron. As it happens, this picture is misleading. Albert Gidon, Timothy Adam Zolnik, et al., Dendritic action potentials and computation in human layer 2/3 cortical neurons Science 367, p 83, 3 January 2020 looked at electrical currents inside dendrites themselves "These [...] enabled the dendrites of individual human neocortical pyramidal neurons to classify linearly nonseparable inputs - a computation conventionally thought to require multilayered networks.” That is, a single dendrite may be pre-processing the electrical impulses it receives from other neurons. This would make the neuron itself less like a circuit element and more like a small computer, deciding to enhance certain inputs, alter or suppress others, and, possibly, altering its own programming as it goes along.

The same is true for the glial cells, the blobs:

*Originally, scientists didn’t think they did anything. Until the last 20 years, brain scientists believed ... that glia were kind of like stucco and mortar holding the house together. They were considered simple insulators for neuron communication.*


Recent research shows glia have their own energy and signaling systems and can direct or destroy the growth of neurons (see Darran Yates, Glial messaging, Nature Reviews Neuroscience 18, 2017).
Of course ‘brain cells’ don’t float around in splendid isolation, making occasional connections with other brain cells. The figure below is from Kasthuri et al., *Saturated Reconstruction of a Volume of Neocortex 2015*, Cell 661 July 30, 2015. It demonstrates the complex mesh of support systems for cortical cells; some of the support cells are shown in isolation below.

Not only is the ‘brain cell’ surrounded by support cells, it interacts with them and is influenced in turn; the ‘wiring diagram’ idea is far too simplistic, as the figure above illustrates. On the left, a section of the cortex of a rat brain; all kinds of cells, indicated by colors, are packed tightly together; on the right, the cell types. They interact with the neuron, and other neurons, and with the other kinds of cells, in ways that are not understood. Analogies like channels, wiring and signals may be too simple for understanding the mutual interactions of this system.

Returning the discussion to the retina, recall Franki’s comment p131 "Amacrine cell circuits help to decompose the visual scene into the parallel channels carried by the bipolar cell."

We could read Franki and think of simple channels that can only process certain fixed kinds of information. Then what the eye can see is limited by the channel, and there are kinds of things we simply can’t see. Again, rather like a television news channel, where see only what the politics of that station wants us to know. But we could think instead of more complex channels, which are self-adapting. In
that case, the retina itself might be capable of learning, of seeing new things.

Recent work on the retina and its interactions with adjacent cells suggests something like that. Cells in the retina require energy to convert light to changes in cell potential; that energy is carried in the highly reactive element, oxygen, which is supplied to the cells by blood vessels. As it happens, the network of blood vessels is partly controlled by the amacrine cells. Yoshihiko Usui et. al. examined the relationship between these. Their results:

i) "Amacrine and horizontal cells form neurovascular units with capillaries . . ."

ii) "Amacrine cell and horizontal cell derived VEGF is essential for neurovascular-unit formation . . ."

(VEGF is Vascular Endothelial Growth Factor, a protein promoting the growth of blood vessels) See Yoshihiko Usui et. al. Neurovascular crosstalk between interneurons and capillaries is required for vision, The Journal of Clinical Investigation, Volume 125 Number 6 June 2015.

It seems the neurons construct their own environment. It stands as a warning: words like pipes, channels, signals and information suggest a fixed system, which may not exist. The kinds of metaphors we choose influence our science – another example of what we can and cannot see! All this brings us back to epistemology and the discussion on p137, as well as Plato’s remark that all our senses can tell us about the world is a kind of myth.

p127 The number of electrons that a photodiode captures could be anywhere from one to millions; much of this is meaningless, as the eye can’t perceive the difference between 1,000 versus 1,001 electrons. Circuitry assigns data from the photodiode to one of $2^{10}$ different levels; the process is called quantization. We won’t estimate the exact number of electrons, but we’ll be off by at most $2^{-10}$.

p132 The discussion above, and that on p132, asks us to think:

**can** we understand what the different channels of the retina are telling the brain? As we saw, the ‘channels’ may be complex self-constructing systems of their own. They also appear to compress visual signals, before passing them to the brain. The engineering equivalent would be jpeg or mp3 compression. But those were designed by engineers, using well-known techniques from signal processing. jpeg, mp3, are cross-platform, designed for use in many different kinds of electronics. So engineers need to know, in advance, how well they work. This is typical engineering: there are specifications and you design to meet them. And, if there’s a failure, it can be traced.
Vision is radically different: it evolved, over hundreds of millions of years, and performance was shaped by reproductive success (including survival). System failures are diseases and blindness, but even today, after years of medical progress, there are kinds of blindness we don’t understand and can’t treat.

Now we’re trying to look inside, see what makes it go. Should we expect to understand it? We suspect the very simplest mathematics of signal processing and decorrelated channels might not apply. What else do we have?

In contemporary technology, the closest analogy to evolution is artificial intelligence, AI, which uses non-classical engineering. An example is Google’s work on what it calls deep learning. Google engineers taught a computer to play the game of Go as well as the other video games on early Atari machines (see Figures 147 and 148).

The engineers

*applied deep learning in neural networks – brain-inspired programs in which connections between layers of simulated neurons are strengthened through examples and experience. It first studied 30 million positions from expert games, gleaning abstract information on the state of play from board data, much as other programmes categorize images from pixels (see Nature 505, 146-148; 2014). Then it played against itself across 50 computers, improving with each iteration, a technique known as reinforcement learning.


These techniques were evolved: the computer reprogrammed itself through its experience of what worked and what didn’t. Engineers can experiment with the program, but can’t look at the code to ‘fix’ anything.

Getting computers to play games isn’t the real goal of deep learning; the goal is to program self-driving cars, web browser page-ranking schemes and speech recognition. And to build ‘an M.D. in a box’. Will we trust computer doctors?

Paul Voosen writes of a programmer who worked on deep learning to help diagnose pneumonia (The AI Detectives, Science 357, 9 July 2017 Issue 6346). ”In general, sending the hale and hearty home is best, so they can avoid picking up other infections in the hospital. But some patients, especially those with complicating factors such as asthma, should be admitted immediately. […] disturbingly, he saw that a simpler, transparent model […] suggested sending asthmatic patients home.” The programmer wonders what other mistakes the computer might be making.
Apple’s Siri, Amazon’s Alexa, and Netflix’s movie recommendations all use deep learning. Some credit companies also use it to determine who is worthy of a loan. What if there’s a mistake?

In 2018, Article 13 of the European Union’s General Data Protection Regulation law will take effect; it offers specific legal protections against deep learning:

Paragraph 1 […] the controller shall, at the time when personal data are obtained, provide the data subject with the following further information necessary to ensure fair and transparent processing.

Subparagraph (f) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

That is, the computer can be required to explain its decisions. This is an active field of research; see the above article by Voosen in Science.

Our question is whether we can use these techniques to understand vision, or more generally, use deep learning to do science.

Eventually, some researchers believe, computers equipped with deep learning may even display imagination and creativity. “You would just throw data at this machine, and it would come back with the laws of nature,” says Jean-Roch Vlimant, a physicist at the California Institute of Technology in Pasadena.

The Black Box of AI Nature, 538, October 6, 2016.

Here’s a thought experiment: what if we’d invented statistics and deep learning before we invented calculus? A deep learner has centuries of data on the positions of all the planets. Would we get all the laws of physics? Or perhaps, there wouldn’t be concepts or laws, just statistical correlations.

Instead of a thought experiment Raban Iten, Tony Metger, et. al., Discovering physical concepts with neural networks in arXiv.org > quantph > arXiv:1807.10300 took an actual example and used AI on it.

The example concerned the motion of the planet Mars, as seen from the earth. As shown in Figure 149, the apparent motion of Mars, as seen from the earth, is complex: Mars appears to stop and turn around. This is called retrograde motion. The figure gives a modern interpretation: as the Earth and Mars circle the sun, the difference in their speeds makes one catch up, then surpass the earth, giving the
illusion of retrograde motion. This interpretation relies on the theory that the planets circle the sun: the heliocentric model of planetary motion.

Raban Iten, Tony Metger, et. al. took the observed position of the Sun and Mars, gave it to the AI, and

...our network learns to compress experimental data to a simple representation and uses the representation to answer questions about the physical system. [...] For example, given a time series of the positions of the Sun and Mars as observed from Earth, the network discovers the heliocentric model of the solar system - that is, it encodes the data into the angles of the two planets as seen from the Sun. Our work provides a first step towards answering the question whether the traditional ways by which physicists model nature naturally arise from the experimental data without any mathematical and physical pre-knowledge, or if there are alternative elegant formalisms.

The following critique is from the article Are Neural Networks About to Reinvent Physics? by Gary Marcus and Ernest Davis, in NAUTILUS November 21 2019. Marcus is a Professor in the Department of Psychology at New York University; Davis is a Professor of Computer Science at the Courant Institute for Mathematical Sciences. They write:

The trouble is that it is entirely misleading to say that their neural network infers that 'the Earth and Mars revolve about the sun.' The neural network doesn't actually understand that anything is revolving around anything, in a geometric sense; it has no sense of geometry and no idea what it would mean to revolve. All the neural network does is to extract the two numerical parameters involved; it has no idea that these represent angles from some fixed central point. As far as the network is concerned, these could be time-varying masses, or electric charges, or angles from two different central points. Correlations between data sources were extracted, but the system made no inferences about how those data sources related to the world; it is the human scientists who identify these as the angles of the Earth and Mars as measured from the sun, and who abstract the facts that such parameters are best interpreted as orbits. All of the real work of Copernicus' actual discovery is done in advance; the system was a calculator, not a discoverer.

Further, in the synthetic data the authors generated, the Earth and Mars move in constant-velocity circular orbits in the same plane. In the real solar system, things are notably trickier...

To quote Andrew Saxe, at Oxford:
A strong intuition among many neuroscientists is that individual neurons are exquisitely complicated: They have all of these back-propagating action potentials, they have dendritic compartments that are independent, they have all these different channels there. And so a single neuron might even itself be a network. To caricature that [...] the simple mathematical model of a neuron in [Digital Neural Networks] is clearly missing out on so much.


There are other issues: data on the relative positions of Mars and the Sun were chosen out of the vast collection of data on the positions of celestial objects visible in the sky. Why would an AI pick just these? The Sun, Mars, planets, orbits, angles are all human constructs, and are not data at all.

How much does it matter that ‘the system was a calculator, not a discoverer’? Calculators are great tools, but, as we saw in Section 8 on computer numbers, one has to have an intuition about where calculators might get into trouble. That kind of intuition is often lacking in AI: we don’t know what the calculator is doing.

There’s a second issue: humans have reasons for their computations. The planets, their positions, their risings and disappearances, were part of religions and astrologies (see p15). The position of the moon is also used to determine religious holidays: Easter, Holi, Yom Kippur, Chinese New Year, Ramadan and Eid-al-Fitr (see p17). All of these depend on the phases of the moon.

Presumably an AI could predict the orbit of the moon around the earth (as opposed to the prediction of the orbit of Mars around the sun). This is important to us because the moon determines the tides, which affects fishing and shipping. But how would an AI know to throw tidal data into the mix of celestial data?

Figure 150 shows a collection of words using one of the three writing systems used in modern Japan. An AI might ‘learn’ to identify these words after being exposed to thousands of variants. It might even pronounce the word correctly. But then what?

A human, in this case, the author, could use these words to read, or converse, to write, to purchase items from a grocery or clothing store. And the human would notice very quickly that the two characters at the end of each word are identical, are pronounced ‘mono’, and mean, roughly, ‘things’. Thus, drink things, things cooked in a pot, clothing things, pickled things. The human could then go on to read, and use, new words.
It brings us back to Wittgenstein (p37): the meaning of a word is embedded in the cultural, human context of the uses of that word; the meaning is in our interactions.

 [...] two aspects of human conceptual knowledge have eluded machine systems. First, for most interesting kinds of natural and man-made categories, people can learn from just one or a handful of examples, whereas standard algorithms in machine learning require tens or hundreds of examples to perform similarly ...

Second, people learn richer representations than machines do, even for simple concepts, using them for a wider range of functions, including creating new exemplars, parsing objects into parts and relations, and creating new abstract categories of objects based on existing categories.

Brenden M. Lake et. al., Human-level concept learning through probabilistic program induction Science 350, 11 December 2015 Issue 6266.

As with Plato, AI challenges what we mean by terms like "understanding", "meaning", "explanation." And, with Plato, we have to ask: if these don’t come from our senses, from data alone, where do they come from?