The Neocortex

In the last lecture we studied a simple type of neural representation: PCA. PCA can be done using a Hebb rule, but the Hebb rule can also be used to do more elaborate types of representation. It seems likely that the neocortex is basically a Hebbian learning machine that creates particularly efficient representations of its inputs, and then generates outputs that contribute to useful behavior. However, although we know an enormous amount about the structure and function of the neocortex, and we are starting to better understand representational learning, we are still a long way from putting the 2 together.

The neocortex is a crumpled sheet of neurons that in humans is about a quarter square meter broad and several millimeters thick. The sheet is composed of 6 layers, numbered 1 through 6 from the top down. Layer 1 has no neuron somata; layers 2 and 3 (which are often lumped together) have small excitatory pyramidal cells; layer 4 has very small excitatory spiny star-shaped or granule cells; layer 5 has large spiny excitatory pyramidal cells; layer 6 has pyramidal and multiform cells. Layers 2-6 also have a variety of nonspiny inhibitory interneurons. Almost all inputs to cortex arrive from thalamus, as the glutamatergic axons of thalamic relay cells, which synapse profusely in layer 4, and sparsely in layer 6. However, there are some additional largely neuromodulatory inputs, such as cholinergic axons arriving from the nucleus basalis of Meynert and the diagonal band of Broca, serotonergic axons arriving from the Raphe nucleus, and adrenergic axons arriving from the locus ceruleus. These axons serve at least 2 important roles (1) they convey “reward” signals that measure how successful or useful behavior is and (2) they modulate the brain’s state between wakefulness (high release of Ach and Norepinephrine), slow wave (dreamless) sleep (lower release of Ach and norepi) and rapid-eye movement (REM or dream sleep) high release of Ach; no norepi release).

Axons of layer 4 cells have 2 major destinations: they feedforward to layer 2/3 and they form recurrent synapses on other, nearby, layer 4 cells. Layer 2/3 cells form recurrent connections on each other, they feedforward to layer 5, and they feedforward to other cortical areas (where they typically innervate other layer 2/3 cells). Massive numbers of these layer 2/3 axons cross to the contralateral neocortex in the huge fiber bundles the corpus callosum and anterior comissure. Layer 5 cells make recurrent synapses on each other, may innervate more superficial cortical layers, and often send axons subcortically. These layer 5 axons thus convey the “output” of cortex. These output axons typically send a branch to thalamus and a branch to other subcortical structures. Thus large layer 5 cells in motor cortex send axons straight to motoneurons, where they trigger movement commands.

The neocortex is divided into hundreds of different areas which are specialized for handling different types of incoming information. Some cortical areas receive input from “first order” thalamic nuclei, which in turn receive input from *subcortical* structures. Thus the lateral geniculate nucleus (lgn) of the thalamus gets input from the retina and supplies the primary visual cortex (called V1 or striate cortex) in the occipital lobe at the back of the head; the medial geniculate nucleus supplies auditory cortex in the temporal lobe; the ventrobasal nucleus of the thalamus supplies somatosensory cortex just behind

the central sulcus at the front of the parietal lobe; the ventrolateral nucleus of the thalamus supplies cerebellar information to the motor cortex (in front of the central sulcus); the anterior thalamic nuclei supply hippocampal information to the cingulate cortex (medial cortex that is hidden on the inner surface of the temporal lobe). Many other cortical areas receive instead input from "higher order" thalamic nuclei, such as the pulvinar nucleus, which relay information from layer 5 cells of other *cortical* areas. For example, in humans, the pulvinar (which is the largest thalamic nucleus) transfers information from striate cortex (via the axons of layer 5 cells) to higher order visual cortices (such as V2, V3, V4 and V5 – the last 2 are particularly concerned with color and motion respectively).

We cannot in this course properly consider the incredible richness, complexity and diversity of these cortical circuits. We will merely highlight a few topics in the visual system.

The Retina

First we should briefly consider the properties of retinal ganglion cells, which provide the sole output of the eye. What visual stimuli are ganglion cells designed to detect? For what visual prototype are these cells tuned? In other words, what visual stimulus provokes the biggest response from these cells? Why? – in both senses of the word: what circuits generate these responses (and how do these circuits develop?) and why (from an information processing perspective) do they respond in this way?

The first question however, is, how do we measure to which visual stimuli the cells are tuned? One way would be to try all possible visual stimuli, and see which gave the biggest response. This can be done by applying random visual stimuli, and remembering (in a computer) those that give the biggest responses. Actually, since we are really interested in the stimuli that are most likely to trigger spikes, we can store just the average of any stimulus that triggered a spike (together with the time between that stimulus and the spike). This procedure is called spike triggered averaging or reverse correlation.

For the most part retinal ganglion cells only respond to visual stimuli that fall quite close to the position of the ganglion cell on the retina. This region is called its receptive field. Typically, the receptive fields have a concentric center-surround organization. They fall into 2 classes. "On" cells have RF centers where a brief light increase causes, after a brief delay, a brief increase in spike probability followed by a decrease. Off cells do the reverse.

The circuitry involving these responses is well understood. They are generated by input from "on" or "off" bipolar cells that gather signals from a small patch of photoreceptors . The central photoreceptors in this patch (in the fovea in primates, just 1 bipolar and 1 photoreceptor) create the center, and the surrounding photoreceptors create the surround, by subtracting from the net input to bipolars (the subtraction is done at the level of the

photoreceptor synaptic terminals, by feedback from horizontal cells that collect from the surround).

The basic reason for this arrangement is “decorrelation” or “whitening” – basically a form of PCA. If one examines different pixels in different visual images (i.e. the statistics of natural scenes) one finds that if they are close together they tend to be similar but this positive correlation dies off the further apart the pixels are. Also, the optics of the eye blur the image to some extent, further increasing the correlations between neighboring pixels. Furthermore, pixel values in different successive snapshots of the world are also correlated (e.g. successive frames of a movie), to an extent that dies off with increasing temporal separation. Just as PCA identifies and removes correlations in inputs, so the center-surround receptive field structure removes these visual correlations. If there were no local visual correlations, the best receptive field structure would be a pure narrow center (it could be as narrow as a single photoreceptor if there was no photon noise – this is basically the situation for foveal cones). But because surrounding pixels are typically similar to the central pixel, their signal should be subtracted, which will emphasize the unusual but significant cases where neighboring pixels have very different values. Similarly, because pixel values are temporally positively correlated, the receptive field structure at early times should be opposite to the structure at later times. This is why the RFs reverse from early to late times (from on center to off center, etc).
If there is a great deal of photon noise (for example in dim light), the need to average signals from neighboring pixels outweighs the need to decorrelate the ganglion cell signals, and the functional circuitry changes, eliminating the surrounds and broadening the centers (and likewise, eliminating the temporal RF reversal). Finally, there are 2 types of ganglion cell (on and off) because they each efficiently represent the positive and negative parts of the image. Thus the spikes of the ganglion cells provide an optimal encoding of visual images, without making any assumptions about image statistics (other than the radially symmetric fall-off of pixel correlations). It should be noted that the overall aim is not for the ganglion cells to send a perfect “image” of the visual world to the rest of the brain, but to efficiently encode the visual information using the limited capacity of their axons. If the brain wanted, it could then use that encoded information to “reconstruct” the visual scene, but, since there is no “homunculus” in the brain to view that reconstructed scene, it does not do so. Instead, it further manipulates that information so that the aspects that were merely ‘implicit” in the retinal encoding become “explicit” in the new cortical re-encoding, and easily available to guide decisions.

[There is one slight complication. There are in fact 2 major types of ganglion cells, small linear ones called midgets or P-cells and large nonlinear ones called parasol or M-cells. This arises because the simultaneous requirements for high spatial and temporal resolution are contradictory: high temporal resolution inevitably requires more photons, which can be achieved only by broadening the collecting area for ganglion cells. This P/M distinction is maintained at many stages of visual processing. Also, of course, resolution varies greatly from fovea to peripheral retina.]

The Thalamus

Relay cells in the visual thalamus (lgn) have very similar receptive field properties to those of retinal ganglion cells. This is reflected in the circuitry – typically, only one or a few ganglion cells provide the major input to a relay cell. These inputs are concentrated on the proximal dendrites of relay cells, so that they can control relay cell spikes with great temporal fidelity.

[There have been suggestions that lgn relay cells do differ in aspects of their spatiotemporal RFs from ganglion cells, possibly because they do further whitening. It is interesting to note that there will in general be tradeoffs between optimal spatial and temporal whitening, but this tradeoff point cannot be correctly set in the absence of particular information processing goals. For example, if high temporal resolution is more important than high spatial resolution, a different spatiotemporal whitening filter will be needed from the converse case. However, this cannot be decided without preliminary cortical processing, i.e some decision about whether moving lines, corners, objects are present in the visual scene. This may relate to the layer 6 corticothalamic feedback discussed below).

However, the relay cells also get massive input (on their distal dendrites) from 2 other sources (1) brainstem and (2) layer 6 of striate cortex. This creates a bit of a puzzle. If the response properties of lgn relay cells are largely unchanged from those of retinal ganglion cell, what role do these additional inputs (which numerically far exceed the retinal inputs) play?

A key insight into this problem comes from the observation that relay cells have 2 firing modes: “burst” and “tonic”. In tonic mode the cells fire individual spikes rather randomly spaced, but in burst mode they fire extremely brief bursts of spikes followed or preceded by silent gaps. The modal state of the relay cells depends on slow changes in membrane potential, which are probably mostly caused by the cortical and brainstem inputs. We do not know what role this burst/tonic transition plays in neocortical processing, but it appears to correspond to a “label” applied to the message being relayed, rather like a highlighting operation. There is recent evidence that burst or tonic spikes arriving in cortex are handled differently, even though the message they carry is the same. One possibility is that the highlight corresponds to increased local cortical attention. Clearly however, the application of the highlight is done on the basis *of preliminary cortical analysis*, via layer 6 feedback.

The role of the brainstem input is a little clearer. It seems to reflect more global aspects of attention and sleep/wakefulness. In an awake attentive state relay cells are generally in tonic mode, but frequently but unpredictably slip into burst mode. In slow wave (dreamless) sleep relay cells are mostly in burst mode, and these bursts tend to occur rhythmically. The circuitry involved in generate these rhythmic bursts, and the associated “sleep spindles” (features of slow wave sleep EEG) is largely understood – it involves cyclic reciprocal feedback of relay cells and inhibitory interneurons. However, we do not understand why these rhythmic bursts, which are sent to cortex, are important (just as we

do not understand sleep itself). Progress in understanding sleep (one of the greatest neurobiological riddles) will surely come from understanding this question. In dream sleep, the relay cells switch to random tonic activity, again probably because this type of activity is necessary to "recalibrate" the neocortex. However, we do not understand what type of "recalibration" is being done. Note however that the linkage between the unconscious state of slow wave sleep and the hallucinatory consciousness of dream sleep and the corresponding thalamic activity patterns underscores the possible linkage between burst/tonic transitions and awareness. Despite the fact that understanding the neural basis of consciousness is one of the ultimate goals of neuroscience, almost no work is being currently done in this area where the enigma seems most vulnerable.

[P and M cells information is kept separate in different thalamic layers. In development, retinal axons initially branch promiscuously in lgn, but Hebbian mechanisms then restrict their terminals to appropriate layers. This process is driven by spontaneous activity waves that move across the fetal retina, ensuring that axons from the same eye have highly correlated activity; it is largely complete at eye-opening. Relay cells that respond to adjacent parts of visual space tend to lie together within an lgn layer, a so-called "topographic map". Furthermore, the maps in different layers are in register, probably to facilitate correct innervation by layer 6.]

Striate Cortex

What happens to visual information after it reaches cortex? Let us first consider the first level of analysis, layer 4, specifically in cats, where the pioneering work was done. (Primates are in principle similar but some of the analysis is deferred to slightly later layers, and color is much more important). In other words, what are the receptive field properties of layer 4 cells, what circuitry underlies these responses, and why do these cells have these properties?

Perhaps the most striking observation is that, while the visual scene (in binocular animals) is represented on both retinas and in both thalami, only the left part of the scene is represented in the right cortex, and vice versa. This roughly topographic map is probably established by Hebbian mechanisms – neighboring cells in retina tend to innervate (via the thalamus) neighboring cells in cortex because (1) retinal neighbors tend to be correlated (2) excitatory short-distance lateral connections between cortical layer 4 cells ensure that neighboring cortical cells are correlated (3) therefore synapses between 2 retinocortical pairs of cells that are neighbors in retina and also neighbors in cortex tend to mutually reinforce each other (see below).

Another striking observation is that superimposed on this topographic map is an "ocular dominance" map. Layer 4 cells tend to respond either to one eye or the other, but not both – they are monocular not binocular. Ocular preference is not random – cells that prefer the same eye tend to lie together, in stripes. This can be seen both by recording electrical responses of layer 4 cells, and by injecting a radioactive marker in to one eye. The marker crosses the retinogeniculate synapses and labels the axonal endings in layer 4, and in the adult these are segregated into zebra-like stripes. Ocular dominance is strongly influenced by visual experience, but only during a "critical period" of a few days that occurs a few

weeks postnatally in kittens. If one eye is briefly deprived of patterned visual experience during this critical period, few cells will subsequently respond to that eye (and the corresponding axonal arborisations retract) . However, if both eyes are deprived, normal ocular dominance occurs, and normal terminal segregation. It appears that axons from the 2 eyes compete for cortical space and targets. Ocular dominance segregation is prevented by NMDA receptor blockers applied locally during the critical period, implicating Hebbian processes.

This process can be modeled as follows. Consider a single cortical cell receiving inputs from both left and right eyes. The strengths of these inputs are w_l and w_r respectively. Suppose that both left and right eyes view the same world, so the variances (v) of the activities of the 2 input cells will be the same. However, the 2 cells will not always be looking at exactly the same point in visual space, so their activities will not be perfectly correlated. Let us call their covariance c (which will always be positive) . The covariance matrix C for the 2 cells can thus be written

v c
c v

We have seen that under the modified Hebbian Oja rule the outcome will be that w_l,w_r evolves to lie parallel to the leading eigenvector of C, which we showed was 1,1 (or –1,-1). Thus using Oja's rule the predicted outcome is that the neuron will be perfectly binocular – not the observed outcome. The remaining eigenvector of **C** is 1,-1 (or –1,1), which, bearing in mind that geniculate axons can only excite, can be interpreted as corresponding to the monocular outcome. In order to obtain the observed outcome we must therefore destabilize the binocular outcome relative to the monocular outcome. One way to do this is to make the competition between the eyes stronger than in the Oja rule. The Oja rule ensures that the weight vector does not grow indefinitely by dividing each weight change by the average squared weight change. A much stronger form of competition is obtained by subtracting the average weight change, since this affects weak connections much more than strong ones. Subtractive competition results monocular rather than binocular outcomes. It can be shown that with subtractive weight normalization, the Hebb rule finds the eigenvectors of the matrix

v -c
-c v

These are the same as the eigenvectors of **C**, but their eigenvalues are reversed, so that that now 1,-1 is preferred.

Although strongly competitive Hebbian learning promotes monocularity, by itself it would result in a salt-and-pepper pattern in which ocular preference varies randomly from cell to cell. The observed striped pattern probably arises because, as already noted, neighboring layer 4 cells tend to excite each other via their recurrent local collaterals. (To balance this local excitation, longer range inhibition is required if the recurrent network is

not to become unstable; the thickness of the stripes is controlled by the range of the local excitation).

Obviously, to account for the observed retraction of nonpreferred axons (and the expansion of preferred eye axons) in deprivation experiments, it is necessary to assume that axon branches that supply weakened synapses eventually withdraw (and that branches providing strong synapses grow). There is evidence to support this view. This raises an interesting question about the basis of the critical period: does it arise because afterwards synapses lose their Hebbian plasticity? Or does it arise simply because there are few synapses left? Experiments suggest that the first view is correct, since LTP of thalamocortical connections seems to be restricted to the critical period.

It is easy to see that the same model can be extended to account for the formation of topographic maps from a given eye to cortex. Now the covariance matrix will be much larger, so that each retinal cell corresponds to a column and row. As the eye scans the world, each retinal cell will fire in turn, and each cell will have roughly the same variance. As previously noted, neighboring retinal cells will covary more strongly than widely separated cells, generating a banded bisymmetric covariance matrix, whose leading eigenvector corresponds to uniform innervation by all retinal cells. Again, subtractive normalization (or some other nonlinearity) is required to "break" the symmetry, resulting in a point to point representation, which is topographically organized by the lateral excitatory connections.

However, the layer 4 map is only crudely point to point, and layer 4 cells do not respond to the spots of light (or darkness) that are so effective for retinal or geniculate cells. Instead, layer 4 cells respond to oriented edges or bars located at specific positions in visual space. A given cell responds best to a bar or edge with a specific orientation, and responds much less vigorously to bars of slightly different orientation (some cells also respond to direction of movement of a bar). This is referred to as "orientation tuning". It is found that all the orientation-tuned cells that lie in a narrow vertical "column" of neocortical gray matter (running from layer 1 down to layer 6) respond to the same orientation. Cells that lie in nearby columns typically have similar but slightly different preferred orientations, and as the distance across the cortical surface the preferred orientation typically gradually changes, going through 180 degrees in about 1 mm (though there are spots where there are abrupt changes in preference). Thus 3.5 different variables are mapped onto the 2 D surface of the primary visual cortex maps: retinal position (2 dimensions) , eye origin (which is just half a dimension since it is just left or right) and local retinal orientation. Just as the 3D surface of the globe cannot be mapped onto flat paper without distortion, there are inevitably distortions in the retina- to-cortex mapping. However, these distortions do not matter because the goal of vision is not to project an accurate image onto the cortex, but for the cortex to analyse the retinal information in useful ways.
How precisely positioned does the optimally-oriented retinal bar have to be? Oriented cells fall into 2 classes, simple and complex. The simple cells respond only to precisely located bars, whereas the complex cells respond to bars that can have slightly variable locations, as long as they are correctly positioned. The simple cells are linear, in the sense

that the effects of bars of light falling within their receptive fields are just summed versions of the individual responses to small spots of light, whereas the complex cells are nonlinear (their responses are not just the sum of the responses to constituent lights).

Once again, 3 questions arise. What are the circuits that underlie orientation selectivity? How do these circuits originate? Why is the initial visual representation in striate cortex done in this way?

What are the circuits that underlie orientation selectivity?

Orientation tuning was first discovered by Hubel and Wiesel, who got a Nobel prize for this work. They proposed that simple cells receive input from a set of geniculate cells that in turn receive inputs from ganglion cells whose receptive fields are colinear, and that the simple cells sum their inputs. They proposed that complex cells receive inputs from a set of similarly-oriented simple cells that "cover" a patch of retina, and that complex cells fire when any of their simple cells inputs fire (a logical "or" operation rather than an "and" operation). In the cat, simple cells are more numerous in layer 4 (which receives most of the geniculate input) and the complex cells in layer 2 and 3 (which receive input from layer 4 cells), supporting this suggestions. Various experiments support this view, but this "feedforward" picture of orientation tuning cannot be the whole story, because the tuning is sharper than the feedforward model predicts. This has led to the view that recurrent (or feedback) circuitry (like the Hopfield network) also plays a role. There is evidence that layer 4 (and also 2/3 cells) form "lateral" excitatory connections on each other, and models show that such circuitry can "amplify" tuning. It is likely that feedforward inhibition (i.e. from geniculate afferents onto cortical inhibitory interneurons and thence to excitatory layer 4 cells) also plays a role. The issue remains controversial.

How do these circuits originate?

This issue is also controversial. There is evidence that orientation tuning is affected by visual experience, but also that there is some tuning even before the eyes open. In the latter case, this tuning could arise in 2 different ways: the circuits are genetically programmed (like glomerular "tuning" in the olfactory bulb) or are created by spontaneously occurring activity. Neural network models show that Hebb rules can lead to the development of orientation tuning.

Why is the initial visual representation in striate cortex done in this way?

Our discussion of PCA might lead one to expect that the striate cortex would reanalyze the retinal information as Principle Components rather than as locally oriented patches, since in this way it could create a much more compact description of the current visual scene. However, though the basic reasoning is correct, we saw that PCA has severe limitations: it is only optimal for Gaussian input statistics, and (because PCs are eigenvectors of the covariance matrix) it is only sensitive to pairwise statistical dependencies of input pixels. However the really interesting aspects of the visual world are contained in higher-order correlations: the fact that certain pixels tend to be bright if

certain other pixels occur in a particular combination (noses combine in particular ways with eyes and mouths, to form the higher-order regularity we call a “face”). A moment’s thought reveals that even if the image is composed of a small number of pixels, there are astronomically large numbers of ways these pixels could occur, and efficient representation requires that the cortical neurons identify which of these astronomical possibilities actually do occur with high probability.

In everyday speech we use the terms “uncorrelated” and “independent” to mean much the same thing: that there is no relationship between the observations under consideration. However, in statistics there is an important difference: correlation refers to a relationship between 2 variables, whereas “dependence” refers to a relationship between more than 2 variables. Correlation is measured in terms of the average value of the products of (zero-mean) pair of variables. However, the 2 variables could be dependent (in the sense that knowing 1 value allows one to predict the other) but uncorrelated. Independence is therefore a more restrictive (and interesting) condition that uncorrelatedness. Thus in the following figure, the uniformly scattered data points in both graphs are uncorrelated, but only in the first graph are they independent (because knowing the x value of an observation narrows the range of likelihoods of the y value – in fact if one observes an x value corresponding to a left or right corner, one knows the y value precisely).

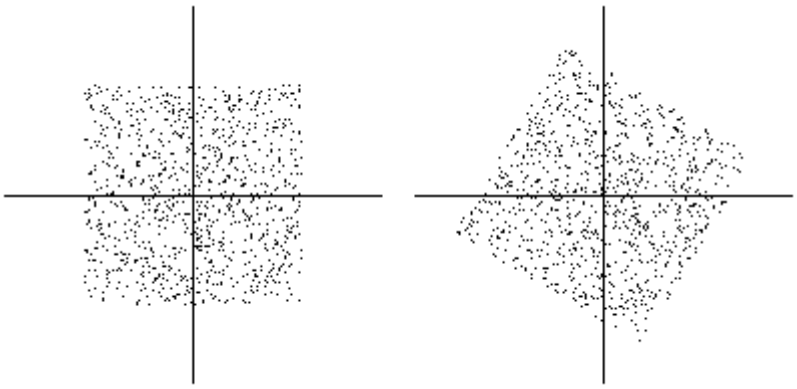

Fig: Independent (left) and dependent (right) pairs of random variables.

Of course in this case the variables show *uniform* not *Gaussian* scatter. If the variables are Gaussian, then the distinction between independence and uncorrelatedness disappears.

PCA takes correlated variables and linearly transforms them (by an appropriate rotation) to new, uncorrelated, variables . A recent powerful statistical technique called Independent Component Analysis linearly transforms dependent variables to new, independent variables. Typically, in PCA only the first few principle components are used, since almost all the variability of the observations is captured in these measurements (after all, it is a variance maximization technique). However, in ICA, one usually uses as many ICs as there are original variables (and the more one uses the more information one captures about the original measurements).

In ICA a suitable index of independence (involving products of triples etc of measurements) is made on sets of some linear combination of original measurements, and then the weights that define the linear combination are iteratively adjusted so as to maximize this index.

Another, very instructive, way to view ICA is in terms of the underlying causes of complex phenomena. If these phenomena have underlying causes, these causes should be independent of each other (since if they were not, they would be effects, rather than causes). So by seeking a description of complex data in terms of independent variables, one is likely identifying causes, and hence "understanding" the data. For example, ICA can be used to separate the voices at a crowded party, by using data obtained from several microphones placed at various points in the room. Each microphone picks up a differently weighted mixture of the underlying voices, but the different microphone signals are not independent of each other (because each microphone detects several voices). ICA unmixes the microphone signals in a way that maximizes the independence of the unmixed signals, and captures the original (presumably independent) voices.

What has this got to do with orientation tuning? It turns out that the Independent Components of "natural scenes" (i.e. typical pictures of the world) are oriented bars and edges, at various scales and positions. In other words, if one analyses typical scenes in terms of these underlying descriptors, one can accurately (and optimally) reconstruct the original scenes. Of course this only works because the scenes one typically sees are only a small subset of all possible scenes: the world is not random, but follows certain underlying "causes". At an intuitive level this corresponds to the observation that oriented edges and lines are very common in natural scenes (edges of objects, twigs, horizons etc), so it is economical to identify these edges and represent them by the activity of individual neurons rather than the collections of neurons composing the original format (the retinal representation).

It is important to realize that there are likely to be many steps in the identification of underlying "causes", and that edge analysis does not "throw away" the information contained in the original image, it merely re-represents it in a way that is more explicit. In particular if the task faced by the animal is actually identification of orientation of small patches, then the output of these edge detectors could be directly used (and indeed some of the layer 5 neurons send their axons to subcortical targets that are involved in controlling eye movements). However, most likely the task the animal faces will require identification of higher-level "causes" (such as the presence of objects of certain shape and size, or movement, or texture, or color), information which is present but only implicit in the low level "edge" description.

Because there are many more possible "lines" than there are "dots" (after all, if a line is defined as connecting 2 dots, there are n^2 possible lines in a n-pixel image), there have to be many more striate cortical neurons than there are retinal ganglion cells. This answers the obvious objection "if the goal of the cortex is to compactly represent the world, why does the "compact" neocortical description require far more neurons than the original

retinal description? (there are a 100 million layer 4 cells but only 1 million retinal ganglion cells). The answer is that at least some aspects of the scene have been made explicit in the cortical representation which were only implicit in the retinal representation.

We have seen neurons are very good at rerepresenting inputs as linear combinations, and we saw that the particular linear combination required by PCA can readily be achieved by Hebbian “learning”. How about ICA? Although we do not yet completely understand the biological mechanism, it is interesting that a Hebb rule also allows the extraction of ICs. However, during the learning process the postsynaptic neuron has to have a nonlinear input-output relation (for example a sigmoid). Since ICs are linear, this implies that the state of the neuron has to be different during “learning” and “information-processing”, which poses some interesting biological problems and possibilities.

Enormous efforts have been devoted to understanding these early steps in cortical vision, and it may seem as if the fruits have been meager, and obscure. However, if we are to understand how the human brain works, and ultimately to understand “understanding”, we have to grapple with these details, and we have to develop theories that account for these details. It might once have seemed absurd to study the stars and planets by rolling balls down ramps and dropping stones from leaning towers, but Newton and Galileo showed otherwise.