Chapter V Adventures into *terra incognita*: probing the neural
circuits along the ventral visual stream

Lesions studies had provided a compelling case that damage to
circumscribed brain regions leads to specific visual processing deficits (*Chapter IV*) and behavioral experiments had characterized many phenomenological aspects of visual perception that begged for a mechanistic explanation (*Chapter III*). The time was ripe to open the black box of the brain and begin to think about how vision emerges from the spiking activity of neurons in cortex.

Retinal ganglion neurons project to the lateral geniculate nucleus (LGN) in the thalamus and the main output projection from the LGN conveys visual information to primary visual cortex (V1) (*Chapter II*), the first stage of cortical processing for visual information. From V1, information is propagated into a large number of visual cortical areas which are responsible for transforming a pixel-like representation of sensory information into the rich and complex visual percepts (*Chapter I, Figure I-5*). The exploration and computational modeling of visual cortex is an ongoing adventure, where courageous conquistadores dare to peek inside the inner workings of the most complex system ever examined by science. Basic structural and functional principles of computation are beginning to emerge out of the sometimes seemingly enigmatic *terra incognita* of visual cortex. These basic principles are introduced in this Chapter and form the basis of the computational models of vision discussed in *Chapters VII-IX*.

### V.1. About neocortex

The neocortex is the outer layer of the neural tissue in the brain, and is thought to be responsible for cognition. The human neocortex is about 2-4 mm thick, comprises about 40% of the brain mass and contains on the order of $10^{10}$ neurons. Cortex shows a large number of folds such that it can fit about 2600 cm$^2$, approximately half a basketball court, into the size of the brain. Because of its extensive surface and relatively shallow depth, many investigators think of neocortex as a quasi-2D structure, though there are important computations that take place along the third dimension. The biggest fold is the longitudinal fissure separating the right and left hemisphere. The human neocortex has more folds than that of many other mammals; for example, the mouse cortex appears relatively smooth in comparison to the human cortex. Mechanical tension, combined with the strong constraint to save wiring and space are likely to have been important factors in determining the shape and folds of cortex {Van Essen, 1997 #193}.

To a pretty reasonable approximation, cortex is cortex: staining of cortical tissue appears at a gross level to be very similar both across different parts of the brain, and even across different species. It takes a connoisseur to distinguish a
piece of cortical tissue from a mouse from a human one. This similarity is perhaps remarkably to some people. Egomaniac or anthropomorphic considerations might lead some to think that human cortex might be substantially different; after all, mice don’t place chess or read Shakespeare. These coarse similarities suggest that approximately the same pieces of hardware can be combined in different and exciting ways to account for the cognitive capacities of different species.

Upon further scrutiny, specialists can distinguish different species by examining cortical tissue. Furthermore, it is also possible to demarcate different brain regions by examining cortex. The German neuroanatomist Korbinian Brodmann (1868-1918) devised a parcellation of the human and monkey brains – as well as many other species – based on morphological cytoarchitectonic criteria. Many parts of neocortex are still referred to by their Brodmann area number (Figure V-1) {Brodmann, 1909 #1031}. For example, primary visual cortex corresponds to Brodmann’s area 17. Biologists are rather fond of assigning names to structures and sometimes they are not content with only one name. Primary visual cortex is described in the literature as area 17, or area V1, or even striate cortex because of its distinctive stripe patterns. Subsequent physiological and lesion studies have shown that several of the structural subdivisions proposed by Brodmann and other neuroanatomists correlate with functional differences. Attempts to separate cortical regions, particularly combined with attempts to attach cognitive functions to different regions have a long and rich history that continues to current days (Finger, 2000).

V.2. Connectivity to and from primary visual cortex

Primary visual cortex is the first stage where information from the two eyes converges onto individual neurons. Each hemisphere in V1 represents the contralateral visual field. The part of the retina that is closer to the nose is called *nasal* while the other half of the retina is called *temporal*. The left visual field (left...
of where the eyes are fixating on) is represented by the nasal part of the retina on the left eye and by the temporal part of the retina on the right eye. Information from the nasal retina on the left eye will cross the brain and end up represented in the right hemisphere in primary visual cortex. Information from the temporal retina on the right eye will turn at the optic chiasm and also end up represented in the right hemisphere in primary visual cortex.

Like most other aspects of neuroanatomy, the first drawings of primary visual cortex were made by Santiago Ramon y Cajal, who was introduced in Chapter II. The basic architecture of primary visual cortex turned out to be approximately similar to that of other parts of visual neocortex. The neocortical sheet is characterized by six layers that can be distinguished in a Nissl staining, a technique used to sparsely introduce a dye into many neurons in a given area for visualization. The six layers are characterized by a stereotypical connectivity pattern that is often referred to as the canonical microcircuit. With some exceptions -- it is Biology after all -- this canonical connectivity pattern is shared across different visual areas and also across different sensory modalities.

Connections among different areas of cortex are often described as “bottom-up”, “top-down” or “horizontal” connections, a nomenclature that is also used to describe connectivity in neural network architectures (Chapter VII). These different types of connections can be defined based on the specific layer of the pre- and post-synaptic neurons. The connections between and within visual cortical areas follow a typical pattern that has been used to define what area is “upstream” or “downstream”, and which connections are bottom-up or top-down (Felleman and Van Essen, 1991) (Figure V-2). Bottom-up connections arrive at layer 4. The LGN projects to pyramidal neurons in layer 4 in primary visual cortex, perhaps the most studied layer. Layer 1 is the most superficial layer and contains mostly dendrites and few neuron bodies; the neuron bodies for those dendritic arbors are mostly located in layers 2 and 3. Top-down connections from other visual cortical areas typically end in the deep layers 5 and 6, and also to a lesser
degree in layers 2 and 3. After LGN input (or input from a “lower area”) arrives onto
layer 4, information flows from layer 4 to layers 2/3 and then onto layer 5 and layer
6. Information from layer 6 provides back projections to the LGN (or to a “lower”
visual area) and is also fed back to layer 4. An important aspect of connectivity in
visual cortex is that connections between areas are almost invariably reciprocal in
that if area A provides bottom-up input into area B, area B provides top-down inputs
to area A {Felleman, 1991 #641; Markov, 2014 #6816}.

By scrutinizing the connectivity patterns across layers in multiple brain areas,
investigators have come up with an approximate map of the anatomical paths
through which different visual areas communicate with each other (Chapter I,
Figure I-5). Based on the separation of connections into bottom-up and top-down,
it is possible to arrange the multiple different visual brain areas into an
approximately hierarchical structure {Felleman, 1991 #641}. The Felleman and
Van Essen diagram provides a semi-hierarchical description of the anatomical flow
of information in the visual system.

The more we study connectivity in visual cortex, the more we realize that this
stereotypical pattern is full of exceptions. There are differences across species,
differences between visual cortex and motor cortex, even differences between
different visual cortical areas. To make matters more complicated, these layers
can in turn be subdivided into sub-layer structures and the connectivity patterns
may be different depending on the types of neurons being considered.
Additionally, the hierarchical nature of visual cortex should not be interpreted too
strictly; for example, there are numerous “bypass” connections that send
information from an area A to area D, without going through the intermediate areas
B and C. Despite the subdivisions, exceptions, and refinements, the basic
principles of connectivity in visual cortex have played an important role in imposing
method to the apparent madness, and have inspired the development of the best
computational models that we have today.

A word of caution about nomenclature is pertinent, particularly for computer
scientists used to thinking about neural networks. Biologists talk about different
cortical areas, such as V1, V2, etc. Each of these areas has 6 layers, as described
in this section. In Chapters VII-VIII, we will discuss computational models of visual
processing, which often refer to computational steps instantiated in different
“layers”. Those computational layers should not be confounded with the cortical
layers described in here. A layer in a neural network is not necessarily directly
linked to one of the 6 layers in neocortex in any given brain area. In fact, in many
cases, people think about a layer in a neural network as potentially equivalent to a
whole brain area in cortex. We will come back to the issue of making a commitment
in the mapping between computational models and biological anatomy. For the
moment, here we refer to layers in the biological sense discussed in the previous
paragraph and in Figure V-2. In addition to information flowing from one layer to
another layer within a visual area, and also information flowing between brain
areas, there are extensive horizontal connections whereby information flows within
a layer. Some investigators use the term \textit{recurrent} connections to refer to both horizontal and top-down connections.

V.3. How to study neuronal circuits

Every problem has an appropriate scale of study, a Goldilocks scale, not too coarse, not too fine. For example, it is particularly tedious and difficult to attempt to read the newspaper using a microscope (too fine a resolution). It is also extremely challenging to read a newspaper from a distance of 200 meters away (too coarse). A plethora of methods are available to study the brain, ranging from elucidating the three-dimensional structure of specific types of ion channels, all the way to indirectly measuring signals that show some degree of correlation with blood flow, averaged over coarse spatial scales.

In the case of neocortical circuits, this Goldilocks scale is the activity of individual neurons. Studying the three-dimensional structure of each protein inside a neuron is equivalent to trying to read the newspaper with a microscope -- but it can be extremely useful for other questions such as understanding the kinetics and properties of ion channels in the neuronal membrane. Studying blood flowing through half a cubic centimeter of cortex averaged over several seconds is equivalent to attempting to extract ink tones from the newspaper from 200 meters away (but it can be extremely useful for other questions such as differentiating general properties of a part of cortex). In addition to this spatial scale, there is also a natural time scale to examine neuronal activity. Neurons communicate with each other by sending electrical signals called action potentials (Kandel et al., 2000) lasting about 2 milliseconds. For most purposes, it is sufficient to study neuronal activity at the millisecond level. With a few exceptions (e.g. small differences in timing between signals arriving at the two ears), microsecond resolution timescales do not provide additional information (one day has 1,440 minutes and therefore the analogy for studying brain at the microsecond scale would be to reread the same newspaper every minute). At the other end of the spectrum, techniques that average activity over many seconds are too coarse to elucidate cortical computations (one second is 1,000 milliseconds and therefore the analogy for studying brain at the scale of several seconds would be to average the newspaper over 2 years).

Studying the activity of neocortical circuits at neuronal resolution is not trivial. The gold standard is to examine the activity of individual neurons at millisecond resolution by inserting thin microelectrodes. Neuronal action potentials lead to changes in the electrical potential in the extracellular milieu. It is possible to amplify and measure this electrical potential in the extracellular space and measure the action potentials emitted by individual neurons. The methodology was established by Edgar Adrian (Adrian, 1926) and we already introduced example measurements of single neuron activity in the retina in Chapter II.
V.4. Neurons in primary visual cortex respond selectively to bars shown at specific orientations

Primary visual cortex consists of about 250 million neurons arranged in a 2-mm-thick sheet encompasses a few square inches in surface. There are probably more papers examining the neurophysiology of primary visual cortex than the rest of the visual cortex combined. Neurons in primary visual cortex -- as well as those in the retina and LGN (Chapter II), and also neurons in other parts of visual cortex -- show spatially restricted receptive fields, that is, they respond only to a certain part of the visual field. The ensemble of all the neurons tiles the visual field. The receptive field size of neurons in primary visual cortex is larger than the ones in the retina and LGN and can typically encompass about 0.5 to 1 degree of visual angle. A typical neurophysiology experiment often starts by determining the receptive field location of the neuron under study. After determining the location of the receptive field, a battery of stimuli is used to probe the neuron’s response preferences.

The initial and paradigm-shifting strides towards describing the neurophysiological responses in primary visual cortex were introduced by Torsten Wiesel and David Hubel (1926-2013). It is said that, the history of visual neuroscience is the history of visual stimuli. Typically, before the Hubel-Wiesel era, investigators had examined the responses in primary visual cortex using diffuse light or the type of point sources used to elicit activity in the retina and LGN. By a combination of inspiration, perspiration, and careful observation, Hubel and Wiesel realized that neurons in primary visual cortex responded most strongly when a bar of a

Figure V-3. Example responses of a neuron in monkey primary visual cortex. Physiological responses of a neuron in primary visual cortex to bars of different orientation. In these examples, the bar was moved in a direction perpendicular to its orientation. The dashed lines on the left indicate the receptive field, the black rectangle is the oriented bar and the arrows indicate the direction of motion. The neuronal response traces are shown on the right. Reproduced from (Hubel, 1968 #2321).
A particular orientation was presented within the neuron’s receptive field (Hubel, 1959 #1045) (Hubel and Wiesel, 1998). The story of how this discovery came about is particularly fascinating and recounted in David Hubel’s Nobel Lecture (Hubel, 1981 #6722). Hubel and Wiesel did not have particularly grandiose hypotheses about the function of neurons in visual cortex before they embarked on these investigations, but rather intuited that interesting principles would emerge by courageously placing electrodes in V1. After a particularly long day recording the activity of a V1 neuron, they were frustrated by how little the neuron seemed to care about the presence of a light or dark annulus inside the receptive field. In those days, they did not have computers to present stimuli; instead they used slides inserted into a projector. Their careful power of observation led them to realize that the neuron would show a burst of activity every time they inserted the slide into the projector. It was the edge of the slide moving into and out of the projector that triggered activation much more than its content. Excited by this finding, they went on to discover that the orientation of the edge mattered quite a lot, certain orientations led to much larger activation than others.

A typical pattern of responses obtained from V1 recordings is illustrated in Figure V-3. In this experiment, an oriented bar was moved within the receptive field of the neuron under study. The direction of movement was perpendicular to the bar’s orientation. Different orientations elicited drastically distinct numbers of action potentials. While the number of action potentials (or spike count) is not the only variable that can be used to define the neuronal response, it provides a simple and good starting point to examine neuronal preferences. When the bar was approximately at a -45-degree angle (panel D in Figure
V-3), the neuron emitted more spikes than for any other orientation. Moreover, the activity of this neuron was also dependent on the direction of motion. When the bar was moving towards the upper right, the neuron was vigorously active whereas there was minimal activation in the opposite direction of motion.

Hubel and Wiesel went on to characterize the properties of V1 neurons in terms of their topography, orientation preference, ocular preference, color, direction of motion, and even how those properties arise during development. Their Nobel-prize winning discovery inspired generations of neurophysiologists to examine neuronal responses throughout the visual cortex.

V.5. Complex neurons achieve tolerance to position changes

In the example shown in Figure V-3, the V1 neuron responds preferentially to a moving bar. Neurons in V1 also respond to flashes of static stimuli. How precise does the position of the oriented bar have to be within the neuron’s receptive field to trigger a response? A distinction has been observed between two types of responses in V1: simple and complex V1 neurons. The latter are less sensitive to the exact position of the bar. When using gratings containing multiple oriented bars arranged at a certain spatial frequency, complex neurons tolerate larger changes in those spatial frequencies. Simple and complex neurons are often distinguished by the ratio of the “DC” maintained response to their “AC” response elicited by a moving grating (De Valois et al., 1982). Complex neurons show a small AC/DC ratio (typically <10) whereas simple neurons have a larger AC/DC ratio (typically >10). In other words, complex neurons show a higher degree of tolerance to the exact position of a bar with the preferred orientation within the receptive field compared to a simple neuron whose response magnitude decreases when the bar is shifted away from the preferred position within the receptive field (Figure V-4). This progression from a simple neuron to a complex neuron showing increased tolerance has inspired the development of hierarchical computational models of object recognition that concatenate simple and complex-like operations as a way of keeping selectivity while achieving tolerance to transformations in the stimulus.

Some complex neurons also show “end-stopping”, meaning that their optimum stimulus includes an end within the receptive field, as opposed to very long bars whose ending is outside of the receptive field. This end-stopping phenomenon can be understood as a form of contextual modulation where the light patterns in the region surrounding the receptive field influence the responses to the stimulus inside the receptive field. Such influences from outside the receptive field are not restricted to end-stopping. V1 neurons also show surround suppression, similar to the suppressive effects of light around the receptive field center for on-center retinal ganglion neurons described in Chapter II. In sum, V1 neurons are particularly sensitive to spatial changes, to detect an edge indicative of a discontinuity in the visual field, to detect when the line stops, and complex
neurons can further detect the presence of their preferred features irrespective of small changes in position or scale.

V.6. Quantitative phenomenological description of the responses in primary visual cortex

The receptive field structure of orientation-tuned simple V1 neurons \( D(x,y) \) describing the responses at position \( x, y \) with respect to the fixation point is often mathematically characterized by a Gabor function, that is, the product of an exponential and a cosine function:

\[
D(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \cos(kx - \phi)
\]

Equation V.1

where \( \sigma_x \) and \( \sigma_y \) control the spatial spread of the receptive field, \( k \) controls the spatial frequency, and \( \phi \) the phase (Dayan and Abbott, 2001). An example illustration of a Gabor function is shown in Figure V-5. The Gabor function is characterized by an excitatory region as well as a surrounding inhibitory region.

In addition to the spatial aspects of the receptive field, it is important to characterize the temporal dynamics of responses in V1. In most cases, the spatial and temporal aspects of the receptive fields in V1 can be considered to be approximately independent or separable. The temporal aspects of the receptive field can be described by the following equation:

\[
D(\tau) = \alpha \exp(-\alpha \tau) \left[ (\alpha \tau)^5 / 5! - (\alpha \tau)^7 / 7! \right]
\]

Equation V.2

for \( \tau \geq 0 \) and 0 otherwise.

V.7. Nearby neurons show similar properties

Neurons in primary visual cortex are topographically organized, in a similar fashion to the situation described in the retina in Chapter II. The V1 topography is
inherited from the LGN: the connections from the LGN to primary visual cortex are topographically organized, meaning that nearby neurons in the LGN map onto nearby neurons in primary visual cortex. V1 neurons cover the visible visual field, with a much higher density of neurons covering the foveal region. These neurophysiological observations are consistent with the types of scotomas described for cases of V1 lesions described in Chapter III.

Another aspect of the topographical arrangement of neurons in V1 was discovered by Hubel and Wiesel by comparing the preferences of different neurons recorded during the same electrode penetration. In addition to sharing properties with their two-dimensional neighbors along the cortical sheet, neurons also share similar response patterns with their neighbors in the third dimension representing cortical depth. Advancing the electrode in a direction approximately tangential to the cortical surface, different neurons along a penetration share similar orientation tuning preferences. This observation led to the notion of a columnar structure: neurons within a column have similar preferences, neurons in adjacent columns show a continuous variation in their orientation tuning preferences. This smooth map of tuning properties onto V1 is not a necessary requirement for V1 computations. In fact, recent work has shown that this level of organization may not be a universal property. The mouse primary visual cortex does not have such a clear topographical mapping and the map of tuning preferences is described as salt-and-pepper, or at best, showing having a very weak topographical arrangement. Topography may thus be dissociated from function.

V.8. A simple model of orientation selectivity in primary visual cortex

![Figure V-6. Building orientation tuning by combining circular center-surround neurons. Schematic diagram showing how multiple LGN neurons with a circular center-surround receptive field structure can be combined to give rise to a V1 simple neuron that shows orientation tuning when those receptive field centers are properly aligned (modified from {Hubel, 1962 #1852})](image-url)
In a remarkable feat of intuition, Hubel and Wiesel proposed a simple and elegant biophysically plausible model of how orientation tuning could arise from the responses of LGN-type receptive fields (Figure V-6). The input to a V1 neuron in the model comes from LG.

In their model, multiple LGN neurons with circularly symmetric center-surround receptive fields oriented along a line were made to project and converge onto a single V1 neuron. Subsequent work gave rise to a plethora of other possible models and there is still ongoing debate about the extent to which the Hubel-Wiesel purely feed-forward model represents the only mechanism giving rise to orientation selectivity in area V1 (e.g., (Carandini et al., 2005)). Still, this simple and elegant interpretation of the origin of V1 receptive fields constitutes a remarkable example of how experimentalists can provide reasonable and profound models that account for their data. Furthermore, the basic ideas behind this model have been extended to explain the build-up of more complex neuronal preferences in other areas (e.g., (Serre et al., 2007)).

In spite of significant amounts of work investigating the neuronal properties in primary visual cortex, investigators do not agree in terms of how much still remains to be explained (Carandini et al., 2005). Biases in the recording procedures, stimuli, theories and ignorance of contextual effects and internal expectations may have an effect on the responses of neurons in V1. Yet, there has been significant progress over the last several years. Deciphering the neuronal preferences along the human ventral visual cortex is arguably one of the greatest adventures of Neuroscience.

Extending their model for orientation selectivity in simple neurons by combining the output of LGN neurons, Hubel and Wiesel proposed that the responses of a complex neuron could originate by the combination of responses from multiple simple neurons with similar orientation preferences but slightly shifted receptive fields.

V.9. Divide and conquer

Ascending through the hierarchy of cortical computations, leaving primary visual cortex, we reach the fascinating and bewildering cortical areas that bridge the pixel-percept boundary (Figure V-7). How does cortex convert pixels to percepts? Through the cascade of computations along the ventral visual stream, the brain can convert preferences for simple stimulus properties such as orientation tuning into complex features such as faces.
low-level visual features into the building blocks of perception. In primary visual cortex there are neurons that respond selectively to lines of different orientation (Hubel & Wiesel 1959, Hubel & Wiesel 1968). At the other end of the visual hierarchy, there are neurons that respond selectively to complex shapes such as faces. In between, there is a large expanse of cortex involved in the magic transformations that take oriented lines into complex shapes. How do we go from oriented lines to recognizing faces and cars and other fancy shapes Figure V-7.? Despite heroic efforts by a talented cadre of investigators to scrutinize the responses between primary visual cortex and the highest echelons of inferior temporal cortex, this part of cortex remains terra incognita in many ways. Visual information flows along the ventral visual stream from primary visual cortex into areas V2, V4, posterior and anterior parts of inferior temporal cortex. The cortical real estate between V2 and inferior temporal cortex composes a mysterious, seductive, controversial and fascinating ensemble of neurons whose functions remain unclear and are only beginning to be deciphered.

To solve the complex task of object recognition, the visual system seems to have adopted a “divide and conquer” strategy. Instead of trying to come up with a single function that will transform lines into complex shapes in one step, the computations underlying pattern recognition are implemented by a cascade of multiple approximately sequential computations. Each of these computations may be deceptively simple and yet the concatenation of such steps can lead to interesting and complex results. As a coarse analogy, consider a factory making cars. There is a long sequence of specialized areas, departments and tasks. One group of workers may be involved in receiving and ordering different parts, others may be specialized in assembling the carburetor, others in painting the exterior. The car is the result of all of these sequential and parallel steps. To understand the entire mechanistic process by which a car is made, we need to dig deeper into each of those specialized sub-steps. To understand the mechanisms orchestrating visual object recognition, we need to inspect neuronal ensembles along the ventral visual stream.

V.10. We cannot exhaustively study all possible visual stimuli

One of the challenges to investigate the function and preferences of neurons in cortex is that we have a limited amount of recording time for a given neuron. Given current techniques, it is simply impossible to examine the large
number of possible combinations of different stimuli that might drive a neuron.

Consider a simple scenario where we present a 5x5 image patch where each pixel is either black or white. There are $2^{25}$ such stimuli. If we present each stimulus for 100 ms and do not allow for any intervening time in between stimuli, it would take more than 5 weeks to present all possible combinations. There are many more possibilities if we allow each pixel to have three colors (Red, Green, Blue) with an intensity between 0 and 255. We can typically hold extracellular recordings with single (non-chronic) electrodes for a couple of hours. Recent heroic efforts have managed to track the activity of presumably the same neuron for months (Bondar et al 2009, McMahon et al 2014). Yet, even with such chronic electrodes, it is difficult to keep an animal engaged in a visual presentation task for more than a few hours a day. Thus, investigators often recur to a number of astute strategies to decide which stimuli to use to investigate the responses of cortical neurons. These strategies typically involve a combination of: (i) inspiration from previous studies (past behavior of neurons in other studies is a good predictor of how neurons will behave in a new investigation); (ii) intuitions about what might matter for neurons (for example, many investigators have argued that real world objects such as faces should be important); (iii) statistics of natural stimuli (as discussed in Chapter II, it is reasonable to assume that neuronal tuning is sculpted by exposure to images in the natural world); (iv) computational models (to be discussed in more detail in Chapters VII–IX); (v) serendipity (the role of rigorous scrutiny and systematic observation combined with luck should not be underestimated). Combining these approaches, several investigators have begun probing the neural code for visual shapes along the ventral visual cortex.

V.11. We live in the visual past: response latencies increase along the ventral stream

Figure V-8. The curse of dimensionality in vision. With current techniques, we cannot exhaustively sample all possible stimuli. Here we consider a 5x5 grid of possible binary images (top) or possible grayscale images (bottom). Even for such simple stimuli, the number of possibilities is immense.
Vision seems to be instantaneous. You open your eyes and the world is out there, apparently immediately. Visual processing is indeed very fast. Indeed, we argued in Chapter I that the speed of vision is likely to have conferred critical advantages to the first species with eyes and may well constitute one of the key reasons why evolution led to the expansion of visual capabilities. Yet, the intuition that vision is instantaneous is nothing more than an illusion. It takes time for signals to propagate through the brain. A small fraction of this time has to do with the speed of propagation of signals within a neuron, along dendrites and axons. Yet, the within neuron delays are relatively small. In particular, action potential signals within axons that are insulated by myelin can propagate with speeds of about 100 meters per second. Dendrites tend to be shorter than axons and propagation speeds within dendrites is also quite fast. The main reason why vision is far from instantaneous is the multiple computations and integration steps in each neuron combined with the synaptic hand-off of information from one neuron to the next.

At each processing stage in the visual system, it is possible to estimate the time it takes for neurons in that area to realize that a flash of light was presented. Response latencies to a stimulus flash within the receptive field of a neuron increase from ~45 ms in the LGN to ~100 ms in inferior temporal cortex (Hung et al, 2005, Schmolesky et al 1998) (Table 5.1). There is an increase in the average latency within each area from the retina to the LGN to V1, to V2, to V4, to ITC. This progression of latencies has further reinforced the notion of the ventral processing stream as an approximately hierarchical and sequential architectures. Each additional processing stage along the ventral stream adds an average of ~15 ms of computation time. It should be emphasized that these are only coarse values and there is a lot of neuron-to-neuron variability within each area. For example, an analysis in anesthetized monkeys by Schmolesky and colleagues showed latencies ranging from 30 ms all the way to 70 ms in primary visual cortex. Because of this heterogeneity, the distribution of response latencies overlap and the fastest neurons in a given area (say V2) may fire before the slowest neurons in an earlier area (say V1). The notion of sequential processing is only a coarse approximation. The response latencies constrain the number of computations required to perform computations along the visual hierarchy.

Because of these latencies, we continuously live in the past in terms of vision. The notion that we only see the past events is particularly evident when we consider distant stars. The light signals that reach the Earth have left those stars a long time ago. Although much less intuitive, the same idea applies to visual processing in the brain. Of course, the time it takes for light to bounce on a given object and reach the retina is negligible, yet signal propagation in the brain takes on the order of a hundred milliseconds as discussed above. In several cases, through learning, the brain might be able to account for these delays by predicting what will happen next. For example, how is it possible for a ping-pong player to respond to a smash? The ball may be moving at about 50 km/h (apparently, the world record is about 112 km/h) and thus traverses the ~3 m distance in about 200 ms. By the time the opponent has to hit the ball back, his or her visual cortex are
processing sensory inputs from the time when the ball was passing the net in the best-case scenario. Not to mention that to orchestrate a movement also takes time (signals need to travel from the decision centers of the brain all the way to the muscles). The only way to play ping-pong and other sports is to use the visual input combined with predictions learnt through experience. Because of these predictions, players not only capitalize on smashing speed but also recur to other strategies such as embedding the ball with spinning effects to confuse the opponent.

**V.12. Receptive field sizes increase along the ventral visual stream**

Concomitant with the prolonged latencies, as we ascend through the visual hierarchy, receptive fields become larger (Error! Reference source not found.). Receptive fields range from below one degree in the initial steps (LGN, V1) all the way to several degrees or even in some cases tens of degrees in the highest echelons of cortex (Kobatake & Tanaka 1994, Rolls 1991). Each area has a complete map of the visual field; thus, the centers of the receptive fields go from the fovea all the way to the periphery. As discussed for primary visual cortex, within each area, the size of the receptive field increases as we move farther away from the periphery. There is always better resolution in the fovea, across all visual areas. The range of receptive field sizes within an area also increases with the mean receptive field size. The distributions are relatively narrow in primary visual cortex but investigators have described a wide range of receptive field sizes in V4 or inferior temporal cortex. The scaling factor between receptive field size and eccentricity is more pronounced in V4 than in V2 and in V2 compared to V1.

**V.13. What do neurons beyond V1 prefer?**

There have been a few systematic parametric studies of the neuronal preferences in areas V2 and V4. These studies have clearly opened the doors to
investigate the complex transformations along the ventral visual stream. Despite multiple interesting studies comparing responses in V1, V2 and V4, there isn’t yet a clear unified theory of what neurons “prefer” in these higher visual areas. Of course, the term “prefer” is an anthropomorphism. Neurons do not prefer anything. They fire spikes whenever the integration of their inputs exceeds a given threshold. Investigators often speak about neuronal preferences in terms of what types of images will elicit high firing rates.

The notion that V1 neurons show a preference for orientation tuning is well established, even if this only accounts for part of the variance in V1 responses to natural stimuli (Carandini et al 2005). There is significantly less agreement as to the types of shape features that are encoded in V2 and V4. There have been several studies probing responses with stimuli that are more complex than oriented bars and less complex than everyday objects. These stimuli include sinusoidal gratings, hyperbolic gratings, polar gratings, angles formed by intersecting lines, curvatures with different properties, among others (Hegde & Van Essen 2003, Hegde & Van Essen 2007, Kobatake & Tanaka 1994, Pasupathy & Connor 2001). Simple stimuli such as Cartesian gratings can certainly drive responses in V2 and V4. As a general rule, neurons in V2 and V4 can be driven more strongly by more complex shapes. As discussed above in the context of latency, there is a wide distribution of stimulus preferences in V2 and V4.

Perhaps one of the challenges is that investigators seek an explanation of neural coding preferences in terms of colloquial English expressions such as orientation, curvature, etc. An attractive idea that is gaining momentum is the notion that neurons in these higher visual areas filter the inputs from previous stages to produce complex tuning functions that defy language-based descriptions. A neuron may be particularly activated by a patch representing complex shapes and textures that is not simply defined as an angle or a convex curve. Ultimately, the language of science is mathematics, not English or Esperanto. Neuronal tuning properties do not have to map in any direct way to a short language-based description.

V.14. Brains construct an interpretation of the world: the case of illusory contours

Another pervasive illusion is that our senses contain a veridical representation of exactly what is out there in the world. This notion can be readily debunked through the study of visual illusions (Chapter I). Let us revisit the Kanizsa triangle (Error! Reference source not found.) where we have the strong illusion of perceiving an equilateral triangle in the midst of the three pacman icons. The sides of the triangle near the vertices are composed of real black/white contours. However, the center of each side is composed of a line that does not really exist. These lines represent illusory contours, edges created without any change in luminance. It is relatively easy to “trick the eye”. Except that the eye is typically not tricked in most visual illusions. Visual illusions represent situations
where our brains construct an interpretation of the image that is different from the pixel level content. In most such illusions, retinal ganglion neurons do follow the pixel level content in the image relatively well. If we record the activity of a retinal ganglion neuron whose receptive field is right in the center of the illusory contour, nothing would happen upon flashing the Kanizsa figure. In other words, the activity of retinal ganglion neurons does not correlate with our perception. But if the retina does not reflect perception, then who does? It seems reasonable to conjecture that there must be neurons somewhere that explicitly represent the contents of our perception, in this case the illusory contours (this is a critical postulate that we will discuss again in more depth when we take up the question of the neuronal correlates of consciousness in Chapter X).

Indeed, neurons in area V2 respond vigorously to illusory contours (Error! Reference source not found.). These V2 neurons respond almost equally well to an illusory line or to a real line (Lee 2003, Lee & Nguyen 2001, von der Heydt et al 1984). The responses to illusory contours are remarkable because there is no contrast change within the neuron’s receptive field. Hence, these responses must indicate a form of context modulation that is consistent with the subjective interpretation of borders. There are also neurons in V1 that respond to illusory contours but there are more such neurons in V2.

V.15. A colorful V4

Neurons in area V4 are particularly sensitive to stimulus color (Zeki 1983). Neurons in area V4 demonstrate sensitivity to color properties that are more complex than those observed in earlier areas such as LGN parvocellular neurons or V1 blobs. Neurons in V4 have been implicated in the phenomenon of color constancy whereby an object’s color is relatively insensitive to large changes in the illumination, in contrast to the responses earlier in the visual system.
V.16. Attentional modulation

As discussed for V1, neurons along the ventral visual cortex receive massive top-down signals in addition to their bottom-up inputs (Markov et al. 2012). Presumably through these top-down signaling mechanisms, the activity of neurons along ventral visual cortex can be strongly modulated by signals beyond the specific content of their receptive fields including spatial context and higher level cognitive influences such as task goals.

A prime example of this type of modulation involves spatial attention (Desimone & Duncan 1995, Reynolds & Chelazzi 2004). Importantly, spatial attention effects can be demonstrated outside of the fixation focus. That is, a subject can be looking at one place and paying attention to another place, a phenomenon known as covert attention (as opposed to overt attention which is the more common scenario where attention is allocated to the fixated area). Through a series of astute training paradigms, investigators have been able to train animals to deploy covert attention, thus enabling them to investigate the consequences of spatial attention on neurons with receptive fields outside the fovea. Neurons typically show an enhancement in the responses when their receptive field is within the locus of attention. The magnitude of this attentional effect follows the reverse hierarchical order, being significantly stronger in area V4 compared to area V1.

V.17. Summary

• Visual computations transpire in the six-layered neocortical structure
• Cortex is characterized by stereotypical connectivity patterns from one area to the next forming approximately canonical microcircuits
• The gold standard to study cortical function is to scrutinize the activity of individual neurons
• Neurons in primary visual cortex show orientation tuning, responding more strongly to a bar in a certain orientation within the receptive field
• Complex neurons in primary visual cortex show tolerance to the exact position of the preferred stimulus within the receptive field
• The responses of neurons in primary visual cortex can be phenomenologically fitted by Gabor functions
• A model posits that V1 simple cell receptive fields can be created by adequately combining the outputs of center-surround neurons from the lateral geniculate nucleus positioned to create the desired orientation
• A model posits that V1 complex cell receptive fields can be created by adequately combining the outputs of V1 simple cells with the same orientation preferences but slightly shifted receptive fields.

• Visual cortex uses a divide and conquer strategy, subdividing visual processing into a sequence of computations in tens of different brain areas arranged into an approximate hierarchy.

• Ascending through the visual hierarchy, neurons show increased receptive field sizes, more complex tuning preferences, and longer latencies.

V.18. References


