Chapter I. Introduction to visual recognition

The greatest challenge of our times is to understand how the brain works. The conversations and maneuvers of a hundred billion neurons in our brains are responsible for our ability to interpret sensory information, to navigate, to communicate, to feel and to love, to make decisions and plans for the future, to learn. Understanding how neural circuits give rise to these functions will transform our lives: it will enable us to alleviate the ubiquitous mental health conditions that afflict millions, it will lead to building truly artificial intelligence machines that are as smart as or probably smarter than we are, and it will open the doors to finally understand who we are.

As a paradigmatic example of brain function, we will focus here on one of the most exquisite pieces of neural machinery ever evolved: the visual system. In a small fraction of a second, we can get a glimpse of an image and capture a large amount of information. For example, we can take a look at the picture in **Figure 1.1** and answer an infinite series of questions including *Who is there*, *What is there*, *Where is this place*, *What is the weather like*, *How many people are there*, *What are they doing*, *What is the relationship between people in the picture?* We can even make educated guesses about a potential narrative including answering questions such as *What happened before*, *What will happen next*. At the heart of these questions is our capacity for visual recognition and

Figure 1.1: We can visually interpret complex images at a glance
Who is there? What are they doing? What will happen next? These are among the sets of questions that we can answer after a few hundred milliseconds of exposure to a novel image.

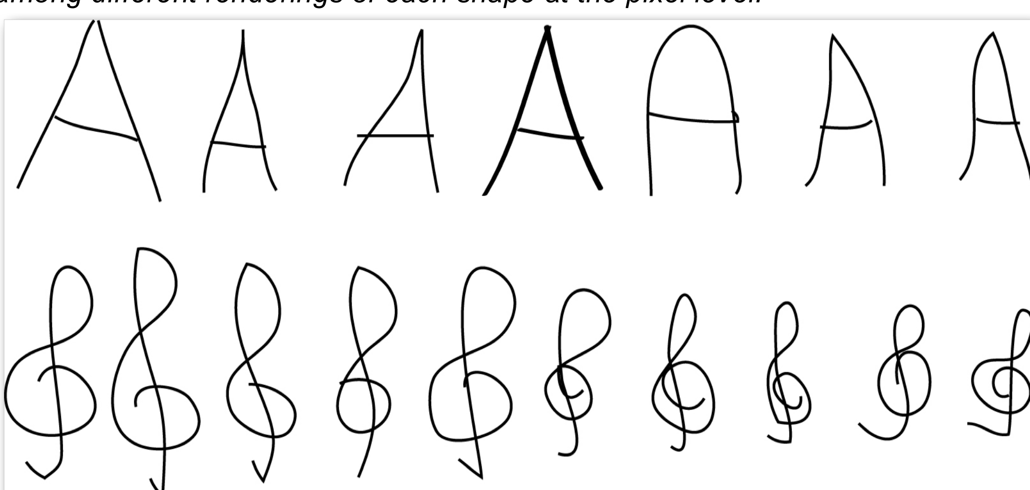
intelligent inference based on visual inputs.

Our remarkable ability to recognize complex spatiotemporal input sequences, which we can loosely ascribe to part of “common sense”, does not require us to sit down and solve complex differential equations. In fact, a 5-year old can answer most of the questions outlined above quite accurately, younger kids can answer a large fraction of them and many non-human animal species can also be trained to correctly describe many aspects of a visual scene. Furthermore, it takes only a few hundred milliseconds to deduce such profound information from an image. Even though we have computers that thrive at tasks such as solving complex differential equations, computers still fall short of human performance at answering common sense questions about an image.

1.1. Evolution of the visual system

Visual recognition is essential for most everyday tasks including navigation, reading and socialization. Reading this text involves identifying shape patterns. Driving home involves detecting pedestrians, other cars and routes. Vision is critical to recognize our friends and their emotions. It is therefore not much of a strain to conceive that the expansion of visual cortex has played a significant role in the evolution of mammals in general and primates in particular. The evolution of enhanced algorithms for recognizing patterns based on visual inputs is likely to have yielded a significant increase in adaptive value through improvement in navigation, through discrimination of friend versus foe, through differentiating food from poison, and through deciphering social interactions. In contrast to tactile inputs and, to some extent, even auditory inputs, visual signals provide information from large and far away areas. While olfactory signals can

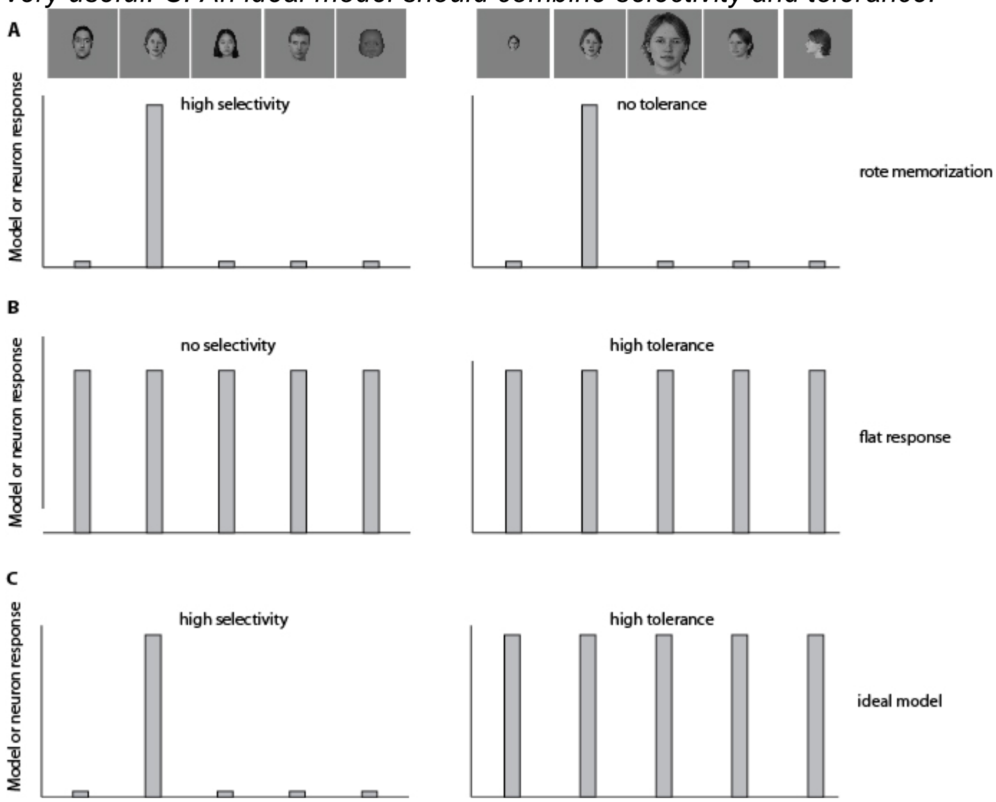
Figure 1.2: The same pattern can look very different…
Even though we can easily recognize these patterns, there is considerable variability among different renderings of each shape at the pixel level.



also propagate long distances, the speed of propagation is significantly lower than that of photons. The potential selective advantage conveyed by visual processing is so large that it has led some investigators to propose the so-called "Light switch" theory stating that the evolution of visual recognition was a key determinant in triggering the Cambrian explosion that led to a rapid renewal and expansion of the number and diversity of life on Earth (Parker, 2004).

The history and evolution of the visual system is only poorly understood and remains an interesting topic for further investigation. The future of the visual system is arguably equally fascinating. It is easier to speculate on the technological advances that will become feasible once we understand more about the neural circuitry involved in visual recognition. One may imagine that in the not-too-distant future, we may be able to build high-speed high-resolution video sensors that convey information to computers implementing sophisticated simulations of the visual cortex in real time. So-called machine vision applications may reach (or even surpass) human performance levels in multiple recognition tasks. Computers may excel in face recognition tasks to a level where an ATM machine will greet you by your name without the need of a password. Self-driving vehicles propelled by machine vision algorithms have escaped the science fiction pages and entered our streets. Computers may also be able to analyze images

Figure 1.3: A naïve approach to a model of visual recognition
A, B. *Two simple models that are easy to implement, easy to understand and not very useful.* **C**. *An ideal model should combine selectivity and tolerance.*



intelligently to search the web by image content (as opposed to image names). Doctors may rely more and more on artificial vision systems to screen and analyze clinical images. Robots may be able to navigate complex cluttered terrains. And this is only the beginning.

When debates arose about the possibility that computers could one day play competitive chess against humans, most people were skeptic. Yet, computers today can surpass even sophisticated chess aficionados. Recently, computers have also thrived in the ancient and complex game of Go. In spite of the obvious fact that most people can recognize objects much better than they can play chess or Go, visual shape recognition is actually more difficult than these games from a computational perspective. However, we may not be too far from accurate approximations where we will be able to trust "computers' eyes" as much as we trust our own eyes.

1.2. Why is vision difficult?

Why is it so difficult for computers to perform pattern recognition tasks that appear to be so simple to us? The primate visual system excels at recognizing patterns even when those patterns change radically from one instantiation to another. Consider the simple line schematics in **Figure 1.2**. It is straightforward to recognize those handwritten symbols in spite of the fact that, at the pixel level, they show considerable variation within each row. These drawings have only a few traces. The problem is far more complicated with real scenes and objects. Imagine all the possible variations of pictures taken at Piazza San Marco in Venice (**Figure 1.1**) and how the visual system can interpret them with ease. Consider the enormous variation that the visual system has to be able to cope with to recognize a tiger camouflaged in the dense jungle. Any object can cast an infinite number of projections onto the retina. These variations include changes in scale, position, viewpoint, illumination, etc. In a seemingly effortless fashion, our visual systems are able to map all of those images onto a particular object.

1.3. Four key features of visual object recognition

In order to explain how the visual system tackles the identification of complex patterns, we need to explain four key features of visual recognition: *selectivity, robustness, speed and capacity*.

Selectivity involves the ability to discriminate among shapes that are very similar at the pixel level. Examples of the exquisite selectivity of the primate visual system include face identification and reading. In both cases, the visual system can distinguish between inputs that are very close if we compare them side-by-side at the pixel level. A trivial and useless way of implementing *Selectivity* in a computational algorithm is to memorize all the pixels in the image

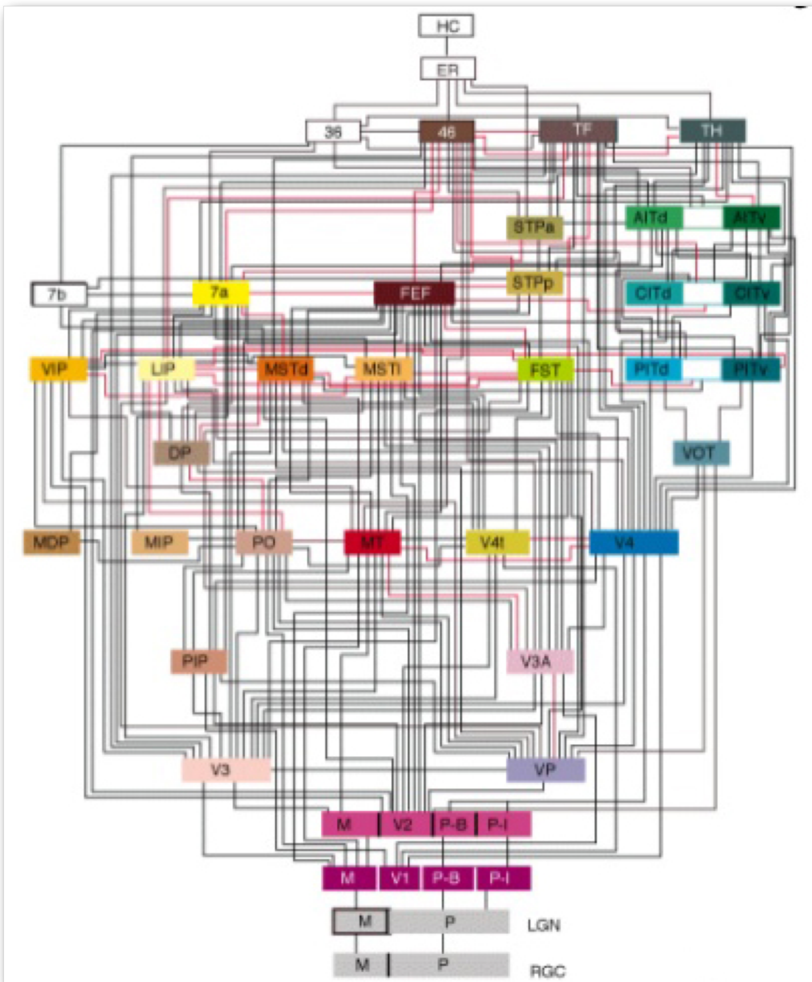
(**Figure 1.3A**). Upon encountering the exact same pixels, the computer would be able to “recognize” the image. The computer would be very selective because it would not respond to any other possible image. The problem with this implementation is that it lacks *Robustness*.

Robustness refers to the ability of recognizing an object in spite of multiple transformations of the object’s image. For example, we can recognize objects even if they are presented in a different position, scale, viewpoint, contrast, illumination, colors, etc. We can even recognize objects where the image undergoes non-rigid transformations such as the one a face goes through upon smiling. A simple and useless way of implementing robustness is to build a model that will output a flat response no matter the input. While the model would show “robustness” to image transformations, it would not show any selectivity to different shapes (**Figure 1.3B**). Combining *Selectivity* and *Robustness* (**Figure 1.3C**) is arguably the key challenge in developing computer vision algorithms.

Given the combinatorial explosion in the number of images that map onto the same “object”, one could imagine that visual recognition is a very hard task that requires many years of learning at school. Of course, this is far from the case. Well before a first grader is starting to learn the basics of addition and subtraction (rather trivial problems for computers), he is already quite proficient at visual recognition. In spite of the infinite number of possible images cast by a given object onto the retina, recognizing objects is very fast. Objects can be readily recognized in a stream of objects presented at a rate of 100 milliseconds per image (Potter and Levy, 1969) and there is behavioral evidence showing that subjects can make an eye movement to indicate the presence of a face about 200 milliseconds after showing the visual stimulus (Kirchner and Thorpe, 2006). Furthermore, both scalp as well as invasive recordings from the human brain reveal signals that can discriminate among complex objects as early as ~150 milliseconds after stimulus onset (Liu et al., 2009; Thorpe et al., 1996). The *Speed* of visual recognition constrains the number of computational steps that any theory of recognition can use to account for recognition performance. To be sure, vision does not “stop” at 150 ms. Many important visual signals arise or develop well after 150 ms. Moreover, recognition performance does improve with longer presentation times (e.g. (Serre et al., 2007)). However, a basic understanding of an image or the main objects within the image can be accomplished in ~150 ms. We denote this regime as “*rapid visual recognition*”.

One way of making progress towards combining selectivity, robustness and speed has been to focus on object-specific or category-specific algorithms. An example of this approach would be the development of algorithms for detecting cars in natural scenes by taking advantage of the idiosyncrasies of cars and the scenes in which they typically appear. Some of these specific heuristics may be extremely useful and the brain may learn to take advantage of them (e.g. if most of the image is sky blue, suggesting that the image background may represent the sky, then the prior probabilities for seeing a car would be low and

Figure 1.4: The travels of a photon.
Schematic diagram of the connectivity in the visual system (adapted from (Felleman and Van Essen, 1991)).



the prior probabilities for seeing a bird would be high). We will discuss some of the regularities in the visual world (statistics of natural images) in **Chapter 2**. Yet, in the more general scenario, our visual recognition machinery is capable of combining selectivity, robustness and speed for an enormous range of objects and images. For example, the Chinese language has over 2,000 characters. Estimations of the capacity of the human visual recognition system vary substantially across studies. Several studies cite numbers that are well over 10,000 items (e.g. (Biederman, 1987; Shepard, 1987; Standing, 1973)).

In sum, a theory of visual recognition must be able to account for the high selectivity, robustness, speed and capacity of the primate visual system. In spite of the apparent simplicity of "seeing", combining these four key features is by no means a simple task.

1.4. The travels and adeventures of a photon

We start by providing a global overview of the transformations of information carried by light to the brain signals that support visual recognition (for reviews, see (Felleman and Van Essen, 1991; Maunsell, 1995; Wandell, 1995).

Light arrives at the retina after being reflected by objects. The patterns of light impinging on our eyes is far from random and the natural image statistics of those patterns play an important role in the development and evolution of the visual system (**Chapter 2**). In the retina, light is transduced into an electrical signal by specialized photoreceptor cells. Information is processed in the retina through a cascade of computations before it is submitted to cortex. Several visual recognition models treat the retina as analogous to the pixel-by-pixel representation in a digital camera. This is a highly inaccurate description of the computational power in the retina[1]. The retina is capable of performing multiple and complex computations on the input image (**Chapter 2**). The output of the retina is conveyed to multiple areas including the superior colliculus and the suprachiasmatic nucleus. The pathway that carries information to cortex goes from the retina to a part of the thalamus called the lateral geniculate nucleus (LGN). The LGN projects to primary visual cortex, located in the back of our brains. Primary visual cortex is often referred to as V1 (**Chapter 3**). The fundamental role of primary visual cortex in visual processing and some of the basic properties of V1 were discovered through the study of the effects of bullet wounds during the First World War. Processing of information in the retina, LGN and V1 is coarsely labeled "early vision" by many researchers.

Primary visual cortex is only the first stage in the processing of visual information in cortex. Researchers have discovered tens of areas responsible for different aspects of vision (the actual number is still a matter of debate and depends on what we mean by "area"). An influential way of depicting these multiple areas and their interconnections is the diagram proposed by Felleman and Van Essen, shown in Figure 1.4 (Felleman and Van Essen, 1991). To the untrained eye, this diagram appears to show a bewildering complexity, not unlike the type of circuit diagrams typically employed by electrical engineers. In subsequent Chapters, we will delve into this diagram in more detail and discuss some of the areas and connections that play a key role in visual recognition. In spite of the apparent complexity of the neural circuitry in visual cortex, the scheme in Figure 1.4 is an oversimplification of the actual wiring diagram. First, each of the boxes in this diagram contains millions of neurons and it is well know that there are many different types of neurons. The arrangement of neurons can be described in terms of six main layers of cortex (some of which have different sublayers) and the topographical arrangement of neurons within and across layers. Second, we are still very far from characterizing all the connections in the visual system. It is likely that major surprises in neuroanatomy will come from the usage of novel tools that take advantage of the high specificity of molecular biology. Even if we did know the connectivity of every single neuron in visual

[1] As of June 2015, some computers boasted a "retinal display" of 5120 by 2880 pixels and there are commercially available digital cameras with tens of millions of Megapixels (and even more than this in professional devices). While this number may well approximate the numbers of photoreceptor cells in some retinas (~5 million cone cells and ~120 million rod cells in the human retina), the number of pixels is not the only variable to compare. Several digital cameras have more pixels than the retina but they lag behind in important properties such as luminance adaptation, motion detection, focusing, speed, etc.

cortex, this knowledge would not immediately reveal the functions or computations (but it would be immensely helpful). In contrast to electrical circuits where we understand each element and the overall function can be appreciated from the wiring diagram, many neurobiological factors make the map from structure to function a non-trivial one.

1.5. Lesion studies

One way of finding out how something works is by taking it apart, removing parts of it and re-evaluating function. This is an important way of studying the visual system as well. For this purpose, investigators typically consider the behavioral deficits that are apparent when parts of the brain are lesioned in either macaque monkey studies or through natural lesions in humans (**Chapter 5**).

An example mentioned above is given by the studies of the behavioral effects of bullet wounds during World War, which provided important information about the architecture and function of V1. In this case, subjects typically reported that there was a part of the visual field where they were essentially blind (this area is referred to as a visual *scotoma*). Ascending through the visual hierarchy, lesions may yield more specific behavioral deficits. For example, subjects who suffer from a rare but well-known condition called *prosopagnosia* typically show a significant impairment in recognizing faces.

One of the challenges in interpreting lesions in the human brain and localizing visual functions based on these studies is that these lesions often encompass large brain area and are not restricted to neuroanatomically- and neurophysiologically-defined areas. Several more controlled studies have been performed in animal models including rodents, cats and monkeys to examine the behavioral deficits that arise after lesioning specific parts of visual cortex.

Are the lesion effects specific to one sensory modality or are they multimodal? How selective are the visual impairments? Can learning effects be dissociated from representation effects? What is the neuroanatomical code? Lesion and neurological studies are discussed in **Chapter 5**.

1.6. Functions of circuits in visual cortex

The gold standard to examine function in brain circuits is to implant a microelectrode (or multiple microelectrodes) into the area of interest (Figure 1.5). These extracellular recordings allow the investigators to monitor the activity of one or a few neurons in the near vicinity of the electrode (~200 μm) at neuronal resolution and sub-millisecond temporal resolution.

Recording the activity of neurons has defined the receptive field structure (i.e., the spatiotemporal preferences) of neurons in the retina, LGN and primary visual cortex. The receptive field, loosely speaking, is defined as the area within the visual field where a neuronal response can be elicited by visual stimulation. The size of these receptive fields typically increases from the retina all the way to inferior temporal cortex. In a classical neurophysiology experiment, Hubel and Wiesel inserted a thin microwire to isolate single neuron responses in the primary visual cortex of a cat (Hubel and Wiesel, 1962). After presenting different visual stimuli, they discovered that the neuron fired vigorously when a bar of a certain orientation was presented within the neuron's receptive field. The response was significantly less strong when the bar showed a different orientation. This orientation preference constitutes a hallmark of a large fraction of the neurons in V1 (**Chapter 3**).

Recording from other parts of visual cortex, investigators have characterized neurons that show enhanced responses to stimuli moving in specific directions, neurons that prefer complex shapes such as fractal patterns or faces, neurons that are particularly sensitive to color contrasts. **Chapter 5** begins the examination of the neurophysiological responses beyond primary visual cortex. How does selectivity to complex shapes arise and what are the computational transformations that can convert the simpler receptive field structure at the level of the retina into more complex shapes?

Figure 1.5: Listening to the activity of individual neurons with a microelectrode. *Illustration of electrical recordings from microwires electrodes (adapted from Hubel).*

Rapidly ascending through the ventral visual stream, we reach inferior temporal cortex, usually labeled ITC (**Chapter 7**). ITC constitutes one of the highest echelons in the transformation of visual input, receiving direct inputs from extrastriate areas such as V2 and V4 and projecting to areas involved in memory formation (rhinal cortices and hippocampus), areas involved in processing emotional valence (amygdala) and areas involved in planning, decisions and task solving (pre-frontal cortex). As noted above, it is important to combine selectivity with robustness to object transformations. How robust are the visual responses in ITC to object transformations? How fast do neurons along the visual cortex respond to new stimuli? What is the neural code, that is, what aspects of neuronal responses better reflect the input stimuli? What are the biological circuits and mechanisms to combine selectivity and invariance?

There is much more to vision than filtering and processing images in interesting way for recognition. **Chapter 8** will present some of the interactions between recognition and important aspects of cognition including attention, perception, learning and memory.

1.7. Moving beyond correlations

Neurophysiological recordings provide a correlation between the activity of neurons (or groups of neurons) and the visual stimulus presented to the subject. Neurophysiological recordings can also provide a correlation with the subject's behavioral response (e.g. image recognized or not recognized). Yet, as often stated, correlations do not imply causation.

In addition to the lesion studies briefly mentioned above, an important tool to move beyond correlations is to use electrical stimulation in an attempt to bias the subject's behavioral performance. It is possible to inject current with the same electrodes used to record neural responses. Combined with careful psychophysical measurements, electrical stimulation can provide a glimpse at how influencing activity in a given cluster of neurons can affect behavior. In a classical study, Newsome's group recorded the activity of neurons in an area called MT, located within the dorsal part of the macaque visual cortex. As observed previously, these neurons showed strong motion direction preferences. The investigators trained the monkey to report the direction of motion of the stimulus. Once the monkeys were proficient in this task, they started introducing trials where they would perform electrical stimulation. Remarkably, they observed that electrical stimulation could bias the monkey's performance by about 10 to 20% in the preferred direction of the recorded neurons (Salzman et al., 1990).

There is also a long history of electrical stimulation studies in humans in subjects with epilepsy. Neurosurgeons need to decide on the possibility of resecting the epileptogenic tissue to treat the epilepsy. Before the resection procedure, they use electrical stimulation to examine the function of the tissue that may undergo resection. Penfield was one of the pioneers in using this technique to map neural function and described the effects of stimulating many locations and in many subjects (Penfield and Perot, 1963). Anecdotal reports provide a fascinating account of the potential behavioral output of stimulating cortex. For example, in one of many cases, a subject reported that it felt like "… being in a dance hall, like standing in the doorway, in a gymnasium…"

How specific are the effects of electrical stimulation? Under what conditions is neuronal firing causally related to perception? How many neurons and what types of neurons are activated during electrical stimulation? How do stimulation effects depend on the timing, duration and intensity of electrical stimulation? Is visual awareness better modeled by a threshold mechanism or by gradual transitions? **Chapter 9** is devoted to the effects of electrical stimulation in the macaque and human brains.

1.8. Towards a theory of visual object recognition

Ultimately, a key goal is to develop a theory of visual recognition that can explain the high levels of primate performance in rapid recognition tasks. A successful theory would be amenable for computational implementation, in which case, one could directly compare the output of the computational model against behavioral performance measures (Serre et al., 2005). A complete theory would include the information from lesion studies, neurophysiological recordings, psychophysics, electrical stimulation studies, etc. **Chapters 10-11** discuss multiple approaches to building computational models and theories of visual recognition.

In the absence of a complete understanding of the wiring circuitry, only sparse knowledge about neurophysiological responses and other limitations, it is important to ponder upon whether it is worth even thinking about theoretical efforts. My (biased) answer is that it is not only useful; it is essential to develop theories and instantiate them through computational models to enhance progress in the field. Computational models can integrate existing data across different laboratories, techniques and experimental conditions, explaining apparently disparate observations. Models can formalize knowledge and assumptions and provide a quantitative, systematic and rigorous path towards examining computations in visual cortex. A good model should be inspired by the empirical findings and should in turn be able to produce non-trivial (and hopefully experimentally-testable) predictions. These predictions can be empirically evaluated to validate, refute or expand the models.

How do we build and test computational models? How should we deal with the sparseness in knowledge and the large number of parameters often required in models? What are the approximations and abstractions that can be made? Too much simplification and we may miss the crucial aspects of the problem. Too little simplification and we may spend decades bogged down by non-essential details. Consider as a simple analogy, physicists in the pre-Newton era, discussing how to characterize the motion of an object when a force is applied. In principle, one of these scientists may think of many variables that might affect the object's motion including the object's shape, its temperature, the time of the day, the object's material, the surface where it stands, the exact position where force is applied and so on. We should perhaps be thankful for the lack of computers in that time: there was no possibility of running simulations that included all these inessential variables to understand the beauty of the linear relationship between force and acceleration. At the other extreme, oversimplification (e.g. ignoring the object's mass in this simple example) is not good either. Perhaps a central question in computational neuroscience is to achieve the right level of abstraction for each problem.

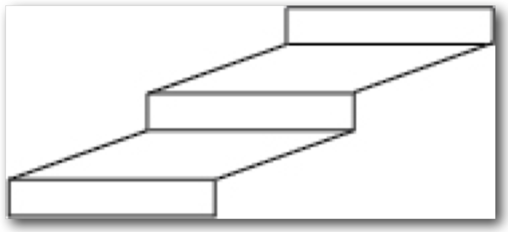
Chapter 12 will provide an overview of the state-of-the-art of computer vision approaches to visual recognition, including biologically inspired and non-biological approaches. Humans still outperform computers in mostly every recognition task but the gap between the two is closing rapidly. We trust computers to compute the square root of 2 with as many decimals as we want but we do not have yet the same level of rigor and efficacy in automatic pattern recognition. However, many real-world applications may not require that type of precision. Facebook may be content with being able to automatically label 99.9% of the faces in its database. Blind people may recognize where they are even if their mobile device can only recognize a fraction of the buildings in a given location. We will ask how well computers can detect objects, segment them and ultimately recognize them. Well within our lifetimes, we may have computers passing some basic Turing tests of visual recognition whereby you present an image and out comes a label and you have to decide whether the label was produced by a human or a(nother) machine.

1.9. Towards the neural correlates of visual consciousness

The complex cascade of interconnected processes along the visual system must give rise to our rich subjective perception of the objects and scenes around us. Most scientists would agree that subjective feelings and percepts emerge from the activity of neuronal circuits in the brain. Much less agreement can be reached as to the mechanisms responsible for subjective sensations. The "where", "when", and particularly "how" of the so-called neuronal correlates of consciousness constitutes an area of active research and passionate debates (Koch, 2005). Historically, many neuroscientists avoided research in this field as a topic too complex or too far removed from what we understood to be worth a serious investment of time and effort. In recent years, however, this has begun to change: while still very far from a solution, systematic and rigorous approaches guided by neuroscience knowledge may one day unveil the answer to one of the greatest challenges of our times.

Due to several practical reasons, the underpinnings of subjective perception have been particularly (but not exclusively) studied in the domain of vision. There have been several heroic efforts to study the neuronal correlates of visual perception using animal models (e.g. (Leopold and Logothetis, 1999; Macknik, 2006) among many others). A prevalent experimental paradigm involves dissociating the visual input from perception. For example, in multistable percepts (e.g. Figure 1.6) the same input can lead to two distinct percepts. Under these conditions, investigators ask which neuronal events correlate with the alternating subjective percepts. It has become clear that the firing of neurons in many parts of the brain may not be correlated with perception. In an arguably trivial example, activity in the retina is essential for seeing but the perceptual experience does not arise until several synapses later, when activity reaches higher stages within visual cortex. Neurophysiological, neuroanatomical and

Figure 1.6: A bistable percept. *The image can be interpreted in two different ways.*



theoretical considerations suggest that subjective perception correlates with activity occurring after primary visual cortex (Koch, 2005; Leopold and Logothetis, 1999; Macknik, 2006). Similarly investigators have suggested an upper bound in terms where in the visual hierarchy the circuits involved in subjective perception could be. Although lesions restricted to the hippocampus and frontal cortex (thought to underlie memory and association) yield severe cognitive impairments, these lesions seem to leave many aspects of visual perception largely intact. Thus, the neurophysiology and lesion studies seem to constrain the problem to the multiple stages involved in processing visual information along the ventral visual cortex. Ascending through the ventral visual cortex several neurophysiological studies suggest that there is an increase in the degree of correlation between neuronal activity and visual awareness (Koch, 2005; Leopold and Logothetis, 1999; Macknik, 2006).

How can "visual consciousness" be studied using scientific methods? Which brain areas, circuits and mechanisms could be responsible for visual consciousness? What are the functions of visual consciousness? **Chapter 13** will provide some glimpses into what is known (and what is not known) about these fascinating questions.

1.10. References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review 24, 115-147.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral cortex 1, 1-47.

Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of physiology 160, 106-154.

Kirchner, H., and Thorpe, S.J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. Vision research 46, 1762-1776.

Koch, C. (2005). The quest for consciousness, 1 edn (Los Angeles: Roberts & Company Publishers).

Leopold, D.A., and Logothetis, N.K. (1999). Multistable phenomena: changing views in perception. Trends in Cognitive Sciences 3, 254-264.

Liu, H., Agam, Y., Madsen, J.R., and Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62, 281-290.

Macknik, S. (2006). Visual masking approaches to visual awareness. Progress in Brain Research 155, 177-215.
Maunsell, J.H.R. (1995). The brain's visual world: representatioin of visual targets in cerebral cortex. Science 270, 764-769.
Parker, A. (2004). In the blink of an eye: how vision sparked the big bang of evolution (New York: Basic Books).
Penfield, W., and Perot, P. (1963). The brain's record of auditory and visual experience. A final summary and discussion. Brain : a journal of neurology 86, 595-696.
Potter, M., and Levy, E. (1969). Recognition memory for a rapid sequence of pictures. Journal of experimental psychology 81, 10-15.
Salzman, C., Britten, K., and Newsome, W. (1990). Cortical microstimulation influences perceptual judgments of motion direction. Nature 346, 174-177.
Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., and Poggio, T. (2005). A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. (Boston, MIT), pp. CBCL Paper #259/AI Memo #2005-2036.
Serre, T., Oliva, A., and Poggio, T. (2007). Feedforward theories of visual cortex account for human performance in rapid categorization. PNAS 104, 6424-6429.
Shepard, R.N. (1987). Toward a universal law of generalization for psychological science. Science 237, 1317-1323.
Standing, L. (1973). Learning 10,000 pictures. Q J Exp Psychol 25, 207-222.
Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. Nature 381, 520-522.
Wandell, B.A. (1995). Foundations of vision (Sunderland: Sinauer Associates Inc.).