Chapter 6

The Brain

The human brain is an information processing system, which can be analysed in terms of input, output, calculation and storage. At this level of abstraction the brain is like a computer. The smallest information processing components in the brain are the **neurons**. Each individual neuron can be considered to be an information processing system, with its own input, output, calculation and storage.

In between the whole brain considered as an information processing system, and individual neurons considered as information processing systems, it is possible to some extent to identify subsystems in the brain (variously known as **maps**, **functional maps** or **modules**), consisting of groups of neurons that perform a particular information processing function.

The fundamental problem of brain research is to determine how and where meaning is represented in the brain.

6.1 An Information Processing System

To understand the brain it is easiest to see it as being part of the **nervous system**. Taken as a whole, the nervous system and brain constitute a very sophisticated **information processing system**.

The functions of any information processing system can be divided roughly into four components:

- Input of information from external sources.

- Output of information to external destinations.
- Calculation: using available information to create new information.
- Storage of information, so that it can be retrieved and used again at some later time. Some information processing systems do not have any storage. Such systems can be described as **stateless**, because they do not have any **state** that represents information stored in the system. Other information processing systems have a very limited amount of state.[1]

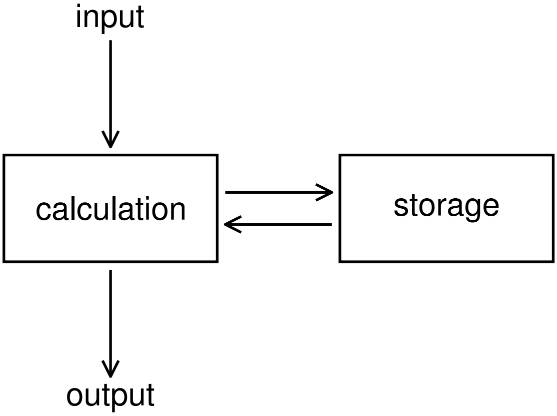

Figure 6.1. Basic components of an information processing system.

6.1.1 Analogy with Computers

Is the brain like an electronic computer? The best answer to this question is "yes" at an abstract level, but "no" when you look at the particulars. An electronic computer is certainly an information processing system. We can identify aspects of its functionality according to the list above. For example, considering the personal computer I am using to write this book, its information inputs include the keyboard, my Internet connection (when receiving data), the mouse, the microphone, and the scanner. Its information outputs

---

[1] Any type of calculation other than simply passing the input to the output requires temporary storage of the current state of the calculation, so the concepts of calculation and storage cannot be completely separated from each other. The notion of a **Turing Machine**, devised by Alan Turing to describe the capabilities of any physically plausible information processing system, divides storage into a **state machine** allowing a finite number of states, and an infinitely long **tape** which is read, written and moved along according to the operation of the state machine.

include the monitor, my Internet connection (when sending data), the speakers, and the printer. The computer performs calculations on the information it has available to it to create new information, and it can store information, either temporarily in main memory, or more permanently on the hard disk.

There are some ways that the fine details of the nervous system look like a computer: individual components are connected to each other through connections that look a bit like wires, and electrical signals of a sort travel through these connections. But we will also see many ways that the human brain is not at all like a computer. Computer components and connections between components are almost always driven by regular clock signals, whereas no such thing exists in the brain. Computers are very fragile: a failure in even one tiny hardware component can render the whole system unusable. Brains tend to be more robust than that.

Arguing about whether we can prove that the brain is or isn't like a computer is not a useful end in itself, but the analogies between the two are often illuminating. Some information processing tasks can be better performed by electronic computers, and others are better performed by the brain. Understanding the reasons for these differences in performance can help us understand why certain things in the brain happen the way they do.

6.2 The Neuron

The fundamental information processing component of the brain and nervous system appears to be the **neuron**, which is a particular type of cell[2] found in the brain and nervous system. I say "appears to be" because there is enough mystery and uncertainty about how the brain works that some scientists believe there must be more to it than just neurons and the connections between them.

Informally people often talk about "brain cells" as being the cells in our brain that do the thinking, but neurons are not the only brain cells. Other types of cell found in the brain include the **glial cells**, which are in fact more numerous than neurons. The evidence is that glial cells play a supporting role, which includes controlling ionic concentrations around neurons and recycling neurotransmitters released from synapses.

The **neuron doctrine** says that neurons are the fundamental information processing components of the brain and nervous system, and that the flow of information through the nervous system occurs via the physical connections between neurons. This is a "doctrine" in the sense of a useful working as-

---

[2] **Cells** are the basic components of all living things. Some living things, like germs, consist of only one cell. Other organisms (including us) are **multi-cellular**. Almost all cells are created by one cell splitting up into two cells (the main exception being that sometimes cells merge, like the sperm and the egg at the moment of conception). Different body tissues are formed from conglomerations of different types of cells.

sumption.[3] It is a working assumption accepted by most but not all working neuroscientists. (**Neuroscience** is the study of brains and nervous systems.)

So what do neurons look like, and how are they connected to each other? A neuron consists of a **soma**, which is its central cell body, and an **axon** and **dendrites**. The axon and dendrites are thin branching tubes that form tree-like structures coming out of the soma.

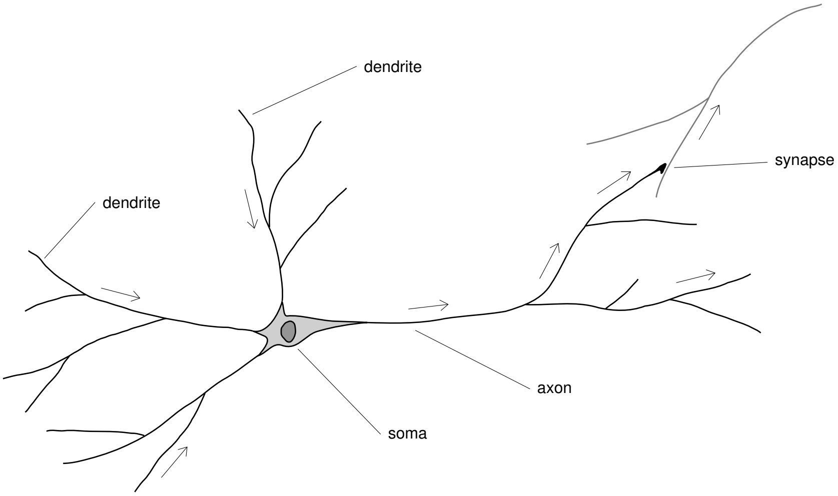

Figure 6.2. A simplified picture of a neuron, including a synaptic connection to another neuron. The arrows show the direction of the flow of information through the neurons.

The primary activity of a neuron is the generation and propagation of **action potentials** that start from the soma and propagate along the axon. The action potential is the signal that neurons use to communicate with each other. It is a type of electrical signal, but it is not a current flowing through a wire as in a computer: it is a complex transfer of sodium and potassium ions between the outside and inside of the axon. The ion transfer becomes self-propagating once initiated from the **axon hillock**, which is the point of the axon where it starts on the soma.

The branches of the axon are called **axon collaterals**. These axon branches have well defined end-points called **terminal boutons**. The boutons form connections to other neurons (and occasionally back to the same neuron). These connections are known as **synapses**.

---

[3] Other famous scientific doctrines include the **cell doctrine**, which says that living organisms are completely constructed from cells, and the **central dogma of molecular biology**, which says that DNA encodes for RNA which encodes for protein. These doctrines have turned out to have various caveats and exceptions, but they nevertheless continue to provide the major framework for understanding the phenomena that they describe.

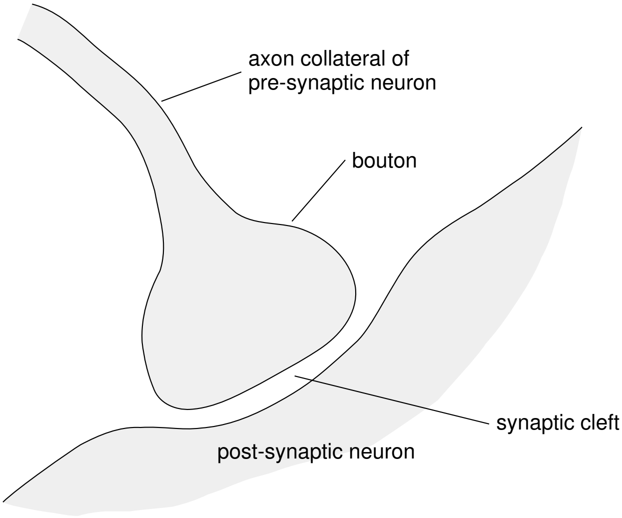

Figure 6.3. A neural synapse.

When referring to a particular synapse, the neuron that sends information into the synapse is the **pre-synaptic** neuron, and the target neuron that receives information from the synapse is the **post-synaptic** neuron.

In effect the synapse is a point of attachment, but there is actually a gap between the bouton and the post-synaptic neuron which is called the **synaptic cleft**. When an action potential arrives at a bouton, it is not transmitted as such to the post-synaptic neuron. Rather certain chemicals called **neurotransmitters** are released across the synaptic cleft. Different synapses release different types of neurotransmitter. Common neurotransmitters include **glutamate**, **GABA** (**gamma-aminobutyric acid**), **norepinephrine**, **dopamine** and **serotonin**.

On the other side of the synaptic cleft, on the post-synaptic neuron, are the **receptors**, which receive the neurotransmitter. Different neurotransmitters have different effects on the neuron that they target, but the intention is the same in all cases: the release of neurotransmitters at a synapse affects the tendency of the target neuron to propagate an action potential on its axon.

A neuron is said to **fire** when an action potential is propagated. Action potentials are an all-or-nothing affair: once one starts it continues until it reaches the terminal boutons of the axon. An action potential propagates in a manner somewhat analogous to the burning of a fuse, in that there is an advancing front of activity (which consists of an exchange of sodium and potassium ions), such that the activity at one point initiates activity at neighbouring points that have not yet been activated. A major difference between action potentials and burning fuses is that the axon is not permanently used up by the action potential: there is a gradual recharging process that makes it ready to propagate a new action potential on the next occasion.

Some neurotransmitters transmitted via a synapse make the target neuron more likely to fire; these are called **excitatory** neurotransmitters. Other **inhibitory** neurotransmitters make the target neuron less likely to fire. The terms "excitatory" and "inhibitory" are used to describe both the neurotransmitters and the synapses that transmit them.[4] The effects of neurotransmitters also depend on the locations of synapses: synapses on the central soma have a more immediate effect than synapses on distant branches of the dendrites, and some synapses on the dendrites act only to cancel out the effects of synapses further away from the soma on the same dendrite. Another thing that alters the tendency of a neuron to fire is how long it was since the last time it fired. As already stated, there is a recharging system, and the more time this has had to act since a previous firing, the more readily the neuron will fire again.

There is considerable complexity in the workings of each neuron, and scientists do not yet understand everything that goes on in individual neurons. As well as neurotransmitters transmitted across synaptic clefts, there are other neurotransmitters that leak somewhat into the surrounding medium, and act as broadcast messages that can be delivered to multiple neurons. There is also so-called **retrograde transmission** of **nitric oxide** (**NO**) from the post-synaptic neuron back to the pre-synaptic neuron that activated it. Some type of retrograde transmission of information is needed if neurons are to provide feedback about the value of information received to the neurons that sent them the information—this may be the function that NO transmission performs.

The description of a neuron I have given here illustrates the basic concept of the neuron as an information processing component. In fact we can readily identify three out of the four information processing functions:

1. The inputs of the neuron are the neurotransmitters received by receptors on its dendrites.
2. The outputs of the neuron are the action potentials propagated along its axon.
3. The calculation performed by the neuron is determined by the effect that input signals have on its tendency to fire.

How a neuron stores information is not so obvious. In the first instance, information is stored temporarily according to the neuron's firing state: whether or not it is currently firing, or if it is not firing, how much excitation would

---

[4] And the neurons, because, as it happens, many neurons primarily release one particular neurotransmitter across the synapses that they form with other neurons.

be required to make it fire. Secondly, information may be stored by changes to the long-term state of the neuron, which will mostly consist of:

- Changes in the strength of synaptic connections, i.e. how much effect an incoming signal has on the target neuron.
- Growth and formation of new connections between neurons, and the disappearance of existing connections.

6.2.1 Comparison to Computer Components

Circuits in computer components such as CPUs (central processing units) primarily process information in the form of currents flowing through wires, or voltages between pairs of points. In either case, there are generally only two states: either there is a current or voltage, or there isn't. Mathematically, these two states can be understood to represent the numbers 1 (for "on") and 0 (for "off"). Most computer circuits are driven by a regular clock signal. Thus the value of a current or voltage is determined for each interval between two clock ticks.

The smallest units of functionality within computer circuits are **logic gates** and **flip-flops**. These normally have only one or two inputs and one or two outputs. Logic gates have output values that are an immediate function of their input values.[5] Flip-flops have their output values in each clock cycle determined by their input values in the previous cycle.[6] For example, a logic gate with two inputs and one output might determine its output value according to the following logic table:

Input 1	Input 2	Output
0	0	0
1	0	1
0	1	0
1	1	0

To state this table in a sentence, the output is on (i.e. equal to 1) only if input 1 is on (i.e. equal to 1) and input 2 is off (i.e. equal to 0). If we want to use our neural terminology, we could say that input 1 is an excitatory input and input 2 is an inhibitory input. Also we note that the inhibitory effect of input 2 overrides the excitatory effect of input 1. So we can see some resemblance between the operation of a neuron and the operation of a logic gate in a computer circuit. We can even identify inputs as being either inhibitory or excitatory.

---

[5] There is necessarily some delay, and circuits must be designed so that any accumulated delays do not extend from the beginning of a clock cycle into the next clock cycle.

[6] Most types of flip-flop have their output as an implicit input, so that when a clock tick occurs, the values of the other inputs determine whether or not the current output "flip-flops" to the opposite value.

But a major difference with a neuron is that it is not controlled by clock cycles. The input signals and output signals in neural circuits are discrete events that can happen at any time. We will see that this has implications for understanding and comparing the **representation of meaning** in computers and in the brain. A set of electronic components in a computer can have one set of meanings for one clock cycle, and then have a completely different set of meanings in the next clock cycle. The lack of such a precise and global control of time periods in the brain means that the meanings represented by signals in neurons must be fairly independent of time (at least in the short term: processes of learning and cortical plasticity can cause meaning to change in the longer term).[7]

6.2.2 How Many Connections?

Another big difference between computers and brains is the number of connections between components. Neurons don't have one or two input and one or two output connections; they have *thousands* of connections to other neurons. The average is about 10,000 inputs and 10,000 outputs. Some neurons have more than 100,000 connections to other neurons.

There are about 100,000,000,000 (one hundred thousand million) neurons in the human brain. You can do the arithmetic, and see that this means there are about 1,000,000,000,000,000 synapses. (This number is so big that it has no common name, so we can just call it a thousand million million.) In some ways it might be more realistic to compare synapses (rather than whole neurons) to the individual components that occur in computer circuits.

We can compare the human brain to a personal computer, comparing numbers of components, numbers of connections and speed of operation:[8]

- 1,000,000,000,000,000 synapses in the brain compares to 100,000,000 transistors in a modern CPU, maybe 8,000,000,000 bits stored in RAM, and 1,000,000,000,000 bits stored on a typical hard disk.

- Individual components in computers do things much faster than anything in the brain: a 2GHz CPU is performing 2,000,000,000 operations per second. Very few neurons fire more than 1000 times a second, and most fire less than 100 times a second.

- Computers are terrible at making full and continuous use of their circuitry: your computer's RAM might have 8,000,000,000 bits, and operate at 500MHz, but you will be lucky if more than 128 bits of that memory are in use at any one time.

---

[7] There can be meaning in the actual timings of action potentials—this is **temporal coding** which is explained later in this chapter. The concept of temporal coding is distinct from the concept of the meaning of action potentials changing over time.

[8] The figures given are all very approximate, and the values for computers change as computer technology develops.

A lot of design effort has gone into making the CPU do at least a few things in parallel, but it only contains a small fraction of the overall number of components in the system.[9] Neurons in your brain do not have to wait for some central authority to tell them to do something—each neuron reacts directly at all times to the inputs of the neurons immediately connected to it.

- A computer's hard disk retains information even when the power is turned off. Compared to RAM, hard disks are usually larger but slower, and the processing bottleneck is even more extreme: a typical hard disk might store 1,000,000,000,000 bits of information, stored on several **platters**, with two heads per platter, each head capable of transferring just *one* bit at a time at a rate of perhaps 100,000,000 bits per second.

These differences are revealed in the different abilities of human and computer information processing systems: all the different parts of your brain can operate simultaneously to calculate the relevant consequences of information made available to it, whereas a computer has to work its way through all the potential deductions and conclusions sequentially. On the other hand, if there is a need to multiply a million numbers together—and be sure of getting exactly the right answer—the computer is going to finish the job a whole lot quicker.

6.3 Modularity in the Brain

We can identify the four information processing components—input, output, calculation and storage—for the brain and nervous system as a whole:

- Information is input from sensory cells, also called **receptors**. There are sensory cells that supply the input for the traditional five senses, and also for various internal senses:
    - Sight: neurons in the retina that respond to light. There are four types of retinal receptors: three colour-sensitive types and one "black and white" receptor type for low light conditions.
    - Hearing: cells that receive sound information. These are the hair cells in the **organ of Corti**, which have already been mentioned in the previous chapter.
    - Taste: receptors in the tongue for sweetness, saltiness, sourness, bitterness and "umami".
    - Smell: olfactory receptors in the nose.

---

[9] This problem of not being able to use more than a small portion of the computer's circuitry at any one time is called the **Von Neumann bottleneck**, named after John von Neumann, a famous physicist, mathematician and computer scientist.

  - Touch: various receptors in the skin that detect pressure and temperature.
  - Internal senses include receptors for balance, position and motion of various parts of your body, and other receptors that provide information about the internal state of bodily organs.

- The major output of information from the nervous system is via the **lower motor neurons**. Each motor neuron activates a single muscle fibre. There are two types of lower motor neuron: **alpha motor neurons** activate **extrafusal** muscle fibres which do the real work, and **gamma motor neurons** activate **intrafusal** muscle fibres which play a role in managing feedback to the nervous system about the contracted state of muscles. Other outputs occur via the **autonomic nervous system** which controls such things as heart rate, blood pressure, digestion and the release of various hormones.

- The brain stores information: this is what we call "learning" and "memory".

- The brain calculates: this includes all the processes of perception, where raw sensory information is translated into knowledge and understanding of things in the external world and within ourselves, and the processes of decision-making, which eventually result in us making or controlling muscle movements required to carry out those decisions.

We can analyse the neuron as an information processing system, and we can analyse the whole brain as an information processing system. In both cases we can identify the four components of input, output, calculation and storage. Are there any in-between levels of organisation and functionality that we can analyse?

When we look at an electronic computer system, we can see that it consists of various circuit boards plugged together, and each circuit board consists of integrated chips and other electronic components that have been soldered onto the board and connected by etched connections on the board. There is a lot of modularity in how computer systems are constructed. This partly has to do with the economics of design and manufacture: it is easier to design systems constructed from general purpose components that have already been designed, and it is easier to make profits from manufacturing general purpose components because they can be used in many different systems.

The "economics" of the design and manufacture of the human brain and nervous system is a bit different from that of electronic computers. The "design" has resulted from an accumulation of incremental mutations over millions of years of evolution. The "manufacture" is the process of conception, growth and development. These processes of natural design and manufacture may result in a form of biological modularity, but it is not clear if it is a form of modularity that it going to help us analyse the brain into functional components.

When man-made information processing artefacts are made from components, the components are generally manufactured separately, and then attached to each other by various means to make the final product. It is often easy to pull such an artefact apart into its separate components, especially if we are armed with a screwdriver, or perhaps with a soldering iron that lets us remove components from a circuit board. If a component is general purpose, then it will have a well-defined functionality independent of its role in that particular artefact, and it will be easy to understand that functionality by analysing the design of that individual component.

The "components" of the body of a living organism have to grow and develop **in-place**, i.e. connected as they are to all the other components of the body. And they are also constrained to *evolve* in-place. For example, in all the history of the evolution of lungs and hearts, at no point were the lungs and hearts ever disconnected from each other. Because there is no "assembly" stage in its manufacture, it is not so easy to disassemble the components of a living organism. The boundaries between biological components are not always as sharply defined as in a man-made artefact.

These differences between man-made and biologically-made are most acute when looking at the brain. In a modern computer, the component with the most connections to other components is the CPU, and yet the number of pins on even the latest CPU is no more than a few hundred. Each of these pins has a specific function that is determined at the time the CPU is designed, somewhat independently of the design of any particular computer system that is going to include that CPU.

The design of the human brain (and that of other animals) favours as many connections as possible between components, in as much as components can be identified at all. The functionality of connections between different brain areas is partly genetically pre-programmed and partly determined by the processes of growth, development and learning. The larger scale components of the human brain are not plug-in modules as such; rather they are different areas of functional localisation. For example, the colour-processing component of the brain is an area that contains neurons whose firing is a function of perceived colour, such that processing of colour appears to depend strongly on the presence of that area. And there will be millions of connections between that component and other components that provide its inputs and process its outputs.

This high level of interconnectedness implies that it is not going to be so easy to analyse the brain as an information processing system by breaking it up into a moderate number of smaller information processing components.

6.3.1 The Representation of Meaning

The analysis of signals and components in any information processing system should ultimately result in an understanding of how *meaning is represented in that system.*

Here is a very simple example: a thermostat, as shown in Figure 6.4. In this particular example, the thermostat consists of several components, to make the flow of information more explicit:

- A thermometer, which measures temperature and outputs a signal representing the current temperature.
- A target temperature unit (presumably set by the user), which outputs a value representing that temperature.
- A "comparison" unit, which receives as input the values output from the thermometer and the target temperature unit, and which outputs a signal if the measured temperature from the thermometer is less than the target temperature.
- A relay, which receives the signal from the comparison unit, and switches on when it receives a signal, and switches off when it receives no signal. The relay switch controls a heating circuit which includes a power source and a heating element.

The aim of our analysis is to understand the meaning of the signal travelling from the comparison unit to the relay. In fact there are two meanings, one from the point of view of the comparison unit, and one from the point of view of the relay:

1. Coming out of the comparison unit, the signal means "the temperature is too cold".
2. Going into the relay, the signal means "turn the heater on".

Given these two meanings, we can also assign a meaning to the connection between the comparison unit and the relay:

> "If the temperature is too cold, turn the heater on."

That was an exhaustive analysis of the meaning of just *one* signal travelling through a connection between two active components in a very simple information processing system. We would like to do a similar analysis for every neuron and every synapse in the brain and nervous system. Given the way that the brain works, there are two types of question to ask:

1. For each neuron, what does it mean when it fires?

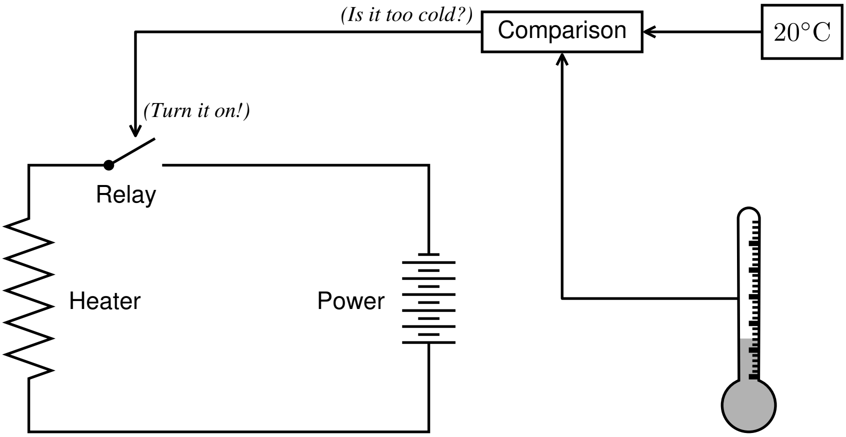

Figure 6.4. Analysis of meaning in a thermostat. The temperature measured in the thermometer is compared to the target temperature of 20°C. If the temperature is too low, a signal is sent to the heating circuit causing the heating circuit to switch on. We can give two interpretations of the signal going from the comparison unit to the relay: coming out of the comparison unit it means "the temperature is too cold", and going into the relay it means "turn the heater on".

2. For each synapse, what is the meaning of the connection between the pre-synaptic neuron and the post-synaptic neuron?

As mentioned above, neurons are also affected by neurotransmitters that are transmitted in a more non-specific manner, and by retrograde messengers like NO. So we can also ask about the meaning of those signals.

The first two types of question cover a lot of ground in themselves. In fact the first question is actually 100,000,000,000 questions, one for each neuron, and the second question is 1,000,000,000,000,000,000 questions, one for each synapse. That's a lot of questions!

And it could get worse. It could be that those are not even the right questions to ask. It may be that we cannot hope to understand the representation of meaning in the brain just by learning the meaning of each neuron and of each synaptic connection between neurons.

It may be that the firing of one neuron has a meaning highly dependent on the firing of other neurons. The meaning may also depend on the relationships between the times at which those other neurons fire, and there may be a complex dependence between the meaning of a neuron firing and the immediate past history of that neuron's own firings.

And when we look at the meaning of a synapse, it may not be sufficient to consider it as just a relationship between the meaning of the pre-synaptic neuron and the meaning of the post-synaptic neuron. I mentioned earlier that the effects of synapses on the post-synaptic neuron can depend on the relationship between the positions of synapses on the dendritic tree. So we have to take into account the locations of synapses on the dendritic tree of the post-synaptic (target) neuron, and separately assign a meaning to the state of each portion of the dendritic tree, and relate the meaning of each synaptic connection to the states of the dendrite on each side of that synapse.

6.3.2 Temporal Coding

The question of how easily we can specify the meaning of a neuron's firing independently of the context of its previous firings relates to the theory of **temporal coding**. Temporal coding refers to the idea that information is encoded in the precise sequence of timings of action potentials in a neuron. It raises the bar on the difficulty of analysing the effects of all the connections between neurons, because for each synapse we must take into account the relationship between the firing times of the pre-synaptic neuron and the firing times of the post-synaptic neuron.

There is one particular type of temporal coding which does occur in the ear, the auditory nerve and auditory processing areas in the brain, which we might call **direct temporal coding**—"direct" because there is a direct relationship between the times of neural firings and the times of the events encoded by those neural firings. (In practice it's just called "temporal coding", as the possibility of temporal coding which represents information symbolically *without* any direct physical relationship to the original information is somewhat more hypothetical.)

This form of temporal coding starts in the ear, in the organ of Corti, where neurons responding to frequencies from 20Hz to about 4000Hz are **phase-locked**. This means that the firings of neurons in a group representing the same (or similar) frequency are locked in phase with the frequency of the original sound. In fact, for the lower range of frequencies, from 20Hz up to about 200Hz, information about frequency is *only* encoded temporally, as location on the basilar membrane does not distinguish between different frequencies in this range. For higher frequencies the frequency being represented is actually faster than the rate at which neurons can fire. This difficulty is solved by having multiple neurons represent the high frequency signal, according to what is known as the **volley principle**, whereby different subsets of neurons within a group of neurons fire signals for each frequency cycle.

We can understand that lower frequencies have to be temporally encoded, because the mechanics of the ear do not allow them to be positionally encoded. Assuming that "higher-level" processing requires positional encoding at some point, we would presume that temporal encoding gets converted to positional encoding somewhere in the auditory cortex, although it appears that current scientific understanding of this part of the brain is not sufficient to say with any certainty where or how (or even if) this actually occurs.

But if frequencies from 200Hz to 4000Hz are positionally encoded, why do they also need to be temporally encoded? A general answer is that the brain will represent information in as many different ways as possible that can help it to process that information. Temporal coding disappears above 4000Hz, because it is not worth the effort for the brain to maintain the quantity and accuracy of neural firings required to represent sounds at those frequencies temporally. A more specific answer is that the perception of the direction of lower frequency sounds depends on the perception of differences between times at which events are perceived in the left ear and the right ear. Temporal coding of sounds at these frequencies must be maintained at least as far as an area called the **superior olive**, where there are neurons that can compare the times of signals received from both ears. (And for higher frequency sounds, the brain uses relative intensities between left and right to determine direction—a secondary problem for determining direction from temporally coded high frequency sounds is that of knowing exactly which individual vibration perceived by the left ear corresponds to which individual vibration perceived by the right ear.)

One might suppose that the perception of music depends on temporal coding. In particular the harmonic relationships between frequencies related by simple integer ratios would give rise to corresponding relationships between neural firings in response to those frequencies. For example, if neuron A is responding to a frequency at 100Hz by firing 100 times a second, and neuron B is responding to a frequency at 200Hz by firing 200 times a second, then there will be exactly 2 firings of B for every 1 firing of A.

Despite this possibility, the theory of music perception developed in this book mostly ignores temporal coding, and indeed defines a general principle of musicality which is entirely a function of unchanging (or not very often changing) *spatial* patterns of activity in cortical maps that process musical information.

Apart from direct temporal encoding of sounds by phase-locked neurons, there are other basic types of temporal encoding that have been recognised as occurring in the brain. The first is simply that many neurons encode information entirely in terms of *frequency* of firing, i.e. frequent firing means that there is a lot of something, and less frequent firing means there is less of it.

The second type of temporal encoding gives meaning to the phase relationships between the firing of different neurons, and one theory supposes that different neurons fire in **synchrony** with each other (i.e. the same phase and frequency), if they are *referring* to information about the same entity. This theory is an attempt to solve the **binding problem** (discussed in more detail in the next section).

There is necessarily some conflict between different types of temporal coding. For example, neurons phase-locked to different frequencies cannot represent a relationship by firing in synchrony, because synchrony would require them to match frequency. And the volley principle mentioned above can be seen as a way to allow phase-locking of a group of neurons representing frequency to coexist with frequency encoding within individual neurons of information about intensity.

6.3.3 Localisation and Functional Maps

Despite the possibility that the massive interconnectedness of neurons makes it impossible to understand how meaning is represented as neural activity in the brain, there are some grounds for optimism that naïve **reductionist** theories may be sufficient for us to understand how the brain works. In particular we hope to *reduce* the problem of understanding how the brain works to the simpler problem of understanding how individual neurons or groups of functionally similar neurons represent and process information:

- The relationship between meaning and neural activity is straightforward enough when we look at the periphery: we can directly describe the meaning of sensory cells in terms of the sensory input that they respond to. For example, the meaning of a retinal receptor firing is that a particle of light has landed on that receptor, most likely of a frequency which the receptor is sensitive to. Similarly, we can understand the relationship between meaning and activity for motor neurons: the meaning of a signal propagated along the axon of a motor neuron is "contract this muscle fibre".

- The history of medical neurology consists mainly of a list of ailments of the mind associated with damage to specific areas of the brain. For example, damage to specific areas of the brain results in specific deficiencies in language: damage to one area reduces fluency, damage to another area reduces comprehension.[10] Damage to areas relating to visual perception results in specific deficiencies in vision, such as inability to perceive motion, or inability to perceive colour. The associations between functional deficiencies and localised damage suggest very strongly that specific functionalities and representations of specific meanings are implemented in specific areas of the brain.

- Work on animals has shown that many neurons can be identified whose firing is a function of specific perceptions that the animal must be making in response to its environment. For example, by recording activity of individual neurons, scientists can do more than observe that one area processes colour—they can determine that each neuron in that area is maximally activated by a colour that is particular to that neuron.

---

[10] It is possible to be fluent without having comprehension. In such cases of **fluent aphasia**, patients speak quickly and easily, but the content of their speech tends towards meaningless nonsense.

A **cortical map** is an area of the **cerebral cortex** where neurons are specialised to perform some particular type of processing, and where there is some identifiable relationship between the position of a neuron in the map and its meaning. The cerebral cortex is the area of the brain which is most developed in mammals and in particular in humans, and it carries out most higher functions. The "map" concept can also apply to areas of the brain outside the cortex, and a general term is **functional map**, or sometimes just **map**. The cortex does, however, have a specific structure that is somewhat flat. The human cortex can be spread out to cover an area of about 0.2 square metres, it is approximately 2mm thick, and it contains 4 to 6 layers of neurons (the layers are fairly fuzzy—one cannot necessarily assign each neuron to a precise layer numbered from 1 to 6). So a cortical map is physically not unlike a real map on a sheet of paper.

In principle the physical position of a neuron has no particular meaning, because what matters is how neurons are connected to each other. However, meanings of signals from some types of sensory cells are necessarily position-dependent: the position of a retinal receptor relates to its position in the image projected onto the retina, the position of a receptor for touch is its actual position on the body, and the position of a receptor for sound in the organ of Corti is a function of frequency. Furthermore, these positional relationships are often preserved by the way that connections are formed travelling from one part of the brain to another. We may presume that the development of the nervous system and brain has evolved in a manner that uses these positional relationships to organise the brain in a way that enables effective processing and use of information from sensory sources.

When scientists look at the **auditory cortex**, which is that part of the cortex that processes sound information, they find many maps that are **tonotopic**, which means that one dimension of the map is correlated with frequency of harmonic components of perceived sound, or with pitch of perceived sound (which more or less corresponds to the frequency of its lowest harmonic). In later chapters, as I present my theory of music, we will have reason to speculate on the existence and purpose of a number of distinct tonotopic cortical maps, and on the relationship between perceived musicality and the patterns of neural activity in those maps.

6.4 Separation and Binding

One general theme that has emerged as scientists have analysed functional localisation in the cortex and elsewhere in the brain is that of *separate processing of different aspects of perception.*

The most studied area of perception is that of vision. Experimenters have used monkeys and other animals to investigate the relationship between brain activity and perceptual functions. Although experimentation on animals is an ongoing ethical controversy, you can get away with inserting probes into monkeys that you couldn't insert into the brains of human subjects.[11] As it happens, there isn't a whole lot of difference between the visual capabilities of monkeys and those of ourselves, and most of our visual capabilities have evolved from the need to be able to climb and jump through the treetops without falling off and getting hurt.[12]

What scientists have found is that different areas of the visual cortex are specialised for different aspects of visual processing. For example, there are areas that specialise in perception of shape, and others that specialise in perception of motion, and yet others that perceive colour. There are about 30 distinct visual processing areas that have been identified in the monkey and/or human brain.[13]

6.4.1 Colour Perception

To give a specific example, there is a colour perception cortical map that encodes information about the colour of an object. The definitive book on this subject is Semir Zeki's *A Vision of the Brain*. This book is specifically about perception of colour, but its underlying themes are cortical mapping and functional localisation.

The colour of an object is quite distinct from the colour of light reflected from that object. The colour of light reflected from an object is a function of both the colour of light falling on the object, and the colour of the object itself. From an information processing point of view, the information about the colour of light is the input, and the information about the colour of the object is the output. One reason it took scientists a while to discover the difference between these two notions of colour is that our colour perception system is so good that we can reliably identify reflective colour of objects under quite extreme variations of lighting conditions. It is so good that we don't realise how good it is: we just take it for granted that we see the colours of objects.

A similar situation occurs with **pitch translation invariance**—our *inability* to perceive absolute pitch when we listen to music (which is analysed in detail in Chapter 9 on symmetries). We take it for granted that a tune sounds much the same if we transpose it into a different key, but actually there must exist a large amount of sophisticated machinery in the brain to convert the raw incoming information into the desired invariant perception.

---

[11] There are occasions where, for the purposes of planning brain surgery, it is necessary to map the functionality of a patient's brain by means of electrode stimulation, so as to discover which portions are acceptable to remove, and which parts should be left alone. Such exploration can provide useful scientific data about the localisation of function in the human brain.

[12] Even though it has been millions of years since our ancestors ceased to be full-time tree dwellers.

[13] In *The Astonishing Hypothesis* (page 149), neuroscientist Francis Crick mentions 20 visual maps and 7 partly visual maps, and suggests that at least one of the visual maps will turn out to be several distinct maps.

And once we realise how much machinery there is performing this task, we will be led to ask ourselves what the purpose of this calculation is, because it must be something important if so many resources are devoted to it. In the case of colour perception, it is important to determine the actual colours of objects, both to identify them reliably, and to determine their properties. To give a simple example, if we are looking for ripe fruit on a fruit tree, we select which fruit to pick and eat based on the colour of the fruit.

6.4.2 The Binding Problem

Returning to the issue of separation, what we find with visual processing is that the earlier stages of visual processing encode information in maps that encode for all aspects of vision, including both position and colour. But as information proceeds to so-called "higher" processing areas, the cortical maps separate out the different aspects of that information. Thus the colour map encodes for colour, *almost without regard to position*, and other maps encode for position *without encoding any information about colour*. This seems a little paradoxical: surely in any scene we see different objects of different colours in different positions, so how does the brain properly track the connections between these aspects of colour and position?

The paradox would perhaps not exist if the brain only processed information about one thing at a time: that one thing would have a certain colour and a certain position, its colour would be encoded by the colour cortical map, its position would be encoded by the position cortical map, and that would be that.

But there are many situations where we perceive (and respond to) multiple characteristics of multiple objects. So neuroscientists are stuck with the problem of how (and where) we actually "see" a scene in which all the aspects of vision, including position, motion and colour, are correctly combined in different parts of the scene. This problem is known as the **binding problem**, referring to the need to "bind together" the different aspects of perception that have been separated.

To give a specific example, we might see a red ball in one position going up and a green ball in another position going down. There will be neurons active in two positions in the cortical map representing positional information, and neurons active in two positions in the cortical map representing motion (one group for "up" and another for "down"), and neurons active in two positions in the cortical map representing colour (one group for "red" and another for group for "green"). How do we know that actually the red ball is going up and the green ball is going down, and not vice versa?

The problem is not just one of *how* binding occurs, but also *where*. The changing retinal images encode information that will be used to calculate all the different aspects of visual perception such as colour and motion. As this information is processed, the different aspects are processed separately in different areas, and there does not seem to be any area where they are *joined* *back together*. In as much as our conscious visual perception (or "seeing") must combine these different aspects, it is apparently distributed in some mysterious manner across different parts of the brain.

Ultimately our high-level perceptions must be made accessible to those parts of the brain that think about the world and make decisions about what to do. For example, if we are playing a game with different balls, and we know that the red ball is the one we need to catch, we need to be able to move appropriately towards the red ball, in response to its perceived position and direction of movement.

Some scientists have felt the binding problem to be so difficult that they have been motivated to provide rather esoteric explanations of how the brain does the binding. For example, quantum mechanical correlations have been invoked to explain the mystery binding. This hypothesis has been advanced by Roger Penrose (a theoretical physicist) and Stuart Hameroff (a professor of anaesthesiology). Most scientists find this combination of quantum mechanics and neuroscience somewhat implausible and perhaps unnecessary. It doesn't help that the quantum components of Penrose's theory depend on as yet undiscovered theories of quantum gravity.

There are two possible solutions to the binding problem that are both simpler and less esoteric than quantum consciousness:

- The first is that different aspects of information are never completely separated: for example, cortical maps encoding for colour still weakly encode for positional information. This weak encoding may be sufficient to enable re-assembly of information in some manner.

- Second is the theory of **synchronous firing**. This says that neurons whose firing is associated with the same object are bound together by firing synchronously (i.e. all at the same time and in phase with each other). So the neurons representing the direction "up" will fire synchronously with the neurons representing the colour "red", and the neurons representing the direction "down" will fire synchronously with the neurons representing the colour "green". The presumption is that information is generally encoded by a neuron as a rate of firing, without regard to particular timing, and that there is therefore the freedom to choose specific firing times in relationship to firing times of other neurons, in order to specify binding. The phase of neural firing can be changed without altering the overall firing rate, and therefore without altering the information value encoded by that neuron. There is an intrinsic plausibility to this theory: if two neurons A and B have inputs to C, and if the activation of C is stronger when its inputs come in repeatedly at almost exactly the same time, then neuron C will be more strongly activated if its inputs A and B are synchronous. Thus C will be activated more strongly by A and B if A and B are referring to the same object, and if this happens then C will also be referring to that object. Synchronised neural firing is observed experimentally to occur, and there is some evidence that it occurs in relation to aspects of a stimulus that either are or need to be bound together.

The concept of separation of aspects would appear relevant to the development of a theory of music perception. For example, following the analogy of how the visual system separates processing of different aspects of vision such as location, colour and motion, we might reasonably expect that the auditory system separates the processing of pitch relationships and temporal relationships. And we would expect that the results of these separated aspects of processing are combined back together again to provide the final conscious percept.

It follows therefore that we should consider the binding problem when analysing how the human brain processes music. On the other hand, whatever solution the binding problem has, it is probably going to be the same solution for all different types of perception, whether visual or aural or anything else. So when the theory requires me to state that certain perceptions are bound together, I am quite happy to state that I don't know for sure how the binding happens, but I know that binding has to happen somehow, and the same "somehow" is how it happens in the case of music.[14]

6.5 Population Encoding

There is another complication in the representation of meaning in cortical maps. As a simplification, we could consider a cortical map which was effectively a one-dimensional map, and which responded to one numerical aspect of a stimulus, for example the frequency of a sound.

The encoded value comes from a continuous range of values: it could be any real number between 20 and 20000 (representing frequency in Hz). But the set of neurons in the cortical map is finite. If we assign a particular frequency to each neuron, then only a discrete number of frequencies can be represented by the map. Some ad-hoc mechanism would be required to deal with the in-between frequencies; for example, we could round to the nearest value that had a representation.

There are a number of reasons why such a simple representation of meaning will not be satisfactory:

- If we consider sensory neurons, it is very unlikely that a neuron is going to have a sharp cutoff in what it responds to, in such a way that there is no overlap between what different neurons in a map respond to.

---

[14] Although there is the difficulty, as previously mentioned, that if different auditory neurons are phase-locked, then whether or not they can or do fire in synchrony is dependent on the relationships between the frequencies that they are firing at.

- If a particular neuron gets damaged or lost, the values it represents will cease to be represented at all.
- If only one signal appears, or only one signal appears within a certain portion of the cortical map, then only one out of all the neurons in that portion will be active, which seems to be a waste of information processing capacity.

Population encoding is a manner in which neurons in a cortical map encode numerical values. Very simply, we can say that for each neuron, and for each possible signal value, the rate of firing of the neuron is a function of the encoded value. Each neuron has its **encoding function**.

This method of encoding would be equivalent to the first method of encoding that we described, *if* the encoding function for each neuron was equal to a maximum value for all the values in the range that the neuron represented, and a minimum (or zero) value for all values outside that range (as in Figure 6.5). But what happens in practice with population encoding is that the encoding function still has a peak value, i.e. an encoded value that results in a maximum firing rate, but this encoding function falls away smoothly as the encoded value moves away from this peak value (as in Figure 6.6).

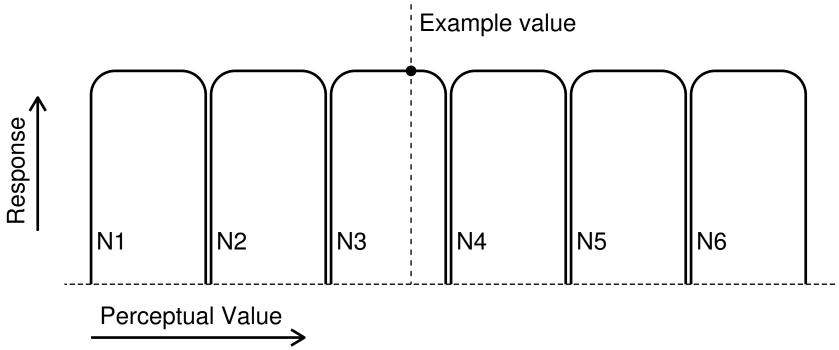

Figure 6.5. Neural response without population encoding. The encoding functions for a perceptual variable are shown for 6 neurons in a hypothetical cortical map. Each neuron has a maximum response to values in the range of values it represents, and the ranges represented by different neurons are all disjoint from each other. An example value is shown, such that only neuron N3 responds to it.

Thus, for any encoded value, the neurons whose peak values are nearest to that value will fire most strongly, and neurons with peak values further away from the encoded value will fire less strongly, or not at all.

Given the observed firing rate of neurons responding to a single encoded value, it is relatively straightforward to determine what the encoded value is.

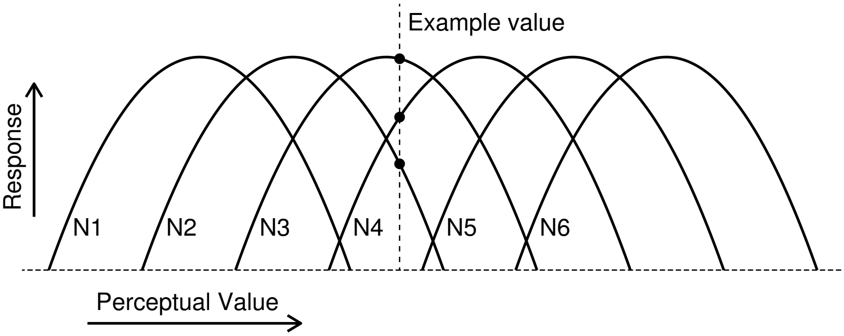

Figure 6.6. Neural response with population encoding. The encoding functions for a perceptual variable are shown for 6 neurons in a cortical map. For each neuron there is a value to which it gives a maximum response, but there is overlap between the ranges of values that different neurons respond to. Neuron N3 responds most strongly to the example value shown, but neurons N4 and N2 also show a response. N3, N4 and N2 constitute the "population" of neurons responding to that perceived value.

Thus the firing of all those neurons accurately represents the encoded value. Population encoding can quite accurately represent encoded values that are in between the peak values of the neurons in the map. For example, to determine the value represented by the firing of a group of neurons, take the average of the neurons' peak values, weighted by their firing rate. (A more accurate procedure is to calculate a **maximum likelihood value**, which is the value for which the current pattern of neural firing would be most probable.)

One problem with population encoding is that if the encoding functions are too broad, then it will be difficult to distinguish two distinct values from one value equal to the average of those distinct values. There is a simple musical example that illustrates this phenomenon: when we hear people singing in chorus. As long as the singers are singing in tune on average, we will hear the singing as being perfectly in tune, even if the individual singers are all slightly off.

In some cases the distinction between one signal and two signals close together may be made by the above-mentioned mechanism of synchronous firing. That is, neurons responding to one signal will fire synchronously with each other, and neurons responding to a second signal will fire synchronously with each other, but not synchronously with those representing the first signal.

It is also possible that the breadth of the encoding functions is itself adjustable by some means, so that neurons in a cortical map can choose broad or narrow encoding depending on which is the most useful in the current circumstance.

Population encoding is pretty much a universal property of cortical maps. So whenever I make a statement like "Cortical map X encodes for values Y and Z", this can be correctly interpreted as "The neurons in cortical map X fire at a rate that is a function of the closeness of their peak values of Y and Z to the observed values of Y and Z".

It's also worth noting that population encoding bears very little resemblance to how numerical values are normally represented in electronic computers. In computers we do not use a linear sequence of components to represent numerical values according to position. Generally we pick a base N (almost always 2), and then write the number as a sequence of digits, where each digit is an integer in the range 0 to $N-1$. So, to represent 1000 possible values, we would need 10 components (i.e. 10 digits in base 2), and to represent 1,000,000 possible values we would need just twice as many components, i.e. 20. To represent 1000 possible values in a cortical map, the brain would need 1000 neurons, although with population encoding this could be reduced by some fixed factor—for instance 2—to 500 neurons, given the ability of population encoding to represent the "in between" values. To encode 1,000,000 possible values would still require 1000 times as many neurons as required to represent 1000 possible values, i.e. 500,000. This contrast between the efficiency of digital and analogue representations appears in the theory of **octave translation invariance** in Chapter 11.