Brain Olympiad Reader
on
Computational Neuroscience
&
Artificial Intelligence

Colophon

Project	The Dutch Brain Olympiad Computational Neuroscience and Artificial Intelligence Task Force
Version	1.1 30-1-2023
Authors	see individual chapters
Editors	Tido Bergmans, Nicolas Rault, Tousif Jamal, Nishant Joshi, Matthew Lennon, Hilal Nizamoğlu, Ildefonso Ferreira Pica, Nicole Vissers, Devin Kellis, Tobio Aarts, Fleur Zeldenrust

Partners

This work has been supported by

Dutch Brain Olympiad Foundation

International Brain Bee

Organization For Computational Neurosciences

NeurotechEU

Donders Institute for Brain, Cognition and Behaviour - Radboud University

Funded by the Horizon 2020 Programme of the European Union under grant agreement No 860949 (SmartNets)

Table of Contents

Introduction

These are exciting times! Artificial Intelligence (AI) applications can now be found everywhere, from smart devices connected to 'the internet of things', to systems assisting doctors making diagnoses to companies trying to sell you things through personalised commercials. You might agree with that or not but it's there, so it is important to understand how these applications work! Many of these applications stem from computational neuroscience work in the '70s and '80s of the last century: computational neuroscientists formed the first ideas on how to compute with networks of abstracted neurons ('perceptrons'). However, at the time computers were not powerful enough to really use these ideas in applications. Only in the last decade or so has this become possible.

We have split the text in two parts: Computational Neuroscience and Machine Learning. These two fields share a history (see above), a lot of methods and a fundamental question, which basically comes down to: How can you make a network do and compute things? However, the aims of the two fields are different. Computational Neuroscience is a natural science that studies the brain, so it is the physics or biology of the brain. On the other hand, Machine Learning is a computing or engineering science, with the aim of making better applications (robots, computers, etc).

Here you find a first draft of a reader written on what we believe are the main topics of computational neuroscience and artificial intelligence. They are aimed as an introduction into these topics for everyone interested in them, in particular for participants of the Dutch and International BrainBee competition.

As this is a first draft, not at all finished or final yet. Therefore, your feedback is much appreciated! Please send it to cnai@hersenolympiade.nl (then we will know your email address) or fill out the form at https://forms.gle/18SLqdDinKb1kAz46 (this is anonymous) .

Have fun reading!

Fleur Zeldenrust

Part 1: Computational Neuroscience

1.1 What is computational neuroscience?

Author: Matt Lennon

Computational neuroscience, also called theoretical neuroscience, is the study area that uses mathematical tools and theories to explore the developmental, structural, physiological and information processing aspects of the nervous system. It takes components of computer science, physics and electrical engineering and applies them to mathematical models of neurobiological systems to better understand the brain. Computational neuroscience can be used to gain understanding at every level of the brain, ranging from modelling single cell firing to whole brain functioning.

We will start with the basic units of the brain: the neurons. How does information travel from neuron to neuron (1.2)? Next, we will explain how these neurons form networks, and how we can study these (1.3). We will dive into what many neuroscientists believe brains are actually doing most of the time: making predictions about the world around us (1.4)[1] and we will conclude with a chapter on how we can use our brains to interact with and control machines (1.5). Hopefully, at the end of this part you will understand more about the computational properties of the brain, including what the brain needs to calculate to function in the world and why.

[1] NB We are still writing a text on "Semantic mapping of the brain - can neuroscientists "read our mind" and how does it actually work?" - this will appear in the next update of this reader

1.2 Membrane potentials - How fast does a signal travel inside the body?

Author: Fleur Zeldenrust

Parts of this work were inspired by the reader 'Leven en Natuurkunde' (www.natuurkunde.nl)

Learning Objectives

1. Understand the importance of the nervous system in rapid signal generation
2. Understand electrochemical gradients and which ions form them in neurons
3. Be able to use the Nernst equation to calculate an ions' equilibrium potential
4. Understand what is meant by the term 'voltage sensitive'
5. Understand key steps of an action potential, including which ions mediate each step
6. Explain how a neuron's cell membrane can be viewed as a capacitor

1.2.1 Introduction

Imagine, you wake up in the morning and you want to get out of bed to start your day. To be able to do this, the muscles in your body need to receive a signal and energy to move. The muscles can use glucose as an energy source. To get glucose, and many other substances in the body, to their destinations, they have to be transported. However, sometimes it is not that the substance itself needs to be transported, but a substance is used in order to send information. A *hormone*, for example, is a chemical messenger that informs an organ that it needs to act, typically by traveling through the blood. In this case, the substance is called a signal. However, to send signals, there are also other ways than using physical substances. So, to use our muscles we need glucose, but also a move-signal that travels the entire distance from our head to our toes, which in humans takes about 0.3 seconds.

An experiment: Let's try to test how fast the motor signals travel to and from our brain. Keep your thumb and index finger a small distance from each other. Someone else holds a banknote between the thumb and forefinger and releases it unexpectedly. You won't always manage to grab it before it has passed your fingers. So signals do not always travel at the same speed. Your reaction time depends on your concentration. When you do the experiment very concentrated - for example because you can keep the banknote if you catch it – your reaction time will be short. It turns out that such rewards are key to learning, see also chapter 2.3 of this reader.

Reaction time is critical in athletics competitions, such as sprinting. Sprinters start from a block with a sensor. The start can be detected. When this is less than 0.3 s after the starting gun was fired, it is declared a false start: the sprinter must then have started before the shot. At athletics competitions, apparently it is assumed that the time needed for a signal to travel from the ears through the brain to the leg muscles cannot be less than 0.3 sec.

In our body, substances can be used to send signals. These substances can be transported via *diffusion*, in which they move from areas of high concentration to lower concentration. They can also be transported by active tractor-like molecules, a process called directed transport. Transport of substances in the body by diffusion is fast enough over very short distances, but very slow over large distances. It can take hours to transport a substance several centimeters through diffusion! However, we will see later in this chapter that diffusion can be very fast for ions crossing a membrane, which is only a few nanometers thick. Molecular machines can transport particles faster, but this too is not fast enough to send a signal from your head to your toes to start running, for example, if you see a lion running towards you, or to step aside if you are about to fall. In these cases, a signal needs to arrive at your legs from the brain in a fraction of a second and feedback must occur similarly fast. For such fast signal transport, your body uses the *nervous system*.

1.2.2 Voltage spikes as signals

How can a signal travel through your body, without physically letting a substance travel the entire distance? Signals can travel fast through a material, faster than the substance that is moving itself can travel. How does that work? For instance, consider a wave that travels through a cord, a sound wave in the air or an electrical signal that goes through a wire. Similarly, the nervous system uses electrical conduction to transmit signals. This works differently from a copper wire between, for example, a smartphone and an amplifier, but the core of the matter is: electrically charged particles attract (similarly charged particles) or repel (oppositely charged particles) each other and thereby can set each other in motion. That movement can be propagated in a medium, such as a copper wire. Within our bodies, however, this medium is known as the nervous system. The individual cells that make up the nervous system, and through which this conduction is made possible, are called *neurons*. Here, we will look at the mechanism of signal transport in a neuron from the point of view of a physicist. This means that we make a model for the basic mechanism.

Neuron

In the previous paragraph, we talked about *neurons*. A neuron is a normal cell, with a cell body, a nucleus containing DNA, and other machinery that makes a cell function. However, neurons also have some special properties: they can send electrical signals throughout their bodies.

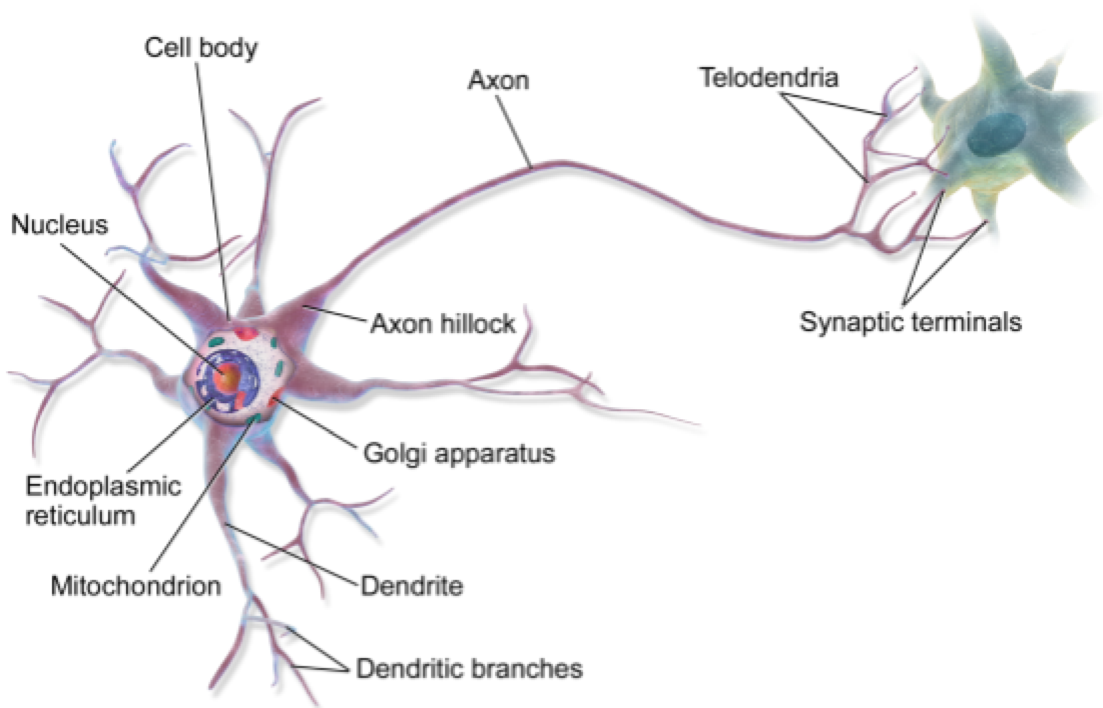


Figure 1: The basic anatomy of a neuron. Image from Wikipedia: Axon (https://en.wikipedia.org/wiki/Axon, accessed 19-11-2021).

From that perspective, neurons consist of 3 important parts: 1) the cell body or *soma*, containing the nucleus and basic machinery of the cell, 2) *dendrites*, the small branch-like parts which receive signals from other neurons, and 3) the long branch-like *axon*, with which a neuron sends signals to other neurons. If we zoom in, we can see the outer layer of the cell: the membrane (Figure 2). This membrane has special properties that makes electrical signal transport possible. We will discuss that here.

Diffusion of ions through the membrane

Like other cells in the body, neurons are separated by their surroundings by a structure called a *cell membrane*. In principle, the membrane of the neuron does not allow water or ions to pass. However, the membrane contains special structures known as *ion channels* that allow specific ions to pass through.

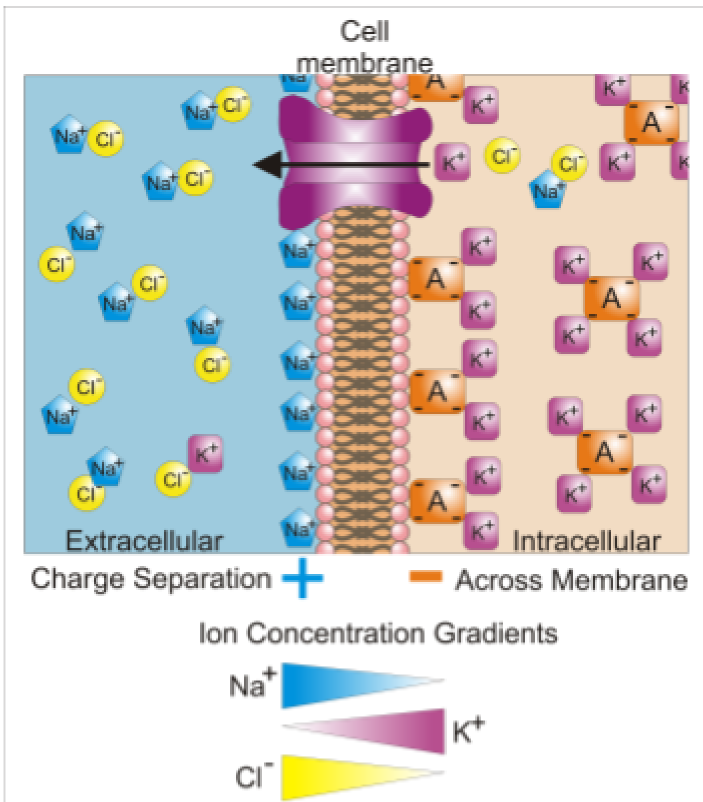


Figure 2: The membrane of a nerve cell. Left is the outside of the cell, containing mostly saline (natrium (blue) and chloride (yellow)). Right is the inside of the cell, containing mostly potassium (purple) and anions bound to other molecules in the cell (oraange). Image from Wikipedia: Membrane potential (https://en.wikipedia.org/wiki/Membrane_potential, accessed 19-11-2021).

The passive flow of ions through these ion channels is influenced by two forces: (1) a concentration difference and (2) a voltage difference between the inside and outside of the cell. Together, these two forces create what is known as an *electrochemical gradient*. For example: (1) inside the neuron the concentration of potassium ions (K^+) is higher than outside (Figure 2). If an ion channel specific for K^+ opens, K^+ ions will be transmitted across the cell membrane due to the concentration difference. So because of the concentration difference, K^+ diffuses out of the neuron. In this way, (2) positive charge will leave the neuron. This results in an electrical voltage difference between the outside and inside of the membrane and forms an electrical current. This voltage difference however, will pull K^+ ions back in: if the inside of the membrane becomes negatively charged, the positive K^+ ions will be attracted. At a certain voltage across the membrane there is an equilibrium between the potassium leaving the neuron by diffusion (through the concentration difference) and potassium entering the neuron by the electrical voltage. In a human neuron, this equilibrium voltage for $K^+$is about -80 mV.

In 1888, the physicist Hermann Walter Nernst first derived an equation for the equilibrium between electrical force and diffusion described above. This equation is shown below. In the equation, you can see the values of the natural constants *k* and

q, and as an example the temperature (T) in the body. If you imagine doubling the valence (i.e. the charge) of the ions, or the ratio between the concentrations, you get an idea of the relationship with the Nernstpotential, also called the *equilibrium potential*.

Nernst potential

$$V_{Nernst} = \frac{k_B \cdot T}{q \cdot z}[ln(C_{out}) - ln(C_{in})] \simeq -58\, log(\frac{C_{in}}{C_{out}})$$

In which

- k_B is the Boltzmann constant ($138 \cdot 10^{-23}$ J/K)
- T is the temperature in Kelvin (typically 310 K, which is equivalent to 37 or °C or 98.6 F)
- q is the charge of a single electron ($1.602 \cdot 10^{-19}$ C
- z is the valence of the ion, i.e. +1 for sodium and potassium, +2 for calcium, -1 for chloride
- C_{in} and C_{out} are the ion concentrations on the inside and outside of the neuron

The potential over the membrane of a neuron is typically around -65 mV. This is slightly higher than the Nernst potential of potassium. This 'resting membrane potential' is higher than the Nernst potential of K^+, because neurons also contain other ions such as sodium, chloride and calcium (see box 'Other ions'). The electrical potential outside the neuron is defined as zero (the 'earth'), and so the resting potential on the inside of the membrane would be -80 mV if the membrane only allowed potassium to pass through, but is in fact around -65 mV. A simple drawing is shown in Figure 3.

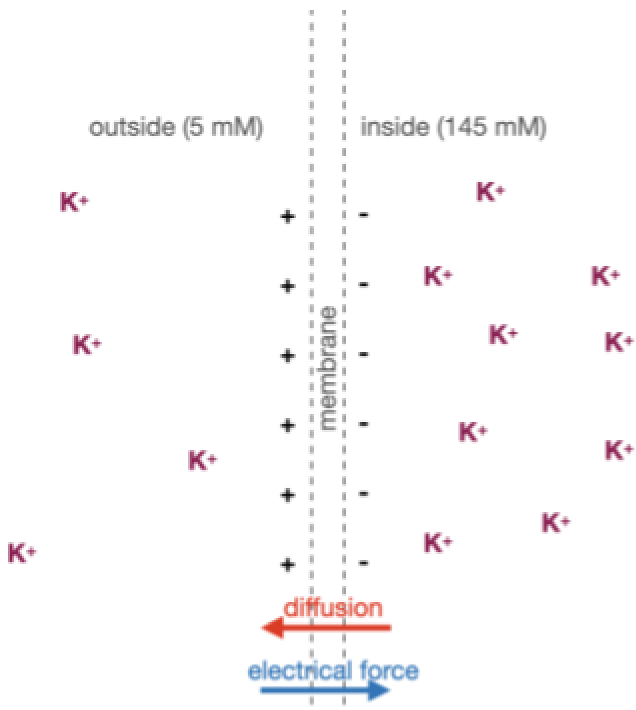


Figure 3 If the membrane allows potassium ions to pass through, and if at the same time there is a difference in potassium concentration between the inside and the outside of the nerve, a potential difference is created across the membrane. The concentration difference drives ions out of the cell, as the concentration of K+ ions is smaller there. The electrical gradient pulls the positively charged ions back into the cell, as the cinside of the cell membrane has a negative charge compared to the extracellular compartment.

Different types of ions

The membrane of a neuron does not only allow potassium ions to pass, but also sodium (Na^{+}) , calcium (Ca^{2+}) and chloride (Cl^{-}) ions. K^{+} and Na^{+} are the most important ions when it comes to building an electrical potential across the membrane. And while K^{+}pulls the membrane potential to around -80 mV, sodium pulls it in the opposite direction to about +70 mV. The equilibrium potential of Na^{+} is positive because Na^{+}has a much higher concentration outside than inside the cell. The combined effect of the potential differences due to sodium and potassium results in a membrane voltage in the resting situation of approximately -65 mV. You can see that it is more towards the potassium than towards the sodium Nernst potential. This is because in the resting state, there are more open potassium channels than sodium channels, resulting in a stronger pull from the potassium potential.

Other ions

The potassium and sodium concentration differences are constant and are maintained by ion pumps. Chlorine does not have such a pump, with the result that the concentration inside the cell adapts to the membrane potential instead of the other way around.

There are also calcium channels, but they only open at very high membrane potentials. The difference in concentration between inside and outside is very high, so when those channels open, a lot of calcium flows into the cell.

To understand the rest and action potential you really only need sodium and potassium, calcium and chlorine have other functions in the cell and contribute to this to a much lesser extent.

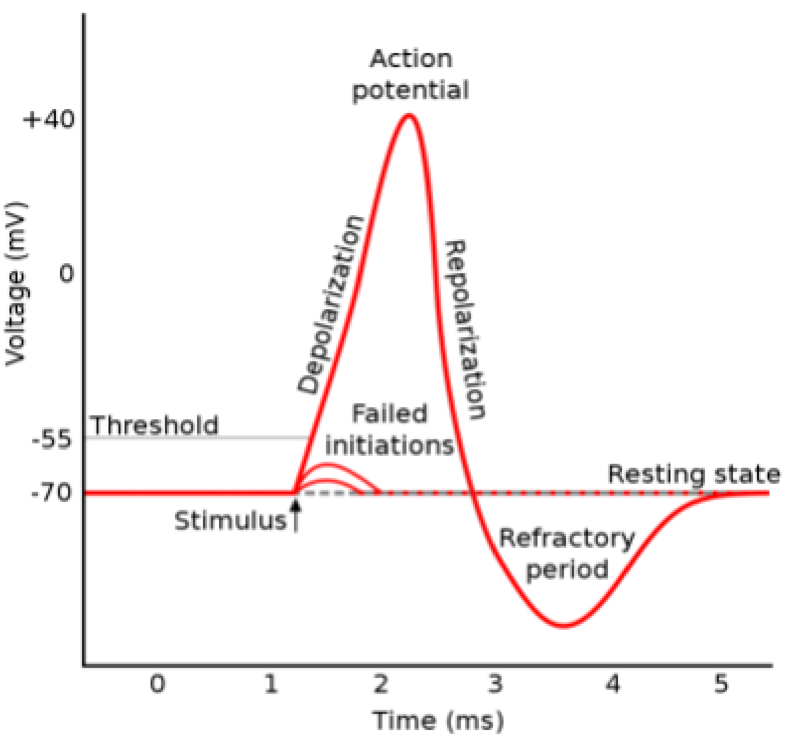


Figure 4: An action potential. Image from Wikipedia: Action potential (https://en.wikipedia.org/wiki/Action_potential, accessed 19-11-2021).

Action potential

Now we are at the point where we can start to understand the concept of signal propagation, which is made possible by the electrochemical differences of Na^{+} and K^{+} across neurons' membranes. A crucial part to understand this is the following: the ion channels for sodium and potassium are *voltage sensitive*. That means that if the membrane potential changes (increases), these ion channels open and let more ions go through. This is essential for understanding how an *action potential* is generated.

A stimulus can be administered to a neuron in the form of a local disturbance of the membrane potential. For example, a substance from another neuron (a *neurotransmitter*) can bind to receptors in the membrane and cause a change in the flow of ions across the cell membrane of ions. This causes the membrane potential to rise slightly — i.e., become less negative — because Na^+ rushes into the cell. When this disturbance is strong enough, the membrane potential might reach and exceed an important **threshold**. At this point, a self-reinforcing feedback effect called *depolarization* occurs: the higher membrane voltage causes more voltage-sensitive sodium channels to open, causing a greater influx of sodium, further increasing the membrane voltage, opening more channels, and so on. This is in fact a chain reaction: when the membrane potential surpasses the threshold, this will result in a further increase in membrane potential, which results in a further increase in membrane potential, and so on. The membrane voltage rises above zero, but in the meantime voltage-gated potassium channels begin to open, causing an outflow of potassium. This outflow of potassium lowers the membrane potential again (*repolarization*). This repolarization brings about a return to the equilibrium situation. This whole process, from small disturbance (stimulus), to upsweep to restoring of the resting membrane potential takes a few milliseconds in total. The peak in the membrane potential is called an action potential (Figure 4). The action potential is produced by the cell itself after the stimulus.

Refractory period
Just after the action potential, there is actually a short period where the potassium channels are still open, but the sodium channels have already closed. During this period the membrane potential is lower than the resting state, and it is more difficult to generate new action potential. This period is called the refractory period.

Propagation of an action potential
We have seen that an action potential can be triggered by an external stimulus at a specific location of the membrane. The membrane potential rises slightly there, causing the cell itself to generate the action potential. Right next to the site where an action potential occurs, the membrane still has the lower resting potential. This electrochemical potential difference leads to the creation of a *current*, which refers to the flow of ions. A current then flows from the site of the action potential initiation to the adjacent site. As a result, the potential also rises there and the entire reaction that we saw above takes place in that neighboring spot, and so on. In this way, the action potential propagates along the cell membrane as shown in Figure 5. This is also a chain reaction, but now in space rather than in time. Moreover, the action potential does not speed up, but rather moves at a constant speed through the axon.

Ion pumps

The resting membrane potential is too high to keep the potassium concentration difference between inside and outside stable. After all, it should be -80 mV, while in practice it is -65 mV (due to the effect of sodium). So potassium continues to flow out. A similar effect keeps sodium flowing in. This would reduce the concentration differences and change the Nernst potential. However, the nerve cell actively maintains them, with ion pumps. The sodium-potassium pump pumps potassium back in and sodium back out. As this transport goes against the electro-chemical gradient it needs energy, in the form of ATP, to function.

After every action potential, the cell has to use energy to restore the resting ion concentrations. To do this, they use active ion pumps, see box 'Ion pumps'.

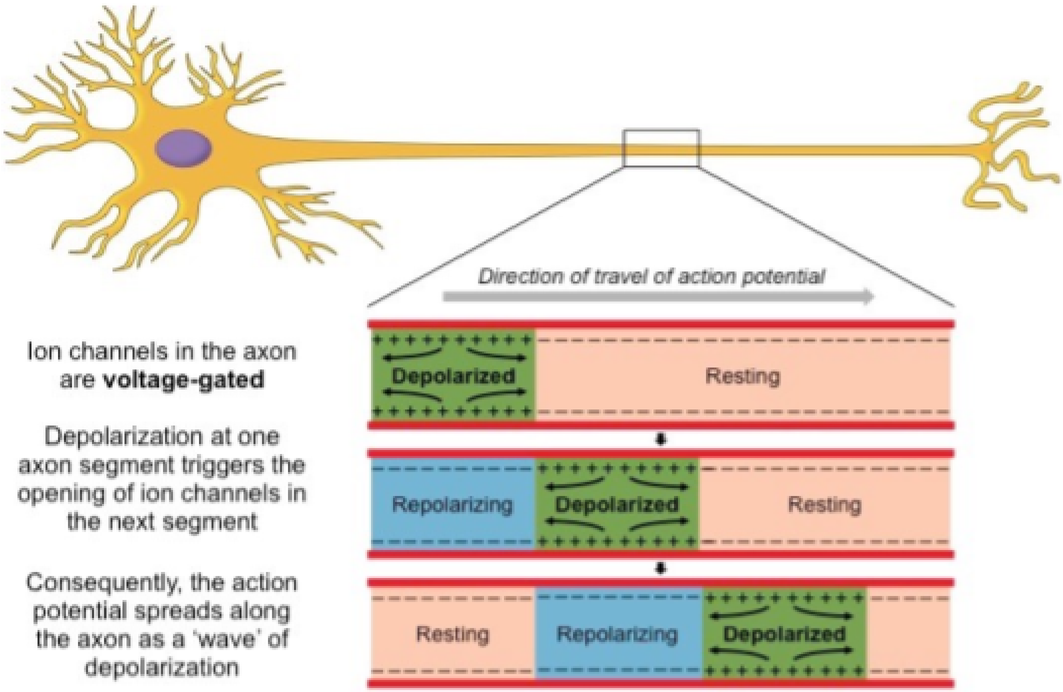


Figure 5: The propagation of an action potential through an axon. Image form BioNinja (Cornell, B., 2016, https://ib.bioninja.com.au/standard-level/topic-6-human-physiology/65-neurons-and-synapses/nerve-impulses.html, accessed 19-11-2021)

1.2.3 The cell membrane as a capacitor

In the previous section we saw how an action potential moves along the membrane of the cell, propagated by the opening and closing of ion channels. We also saw how ion pumps across the membrane maintain and restore the concentration difference of several ion types between the inside and outside of the cell. Remember that during an action potential, electric charge flows *through* the membrane, but the disturbance, the action potential, travels *along* the membrane. Here we will zoom in on the electrical properties of the cell membrane. More precisely, the cell membrane can be seen as a circuit consisting of a resistor and a capacitor in parallel.

A *capacitor* is an electronic device that can store electrical energy, just like a battery can. However, a capacitor is also different from a battery in that it does not give a constant voltage.. You can find them in old-fashioned flashes of cameras (although nowadays these are mostly replaced by led-lights) and in old-fashioned 'tube' televisions. A capacitor works the following way: mostly, a capacitor consists of two parallel metal plates, a short distance from each other and with an insulating layer in between. It can be electrically charged. This happens as follows: if the capacitor is connected to a battery (through a *resistor*), a negative charge flows onto one of the plates. This will push a negative charge of the same magnitude from the other plate.

Thus, a current flows through the capacitor, even though no charge itself crosses the gap between the plates. In time, one plate becomes more negative, the other becomes positive, and between the two plates the electrical voltage increases. The current decreases due to the repulsion from the negative charges at the left-hand plate, until no more charge can be added. The current has ceased and the capacitor is charged (Figure 6). The capacitor can discharge again when the plates are connected to each other (through a resistor).

Capacitance of a capacitor

We can measure how much charge Q (measured in units of Coulomb or C) can be put on the plates and how much electrical potential V (in Volt V) is needed for this. This quantity is called the *capacitance* of a capacitor: $C = \frac{Q}{V}$

Capacitance is measured in the unit Farad (F or C/V). The capacity increases with the size (surface) of the plates, but decreases with the distance between the plates.

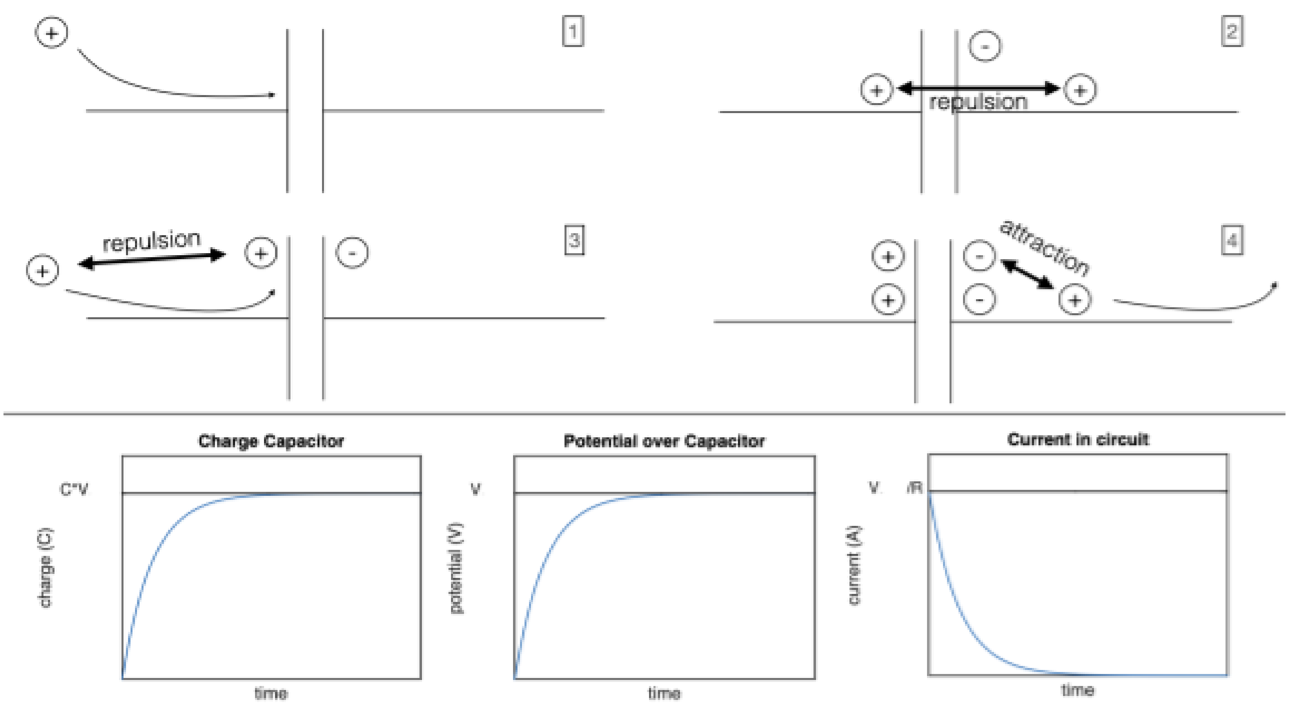


Figure 6: The charging of a capacitor. An electrically charged particle arrives at the plate (1), and pushes away a similarly charged particle at the opposite plate (2). The next particle has a harder time arriving at the plate due to electrical repulsion (3) and similarly the next particle has a harder time leaving the opposite plate (4). Bottom: the charging of a capacitor over time.

The membrane of the cell is a capacitor. The membrane is built up of two layers of lipid molecules: fatty molecules that do not solve in water (both the inside and outside of cells consist of water) and form therefore an effective barrier through which not much can pass. Because the membrane consists of two of such layers (a so-called *lipid bilayer*) it functions the same as the two metal plates of a capacitor: it effectively separates charge (see Figures 2 and 3). The only way ions can go through the membrane is through ion channels. There, they meet a certain amount of resistance: they cannot flow freely. Therefore, the ion channels work as a resistance. So if an action potential propagates through an axon (Figure 5), this can be described as the charging and discharging of a series of coupled (in parallel) capacitors.

Extra explanation: 'RC-time'

The *resistance* (R, measured in the unit Ohm Ω) of the ion channels and capacitance C of the membrane together determine how fast the membrane potential recovers after a disturbance. It is an exponential relationship. The product R·C determines how quickly the disturbance is

$$V(t) = V_{rest} \cdot e^{\frac{-t}{R \cdot C}}$$

When a time equal to the value of R·C has passed, the disturbance is still at 37% of the maximum. This can be seen in Figure 6 in the bottom 3 figures as the 'characteristic' time at which the charge and voltage increase, and the current decreases.

1.2.4 Summary

The transport speed of a bodily substance itself cannot reach the required signal speed of 1 m/s up to 5 m/s to send a signal from your head to your toes. Therefore, the signal is transported electrochemically through the nervous system. This is done by a disturbance of the membrane potential of neurons, which travels along axons. The membrane potential is normally at an equilibrium resting potential where the flow of sodium and potassium ions due to the difference in concentration is exactly as great as (but opposite to) the current due to an electrical potential across the cell membrane.In the resting state there is a potential across the cell membrane of about -65 mV, negative on the inside by definition (the potential outside the cell is defined as zero). A disturbance, caused by mechanisms in the membrane (neurotransmitters binding to receptors and causing an influx of ions) amplifies itself, leading to an action potential. An action potential propagates along the membrane and can act as a signal. That signal travels much faster through the body than a signal could travel using molecular motors ('active transport), let alone using diffusion. Within this system, the cell membrane acts as a capacitor by separating charge (ions) inside and outside of a neuron, whereas ion channels act as resistors by allowing this charge to flow across the membrane.

1.2.5 Key terms

Dendrite
Axon
Soma
Cell membrane
Hormone
Neurotransmitter
Diffusion

Nervous system
Neuron
Ion channel
Electrochemical gradient
Equilibrium potential
Action potential
Threshold

Depolarization
Repolarization
Refractory period
Current
Capacitor
Resistor
Capacitance
Resistance

1.2.6 Want to know more?

- Video: NinjaNerd lecture on youtube: https://www.youtube.com/watch?v=Jk_9IhHVOTk
- Khan Academy on membrane potentials: https://www.khanacademy.org/test-prep/mcat/organ-systems#neuron-membrane-potentials
- Khan Academy on capacitors: https://www.khanacademy.org/science/in-in-class-12th-physics-india/in-in-electrostatic-potential-and-capacitance/x51bd77206da864f3:capacitance-parallel-plate-capacitors/a/capacitors-article

1.3 Network dynamics

Author: Ildefonso Ferreira Pica

Learning Objectives

1. Be able to explain what is meant by the term ‘neural network’ using the terms node, edge, and network dynamics
2. Understand the concept of a steady state within a neural network
3. Understand how the state of a neural network can be represented in state space
4. Understand why building complex models of the nervous system is challenging

1.3.1 Introduction

Our brains consist of billions of neurons, and together these neurons make trillions of connections. Scientists have known for a long time that individual neurons will respond to changes in the environment. For example, while you are reading this text, cells in your retina respond to different wavelengths of light thanks to specialized proteins called photoreceptors. Their response comes in the form of an electrical signal that is passed along the optic nerve into the brain. Cells in other sensory organs such as your nose, ears, and tongue all convert a change in your environment (i.e., a smell, sound or taste) into an electrical signal that is sent towards other neurons in the brain. However, on their own these signals are just electrical pulses. It takes the combined activity of many neurons to interpret these electrical pulses and produce conscious sensations. How neurons do this is an area of intense study. Scientists believe that the way that neurons are connected plays a vital role in how signals in the brain are processed.

To better understand how brain structure and function are related, scientists perform experiments and make mathematical models that can be used for simulations. In this chapter, we will look at how we can make a basic network model of (parts of) the brain. We will call these models neural networks. After we've made a basic neural network model (chapter 1.3.2-1.3.3), we will see what happens when the neurons are stimulated (chapter 1.3.4). When we look at how the neural network (re)acts over time, we are studying its dynamics. This will be the second topic of this chapter.

1.3.2 What are networks?

The world around us is full of networks. Cities are connected to each other through a network of roads, computers are connected to each other through wired and wireless connections to form the internet, and you are connected to your friends, family and other people in a social network. Once you start looking at the world through the lens of networks, you will see them everywhere. But what is a network? Fundamentally, networks consist of two parts: things and the connections between them. In network theory, it is common to refer to things as *nodes* and to the connections between them as *edges*. If we want to study a network of things, it is often useful to make a network model.

1.3.3 Making a basic neural network model

Let's say we have a petri dish with three neurons, and we insert glass pipettes into these neurons so that we can stimulate them and measure their activity. We show this setup in figure 1. We have represented the neurons as circles and the connections between the neurons as arrows. This image already shows us a neural network, albeit a very basic one.

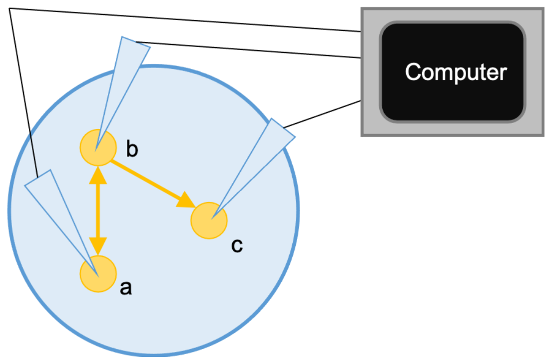


Figure 1: a petri dish with three neurons, labeled a, b and c. Glass pipettes are inserted into the neurons and connected to a computer. The glass pipettes allow us to stimulate the neurons and record their activity with the computer.

To work with our model we will first define three rules.

- Rule 1: Neurons can be in two activity states: a resting state (indicated by a 0) or an active state (indicated by a 1). A neuron is in the resting state unless it receives a stimulating signal. When the stimulating signal ends, the neuron instantaneously returns from the active state to the resting state.
- Rule 2: Neurons that are in the active state will send signals along their axons. The signals will arrive instantaneously. This means that signals take no time to travel from the neuron that sends them to the neuron(s) that receive(s) them. In figure 1, the axons are shown as arrows.
- Rule 3: Signals can only move down the axons, which means that signals must follow the arrows. An arrow that points in two directions, like the arrow between neurons a and b, really means that there are two arrows: one from a to b, and another from b to a. We just summarize this by drawing one arrow that points in both directions.

We can now use this model to make some predictions about how these neurons will behave when we stimulate them through the glass pipettes. For instance, if we stimulate neuron a it will send a signal to neuron b because they are connected. Since neurons b and c are also connected, neuron b will stimulate neuron c. Neurons signal each other using both electrical and chemical signals. The electrical signals, called action potentials, are the result of positively charged sodium ions in the extracellular medium that move into the neuron through channels in the neuronal membrane. However, these electrical signals are typically not directly passed on from neuron to neuron, because this would require a physical channel between the membranes of two neurons. While these physical connections, called gap junctions, do exist, most electrical signals from neuron to neuron are passed on through special structures between neurons called synapses. Synapses consist of a presynaptic terminal on the source neuron, a postsynaptic terminal on the receiving neuron, and the gap between the terminals called the synaptic cleft. Action potentials are transformed into chemical messages, called neurotransmitters, in the presynaptic terminal. The neurotransmitters are released in the synaptic cleft and transformed back into electrical signals in the postsynaptic terminal. In this way, release of neurotransmitters in one neuron, can trigger a second action potential in the neuron on the other side of the synapse, conveying the message onward.

We see that by stimulating neuron a, we can stimulate the entire network. If we would instead stimulate neuron c, the other neurons would not get stimulated. This is because there is no outgoing connection from c to either neurons a or b (neuron c has no axons to neurons a or b). There is an arrow between neurons b and c, but this arrow points from b to c and the stimulation can only follow the arrows (following rule 3). As you can see, by making a network model of our petri dish with neurons, we can explore how these neurons behave when they are stimulated.

What happens when we stimulate neuron b? Try to use the three rules and figure 1 to figure it out.

1.3.4 Investigating neural network dynamics

In the previous section we have seen that the activity states of the neurons in a network can change when we stimulate the neurons. We have also seen that the pattern of activity depends on which neurons are activated. When we stimulate neuron a, neurons b and c also become activated, but when we stimulate neuron c, neurons a and b remain in their resting state. When we examine how the patterns of activity in a neural network change over time, we are examining the **dynamics** of the neural network.

The study of network dynamics is an important scientific discipline with many applications. For example, epidemiologists study how contagious diseases spread through networks of people, and look for ways to slow down and stop a disease from spreading. In the case of neural networks, neuroscientists study neuronal dynamics to understand how signals are spread and changed by the neurons in the network. When studying the dynamics of a neural network, neuroscientists often want to know two things: 1) what happens with the activity of the neurons in the network over long stretches of time? and 2) what happens with the activity of the neurons in the network when neurons are stimulated? We will soon see that these two questions are related to each other.

Let's return to our petri dish with three neurons (figure 1), and recall the 3 rules. Rule 1 said that the neurons could be either in a resting state (indicated by a 0) or an active state (indicated by a 1). When a neuron is stimulated it goes from state 0 to 1, and when it is no longer stimulated it goes back from state 1 to state 0. Now let's think about what happens with the neural network when we don't stimulate any of the neurons. In this case, the three neurons remain in state 0. We show this in figure 2. Each row shows the activity for neuron a, b, and c. The activity state is 0 for the three neurons and remains that way. When the activity in a neural network doesn't change anymore as time passes, we say that the neural network is in a *steady state*. This concept of a steady state, where a system or network does not change its activity over time, is used in many areas of science from physics to economics and biology.

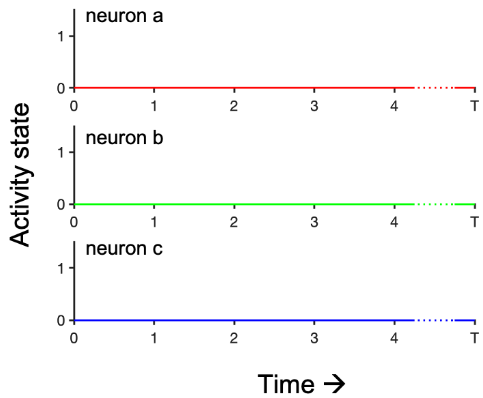


Figure 2: the neurons in the petri dish neural network (figure 1) are not stimulated across an undefined period of time (T).

For the next example, let's recall what happens when we stimulate neuron c. We show this in figure 3. The stimulation of neuron c is shown as a yellow block. We can see that as long as the stimulation is on, the activity state of neuron c is 1. Since neuron c has no outgoing connections to neuron a or b (there are no arrows pointing outwards from neuron c in figure 1), they are not affected by the stimulation. When the stimulation ends at timestep 2, all the neurons in the network have an activity state of 0 and this remains so continuously. This means that after the stimulation, the neural network is in the same steady state as in figure 2. Importantly, our stimulation of neuron c did not cause the network to change from one steady state to another steady state. But can we stimulate the network to change state?

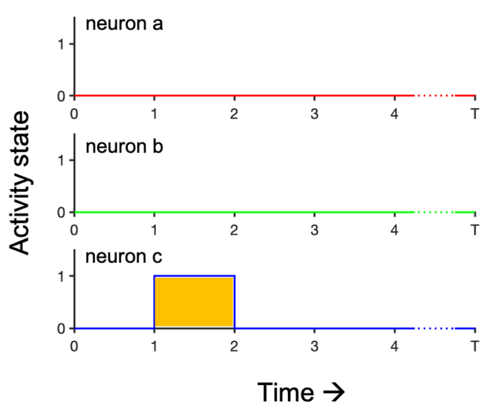


Figure 3: we stimulate neuron c.

We have previously seen that if we stimulate neuron a, the entire neural network becomes stimulated. We show this in figure 4. The yellow block is now positioned in the top row, indicating stimulation of neuron a. But now, since neuron a is connected to neuron b (red arrow), neuron b will become stimulated as well. Neuron b has an outgoing connection to neuron c (lower green arrow), therefore neuron c is stimulated by neuron b. So while the stimulation is on, all neurons are in an active state (activity state 1). But why do the neurons stay active even after we have stopped stimulating neuron a? This is because neurons a and b keep each other activated. As long as neuron a is in an active state, it will send a stimulating signal to neuron b (red arrow), and vice versa (upper green arrow). We can think of neurons a and b as a motor. When either neuron is activated, the motor switches on and stays on. Since neuron b also stimulates neuron c, it too stays active. The activity state is 1 for the three neurons and remains that way. Therefore the system is now in a new steady state, and we caused the system to switch to this new state by stimulating neuron a.

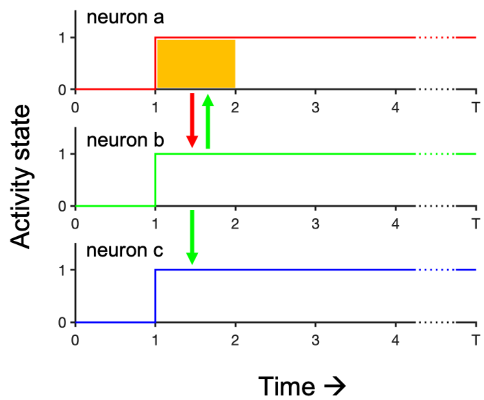


Figure 4: we stimulate neuron a.

When analyzing the dynamics of a system it can be useful to look at the dynamics in the system's *state space*. We show the state space of the petri dish neural network in figure 5.

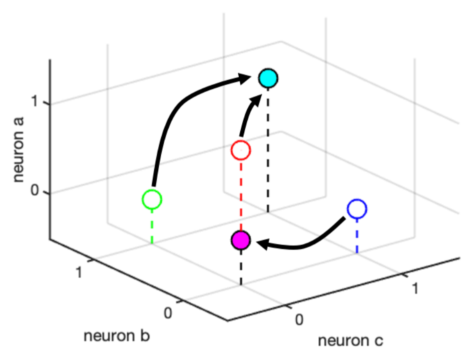


Figure 5: the state space of the petri dish neural network shown in figure 1.

The state space of a neural network is an imaginary space that shows all the activity states that a neural network can be in. The state space can also be used to visualize the activities of all the neurons in a network at once. We do this by thinking of the activity state of the network as a single point in space. How do we determine the position of this point? We let the position of the point depend on the activity states of the neurons. Remember that the neurons in our network can either be in the resting state (activity state = 0) or in the active state (activity state = 1). If all of the neurons are in the resting state, the point marking the activity state of the network will be located at the pink dot. We have marked the position of the point with a yellow triangle in figure 6.

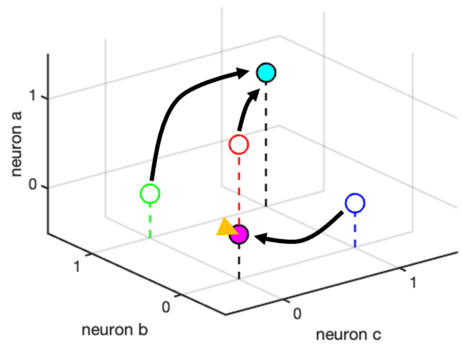


Figure 6: the yellow triangle shows the position of the activity state of the petri dish neural network.

If you look at the three different axes, you can see that they all show the activity of a different neuron. The pink dot shows the position where all the neurons have activity state 0. If we now stimulate neuron c (yellow arrow in figure 7), the point that marks the activity state of the network will move towards the blue ring (see figure 7). Through our stimulation of neuron c, we have forced the neural network to move through state space to a different activity state.

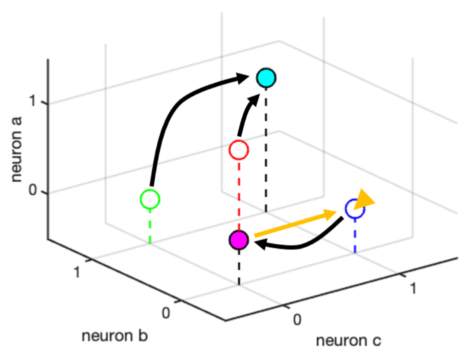


Figure 7: when we stimulate neuron c, the activity state of the petri dish neural network moves from the pink dot to the blue circle. Through stimulation, the neural network moves through state space.

You might be wondering why there are black arrows in figures 5, 6 and 7. These arrows show where the activity of the network will move to once it's in the position where the arrow starts. For example, after stimulating neuron c, the network will move to the blue circle state as we saw in figure 7. When we stop stimulating neuron c however, the yellow arrow is removed and the network (whose position is shown by the yellow triangle) will follow the black arrow back to the pink dot. Now we're back to the situation shown in figure 6. If you recall, we've performed the same experiment a couple of times already. In figure 3, for instance we saw that the activity state of neuron c moved to 1 when we stimulated it, and moved back to 0 once the stimulation ended. We also saw that neurons a and b did not change their activity state when stimulating neuron c. Figure 7 shows the same events, albeit in a different form. Because only neuron c is activated, the yellow triangle only moves in one direction towards the blue circle.

What happens when we stimulate neuron a or b instead? In that case, the activity state of the network moves from the pink dot towards either the red circle (if we stimulate neuron a) or the green circle if we stimulate neuron b. Recall that when either neuron a or b is activated, the rest of the network will be activated as well. The network state where all neurons are activated (their activity state = 1) is shown by the light blue dot in figures 5, 6 and 7. The black arrows from the red and green circles towards the light blue dot tell us that all neurons become active once either neuron a or neuron b is activated. We can also see that there are no black arrows pointing away from the light blue dot. This means that once the activity state of the network reaches the light blue dot, it stays there continuously.

Recall that when the activity of a neural network does not change anymore as time passes, we say that it is in a steady state. Looking at figures 5, 6, and 7 again, you might notice that there are two states indicated with dots (pink and light blue) and three states indicated with open circles (green, red and blue). You might also notice that the dots have black arrows pointing towards them, while the open circles do not. This is no coincidence. The dots show us the steady states! And we have now discovered a way to identify the steady states in state space plots such as in figure 5, even if we do not fill in the circles. Namely by looking at those points that only have incoming black arrows.

1.3.5 Making more detailed neural networks

In the neural network that we have shown here, the neurons and the signals that were passed between them, followed three simple rules. As a result our neural network was a very basic model, and one that cannot perform certain actions that networks of real neurons can. In neuroscience, and science in general, there is a constant trade-off when making and working with models. On the one hand, we want models that are very accurate and that can reproduce the phenomena that we observe in the real world. On the other hand, very accurate models need to include a lot of detailed rules, which makes them hard to work with and difficult to study. With our basic neural network model (figure 1), we only had to follow three rules for three neurons. This made it easy and fast to work with. But now imagine we had a model with one hundred or one thousand neurons, or a model with three neurons but with dozens of complicated rules to check. Working with these models would be a lot harder.

In real life, scientists often use computers to work with more detailed models that would be too time consuming to work with by hand. This is the same for neural network models. But even when using powerful computers, there are limits to the complexity of the models that can be used. You might already have some ideas for how we can make the petri dish neural network more detailed, and as a result more realistic. In the basic model we have used, neurons could only stimulate other neurons. But in reality, there exist different types of neurons some of which can inhibit other neurons. Thus we might define a new rule which says that if a neuron is *excitatory*, it activates other neurons, and that if a neuron is *inhibitory* it deactivates other neurons. Such a model would already match reality a bit better. Another rule in our basic neural network model was that signals travel instantaneously from neuron to neuron, but this is not very realistic. So we might want to change this rule, to make sure that the signals travel at speeds that we find in real neurons. Lastly, we have used a very basic rule to determine how neurons go from a resting state to an active state. If a neuron received a stimulating signal (via a glass pipette or another neuron) it was activated immediately. This rule is not very realistic. Scientists from disciplines ranging from physics to biology have worked on formulating rules and models that realistically describe how neurons change their activities when they are stimulated. For some clues on what these rules look like, we invite you to read the chapter on membrane potentials (1.2).

1.3.6 Summary

In this chapter we've learned about neural networks and network dynamics. We've seen that a neural network is a model of a network of real neurons, which shows how the neurons are connected. Neural networks consist of two parts, the neurons (also called nodes in network theory) and the axons (also called edges in network theory), which are the connections between the neurons.

We saw that the neurons have activity states (a resting state indicated by a 0, and an active state indicated by a 1), and that the axons have a direction (indicated by an arrow). When the neurons are in an active state, they send signals through their axons to connected neurons, activating these neurons.

We have seen that we can determine how neurons change their activity in a network when they are stimulated. The activities of the neurons in the network together form the dynamics of the network. We have seen that a neural network can be in different activity states. Through experimentation, we found that there are states that the network stays in continuously, and we learned that these states are called steady states.

Lastly, we looked at the state space of a neural network. We learned that the state space is an imaginary space and saw that we can use it to visualize the activities of all the neurons in our network at once. We found that we can analyze plots of the state space to learn about the dynamics of the neural network.

NB A future version of this chapter will also contain a section on balanced networks.

1.3.7 Key terms

Neural network	Network dynamics	Excitatory
Node	Steady state	Inhibitory
Edge	State space	

1.4 Predictive coding accounts of brain processing: Bayesian models

Author: Hilal Nizamoğlu

Learning Objectives

1. Describe what is meant by the 'predictive coding' framework of brain function
2. Be able to calculate a posterior prediction using the equation for Bayes rule
3. Explain how decisions might be made through the comparison of posterior predictions with cost functions

1.4.1 Introduction: the predictive coding framework

What would you think if I said that your brain is actually a "prediction machine"? What does that even mean? How can the brain "predict"? And really.. Why does it even matter? If I could grab your attention with these puzzling questions, come along and keep reading :)

Imagine, you are in a rush, running through the crowds to catch the bus. One last corner until you arrive at the bus station, climbing down the stairs, but ouch! You skipped a step and fell over. The bus is gone and so is your once healthy ankle. What happened? The stairway does not seem to lack any stairs for you to skip, but hey.. The gap between two steps seems bigger than the rest. If the step gaps were all equal, you wouldn't fall. But do you know why? Have you ever thought about how humans can so easily climb up or down the stairs without even looking at them? It seems as if there is an ongoing calculator in the brain that keeps track of your actions. It is so automatic that we never think about it until we fail. As in this situation, the brain keeps "guessing" the place of the next step based on what it has experienced with previous ones. As this occurs frequently without any error, the outcome, and thus the behavior, seems perfectly flawless. But as the calculation was fed with an unexpected input — which is the unequal gap between the steps — the prediction fails and the system literally collapses.
Don't you believe me? Well, here is an example:
https://www.youtube.com/watch?v=seieuz__B_g

To explain this phenomenon researchers suggest that the brain works as a prediction machine. Ongoing calculations of the next step from the previous ones through constant feedback loops occur to execute actions of all kinds. How the brain makes sense of the world depends on how accurate the predictions are. The *predictive coding framework* postulates that perception is a constant anticipation of future events based on prior knowledge that is updated subsequently by incoming sensory signals. Its aim is to minimize the mismatch between how it expects the world to be and what the sensory flow reveals about the world evidently. In order for the system to work efficiently, the error must be reduced, thus, the predictions are updated accordingly and the actions are executed based on these updates (Figure 1).

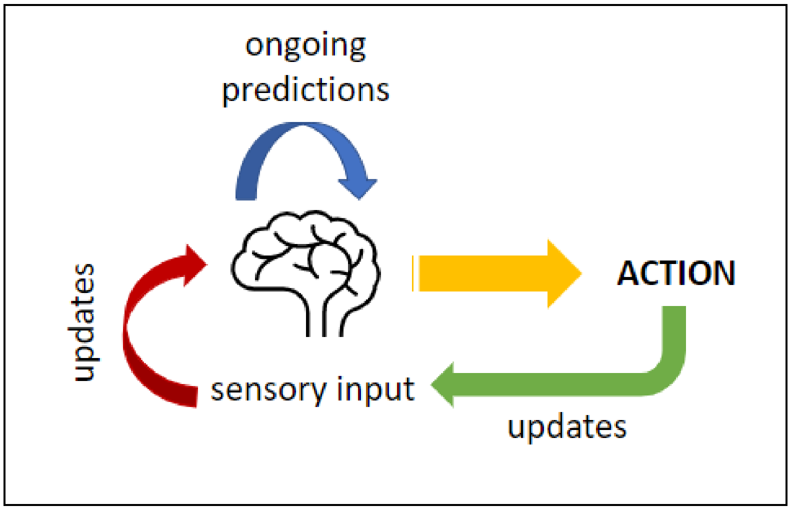


Figure 1. Predictive coding framework scheme. With ongoing predictions happening within the brain, with each incoming sensory input and the action output predictions are updated.

Doesn't that sound like playing a video game? If you have no idea about the plot of the game, you make random decisions. But as you start developing a sense of the story, your guesses -therefore, your actions- become more structured. This generalization is good for the improvement of your character and for winning the game. But it also opens grounds for mistakes in case of uncertain events (Figure 2).

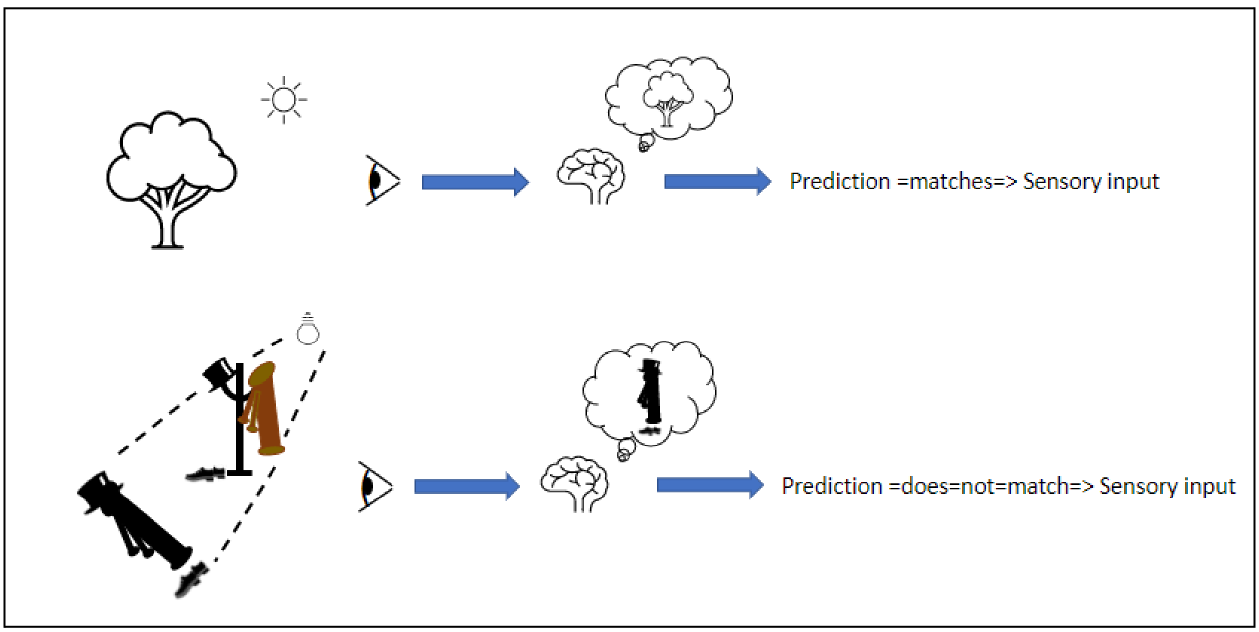


Figure 2. Mismatch between the sensory input and predictive mental representation.

1.4.2 Bayesian Decision Models

Decision making based on predictive coding (or predictive processing, although these terms are used somehow differentially by researchers like David Heeger) involves **Bayesian Decision Models.** These models have two main components: (1) *Bayes rule*, and (2) a cost function. Bayes rule, which was named after 18th century British mathematician Thomas Bayes, is a formula that shows the probability of a hypothesized state given the likelihood and the prior probability of the given observation.

With Bayes rule, one can calculate the probability of their hypothesis being true as a product of the probability of their hypothesis and the likelihood of the current situation, divided by the evidence or the observation at hand. After calculating this posterior probability, the cost function must be defined to make a decision. By minimizing the cost function, the decision is determined for action.

Before moving any further let's work on a scenario. Imagine, on a very hot day of summer, you come back to your house after a very tiring exercise practice with your school team. You are dry-to-your-bone, and looking for something to drink instantly. Considering that the weather is too hot, you casually open the fridge to grab some cold water. To your surprise, your usual jug is nowhere to be seen, but there is a soda bottle with no tag on it hidden at the back of the fridge. You think to yourself rather than water, it could be your brother's favorite sweet soda that he hides for himself to drink. Getting excited for drinking something cold and sweet or at worst sparkling water, you reach up, take a breath, and have a full sip from the bottle. To your surprise, it turns out to be one of your parents' favourite bitter ciders and now your thirst is even worse. How did you end up with the decision of drinking from the bottle and how happy were you with this decision?

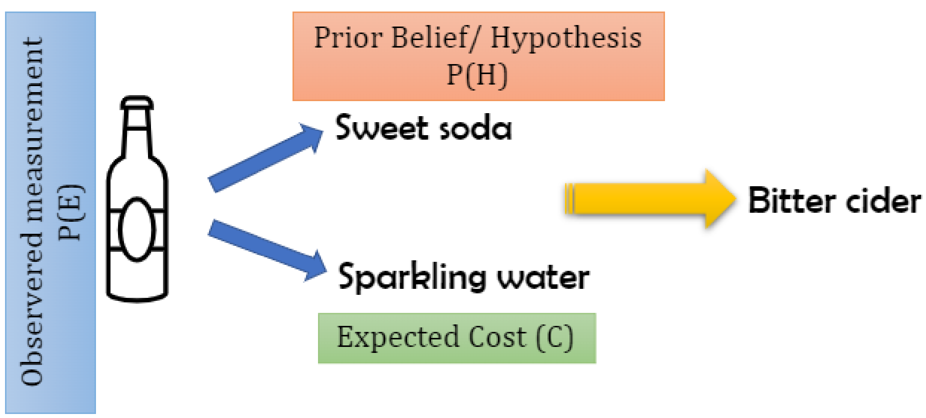


Figure 3. Diagram of the given scenario

Remember, you were thirsty, so you wanted to grab some water in the fridge. But rather than the water jug, the only cup that you think could contain a drink was a leftover bottle hidden at the back (i.e. observation). So, you assumed that the drink in the bottle could be either the sweet soda your brother likes and hides to drink on his own (i.e. belief) or at worst sparkling water (i.e. expected cost). At the end, it turned out to be one of your parents' bitter drinks (Figure 3).

Let’s put all of these into the formula. But please keep in mind, in this example we do not have more than one observation and we do not know the exact probability of each situation. This is just for the ease of explanation, but the real formula is based on the probability of each situation -for a quantitative example please see: https://youtu.be/R13BD8qKeTg

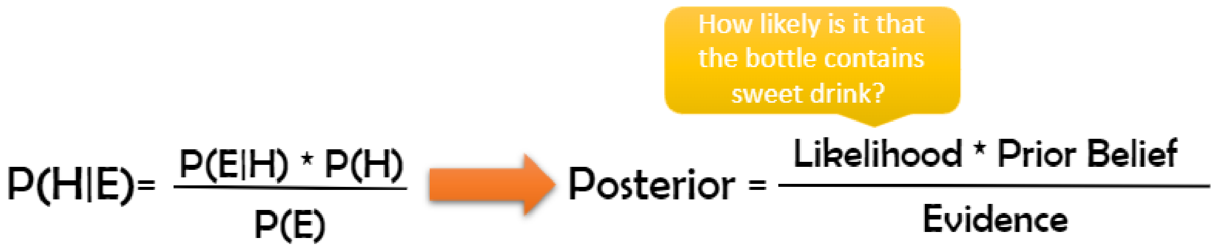


Figure 4. Bayes’ formula. Based on the scenario, here the likelihood is the assumption given the observation. So, how likely it is that the bottle contains sweet drink is the first parameter. It is multiplied by the probability of your prior belief (i.e. probability of the bottle to contain sweet soda). Then these two items are divided by the probability of the overall evidence (i.e. the probability that the bottle contains the sweet drink plus the probability that the bottle does not contain the sweet drink).

After the posterior distribution is calculated with Bayes rule, a cost function (aka. loss function, objective function) that should be minimized must be defined to generate a decision.
There is an expected cost of each action. In our example, the cost of drinking the liquid in an untagged and unknown bottle would vary from being water to being something poisonous/spoiled (considering that it was put hidden in the way back of the fridge).
If you expected the worst thing that could be in that bottle would be something poisonous/spoiled, no matter how likely the bottle could contain a sweet drink would not interest you considering that there is a high risk for you to get sick. But if you expected the worst scenario as the bottle containing sparkling water, you would take a sip after calculating the posterior (i.e. probability of the soda bottle to contain sweet drink).

1.4.3 Want to know more?

Now that you are more familiar with Bayesian decision making, please have a further look at how it is studied in the field of neuroscience and how we can observe the predictive brain at action through the links and descriptions below.

Psychophysics
We can compare the human behavior with the Bayesian model (how the model predicts the behavior and how humans behave for real)
Computational Neuroscience
Neuron firings model can be created by considering the error reduction through giving different weights for each neuron to see its contribution to the population: At the neuronal level, these discrepancies can be translated into changes in synaptic weights using specific learning computational rules, leading to changes in the model and subsequent more accurate predictions (Bubic et al., 2010: 10.3389/fnhum.2010.00025).

Clinical Perspectives

https://www.youtube.com/watch?v=c6OOtzbTXSU

Everyday cases for the predictive brain at action

- Visual Illusions: https://www.brainhq.com/brain-resources/brain-teasers/convex-or-concave/
- Motion sickness: https://www.youtube.com/watch?v=gKhE3CMz7Sk
- Tutorial webpage: https://www.kaggle.com/charel/learn-by-example-active-inference-in-the-brain-1#Bayesian-Inference
- Further blog entries by the key researchers on the topic: https://philosophyofbrains.com/author/hohwyj ; https://philosophyofbrains.com/author/clarka

1.4.4 Key terms

Predictive coding
Posterior prediction
Cost function
Bayes rule

1.5 Human computer interfaces

Author: Nicolas Rault

Learning Objectives

1. Be able to describe how electroencephalography (EEG) measures neural activity
2. Understand the limitations of EEG
3. List two other ways that brain activity can be measured
4. List at least two types of brain-machine interfaces that help patients with paralysis or sensory loss

1.5.1 Introduction

Controlling a machine with our mind, replacing part of our body with machines, improving ourselves with machines, those ideas are not new. In 1839, Edgard Allan Poe wrote a story titled “The man that was used up”. A few years ago, a remake of the movie Robocop was released. Overall, the half man/half machine character (cyborg) has become a classical character in science fiction. The idea of the cyborg seems to have fascinated us for some time now, but how close are we to these results?

First, let’s define a cyborg as presented in science fiction. It appears to be a living being with artificial limbs or components, with better or at least equal abilities to other living beings. The cyborg must be able to efficiently control its artificial components. Depending on the means used to do so, the command can be really slow. Instead of using complex commands requiring the user to translate its intentions into an understandable sequence of actions to perform by a machine, reading from its origins, the brain, seems to be the fastest way we can get access to the information about the command.

In 1973 Jacques J. Vidal published an article saying that it was possible to use the data from an electroencephalogram (EEG) headset as inputs to a machine, giving the first real step toward the cyborg and the first step to the field of brain computer interfaces (BCIs) or brain machine interfaces (BMIs). Until today the field has evolved and found mostly clinical applications for rehabilitation, substituting perception or even helping people with limited movement to communicate as we will see later. What still withholds us from creating cyborgs relies on our understanding of the brain, our ability to analyze its signals in a limited time as well as important ethical issues.

1.5.2 Detecting movement intention from brain activity

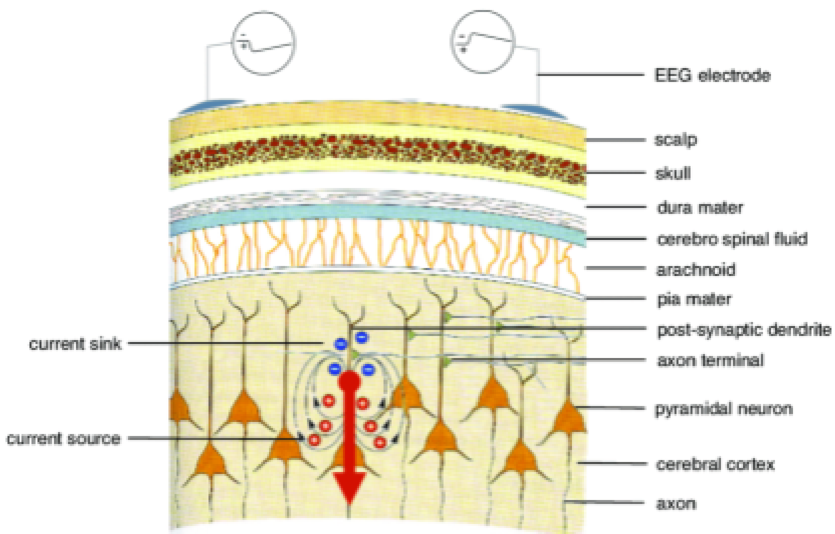


Figure 1 - Electrodes record the electrical current flowing through the pyramidal neurons, organized in such a way that their axons and dendrites are perpendicular to the scalp. Image adapted from 'Advanced forward models for EEG source imaging', Thesis, Gregor Strobbe.

To record neuronal activity, we can attach small sensors to the scalp that record electrical signals. This technique is called *electroencephalography***,** shortened *EEG*. EEG specifically records **synchronized activity of many neurons together**, which can be seen as waveforms when this activity is graphed over time. Those waveforms can be divided into five categories based on their frequency (the speed at which the wave pattern repeats).

Even though our knowledge of what these frequencies mean is limited, neuroscientists have found that some frequencies consistently occur whenever specific actions are made. For instance if we want to know if an individual wants to move their right hand, then we'll take a look at a specific electrode placed on the scalp known as the C3 electrode, in the Alpha and Beta frequencies. The C3 electrode is located above a region of the brain known as the *somatosensory cortex*, which is responsible for processing information related to the sense of touch, body position or muscle tension.

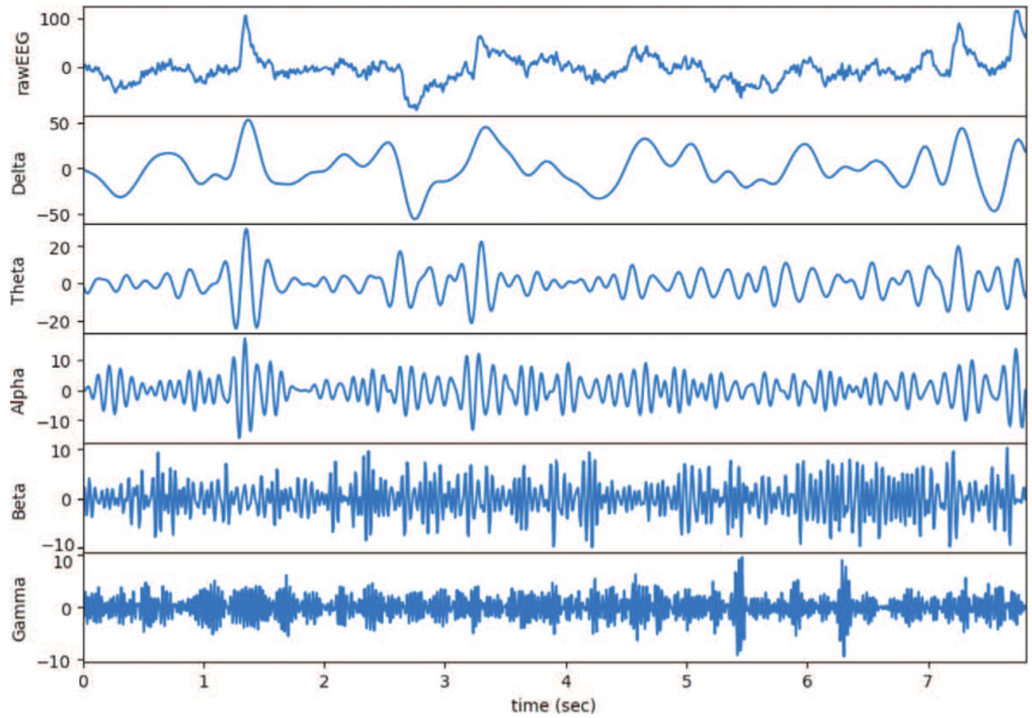


Figure 2 - A signal (Top trace) and its decomposition in different frequency bands : Delta (0.5-4Hz), Theta (4-8Hz), Alpha (8-12Hz), Beta (12-35Hz) and Gamma (>35 Hz). Image from Bajaj, N. (2020, November 12). Wavelets for EEG Analysis. IntechOpen. https://www.intechopen.com/chapters/74032

Frequency band	**Frequency**
Gamma	>35Hz
Beta	12-35Hz
Alpha	8-12Hz
Theta	4-8Hz
Delta	0.5-4Hz

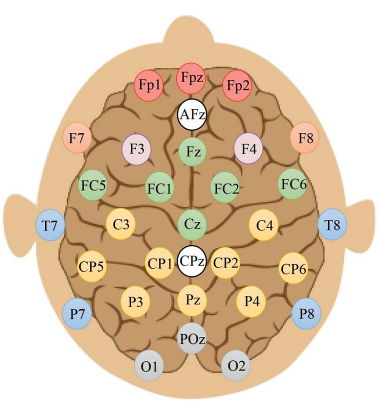
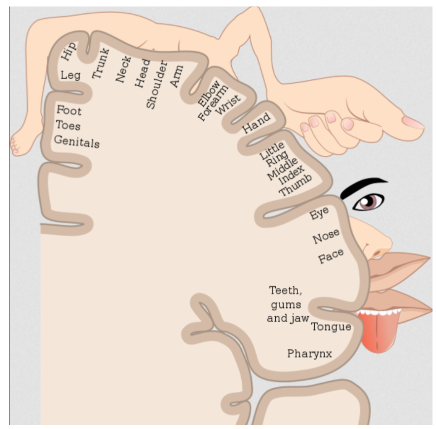


Figure 3 - This figure is a representation of a person's head seen from the top equipped with an EEG headset. Each circle corresponds to the position of one electrode. The electrode C3 in yellow allows the retrieval of information about hand movement. Image from Chai, M. T., Amin, H. U., Izhar, L. I., Saad, M. N. M., Abdul Rahman, M., Malik, A. S., & Tang, T. B. (2019). Exploring EEG Effective Connectivity Network in Estimating Influence of Color on Emotion and Memory. *Frontiers in Neuroinformatics*, *13*. https://doi.org/10.3389/fninf.2019.00066

When we initially look at this activity, we won't see much, since the signal is small and noisy (a lot of irrelevant signals making it difficult to analyze the brain activity of interest). Because of this, we have to **amplify the activity**, **separate it in several frequencies** (through a process known as a Fourier transform, [But what is the Fourier Transform? A visual introduction.]) and **clean it**. In fact, there is no situation in which brain activity is totally silent in normal conditions , hence there is a constant background noise which is different for each individual. We can also measure signals from muscle activity such as eye blinks or jaw moving. Because the brain activity we want to observe is small when measured at the scalp, these muscle movements produce huge perturbations in the signal, known as *muscle artefacts*. Those perturbations are also irrelevant and can be removed before analysing the data.

Figure 4 - The cortical homunculus shown on the right is a coronal slice of the somatosensory cortex which is represented along a body scaled so that each part of the cortex is attached to the body part it represents. For instance, an activity in the brain where "Thumb" is written will correspond to a movement of the thumb or the thumb being touched. Each body part is scaled to the size of the region it is represented in. For instance more neurons will respond to stimulus on the thumb than on the toes therefore the thumb is represented larger than the toes. Image from https://en.wikipedia.org/wiki/Cortical_homunculus, *accessed on 30-11-2021*



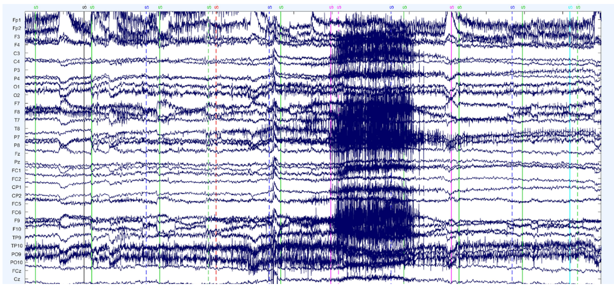


Figure 5 - The picture on the left shows the raw signal extracted from an EEG headset, each line corresponds to the activity in microvolt in one electrode. Image from Rodrigues, J., Weiß, M., Hewig, J., & Allen, J. J. B. (2021). EPOS: EEG Processing Open-Source Scripts. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.660449

To clean the signal from the background brain activity, we'll take the average brain activity during rest (baseline) that we will subtract from the considered signal, which is a technique called **baseline subtraction**. To get rid of artefacts caused by muscle activity, we usually use a reference electrode that we can put on the participants' ear, forehead or nose, which shouldn't have any neural activity. This electrode will be influenced only by electric activity from muscles of the face. With that, we can subtract the activity of this electrode from the overall brain activity. This should allow us to get a clean signal.

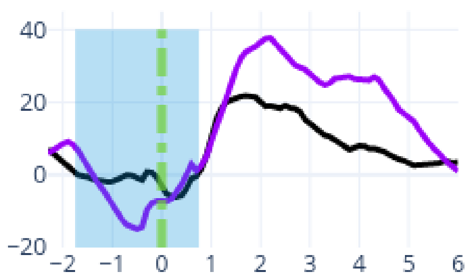


Figure 6 - In this picture we see in purple the activity in the alpha and beta frequency on the C3 electrode. It starts to decrease below 0 (ERD) before rising (ERS). In black is the activity induced in the same electrode by stimulating the median nerve (nerve controlling the muscles of the forearm), by producing an electric impulse on the wrist making the thumbs twitch. From unpublished data, Rimbert et al., 2019

From this cleaned signal, we'll see a particular pattern. Before performing a movement we see that the activity in alpha/beta frequency bands decreases in specific electrodes. That's called an **event-related-desynchronization (ERD)**, and it's thought to be linked to movement preparation. Directly afterwards we observe an increase in activity in the frequency bands, an **event-related-synchronization (ERS)**. Looking for this pattern in a particular zone of the somatosensory cortex can give us information about what part of the body should move and when. We've done that manually, but you can't rely on someone else to inspect your brain activity for every single movement that you want your prosthesis to perform. Hence we record the brain signal for some time during which we ask the subject to perform an action. Then we ask a machine learning algorithm to classify the signal recorded during a trial as presenting an action or not. In order to recognize an action, we have to perform a calibration, which consists of several trials during which brain activity will be recorded without being translated into any action. The goal is to set up patterns corresponding to rest or activity. During the actual task, the computer will compare the signal to activity registered during calibration and will determine based on how close it is to a brain activity emitted to produce a movement whether or not an action has been performed. This needs to be done for every individual using the BCI since the variability between individuals is very high. Hence a calibration for one individual will perform poorly on someone else. Also, as we discussed earlier, the detection of a movement is done by finding a chain of events (ERD/ERS). Hence feeding the classification algorithm one point at a time isn't relevant. As a result, we have to use a time window to classify the signal. All this preprocessing results in a delay of a few seconds between the production of the command in the brain and the reaction of the BCI.

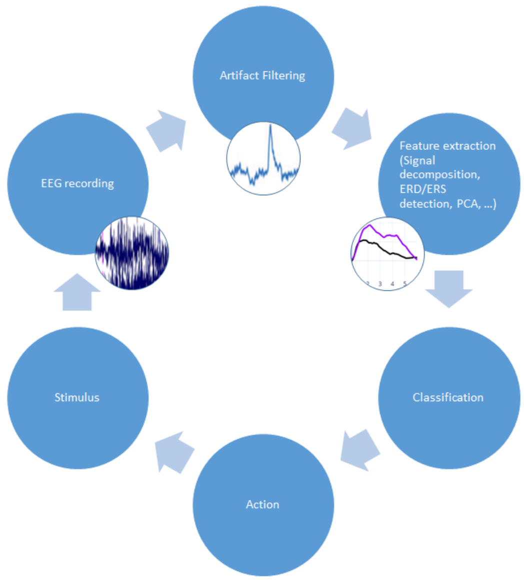


Figure 7 - When the brain receives a stimulus, it produces an activity recorded by the EEG. The recorded signal is filtered, decomposed in different frequency bands and analyzed. The results are sent to a classification algorithm and translated into an action which produces a new stimulus.

1.5.3 Different devices to record brain activity

The **EEG** headset measures the signal every millisecond, creating a high *temporal resolution* which is beneficial for the fast signaling to a machine. A downside of EEG is that you can only place a certain number of electrodes on someone's skull and can only measure activity in superficial cortical structures, creating a small *spatial resolution*. So, with EEG we can get a good idea of the dynamics, but we also have some missing information, such as the type of movement, the speed of the movement or its target. We only know that a movement of a particular limb has been performed. So we can't control a prosthetic limb with that information, we need a more accurate technique and all those techniques fall under the category of invasive BCIs, meaning that they require surgical intervention to be installed inside the head.

As stipulated in the name, **intracranial electroencephalography (iEEG)** is an invasive technique that consists of putting an electrode directly into the brain. Compared to EEG, iEEG has an increased spatial resolution. Some experiments on monkeys ask them to perform action on a screen thanks to iEEG. This technology actually allows you to get more information about the nature of the executed movement but suffers from invasiveness, you need to be operated to have an iEEG. It also requires, as for every BCIs for now, to stay in a controlled environment.
Paralysed woman moves robot with her mind - by Nature Video

Electrocorticography (ECoG) is a grid of electrodes that we put under the skull, on the surface of the brain and is in between EEG and iEEG, in terms of invasiveness and spatial resolution. In a project ongoing at the Clinatec research center in Grenoble, researchers are currently experimenting with ECoG to allow partially paralyzed people to control an exoskeleton. The video below shows a project developed by Neuralink. A monkey is trained to move a cursor on a screen and plays pong with his brain activity.
Monkey MindPong

Endovascular BCIs use sensors that are implanted using the blood vessels alimenting the brain. Contrary to other invasive techniques, their installation doesn't require removing a part of the skull. This technique has been used by synchron to allow paralyzed people to communicate using their brain activity.
[Paralysed man becomes the first person to tweet using only direct thought]

All those techniques allow us to read the top-down signal (initiated in the brain to produce motor command). Actual movement production requires the contribution of low level sensorimotor loops, which gives continuous feedback to the brain. This allows fast adaptation of movement trajectory or stabilization and are crucial for any movement production. With our current knowledge and understanding of the process, using myoelectrics information (electric information from the muscles) is still more efficient to control a prosthesis. Although there is still work to do in order to replace the control of a limb with BCIs, the field nonetheless has found new application.

1.5.3 Other applications

The **P300 speller** while not giving back mobility allows a person suffering from locked-in syndrome to communicate. The P300 is a signal appearing in the brain 300 ms after the sight of an unexpected stimulus. Scientists had the idea to use this brain activity combined with a screen with letters and numbers flashing at fixed intervals. The sight of a flashing letter produces a P300 signal, by comparing the timing of the flash to the timing of the P300, they were able to retrieve the letter seen by the subject. People suffering from locked-in syndrome are only able to move their eyes, hence the P300 speller is their best option to communicate. As a comparison, the device used by the theoretical physicist Stephen Hawking was based on the detection of muscle activity in his cheeks to communicate, so it wouldn't fit a person suffering from locked-in syndrome. The same signal has been used by Marco Congedo (more focused on the issue of calibration and classification algorithms) to create a game called brain-invader.
[Cognionics Dry EEG P300 Speller Demo]

The idea of **Neurofeedback** is that by showing the brain activity to subjects, they can willingly modulate it in a particular way. For instance a project called Grasp'It imagined by the researcher Laurent Bougrain and Sébastien Rimbert from neurosys at INRIA Nancy, aims at helping hemiplegic stroke survivors (who are paralysed on one entire side of their body) during their rehabilitation. Classic rehabilitation for hand movements asks patients to imagine their paralysed limbs moving by performing actions with their healthy hand and looking at the result in a mirror (mirror therapy). The movement being performed by activating the healthy hemisphere of the brain is reflected in the mirror so the visual feedback shows the opposite limb moving, activating *mirror neurons* in the damaged hemisphere. Mirror neurons activate in the motor cortex when we see a movement being performed. This technique helps rehabilitation by transferring the control from one hemisphere to another. Grasp'It based itself on this technique, instead of a mirror it is asked to the subject to look at a screen on which a bottle is displayed, the subject is then asked to imagine his paralysed limb pressing the bottle, based on the strength of the signal, more or less water is expelled from the bottle. Ongoing studies seem to show a comparable application for depression (see : Linden, D. E. J. 2014).

Another type of BCI aims at sending information to the brain, they are called **neuroprosthesis**. Cochlear implants replace the cochlea, a part of the inner ear which transforms sound waves into electrical information. This allows deaf people with a damaged cochlea to retrieve partial hearing. Artificial retina such as the Argus II plays the same role for sight but in a far less

efficient way. Bear in mind that these devices can't be used to solve those problems if the source is directly in the brain.
How the Argus II works

1.5.3 Want to know more?

Sources:
[1] Rimbert, S., Riff, P., Gayraud, N., Schmartz, D., & Bougrain, L. (2019). Median Nerve Stimulation Based BCI: A New Approach to Detect Intraoperative Awareness During General Anesthesia. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00622

[2] Marco Congedo, Matthieu Goyat, Nicolas Tarrin, Gelu Ionescu, Léo Varnet, et al.. "Brain Invaders": a prototype of an open-source P300- based video game working with the OpenViBE platform. BCI 2011 - 5th International Brain-Computer Interface Conference, Sep 2011, Graz, Austria. Pp.280-283. hal-00641412

[3] A. Eliseyev *et al*., "CLINATEC® BCI platform based on the ECoG-recording implant WIMAGINE® and the innovative signal-processing: Preclinical results," *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 1222-1225, doi: 10.1109/EMBC.2014.6943817.

[4] Sébastien Rimbert, Laurent Bougrain, Stéphanie Fleck. Learning How to Generate Kinesthetic Motor Imagery Using a BCI-based Learning Environment: a Comparative Study Based on Guided or Trial-and-Error Approaches. *SMC 2020 - International IEEE Conference on Systems, Man and Cybernetics*, Oct 2020, Toronto, Canada. ⟨hal-02920306⟩

Further reading:
Guger, C., Allison, B., & Ushiba, J. (2017). *Brain-Computer Interface Research: A State-of-the-Art Summary 5 (SpringerBriefs in Electrical and Computer Engineering)* (1st ed. 2017 ed.). Springer.

Linden, D. E. J. (2014). Neurofeedback and networks of depression. *Dialogues in Clinical Neuroscience*, *16*(1), 103–112. https://doi.org/10.31887/dcns.2014.16.1/dlinden

Kevric, J., & Subasi, A. (2017). Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. *Biomedical Signal Processing and Control*, *31*, 398–406. https://doi.org/10.1016/j.bspc.2016.09.007

1.5.4 Key terms

Brain-computer interface
Electroencephalography
Primary motor cortex
Muscle artefacts
Baseline subtraction
Event-related desynchronization
Event-related synchronization
Temporal resolution
Spatial resolution
Intracranial electroencephalography
Electrocorticography
P300 speller
Neurofeedback
Mirror neurons
Neuroprosthesis

Part 2: Machine learning

2.1 What is machine learning?

Author: Fleur Zeldenrust

Machine learning can be taken quite literally: it is the study of machines that learn. So it is about (mostly computer) systems that over time adapt to the task they have to fulfill. Often it is about computer algorithms (an algorithm is a set of rules or instructions that a computer follows, a computer program) that improve themselves by changing, using their experiences so that they can perform tasks better. They are often inspired by the best performing and most adaptive system that we know: the brain. Therefore, machine learning is a field in computer science and artificial intelligence, but shares a common fundamental question with computational neuroscience: How can systems learn?

The probably most well-known of the machine learning algorithms are so-called 'deep learning' algorithms, where networks of artificial 'neurons' are connected to one another consecutively. We will discuss this 'deep learning' method in 2.2, and another well-known neuroscience-based algorithm, reinforcement learning, in 2.3. We will also discuss algorithms that are based on evolution and natural selection: genetic algorithms (2.4). Finally, we will include some examples of the ethical implications of the use of these machine learning systems in our everyday lives (2.5).

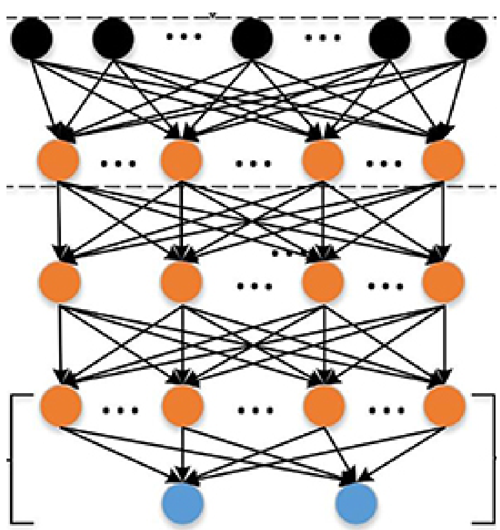


Figure 2.1.1 An example of a 'deep learning network'. Image from Guo X, Dominick KC, Minai AA, Li H, Erickson CA and Lu LJ (2017) Diagnosing Autism Spectrum Disorder from Brain Resting-State Functional Connectivity Patterns Using a Deep Neural Network with a Novel Feature Selection Method. *Front. Neurosci.* 11:460. doi: 10.3389/fnins.2017.00460

2.2 How 'deep networks' roughly work

Author: Tousif Jamal

Learning Objectives

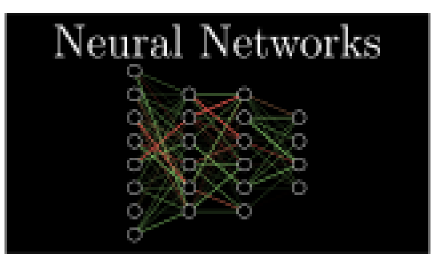
1. Be able to differentiate between machine learning and deep learning
2. Be able to explain the key steps in deep learning
3. Be familiar with biases in deep learning
4. Describe at least three applications of deep learning in healthcare

2.2.1 Introduction

Since the beginning of the first programmable computers, people have wondered if machines can be made more intelligent. It has always been desired to make a machine that can imitate the human brain. Today, Artificial Intelligence is a thriving field for many applications and research topics.

In the early days of artificial intelligence, the field rapidly tackled and solved problems that are intellectually difficult for human beings but relatively straightforward for computers. This is the case when the problems can be described by a list of formal, mathematical rules. The true challenge to artificial intelligence proved to be solving tasks that are easy for people to perform but hard for people to describe formally. These include, for instance, problems that we solve intuitively or that feel automatic, such as recognizing spoken words or faces in images.

To overcome the problem of more intuitive problems, we need to allow the computers to learn from experiences and have a better understanding of the world in terms of a hierarchy of concepts, with each concept defined through its relation to simpler concepts. By gathering knowledge from experience, the machine avoids the need for human operators to formally specify all the knowledge. The hierarchy of concepts enables the computer to learn complicated concepts by building them out of simpler ones. Which means a concept learnt by a machine is by a layer of much simpler concepts. For this reason this approach is called *deep learning*. For more detailed information and visualization you can have a look at these videos:



https://www.youtube.com/watch?v=aircAruvnKk

Deep learning is a subset of machine learning. *Machine learning* is a part of artificial Intelligence focused on using data and algorithms to solve a problem and gradually improve the accuracy, like the way humans learn. Deep learning is a way of classifying, clustering, and predicting things by using a neural network that has been trained on vast amounts of data. Deep learning neural networks attempt to mimic the human brain through a combination of data inputs, weights and biases. These elements work together to accurately recognize, classify and describe objects within the data. Deep learning distinguishes itself from classical machine learning by the type of data that it works with and the methods in which it learns.

2.2.2 The difference between deep learning and Machine learning

To acquire the knowledge for an AI system, it needs to be able to extract patterns from raw data. This capability is known as “Machine Learning”. Machine learning uses different learning algorithms to draw conclusions from the labeled data to make predictions. Which means that specific features are defined from the input data for the model and organise them into tables. This does not mean that it does not use unstructured or unlabeled data. If it uses unstructured data, then it goes through some pre-processing to organise it into a structured format. Deep learning eliminates some of the data pre-processing that is typically involved with machine learning. These algorithms are made to process unstructured data, like text and image. For example, let’s say we have a set of photos of different pets, and we want to categorize them as ”Dog”, ”Guinea pig”, ”Ferret” etc. Deep learning can determine which features(e.g. Nose, Ears) are most important to distinguish each animal from another. In machine learning, this hierarchy of features is established manually by a human programmer. Doing so, deep learning adjusts itself for accuracy, allowing it to make predictions about a new photo of an animal with increased precision.

2.2.3 How deep learning works

A major source of difficulty in many real-world artificial intelligence applications is that many of the factors of variation influence every single piece of data we are able to observe. The individual pixels in an image of a red car might be very close to black at night. The shape of the car’s silhouette depends on the viewing angle. Most applications require us to disentangle the factors of variation and discard the ones that we do not care about. Of course, it can be very difficult to extract such high-level abstract features from raw data. Many of these factors of variation, such as a speaker’s accent, can be identified using sophisticated, nearly human-level understanding of the data. When it is nearly as difficult to obtain a representation as to solve the original problem, representation learning does not, at first glance, seem to help us. Deep learning solves this central problem in representation learning by introducing representations

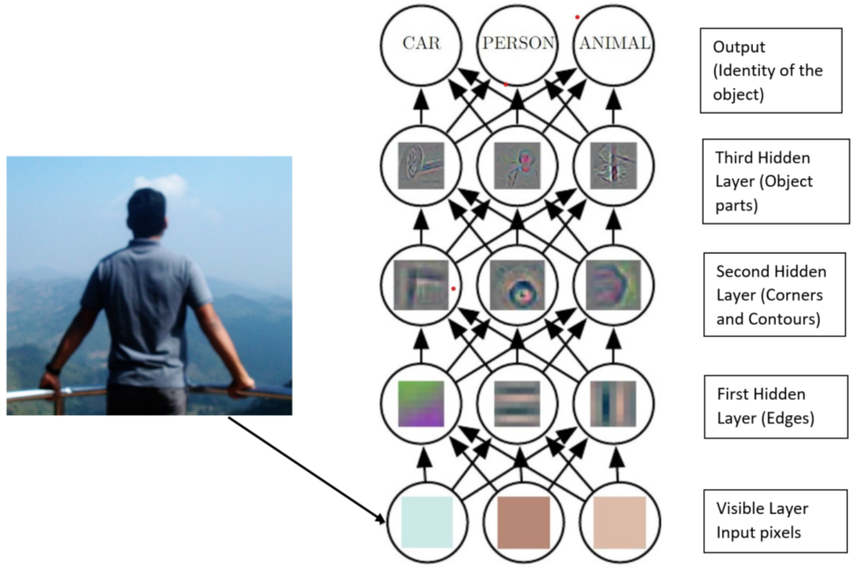


Figure 1: Illustration of a deep learning model

that are expressed in terms of other, simpler representations. Deep learning enables the computer to build complex concepts out of simpler concepts. Figure 1 shows how a deep learning system can represent the concept of an image of a person by combining simpler concepts, such as corners and contours, which are in turn defined in terms of edges. A typical example of a deep learning model is the feed forward deep network, or *multi layer perceptron* (MLP). A multi layer perceptron is a mathematical function mapping some set of input values to output values. The function is formed by composing many simpler functions. We can think of each application of a different mathematical function as providing a new representation of the input.

It is difficult for a computer to understand the meaning of raw sensory input data, such as this image represented as a collection of pixel values. The function mapping from a set of pixels to an object identity is very complicated. Learning or evaluating this mapping seems impossible if tackled directly. Deep learning resolves this difficulty by breaking the desired complicated mapping into a series of nested simple mappings, each described by a different layer of the model. The input is presented at the *visible layer*, so named because it contains the variables that we are able to observe. Then a series of *hidden layers* extracts increasingly abstract features from the image. These layers are called “hidden” because their values are not given in the data; instead the model must determine which concepts are useful for explaining the relationships in the observed data. The images here are visualizations of the kind of feature represented by each hidden unit. Given the pixels, the first layer can easily identify edges, by comparing the brightness of neighboring pixels. Given the first hidden layer’s description of the edges, the second hidden layer can easily search for corners and extended contours, which are recognizable as collections of edges. Given the second hidden layer’s description of the image in terms of corners and contours, the third hidden layer can detect entire parts of specific objects, by finding specific collections of contours and corners. Finally, this description of the image in terms of the object parts it contains can be used to recognize the objects present in the image.

2.2.4 Biases of deep learning

The learning algorithms mostly use a mechanism called “Inductive bias” or “Learning bias”. Inductive biases put some restrictions on either the space of hypotheses or can be said as the underlying model space. Here are some of the biases in deep learning

1. **Structured Perception And Relational Reasoning:**

   This bias was first introduced by researchers at DeepMind into Deep Reinforcement Learning (Deep RL) architectures. This approach improves performance, learning efficiency, generalisation, and interpretability of Deep RL models.

   By introducing structured perception and relational reasoning into Deep RL architectures, the reinforcement learning agents, i.e. the algorithm that draws conclusion from the previous outcomes. The algorithms are able to learn interpretable representations and exceed baseline agents in terms of sample complexity, ability to generalise, and overall performance. This approach can also offer advantages in meeting some of the most challenging test environments in modern artificial intelligence.

2. **Group Equivariance:**

   Equivariance is a good inductive bias for deep convolutional networks, which are special type of CNN. Convolutional neural networks are a specialized type of artificial neural networks that use a mathematical operation called *convolution* in place of general matrix multiplication in at least one of their layers. Convolution is a mathematical operation on two functions (f and g) that produces a third function (f*g) that expresses how the shape of one is modified by the other. Doing so reduces the sample complexity by exploiting symmetries in the networks. Research by the University of Amsterdam showed that the Group equivariant Convolutional Neural Networks (G-CNNs) used G-convolutions, a kind of layer that distributed the weights than the regular convolution layers.

   This layer increases the expressive capacity of the Convolutional Neural Network without increasing the number of parameters. According to its researchers, group convolution layers are easy to use and can be implemented with negligible computational overhead for discrete groups generated by translations, reflections, and rotations.

3. **Spectral Inductive Bias:**

   Spectral bias is an inductive bias or learning bias in deep networks that manifests itself not just in the process of learning but also in therepresentation of the model itself. The research was released in 2019 by Yoshua Bengio and his team. In this bias, the lower frequencies are learned first. According to researchers, this inductive bias’s properties are in line with the observation that over-parameterised networks prioritise learning simple patterns that generalise across data samples.

4. **Invariance and Equivariance Bias:**

   Invariance and equivariance bias can be used to encode the structure of the relational data. This kind of inductive bias notifies the behaviour of a model under various transformations. Equivariant models have been successfully used for various deep learning on data with various structures from translation equivariant image to geometric settings and discrete objects such as sets and graphs.

2.2.5 Applications of deep learning in healthcare:

1. **Identifying Diseases and Diagnosis:**

One of the chief ML applications in healthcare is the identification and diagnosis of diseases and ailments which are otherwise considered hard-to-diagnose. This can include anything from cancers which are tough to catch during the initial stages, to other genetic diseases. IBM Watson Genomics is a prime example of how integrating cognitive computing, which is a computerized models to simulate the human thought process in complex situations where the answers may be ambiguous and uncertain. with genome-based tumor sequencing can help in making a fast diagnosis. Berg, the biopharma giant, is leveraging AI to develop therapeutic treatments in areas such as oncology. P1vital's PReDicT (Predicting Response to Depression Treatment) aims to develop a commercially feasible way to diagnose and provide treatment in routine clinical conditions.

2. Drug Discovery and Manufacturing:

One of the primary clinical applications of machine learning lies in the early stage drug discovery process. This also includes technologies such as next-generation sequencing and precision medicine which can help in finding alternative paths for therapy of multifactorial diseases. Currently, the machine learning techniques involve algorithms that identify patterns in data sets containing unclassified or data points that are not labeled, known as *unsupervised learning*. Project Hanover developed by Microsoft is using ML-based technologies for multiple initiatives including developing AI-based technology for cancer treatment and personalizing drug combinations for a disease known as acute myeloid leukemia or AML.

3. Medical Imaging Diagnosis:

Machine learning and deep learning are both responsible for the breakthrough technology called *computer vision*, a field of AI that enables computers and systems to derive meaningful information from digital images, videos, other visual inputs and make recommendations based on the information. This has found acceptance in the InnerEye initiative developed by Microsoft which works on image diagnostic tools for image analysis. As machine learning becomes more accessible and as they grow in their explanatory capacity, expect to see more data sources from varied medical imagery become a part of this AI-driven diagnostic process.

4. Personalized Medicine:

Personalized treatments can not only be more effective by pairing individual health with predictive analytics but are also ripe for further research and better disease assessment. Currently, physicians are limited to choosing from a specific set of diagnoses or estimate the risk to the patient based on their symptomatic history and available genetic information. But machine learning in medicine is making great strides, and IBM Watson Oncology is at the forefront of this movement by leveraging patient medical history to help generate multiple treatment options. In the coming years, we will see more devices and biosensors with sophisticated health measurement capabilities hit the market, allowing more data to become readily available for such cutting-edge ML-based healthcare technologies.

5. Machine Learning-based Behavioral Modification:

Behavioral modification is an important part of preventive medicine, and ever since the proliferation of machine learning in healthcare, countless startups are cropping up in the fields of cancer prevention and identification, patient treatment, etc.

6. Smart Health Records:

Maintaining up-to-date health records is an exhaustive process, and while technology has played its part in easing the data entry process, the truth is that even now, a majority of the processes take a lot of time to complete. The main role of machine learning in healthcare is to ease processes to save time, effort, and money. Document classification methods using vector machines and ML-based text recognition techniques are slowly gathering steam, such as Google's Cloud Vision API and MATLAB's machine learning-based handwriting recognition technology. MIT is today at the cutting edge of developing the next generation of intelligent, smart health records, which will incorporate ML-based tools from the ground up to help with diagnosis, clinical treatment suggestions, etc.

7. **Clinical Trials and Research:**

   Machine learning has several potential applications in the field of clinical trials and research. As anybody in the pharma industry would tell you, clinical trials cost a lot of time and money and can take years to complete in many cases. Applying ML-based predictive analytics to identify potential clinical trial candidates can help researchers draw a pool from a wide variety of data points, such as previous doctor visits, social media, etc. Machine learning has also found usage in ensuring real-time monitoring and data access of the trial participants, finding the best sample size to be tested, and leveraging the power of electronic records to reduce data-based errors.

8. **Crowdsourced Data Collection:**

   Crowdsourcing is all the rage in the medical field nowadays, allowing researchers and practitioners to access a vast amount of information uploaded by people based on their own consent. This live health data has great ramifications in the way medicine will be perceived down the line. Apple's ResearchKit allows users to access interactive apps which apply ML-based facial recognition to try and treat Asperger's and Parkinson's disease. IBM recently partnered with Medtronic to decipher, accumulate, and make available diabetes and insulin data in real time based on the crowdsourced information. With the advancements being made in the 'Internet of Things' (IoT), the healthcare industry is still discovering new ways in which to use this data and tackle tough-to-diagnose cases and help in the overall improvement of diagnosis and medication.

9. **Better Radiotherapy**:

   One of the most sought-after applications of machine learning in health care is in the field of Radiology. Medical image analysis has many discrete variables which can arise at any particular moment of time. There are many lesions, cancer foci, etc. which cannot be simply modeled using complex equations. Since ML-based algorithms learn from the multitude of different samples available on-hand, it becomes easier to diagnose and find the variables. One of the most popular uses of machine learning in medical image analysis is the classification of objects such as lesions into categories such as normal or abnormal, lesion or non-lesion, etc. Google's DeepMind Health is actively helping researchers in UCLH develop algorithms which can detect the difference between healthy and cancerous tissue and improve radiation treatment for the same.

10. **Outbreak Prediction:**

    AI-based technologies and machine learning are today also being put to use in monitoring and predicting epidemics around the world. Today, scientists have access to a large amount of data collected from satellites, real-time social media updates, website information, etc. Artificial neural networks help to collate this information and predict everything from malaria outbreaks to severe chronic infectious diseases. Predicting these outbreaks is especially helpful in third-world countries as they lack crucial medical infrastructure and educational systems. A primary example of this is the ProMED-mail, an

Internet-based reporting platform which monitors evolving diseases and emerging ones and provides outbreak reports in real-time.

2.2.6 Want to know more?

Sources:

1. The spatial inductive bias of deep learning, B. R. Mitchell, 2017. For Spectral inductive bias.
2. Equivalent entity relationship networks, Anonymous authors, ICLR conference, 2020. for invariance and equivariance bias.
3. Group equivariant convolutional networks, T. S. Cohen and M. Welling, arxiv-1602.07576v3. For group equivarace bias.
4. Deep reinforcement learning with relational inductive biases, V. Zambaldi et al., ICLR conference, 2019. For structured perception and relational reasoning.
5. www.analyticsindiadimag.com/top-5-induictive-biases-in-deep-learning-models/
6. www.towardsdatascience.com/why-we-need-bias-in-neural-networks-db8f7e07cb98

2.2.7 Key terms

Machine learning	Multi layer perceptron	Hidden layer
Deep learning	Visible layer	Unsupervised learning

2.3 Reinforcement learning, classical conditioning and Rescorla Wagner

How the mathematical formulas relate to biological function and have been used in computer science

Author: Matt Lennon

Learning Objectives

1. Define reinforcement learning.
2. Be able to explain the concept of classical conditioning
3. Understand the Rescorla-Wagner model
4. Understand the biological function of dopamine and the mesolimbic system
5. Understand how the biology of reinforcement learning led to the development of reinforcement learning for machines

2.3.1 Introduction

In this chapter we will discuss *reinforcement learning*, which basically means something like 'learning by using rewards and punishments'. Obviously, animals and humans can learn by getting rewards and punishments, something that is explored in section 2.3.2 of this chapter. However, what we know about how animals learn has inspired a set of computer algorithms to make machines learn. This part of machine learning is confusingly also called reinforcement learning, and will be explored in section 2.3.3.

2.3.2 Reinforcement learning in biology

Humans, amongst other creatures, make decisions, plans and goals based on previous experiences. Expectations are set by past experiences and thus each additional experience forms part of the library used to set future expectations. This process is called reinforcement learning. Classical conditioning is one example of reinforcement learning. In brief, classical conditioning involves a subject (e.g. a dog) exposed to an unconditioned stimulus (US) (e.g. a T-bone steak) that produces an unconditioned response (UR) (e.g. salivation). This US is paired with a conditioned stimulus (CS) (e.g. a ringing bell) such that the CS eventually produces a conditioned response (CR) similar to the UR (see figure 1).

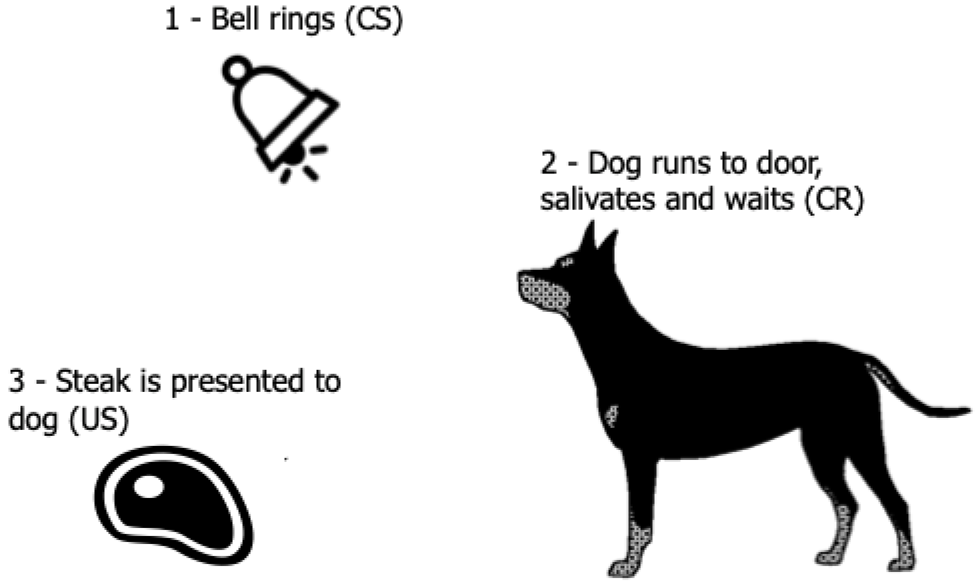


Figure 1 - Illustration of established classical conditioning. CS - Conditioned Stimulus, CR - Conditioned Response, US - Unconditioned Stimulus

Rescorla and Wagner [1] were two psychologists at Yale University who in 1972 published their now famous mathematical model of classical conditioning, where they conceptualised learning by the strength of the association between CS and US. A conditioned stimulus that reliably predicts an unconditioned stimulus will thus generate a strong learned association, whereas a CS that unreliably predicts the US will generate a weaker learned association. The Rescorla-Wagner model considers this learning as a process completed over a number of discrete trials (e.g. every evening when the dog's dinner is served). Within each trial the CS (e.g. ringing bell) may either predict the US (e.g. it is followed by a T bone steak) or it may not predict the US (e.g. no dinner is given). Within each trial the association is either strengthened, weakened or remains unchanged. This is summed up by the following three equations:

$$V_{t+1} = V_t + \eta\left(r_t - V_t\right) \quad (1)$$

- V is the predicted reward (i.e. strength of the association) at time t
- η is the learning constant, a variable that determines how quickly associations are learnt and forgotten
- r is the actual size of the reward at time t

What the above equation tells us is that the strength of the association between the CS and US (V) at time t + 1 will be the same as V at time t but modified by the difference between the reward size and the association strength at time t. What does this mean in practice? Going back to our dog example, say a dog has been trained to salivate at the sound of a bell because the bell predicts one T-bone steak, but on this particular night the owner is feeling generous and gives the dog two T-bone steaks. The size of the reward is unexpected or surprising and thus the difference between the reward and the association strength will be large ($r_t - V_t$). Consequently the change in V will be considerable and the association will be strengthened. The difference between the actual reward (r_t) and the predicted reward (V_t) is called the *reward* *prediction error (RPE)*. Keeping in mind that the 'reward' may not be present or may be negative or aversive causing the learnt positive association to weaken.

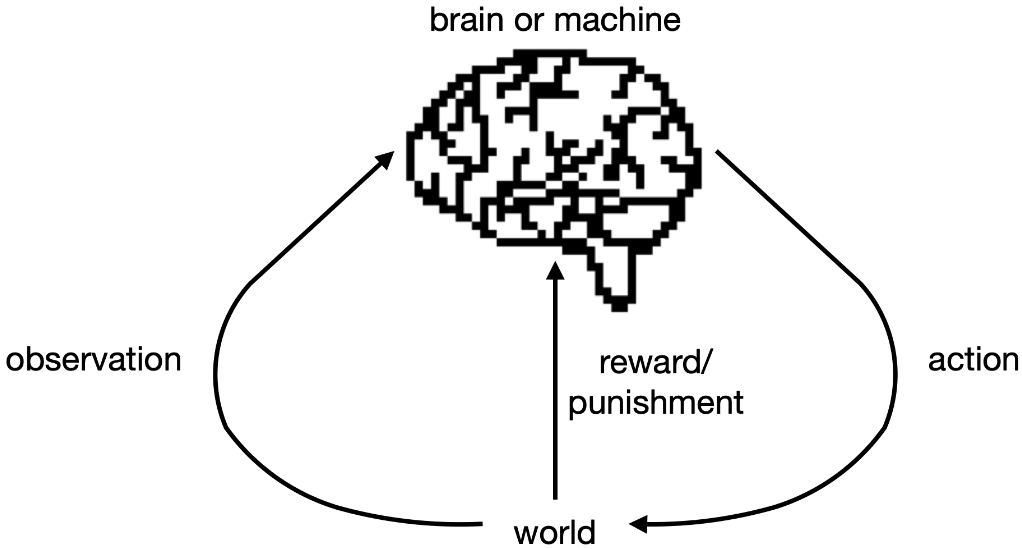


Figure 2 - Diagram of a reinforcement learning process.

If we expand the brackets of equation 1 and then group by V_t we get equation two

$$V_{t+1} = (1 - \eta)V_t + \eta r_t \quad (2)$$

Then if we consider that V_t may be considered as a function of V_{t-1} and V_{t-1} a function of V_{t-2} and so on and so forth eventually the equation 3 is produced.

$$V_t = \eta \sum_{i=1}^{t} (1 - \eta)^{t-i} r_i \quad (3)$$

What equation 3 tells us is that the current predicted reward or strength of the association (Vt) is a weighted sum of all previous rewards with the more recent trials being weighted more heavily than the previous ones.

The beautiful simplicity and elegance of this formula that models learning is what gets computational neuroscientists excited, particularly because it explains the underlying biology reasonably well. This mathematical model predicts behaviour of mice, rats and monkeys, for example Rescorla found in mice the degree of the fear responses (CR) to a tone (CS) were directly proportional to the probability of its association with foot shocks (US) and was accurately predicted by their mathematical model [2],[3].

Dopamine, a neurotransmitter involved in pleasure or reward, is a critical signalling molecule for this type of learning. As seen in the Rescorla-Wagner equations, learning occurs when there is a difference between the expected and actual rewards of an event or action, also called the reward prediction error. In studies of mice and rats they found that dopamine burst firing occurs in the mesolimbic system proportional to the size of the reward prediction error. What this means is the mouse may expect only one pellet with the light turning on (CS) and in fact they receive two, this will be a prediction error of 1 pellet, and will be translated into a future larger dopamine reward spike. Thus the association between the light and the food reward is strengthened. Equally there is a burst suppression of dopamine firing when the subject is surprised about the absence of a reward. Thus a mouse given no pellets, despite expectations, will feel the transient depletion of mesolimbic dopamine and the learnt association will weaken.

In summary, Rescorla and Wagner developed a simple and elegant mathematical model describing how learning occurs over multiple trials and this model has been verified in a number of behavioural and neurophysiological experiments. The challenges of this model of learning include explaining how associations are made across time and the rapid reacquisition of learnt behaviours, more advanced theories including temporal difference learning have continued to explicate the complexities of reinforcement learning.

2.3.3 Reinforcement learning in machines

The biological observations in the previous section inspired a set of algorithms in computer science that are also called 'reinforcement learning'. It is often used for robots or other machines that have to actively interact with an environment. These algorithms mimic what animals also have to do when exploring an environment: a feedback loop of perception (What do I see/feel/hear/smell around me), choosing an action (Do I stay here or move? Where should O go?) based on an expectation of reward or punishment. Once the action has been performed, a new cycle and choosing an action starts, until a reward or punishment is received. The whole system then gets updated (the same way as in figure 2 and 3), so that in the future (hopefully) more reward and less punishment will be received. So in these algorithms, time is divided into small steps, where at every step, the 'agent' (the brain or machine) performs an action and receives an observation and/or an award or punishment from the world. The world on the other hand, receives the action and sends an observation and an award or punishment. The machine or brain keeps track of the history of what happened, by updating a table of what actions lead to what rewards or punishments. Based on this table, it can choose what action to perform at each moment in time. This table, basically the predictions of the amounts of reward related to each action at each point in the world, is called a *'value function'*. The relations between what state the system is in to what actions it should perform, is called a *'policy'*. This type of learning is also called *'actor-critic'* learning: the machine is trying to learn the best policies. The 'actor' in the machines chooses at every step an action based on the value function. The 'critic' evaluates after the chosen action the new state of the system, and whether this is better or worse than expected. It then updates the value function.

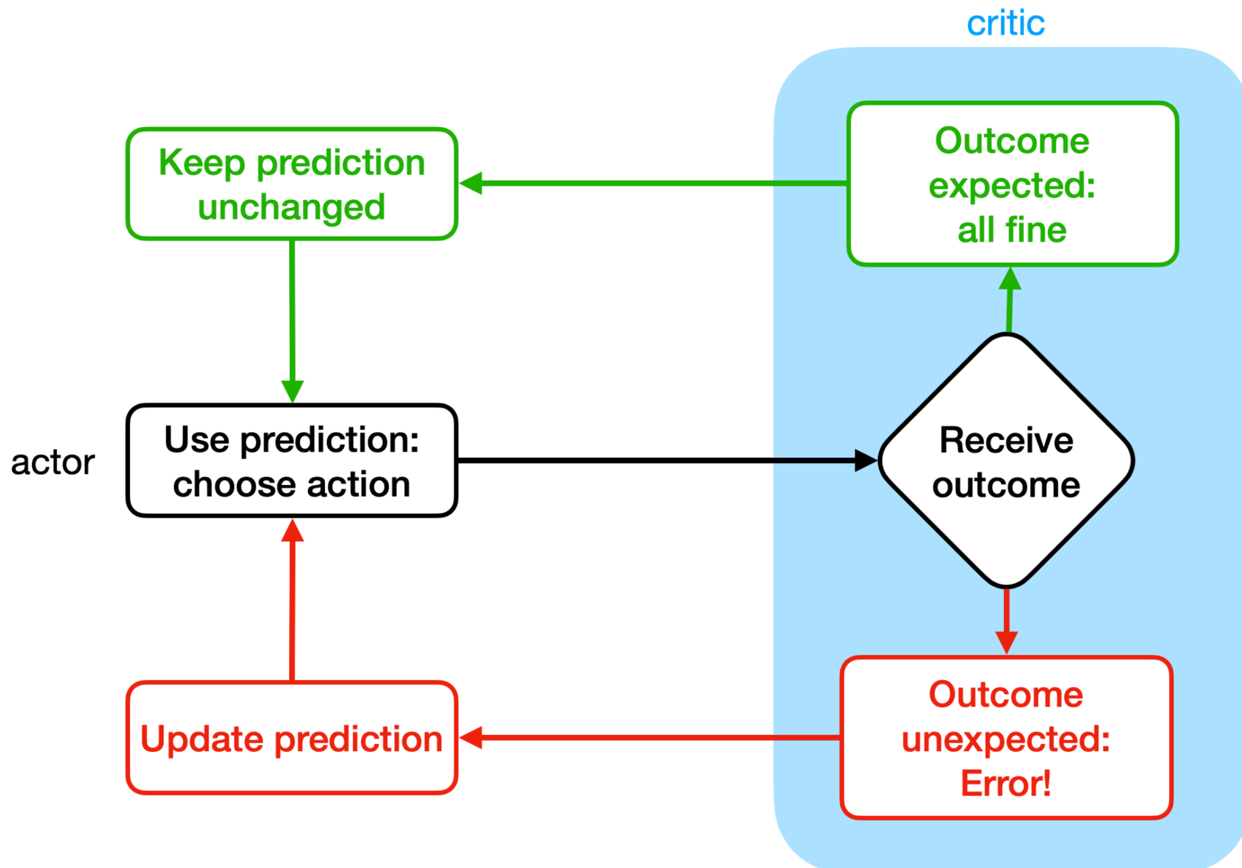


Figure 3 - Diagram of a reinforcement learning process. In red one sees what happens if the expected reward is different from the predicted reward (i.e. the system is updated), in green what happens if the expected and received award is the same (nothing changes). In black the actor, the blue box is the critic.

There is an important dilemma in reinforcement learning, called the *'exploration-exploitation dilemma'*. In order for a system to know where the most reward can be found, it needs to explore its environment, i.e. go to places or do things it has never done before. How else can it know that the actions that it chose were the best? On the other hand, when it has found a rewarding set of actions, it is wise to keep doing this (exploitation), as doing new things might result in less rewards or even punishments. So when does one stop exploring and do only the things that one knows are rewarding? This will depend on a lot of factors, including what the environment looks like (are there harsh punishments expected) and on the system/personality (some people or machines are more risk-aversive than others). In machine learning, machines often start with more exploration, and gradually do more exploitation. This gradual change between exploration and exploitation is called *'annealing'*.

2.3.4 Want to know more?

Sources:

[1] Rescorla RA (1972) “Configural” conditioning in discrete-trial bar pressing. *J. Comp. Physiol. Psychol.* **79**, 307–317.
[2] Rescorla RA (1968) Probability of shock in the presence and absence of CS in fear conditioning. *J. Comp. Physiol. Psychol.* **66**, 1–5.
[3] Schultz W. Updating dopamine reward signals. Curr Opin Neurobiol. 2013 Apr;23(2):229-38. doi: 10.1016/j.conb.2012.11.012. Epub 2012 Dec 22. PMID: 23267662; PMCID: PMC3866681.

Further reading:

- Florentin Woergoetter and Bernd Porr (2008) Reinforcement learning. Scholarpedia, 3(3):1448.
- Robert Rescorla (2008) Rescorla-Wagner model. Scholarpedia, 3(3):2237.

2.3.5 Key terms

Reinforcement learning	Conditioned response	Policy
Classical conditioning	Reward prediction error	Actor-critic
Unconditioned stimulus	Dopamine	Exploration-exploitation dilemma
Unconditioned response	Mesolimbic system	Annealing
Conditioned stimulus	Value function	

2.4 Genetic algorithms

Author: Nishant Joshi

Learning Objectives

1. Differentiate evolution by natural selection from evolution by sexual selection.
2. Understand the key steps in the development of a genetic algorithm

2.4.1 Introduction

We humans have been on a constant quest of understanding how nature works. Especially since we want to reduce the uncertainty regarding the environment we are situated in. Our brains have evolved to understand complex patterns which help in making better models of the nature of things and behavior around us. Evolution in itself has been a topic of interest for scientists since classical antiquity but it was consolidated when Darwin wrote his findings in the famous “On the Origin of Species”. Biological organisms adapt to their surroundings in order to survive, this could be either through becoming better hunters or through becoming better at finding food and evading predators. Most organisms evolve through natural selection and sexual selection.

Natural selection is a crucial step in the process of evolution which allows species with certain traits that are more suitable for survival in the environment to thrive. Each population possesses individuals with variational traits which is a result of random genetic mutation. Some of these traits can make the individuals better suited to the environment that they live in and when it passes on to the next generation, it allows the species as a whole to prosper. With *sexual selection* which is another mode of natural selection, partners choose each other based on the features which would maximize the survival of the offspring.

This idea of evolution has been imbued into computer science as well. Computer scientists have taken the idea of evolution, which is based on natural selection and sexual selection, and tried to create an algorithm (a recipe for solving a particular problem) that adapts to the problem and creates a range of solutions to a particular problem, rather than a single fixed solution. An example problem that might require such an approach might be to teach a computer bipedal figure to walk. A Genetic Algorithm which is an evolutionary algorithm is one such example. In this algorithm, natural selection is viewed as a range of potential solutions that are being scored based on their performance iteratively. In computer science applications, the next step is to eliminate the solutions which score poorly. This elimination takes place to remove the possibilities of creating a weaker solution which might be sub-optimal to a particular problem. Since it is an iterative process we only need a solution to survive which is moving the solution towards optimality and has a high chance of producing a better solution “offspring”.

The real challenge arises when a program needs to “mate” and create offspring(s) that have to be seemingly better than the parents. The problem that occurred was in designing a genetic code for computer programs that resemble the working mechanisms of chromosomes in biological organisms. In the 1960s, Hans J. Bremermann of the University of California at Berkeley came up with a technique to encode the genetic code in a computer program. The technique involved a binary string that represents the possible features/chromosomes of the organism. His solution was simple, just summing up the code from both the parents. And then flipping the bits (aka changing from 0 to 1 or vice versa) for the features which solve the problems more accurately. An example could be, let's say, a program that is given the task of classifying a dolphin. The program would produce (many) potential solutions with different combinations of the following features: 'dorsal fins', 'blow hole', 'long mouth' etc. Such a solution would then consist of a list of zeros and ones (a bit string), indicating the lack or presence of a feature, respectively. The position of each number then corresponds with a particular feature (Figure 1). Now it doesn't necessarily have to start there, a programmer can choose to have any random feature in the bit string such as 'gills', 'hair', 'legs' etc but it has to be modified in such a way that the useless features are eliminated.

['gills', 'fins', 'beak', 'legs']	['gills', 'fins', 'beak', 'legs']
[0 1 0 0]	[1 1 0 0]

Figure 1. An example of Mutation

Fitness function:

The goal of any Genetic Algorithm is to maximize the *fitness function* of the solutions. It is analogous to the ecosystem or environment that the organism is living in. The fitness function should be designed in a way to facilitate the evolution of the previous generation solutions. In our dolphin example, the fitness function could be the level of oxygen in the water so as to facilitate having gills.

2.4.2 Steps of the Algorithm

Let's go through different stages of a Genetic Algorithm

Initial stage: here the program constructs an initial population that can be made to evolve in the next stage.
Initial population: Depending on the problem that needs to be solved, a number of potential solutions can exist. These initial solutions become the initial population that will eventually mutate.

Evolutionary stage: here the program repeats three steps (selection, crossover, and mutation) to produce a new population. With each generation, the fitness of (some) solutions is expected to be greater than the fittest solution of the previous generation. If after some generations the fitness does not (substantially) increase anymore, the program moves onto the final stage.
Selection: In this part of the algorithm, parents are selected from the population pool for breeding purposes. The selection is based on the fitness of the parents, so how well they score on the fitness function. **Crossover:** In this step, information is merged from the parents to produce the offspring. **Mutation:** In this operation, the offsprings have a slight change in their chromosome with a low probability

Final stage: in this stage one or multiple final solutions with the highest fitness are returned by the program.
Termination: This is the final stage in the algorithm. When the chromosomes stop changing compared to the previous generations, it indicates that the algorithm has converged to a particular solution or offspring.

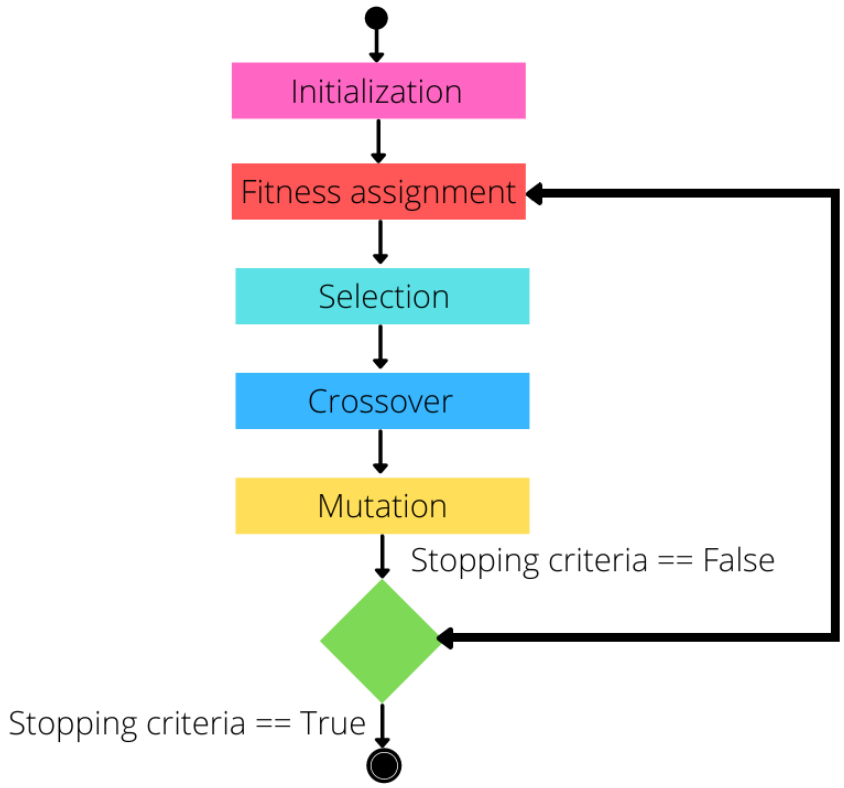


Fig 2. Flowchart representing different stages of the GA

2.4.3 Example

Problem: To maximize the digit-wise sum of a 10-bit string.

?	?	?	?	?	?	?	?	?	?

Since we already know the answer to this problem, we can go through the steps of the genetic algorithm and have a clear understanding.

Initialization:

We start by choosing a random set of 4 10-bit strings.

1.

1	0	0	0	1	1	0	0	1	0

2.

0	0	1	1	1	0	0	0	0	0

3.

0	0	1	1	1	0	0	1	0	1

4.

0	0	0	0	1	1	1	1	1	1

Fitness Function:
The fitness function is simple. Just the individual sum of each string:
String 1 = 4
String 2 = 3
String 3 = 5
String 4 = 6

Selection:
In this step, we select two random strings. We can be more sophisticated than that and design selection criteria based on the sum for example, but we wish to be as crude as possible. We do not always select the best chromosomes since there is a risk of running into local optima.

Let's consider strings 2 and 4 as our random selection.

Crossover:
We can divide both the strings into two same arbitrary-length strings. Let's choose our random point of division to be 4.

String 2:

0	0	1	1	1	0	0	0	0	0

String 4:

0	0	0	0	1	1	1	1	1	1

We cross and connect the two strings with each other to get a new pair of strings.

String 2:

0	0	1	1	1	1	1	1	1	1

String 4:

0	0	0	0	1	0	0	0	0	0

These are the two offspring that are generated.

Now we have to repeat the process and choose 2 random strings for crossover and create a new generation of 4 strings.

Mutation:
After the crossover step, we can randomly flip a value from one of the strings or both the strings. This will be our mutation step. For this example, let's flip the value for position 3 in string 2, so the new strings are following.
String 2:

0	0	0	1	1	1	1	1	1	1

String 4:

0	0	0	0	1	0	0	0	0	0

We then calculate the fitness function for the strings again and repeat the process until it converges to the maximum value.

2.4.4 Applications in Neuroscience

- Fitting model parameters: One of the most important applications of GA in neuroscience is finding the best-fit parameters for the neuron models to fit electrophysiological recordings. An example of this is fitting a conductance-based compartmental model of neurons to electrophysiological recordings, the parameter space for such a model is huge and it requires an extensive search to find the parameters that fit the model. Genetic Algorithm can help in finding the best parameters by making the search more sparse.
- Another application is to model the ion channel conductance and kinetics of pyramidal neurons in monkeys.

2.4.5 Limitations

- One of the limitations of this algorithm is that it is computationally viable for solving hard problems.
- Another limitation of this algorithm is that it is computationally expensive to calculate the fitness function after every iteration.
- Even after a lot of computation optimality is not guaranteed.
- One of the things that might be a hindrance for choosing GA over another optimizing algorithm is that the solution of the algorithms depends on the implementation and design choices associated with it such as selection criteria for mutation, initial conditions, and string representation.

2.4.6 Want to know more?

An interesting visualization of the genetic algorithm at play:
NetLogo Web

Readings:
An Introduction to Genetic Algorithms

Videos:
How algorithms evolve (Genetic Algorithms)

Interesting toolbox for playing around
https://rednuht.org/genetic_walkers/

2.4.7 Key terms

Evolution	Natural selection	Sexual selection
Genetic algorithm	Fitness function	

2.5 AI and Neuroscience applications outside of the brain: ethics and impact

Author: Nishant Johshi

Learning Objectives

1. Be familiar with various applications of artificial intelligence (AI) in society
2. Understand that AI raises ethical questions with real-world societal consequences
3. Be familiar with ways that AI might impact human societies
4. Understand the high energy costs of global computing and need for sustainability

2.5.1 Introduction

The field of Artificial Intelligence draws a lot of inspiration from the findings in Psychology and Neuroscience. The goal of artificial intelligence is to understand human cognition and model it into a machine that has the same or similar kind of intelligent behavior, such as thinking rationally, learning, and problem-solving. Neuroscience aids in modeling these human-intelligence characteristics by explaining how these characteristics emerge from the behavior of neurons. Since these two fields are highly important in the coming future, it is important to understand their application outside the brain, the ethical problems, and the impact it is going to have on our lives.

2.5.2 Applications outside the brain

There could be a wide range of applications for the things we can learn from Neuroscience, especially in the field of Artificial Intelligence, that can be applied outside the brain. Some of the use cases where these two fields merge together are listed below:

Marketing: Marketing is one such use case where Artificial Intelligence and Neuroscience can work in conjunction. As it is commonly understood that consumer behavior is largely the result of neural activities when the consumer searches for a product or service online. The price, colors, discounts, etc. pertain to specific activities which can be tracked through simpler tools, such as eye-tracking, facial expression, or arousal. And some more complicated methods such as fMRI or EEG. Data accumulated through these methods while the consumer engages in the shopping process can help companies filter products and services that fit the needs of the consumers better. It has also been shown that brain data helps to predict the success of the products better than surveys. [1.]

Consciousness: Since conscious human experiences are a result of activities in the brain, it is a challenging endeavor for scientists to explain how conscious experiences arise as a result of neural activities in a detailed fashion. It can be divided into subcategories:

1. *Generic Consciousness*: How do neural activities explain when a state is conscious? [2.]
2. *Specific Consciousness*: What the contents of the conscious state might be? [2.]

Since consciousness is a subjective phenomenon, there is no good way to compare two consciousnesses. Neuroscientists currently rely on introspective reports and brain data to compare and contrast conscious experiences.

Another subject of interest is how consciousness emerges from the elements that make the brain conscious. Such as a single neuron, neural circuits, the central nervous system, and so forth. It is interesting to correlate the phenomenon of consciousness with the activities at different levels of elements such as spikes or firing rate at single neuron level and local field potentials at the circuit level. This connection between activities and experience is not very well understood. [2.]

Sleep: Neuroscientists and AI experts are also interested in sleep. The current state of sleep research involves the effect of stimuli, such as music, white noise, lights, etc., on human sleep. These kinds of studies require the participants to wear a headset with electrodes or a headband to track mental activities while the subject is asleep. The recorded activities are then analyzed using sophisticated AI-based software which helps the researchers to find patterns in the activities while the stimulus is on.

Lucid dreaming is another fascinating area of research where neuroscientists and AI experts join forces. Using well-defined methods, such as Wake Initiated Lucid Dreaming (WILD), the subject is transferred into a lucid dreaming state. The subject is given signals to make them lucid and asked to move their eyes once they become lucid. The brain activity is recorded for the entire duration of the study and then matched with the subject's verbal description. Using AI methods, it is currently possible to classify if the subject is in a REM state or not. [3.]

Entertainment: Cinema is a media, consisting of sound, pictures, and stories interwoven together. It is very well known that cinema can have serious effects on the human brain. It is a powerful media that can incite a range of emotions and behavior. Cognitive neuroscientists are interested in understanding the range of effects cinema can have on the human brain in terms of activities. It is also interesting to understand how varied these effects are between people and different groups of people. Cognitive neuroscientists are also exploring how to use cinema to understand consciousness.

Artificial intelligence is becoming prevalent in the gaming industry. After the famous victory of Deep Blue over world champion chess player Garry Kasparov, AlphaGo [4.], which is a famous Go engine designed by DeepMinds, proved to open new avenues for the gaming industry. Current gaming engines are multiple times faster and better than the best players and thus they provide help for players to learn new strategies and compete better against their opponents. These game engines use Reinforcement Learning algorithms which optimize the strategies for the game in order to win or maximize its points.

Neuro-robotics: Robots are autonomous agents with a limited model of their surroundings, to make robust decisions about their movements and actions, they need to constantly update their models based on their perception. The nervous system in animals serves as a huge inspiration on which these models can be based. Neuro-robotics is an emerging field that applies knowledge from the nervous system in modeling the decision-making ability of robots.

As an example for modeling locomotion and motor control, Kimura et al., 2007 have shown that 4-legged robots can learn to walk based on an embedded pattern recognition system modulated by reflexes. This aspect of neuro-robotics deals with designing motor control in robots which mimics the way animals move and control their bodies. An example of this type of control system is the embedding of a mirror neuron system which makes the robot mimic the action performed by humans while grasping or moving objects. [5.]

Another key aspect of naturally inspired systems is the ability to retain information in memory regarding its environment. Researchers have been designing navigation systems for robots based on rodent brains. Recently, a large-scale system-level model of the hippocampus has been placed in a robotic system to investigate the role of different systems in the collection and retrieval of episodic memory. [5.]

The important characteristic of an autonomous robotic system is the ability to make decisions and act on them. Actions in mammals are value-driven, an example of that is dopamine and other neuromodulators which bias the synaptic connection in a positive way in case of a positive event and bias in the other direction in case of the negative event. Similar systems can be designed in a neuro-robot that motivates an action based on the reward it accumulates. [5.]

Neuromorphic computing

It has been the ultimate goal of AI to reach human-level cognition and beyond which involves thinking and reasoning like humans do. *Neuromorphic computing* is aimed at achieving this goal. These next-generation computers will have processing units that are spiking neurons. With a network of spiking neurons, researchers aim at building a computer that is energy efficient as the brain, could learn from unstructured stimuli or data, and has similar flexibility as the human brain. [6.]

Spiking networks are a simplified model of neurons in the brain. Spikes are events in the neurons when the voltage across membrane potential reaches a certain threshold and slowly fades to a resting state. This event can be mathematically modeled into a matrix-like structure which is an approximation for a neural network in the brain. These spiking networks are important tools for neuroscientists and AI scientists to model neurological phenomena.

2.5.3 Ethics

Developments in the field of neuroscience and artificial intelligence bring forth a lot of ethical challenges. We would like to discuss some of these ethical challenges we face in order to ascertain the risks associated with these developments. This would help us to steer into a direction which would be safer for us and our environment.

Neurotechnology: Neurotechnologies could be invasive such as brain implants or non-invasive such as headbands. With the growing market and an increasing number of these devices, the regulatory agencies are overwhelmed. This has caused many companies to skip the strict regulations and present the products directly to the consumers. The most important ethical concern, hence, is the complete knowledge of how these products might cause harm to the users. The harm could be immediate like a burn caused by electrodes placed under the skin or could be a long term such as cognitive impairment. [7.]

Lie Detection: It has been shown that polygraphs are not always the best method to detect lies. Recently, fMRI has emerged as a tool to detect if a person is lying with >90% accuracy but it is still impossible to tell what the person is lying about which is why most judgments based on these measures of interrogation are incorrect. [7.]

Enhancement through Drugs: It is also possible for one to alter one's cognitive function through drugs while testifying for a crime. Some drugs enhance memory and attention and hence if the testimony is based on the statement of the individual while he/she is under the influence raises a big ethical concern. [7.]

Privacy: Out of all the ethical concerns related to neurotechnology, privacy is the biggest and the most important of all. Since the advent of big-data centers, it is possible to process gargantuan amounts of data, this processing is generally opaque to the individual which raises a big concern. Since clinical data such as fMRI is digitized and is generally kept out of reach by the patients, it is important to regulate what happens to it after the scan is done and analyzed by the medical authorities. Since current AI systems are advanced enough to process this information and create profiles for an individual, regulatory bodies are required to put protocols in place to let the patient or subject know what happens to the scan after the use so that it doesn't end up in the processing pool. The anonymity of the data is another big-ethical issue, due to human error, it is possible to reveal the identity of the subject which might cause trouble for the subject.

Moral Machines: The foreground for artificial intelligence development is increasing rapidly. Due to this rapid pace of development in technology, the implications it might have on our surroundings and our lives are not very well understood. Autonomous cars are one such example where the decision taken by the machine could affect the lives of humans around it. The moral dilemma surrounding autonomous cars is complex. When the machine finds itself in a situation where it has a young human on the left and an old human on the right side of the road, and if the car cannot stop due to a failure, which side of the lane is better to go through? [8.] These decisions are so complex that even humans are not capable of telling what's a good thing to do. Designing AI systems brings forth many such debilitating issues, and it is up to us all to take decisions that are best for us.

Another similar dilemma emerges from the role of AI in reducing recidivism. *Recidivism* is the tendency of a criminal to recommit a crime. Researchers at Purdue University are creating software that could predict the rate of recidivism based on a set of behavior, both physical and psychological, shown by the parolees. Another such system is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) [9.] which helps judges to make better decisions regarding recidivism. Although this software is sophisticated, it has been found that it is no better than a non-expert. Also, these software applications can be biased with the inherent biases of the dataset it is trained on. This raises a serious concern regarding its use by the judges.

2.5.4 Impact

The impact of AI and neuroscience is undeniably large in our society and will eventually grow bigger in the future. From recommendation systems in shopping apps to analyzing CT scans, AI is helping automate jobs that earlier required humans. Although it is important for our society since globalization requires things to move faster than in the past, it also causes great concern. Especially when it comes to creating jobs, it is feared that autonomous systems will reduce jobs for humans in the future and can disrupt the stability of the fabric of our society. Another grave danger of these AI systems is misinformation: the current age is also labeled as the post-truth age by philosophers because our society is overwhelmed with information and the human brain cannot process the authenticity of it all. This causes people to change their cognitive functions based on their interaction with the digital systems. It is required for us to form a symbiotic relationship with the current systems and also take special measures in the development of these technologies.

Humans are central to the AI and neuroscientific systems. The technology developed by AI scientists and neuroscientists directly affects the lives of each and every individual on the planet. We can see the effect of this in this famous documentary called ‘The Social Dilemma’ [10.], which talks about how these AI systems are affecting the cognitive abilities of its users. The design choices made by a handful of engineers affect a large chunk of the population. These systems are designed to grab maximum attention and also play with the dopaminergic response of the users. This calls for action to make these technologies more human-centric. The ethics and large-scale impact of these systems should be assessed before individuals are allowed to use them.

2.5.5 Energy cost of computation

Every single operation on a computer or a smartphone from a simple search query to large scale language models requires a series of computation steps. These computational steps come at the cost of a small amount of energy. Companies like Google, Facebook and Amazon make use of giant datacenters that host large computers that perform computations for running websites. Given the number of people and that demand for computation is growing faster than ever, the energy required to perform these computations is increasing rapidly. It is projected that if the growth for the need of computation increases at a similar pace, by 2040 it will hit the world energy production capacity. Current information and communications (ICT) infrastructure accounts for more than 2 percent of the global energy demand. It is the need of the hour to find innovative solutions to curb this demand and reduce consumption. Some of the suggestions are to change the data center architectures to make them more energy efficient by improving the softwares and hardware simultaneously, also to make use of the computational resource more efficiently by its users. Another alternative is to use non-carbon based energy resources to operate these data centers and supercomputers. [11.]

2.5.6 Want to know more?

Sources:

[1.] Harrell, Eben. “Neuromarketing: What You Need to Know.” Harvard Business Review, 23 January 2019, https://hbr.org/2019/01/neuromarketing-what-you-need-to-know. Accessed 22 November 2021.

[2.] Wu, Wayne, "The Neuroscience of Consciousness", *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/consciousness-neuroscience/>.

[3.] Konkoly KR, Appel K, Chabani E, Mangiaruga A, Gott J, Mallett R, Caughran B, Witkowski S, Whitmore NW, Mazurek CY, Berent JB, Weber FD, Türker B, Leu-Semenescu S, Maranci JB, Pipa G, Arnulf I, Oudiette D, Dresler M, Paller KA. Real-time dialogue between experimenters and dreamers during REM sleep. Current Biology 2021, 31: 1417-1427

[4.] https://deepmind.com/research/case-studies/alphago-the-story-so-far

[5.] http://www.scholarpedia.org/article/Neurorobotics

[6.] https://www.intel.in/content/www/in/en/research/neuromorphic-computing.html

[7.] Illes J, Bird SJ. Neuroethics: a modern context for ethics in neuroscience. Trends Neurosci. 2006;29(9):511-517. doi:10.1016/j.tins.2006.07.002

[8.] https://www.moralmachine.net/

[9.] https://pixelplex.io/blog/artificial-intelligence-criminal-justice-system/

[10.] https://www.thesocialdilemma.com/

[11.] https://news.mit.edu/2022/how-can-we-reduce-carbon-footprint-global-computing-0428

2.5.7 Key terms

General consciousness
Specific consciousness
Lucid dreaming
Neuromorphic computing
Recidivism