

Cognition and Neuroscience

(Module 1)

Last update: 19 May 2024

Academic Year 2023 – 2024
Alma Mater Studiorum · University of Bologna

Contents

1	Introduction	2
1.1	Definitions	2
1.2	Neuroscience history	3
1.2.1	Neuron doctrine	3
1.3	Cognitive science history	3
2	Nervous system anatomy and physiology	4
2.1	Individual cells	4
2.1.1	Glia cells / Neuroglia	4
2.1.2	Neurons / Nerve cells	5
2.2	Information transfer within a neuron	6
2.2.1	Neuron functional regions	6
2.2.2	Neuron transmission signals	7
2.3	Information transfer between two neurons	9
2.3.1	Electrical synapse	9
2.3.2	Chemical synapse	9
2.4	Neural circuit	10
2.5	Neural system	12
2.5.1	Peripheral nervous system (PNS)	12
2.5.2	Central nervous system (CNS)	13
3	Reinforcement learning	17
3.1	Definitions	17
3.1.1	Learning	17
3.1.2	Learning systems	18
3.2	Learning at the neuronal level	18
3.3	Dopamine	19
4	Pavlovian/classical learning	20
4.1	Types of reinforcement	20
4.2	Learning flexibility	21
4.3	Memory	22
4.4	Learning preconditions	23
4.4.1	Contiguity	23
4.4.2	Contingency	24
4.4.3	Surprise	24
4.5	Computational model	25
4.5.1	Rescorla-Wagner model	25
4.5.2	Temporal difference model	26
4.6	Reward prediction error hypothesis of dopamine	27
5	Instrumental/operant learning	31
5.1	Types of schedule	31

5.2	Dopamine	32
5.3	Learning strategies historical evolution	35
5.3.1	Generation 0	35
5.3.2	Generation 1	37
5.3.3	Generation 2	38
5.3.4	Generation 3	40

Acronyms

AP action potential

CNS central nervous system

CR conditioned response

CS conditioned stimulus

EPSP excitatory postsynaptic potential

IPSP inhibitory postsynaptic potential

NR no response

PNS peripheral nervous system

PSP postsynaptic potential

RL reinforcement learning

UR unconditioned response

US unconditioned stimulus

1 Introduction

1.1 Definitions

Neuroscience Study of the nervous system (structure aspects) on various levels of detail: Neuroscience

Molecular Proteins and molecular signaling of the nervous system.

Cellular Morphological and physiological properties of neurons.

Neural system Creation and functioning of networks of neurons.

Cognition Mental processes (function aspects) that react to inputs. It involves processes regarding the acquisition, storage, manipulation, and retrieval of information. Cognition

Perception Information from the environment.

Attention Focus on a specific stimulus in the environment.

Learning Merging new information with prior knowledge.

Memory Encoding, storing, and retrieving information.

Action Interact with the environment using perceived information.

Language Understanding and producing spoken or written thoughts.

Higher reasoning Decision-making and problem-solving.

Biomimicry Solving problems by taking inspiration from elements of nature. Biomimicry

As proof of general intelligence¹, the human brain is taken as the model for artificial intelligence. Moreover, a successful brain-inspired AI application can provide a possibly plausible explanation of the functioning of the brain.

However, a brain differs from a computer in many aspects:

- Hardware and software are distinct while mind and brain are not.
- Machines learn by exploiting the capability of using a large memory while brains have limited capacity but high generalization ability.
- Brains produce both electrical and biochemical signals and have feedforward, feedback, and recurrent connections while machines typically only employ feed-forward connections.

Structure emulation Mimic or reverse engineer the structure of the brain (e.g. Blue Brain Project).

Function emulation Mimic a neural system on the algorithmic level (e.g. Deep Mind).

Cognitive neuroscience Study of the relationship between the physical brain and the intangible mind (thoughts, ideas). In other words, it studies the relationship between structure and function. Cognitive neuroscience

¹  Doubt

Case study (Severed Corpus Callosum video). Normally, the right and left hemispheres of the brain can communicate. Moreover, the left visual field is sent to the right hemisphere and the right visual field is sent to the left hemisphere.

In patients where the hemispheres are split, a text shown on the right visual side is recognized as the speech capabilities are located in the left hemisphere, while a text shown on the left visual side does not trigger any speech reaction.

1.2 Neuroscience history

Two main schools of thought emerged and are still the subject of ongoing debates:

Localizationism Specific regions of the brain are responsible for particular faculties.

Localizationism

Assuming localizationism, 52 distinct regions with different neurons can be identified.

Aggregate field theory The brain works as a whole for mental functions.

Aggregate field theory

1.2.1 Neuron doctrine

The nervous system is made of a discrete amount of individual neurons (and not a continuous tissue).

Neuron doctrine

Principle of dynamic polarization Electrical signals in a neuron flow only in a single direction.

Principle of connectional specificity Neurons do not connect randomly but make specific connections at particular contact points.

Synapse Point of contact of two neurons. A synapse can be chemical or electrical.

Synapse

1.3 Cognitive science history

Rationalism All knowledge can be derived through reasoning, without sensory experiences.

Rationalism

Empiricism The brain starts as a blank slate and knowledge is added through sensory experiences.

Empiricism

Associationism Inspired by empiricism. Learning happens by correlating individual experiences (e.g. actions followed by a reward will be repeated).

Associationism

Behaviorism Inspired by empiricism. Everyone has the same neural basis that is improved through learning. Learning only involves observable behaviors.

Behaviorism

Remark. Associationism and behaviorism are not able to explain all types of learning (e.g. language).

Cognitivism The psychological and biological levels of an individual cannot be separated. Learning is based on the biology of the neurons.

Cognitivism

2 Nervous system anatomy and physiology

Central nervous system Brain and spinal cord.

Peripheral nervous system Nerves that branch off from the brain and the spine.

2.1 Individual cells

2.1.1 Glia cells / Neuroglia

Cells that support neurons. There are 2 to 10 times more glia cells than neurons.

Glia cells/Neuroglia

Microglia Immune system cells located in the central nervous system. They intervene in response to toxic agents or to clear dead cells.

Microglia

- Responsible for antigen presentation (determine the type of external agent).
- Become phagocytes (cells that ingest harmful agents) during injuries, infections, or degenerative diseases.



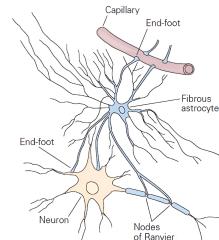
Remark. In patients affected by Alzheimer's disease, microglia may become hyperactive and damage neurons.

Astrocytes Star-shaped cells located in the central nervous system.

Astrocytes

They surround neurons and are in contact with the brain's vasculature.

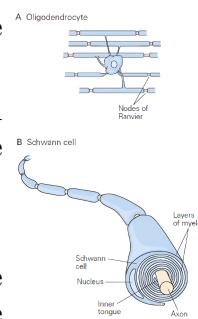
- Provide nourishment to neurons.
- Regulate the concentration of ions and neurotransmitters in the extracellular space.
- Communicate with the neurons to modulate synaptic signaling.
- Maintain the blood-brain barrier that separates the tissues of the central nervous system and the blood.



Oligodendrocytes and Schwann cells Oligodendrocytes are located in the central nervous system, while Schwann cells are located in the peripheral nervous system.

Oligodendrocytes
Schwann cells

- Produce thin sheets of myelin that wrap concentrically around the axon of the neurons. This insulating material allows the rapid conduction of electrical signals along the axon.



Remark. Myelin is white, giving the name to the white matter.

Remark. In multiple sclerosis, the immune system attacks the oligodendrocytes, slowing or disrupting messages traveling along the nerves.

2.1.2 Neurons / Nerve cells

A nervous system has around 100 billion neurons. There are 100 distinct types of neurons varying in form, location, and interconnectivity.

Generally, a neuron does the following:

1. Receives some information.
2. Makes a decision.
3. Passes it to other neurons.

Eukaryotic cell A neuron is an eukaryotic cell. Therefore, it has:

Neurons/Nerve cells

Eukaryotic cell

Cell membrane Membrane that separates the intracellular and extracellular space.

Cytoplasm Intracellular fluid mainly made of proteins and ions of potassium, sodium, chloride, and calcium.

Extracellular fluid Fluid in which the neuron sits. Similar composition of the cytoplasm.

Cell body/soma Metabolic center of the cell.

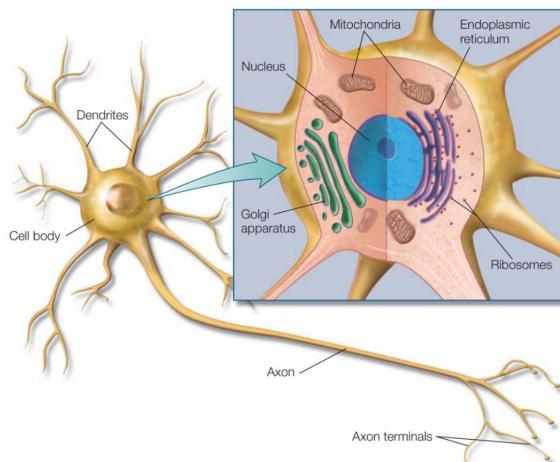


Figure 2.1: Neuron as an eukaryotic cell

Neuron-specific components

Dendrites Receives the outputs of other neurons. A neuron has multiple dendrites with different shapes depending on the type and location of the neuron.

Dendrites

Axon Transmitting zone of the neuron that carries electrical signals from the dendrites to the synapses (from 0.1mm to 2m). A neuron has a single axon.

Axon

Synapses Represents the output zone of the neuron from where electrical or chemical signals can be transmitted to other cells. A neuron has multiple synapses.

Synapses

Presynaptic cell Cell transmitting a signal.

Postsynaptic cell Cell receiving a signal.

Synaptic cleft Narrow space separating presynaptic and postsynaptic cells (i.e. the space separating two neurons).

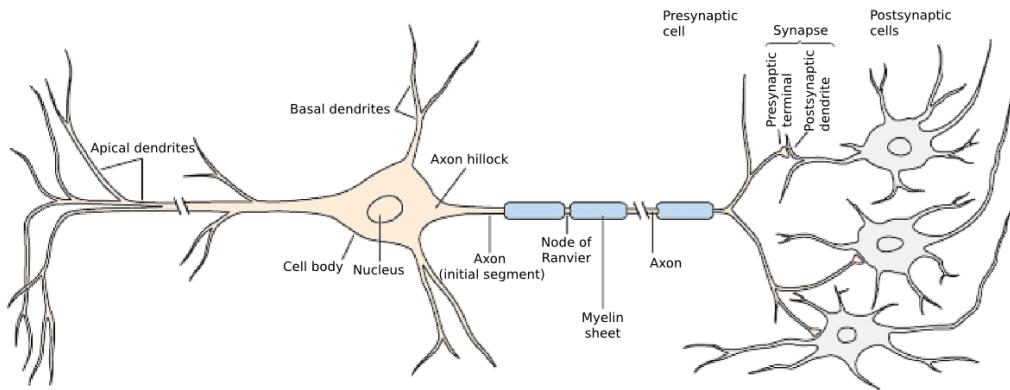
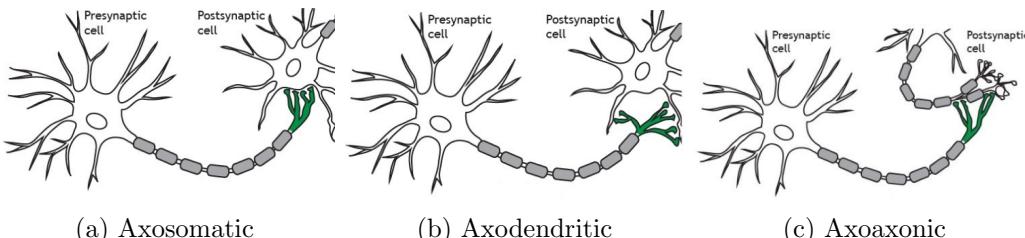


Figure 2.2: Neuron-specific components

There are three types of synapses:

Axosomatic	Synapses that a neuron makes onto the cell body (soma) of another neuron.	Axosomatic
Axodendritic	Synapses that a neuron makes onto the dendrites of another neuron.	Axodendritic
Axoaxonic	Synapses that a neuron makes onto the synapses of another neuron. In this case, the transmitting neuron can be seen as a signal modulator of the receiving neuron.	Axoaxonic



Neurons are divided into three functional categories:

Sensory neurons	Carry information from the body's peripheral sensors into the nervous system. Provides both perception and motor coordination.	Sensory neurons
Motor neurons	Carry commands from the brain or the spinal cord to muscles and glands.	Motor neurons
Interneurons	Intermediate neurons between sensory and motor neurons.	Interneurons
Principle of connectional specificity	Neurons do not connect randomly but rather make specific connections at particular contact points.	Principle of connectional specificity

2.2 Information transfer within a neuron

2.2.1 Neuron functional regions

In a neuron, there are four regions that handle signals:

Input zone	Dendrites collect information from different sources in the form of postsynaptic potentials (PSPs).	Input zone
-------------------	---	------------

Integration/trigger zone PSPs are summed at the axon hillock and an action potential (AP) is generated if a threshold (-55mV) has been exceeded.

Integration/trigger zone

Conductive zone The AP is propagated through the axon.

Conductive zone

Output zone Synapses transfer information to other cells.

Output zone

Chemical synapses The frequency of APs determines the amount of neurotransmitters released.

Electrical synapses The AP is directly transmitted to the next neurons.

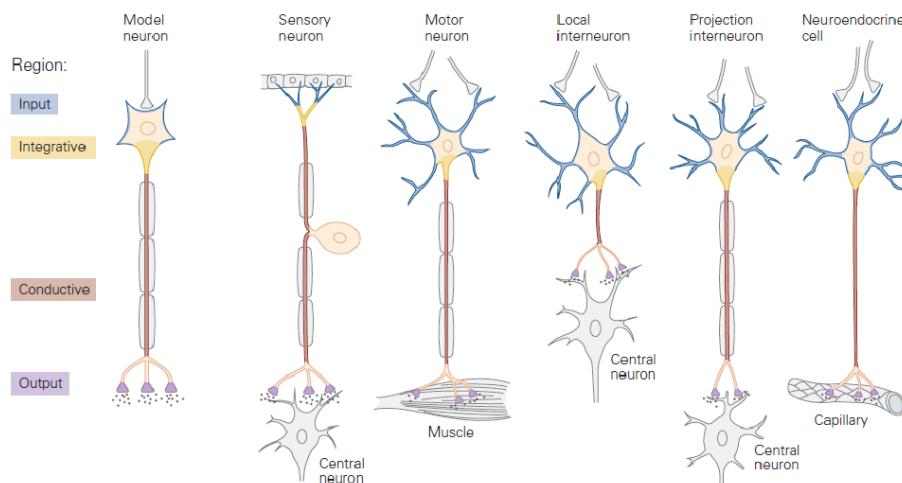


Figure 2.4: Transmitting regions of different types of neurons

2.2.2 Neuron transmission signals

Resting membrane potential In a resting neuron, the voltage inside the cell is more negative (-70mV) than the outside. This allows the creation of an electrical signal when needed.

Resting membrane potential

Postsynaptic potential (PSP) Small change in the membrane potential that alters the resting voltage of the cell.

Postsynaptic potential (PSP)

A PSP can be:

Excitatory PSP

Excitatory PSP (EPSP) Has a depolarizing role: produces a decrease in the membrane potential (i.e. increases voltage inside the cell), therefore enhancing the ability to generate an AP.

Inhibitory PSP

Inhibitory PSP (IPSP) Has a hyperpolarizing role: produces an increase in the membrane potential (i.e. reduces voltage inside the cell), therefore reducing the ability to generate an AP.

A PSP has the following properties:

- The amplitude and duration of the signal are determined by the size of the stimulus that caused it. Overall, the amplitude is small.
- The signal is passively conducted through the cytoplasm, therefore it decays with distance and is able to travel 1mm at most.
- A single EPSP is not enough to fire a neuron. Multiple PSPs are summed at the axon hillock. There are two types of summation:

Spatial summation Sum of the PSPs received at the same time.

Temporal summation Sum of the PSPs received at different time points.

Remark. The fact that a single EPSP is not enough to fire a neuron prevents a response to every single stimulus.

Action potential (AP) Signal generated when the sum of EPSPs exceeds a fixed threshold of -55mV (all-or-none).

Action potential (AP)

Saltatory conduction Mechanism that allows a fast propagation on long distances of APs.

Saltatory conduction

1. Depolarization causes the sodium ion (Na^+) channels located in the nodes of Ranvier of the axon to gradually open.
2. Na^+ flows into the neuron and further depolarizes it until the Na^+ equilibrium potential is reached.
3. With Na^+ equilibrium, Na^+ channels close and potassium ion (K^+) channels open.
4. K^+ flows into the neuron and restores the membrane potential until the K^+ equilibrium potential is reached.
5. With K^+ equilibrium, K^+ channels close and the membrane potential of the neuron is more negative than the resting potential (hyperpolarization). It will gradually return to its resting potential.

Remark. During hyperpolarization, Na^+ channels cannot open (refractory period). This has two implications:

- It limits the number of times a neuron can fire in a given time.
- Guarantees a unidirectional electrical current flow (**Principle of dynamic polarization**).

Principle of dynamic polarization

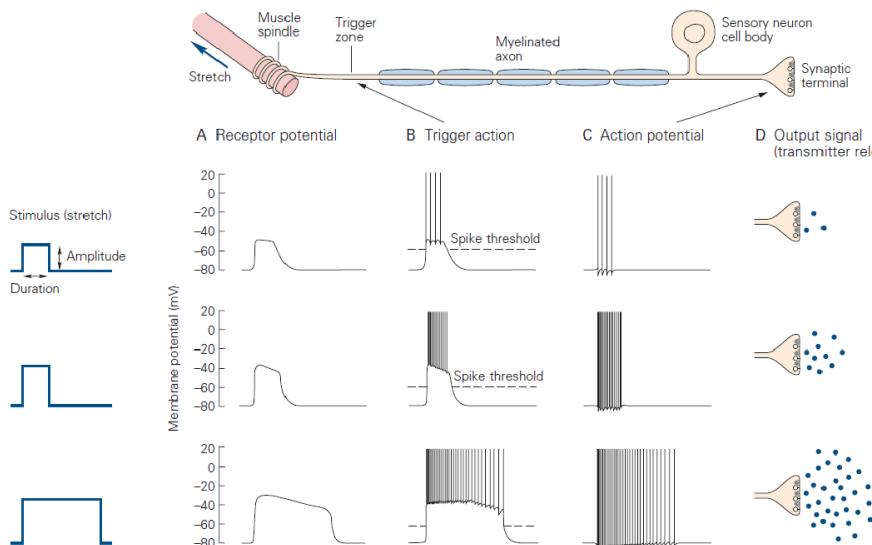
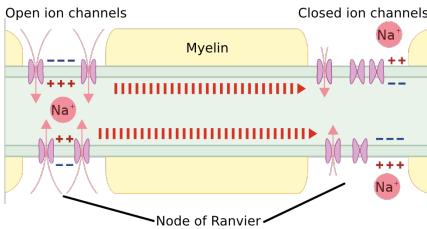
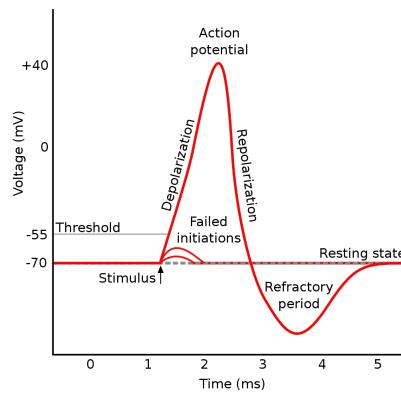


Figure 2.5: Signal from the input to the output zone. The amplitude of the stimulus modulates the frequency of AP.



(a) Ion channels along the axon



(b) Triggering of an action potential

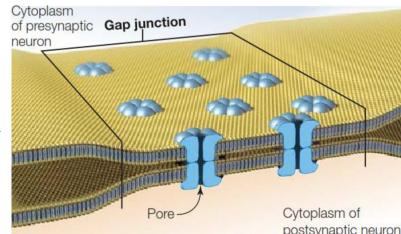
Remark. As the signal is constantly regenerated, APs have similar amplitude and duration in all neurons, regardless of the characteristics of the input PSPs. Therefore, the only way an AP has to carry information is by varying frequency depending on the stimulus intensity, making it a binary signal.

| **Example.** Seizures are caused by misfiring neurons.

2.3 Information transfer between two neurons

2.3.1 Electrical synapse

Structure The neuronal membranes of the presynaptic and postsynaptic neurons are in contact at **gap junctions** and the cytoplasm of the two neurons is virtually continuous through connecting **pores**.



Electrical synapse

Functioning The two neurons are **isopotential** (i.e. they have the same membrane potential) and the ions of the presynaptic neurons are instantaneously transmitted to the postsynaptic neuron.

Properties

- Fast transmission.
- Allows for synchronous operations involving groups of neurons.
- The strength of the signal cannot be modulated.

2.3.2 Chemical synapse

Structure The synaptic cleft separates the presynaptic and postsynaptic neurons.

Chemical synapse

Neurotransmitter Chemical substance received by the receptors of the postsynaptic neuron.

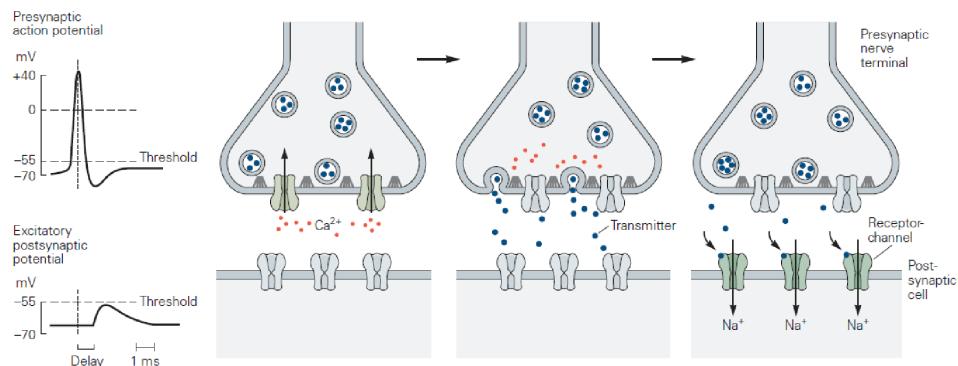
The effect of a neurotransmitter is decided by the receiving receptor and not by the cell transmitting it.

Presynaptic terminals Swellings at the end of the axon that contain synaptic vesicles.

Synaptic vesicles Vesicles containing neurotransmitter molecules.

Functioning The release of neurotransmitter molecules is based on the following steps:

1. An action potential arriving at the terminal of a presynaptic axon causes the calcium ion (Ca^{2+}) voltage-gates to open.
2. Ca^{2+} flow into the cell and cause the synaptic vesicles to bind to the cell membrane to release neurotransmitters into the synaptic cleft.
3. Neurotransmitters cross the synaptic cleft and bind to the receptors of the postsynaptic neuron. Depending on the neurotransmitter and the receiving receptor, there might be a generation of EPSP or IPSP.



When a receptor recognizes the neurotransmitter, it is released back into the synaptic cleft. To avoid a constant stimulation of the receptors, neurotransmitters are inactivated:

- The synaptic terminal can reuptake neurotransmitters through transporter proteins.
- Neurotransmitters might degenerate or be broken down by special enzymes.
- Neurotransmitters can be released far away from the site of the receptors.

Properties

- Slow transmission.
- The signal can be modulated.
- Has specific effects depending on the neurotransmitter and the receptors.

2.4 Neural circuit

Neural circuit Group of interconnected neurons that process a specific kind of information.

Neural circuit

| **Remark.** The behavioral function of each neuron is determined by its connections.

Types of neurons

Sensory neuron Carry information from the peripheral sensors to the nervous system for both perception and motor coordination.

Sensory neuron

Motor neuron Carry information from the nervous system to muscles and glands.

Motor neuron

Interneuron Intermediate neurons between sensory and motor neurons.

Interneuron

Remark. In vertebrates, a stimulus causes multiple neural pathways to simultaneously encode different information. This allows for parallel processing to increase both the speed and reliability of the information transfer.

Neural pathways types

Divergent pathway One neuron activates many target cells. Typically happens at the input stages of the nervous system to ensure that a single neuron has a wide and diverse influence.

Divergent pathway

Convergent pathway Many neurons activate a single target cell. Typically happens at the output stages of the nervous system to ensure that a motor neuron is activated only when a sufficient number of neurons are firing.

Convergent pathway

Neuron firing types

Excitatory neuron Neurons that produce signals that increase the probability of firing of the postsynaptic neurons.

Excitatory neuron

Inhibitory neuron Neurons that produce signals that decrease the probability of firing of the postsynaptic neurons.

Inhibitory neuron

Feed-forward inhibition Excitatory neurons connected to inhibitory interneurons to block other downstream neurons. Allows to enhance the active pathway and to block other antagonist actions.

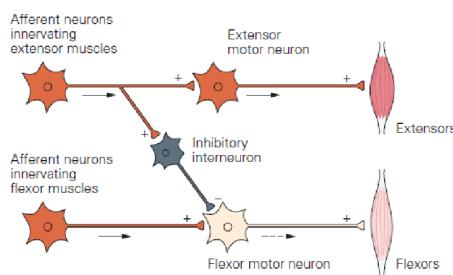


Figure 2.7: Example of feed-forward inhibition

Feed-back inhibition Excitatory neurons connected to inhibitory interneurons that return to the same neurons to inhibit them. Prevents the overload of neurons or muscles.

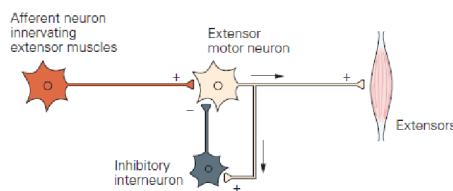


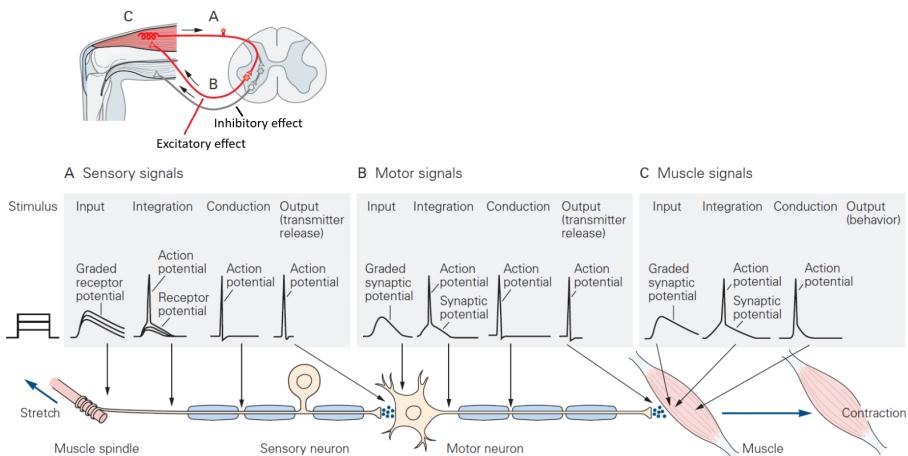
Figure 2.8: Example of feed-back inhibition

Case study (Knee-jerk reflex). By tapping the patellar tendon (below the kneecap), the following happens:

1. The sensory information is conveyed from the muscle to the spinal cord (central nervous system).
2. The nervous system issues motor commands to the muscles which results in the knee

jerk.

- Inhibitory commands are issued to stop antagonist muscles.



2.5 Neural system

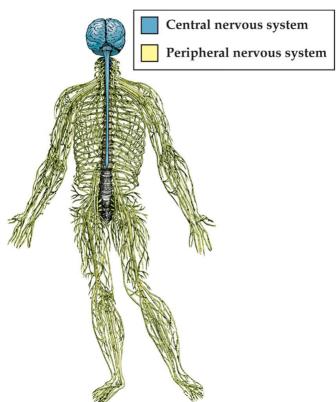


Figure 2.9: Composition of the nervous system

2.5.1 Peripheral nervous system (PNS)

The peripheral nervous system is composed of:

Nerves Groups of axons and glia. Nerves

Ganglia Groups of neuron bodies outside the central nervous system Ganglia

The PNS has the following functions:

- Delivers sensory information to the central nervous system.
- Carries commands from the central nervous system to the muscles.
- Supplies the central nervous system with information regarding both the external and internal environment.

The PNS has the following divisions:

Somatic nervous system

Somatic nervous system

- Sensory neurons that receive information from the skin, muscles, and joints.
- Converts perceived spatial and physical information into electrical signals for the central nervous system to process.
- Controls the voluntary muscles.

Autonomic nervous system

Autonomic nervous system

- Controls internal organs (viscera), the vascular system, and involuntary muscles and glands.
- Divided into three systems:

Sympathetic system Operates antagonistically against the parasympathetic system. Handles the body's response to stress (using norepinephrine).

Sympathetic system

Physically, the sympathetic system originates from the spinal cord. Its ganglia are closer to the spinal cord, therefore the axons from the central nervous system to the ganglia are shorter than the axons from the ganglia to the organs.

Example. Stimulates adrenal glands to prepare the body for action (fight or flight), increases heart rate, diverts the blood from the digestive tract to the somatic musculature, ...

Parasympathetic system Operates antagonistically against the sympathetic system. Acts to preserve the body's resources and restore homeostasis (using acetylcholine).

Parasympathetic system

Physically, the parasympathetic system originates from the base of the brain and from the sacral spinal cord. Its ganglia are outside the spinal cord, sometimes inside the affected organs, therefore the axons from the central nervous system to the ganglia are longer than the axons from the ganglia to the organs.

Example. Slows heart rate, stimulates digestion, ...

Enteric system Controls the involuntary muscles of the gut.

Enteric system

2.5.2 Central nervous system (CNS)

Meninges Three layers of membrane protecting the brain and the spinal cord.

Meninges

Dura mater The outermost and thickest layer.

Arachnoid mater The middle layer.

Pia mater The innermost and most delicate layer. It adheres to the brain's surface.

Cerebrospinal fluid Fluid that allows the brain to float and prevents it from simply sitting on the skull surface. It also reduces the shock to the brain and the spinal cord in case of rapid accelerations/decelerations.

Cerebrospinal fluid

The fluid is located in:

- The space between the arachnoid mater and the pia mater.
- The brain ventricles.

- Cisterns and sulci.
- The central canal of the spinal cord.

Blood-brain barrier Barrier between the brain's capillaries and the brain's tissue. It protects against pathogens and toxins.

Blood-brain barrier

Remark. The effectiveness of the barrier also prevents drugs to treat mental and neurological disorders from passing through.

Spinal cord Acts as a relay for the information coming in and out of the brain. It is enclosed in the vertebral column.

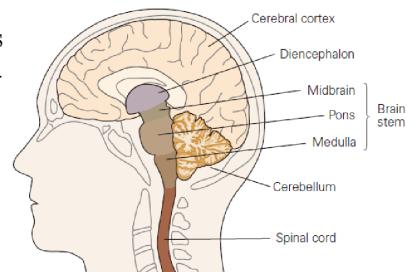
Spinal cord

Remark. Most pathways in the CNS are bilaterally symmetrical: the sensory and motor activities of one side of the body are handled by the cerebral hemisphere on the opposite side.

Brain

Brain stem Regulates basic life functions such as blood pressure, respiration, and sleep/wakefulness. It is divided into three sections:

- Medulla.
- Pons.
- Midbrain.



Brain

Brain stem

Cerebellum Contains lots of neurons and is responsible for:

Cerebellum

- Maintaining posture.
- Coordinating head, eye, and arm movement.
- Regulating motor control (i.e. adjustments to the movement).
- Learning motor skills.

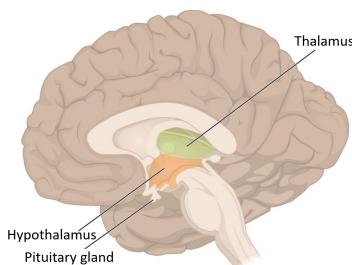
Diencephalon

Thalamus Sorts incoming sensory information (except the sense of smell) of the peripheral nervous system and sends them to the sensory regions of the cerebral hemispheres.

Diencephalon

Thalamus

Hypothalamus Regulates the autonomic nervous system and homeostasis through the pituitary gland (which releases hormones). Handles the motivation system of the brain by favoring behaviors the organism finds rewarding.



Hypothalamus

Telencephalon/Cerebral hemispheres Consists of:

Telencephalon/Cerebral hemispheres

Cerebral cortex Made of gray matter (body of neurons).

White matter (axons and glial cells).

Basal ganglia Receive inputs from sensory and motor areas and mostly send them through the thalamus to the frontal lobe.

Basal ganglia

They have a crucial role in motor control and reinforcement learning. This happens through two pathways:

Direct pathway When active, it causes the disinhibition of the thalamus and has the consequence of initializing movement.

Indirect pathway When active, it causes the inhibition of the thalamus and consequently inhibits movement.

To activate the direct pathway and inhibit the indirect pathway, the substantia nigra pars compacta (SNc) releases the neurotransmitter dopamine.

Example (Parkinson's disease). In patients affected by Parkinson's disease, the dopamine-related neurons in the SNc are lost causing an overactivation of the indirect pathway that inhibits movement.

Amygdala Responsible for recognizing a stimulus and reacting to it. Involved in attention, perception, value representation, decision-making, learning, memory, ...

Amygdala

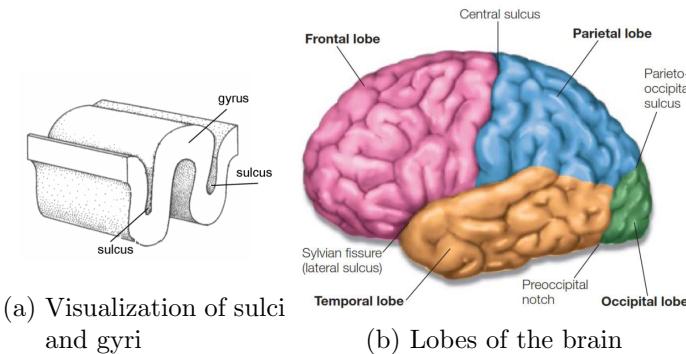
Hippocampus Responsible for long-term memory and spatial memory.

Hippocampus

Cerebral cortex Surface of the brain which covers around 2.2m^2 to 2.4m^2 . To cover more surface, the cortex has infoldings (sulci and gyri) which also allow to connect neurons with shorter axons.

Cerebral cortex

There are two symmetrical hemispheres connected through the corpus callosum and four different lobes.



Frontal lobe

Frontal lobe

Motor cortex

- Planning and execution of movement.
- Contains neurons that directly activate somatic movement neurons in the spinal cord.

Prefrontal cortex

- Long-term planning.
- Decision making.
- Motivation and value.

Parietal lobe Receives and integrates information from the outside world, the body, and memory.

Parietal lobe

Somatosensory cortex Receives information regarding touch, pain, temperature, and limb position.

Remark. Neurons responsible for a specific part of the body are clustered together.

Occipital lobe

Occipital lobe

Visual cortex Responsible for vision. Encodes features like luminance, spatial frequency, orientation, motion, ...

Remark. Neurons responsible for processing a specific feature are clustered together.

Temporal lobe

Temporal lobe

Auditory cortex Responsible for processing sound.

Remark. Neurons responsible for processing a specific sound frequency are clustered together.

Association cortex Portion of the cortex that has neither sensory nor motor responsibility. Receives and integrates inputs from many cortical areas.

Association cortex

Multisensory neuron Cell activated by multiple sensory modalities.

3 Reinforcement learning

3.1 Definitions

Reinforcement learning (RL) methods aim to maximize future reward by mapping the possible states of an environment into actions.

Optimal decision making Aims to maximize rewards and minimize punishments.

Optimal decision making

Remark. This is a difficult task as the outcome might be delayed or depend on a series of actions.

Credit assignment problem Determine how the various factors involved in making a decision contributed to the success or failure of it.

Remark. Multiple competing sub-systems contribute to learning and controlling behavior in animals.

Example (Freud's theory of the mind structure). The mind is composed of three structures:

Ego Mainly works at the conscious level. Rational part of the mind that mediates *id* impulses and *superego* inhibitions.

Superego Mainly works at the preconscious level. Includes one's ideals and morals. Strives for perfection.

Id Mainly works at the unconscious level. Irrational part of the mind based on basic impulses that seek immediate gratification.

3.1.1 Learning

Learning Lasting change in response or behavior originated from experience.

Learning

Non-associative learning Change in response or behavior caused by learning the properties of a single stimulus. It can result in:

Non-associative learning

Habituation A decrease in response to a stimulus that is presented repeatedly.

Example. The first explosion of a firework causes a strong response but the responses to the following ones are much weaker.

Sensitization An increase in response to a stimulus that is presented repeatedly.

Example. When the skin itches, one will start scratching it.

Associative learning Change in response or behavior caused by learning an association of two or more stimuli/events.

Associative learning

Reinforcement learning Learn an association between a neutral stimulus (something the body considers irrelevant) and a reinforcer (something the body considers relevant).

Reinforcement learning

Primary reinforcer Positive or negative stimulus that is biologically relevant and elicits a response.	Primary reinforcer
Example. Food, pain, social interactions, ...	
Secondary reinforcer Positive or negative stimulus that became relevant following associative learning. It elicits a response which usually enables a primary reinforcer.	Secondary reinforcer

3.1.2 Learning systems

Pavlovian/classical system Form of prediction learning. Learns to predict biologically relevant stimuli to trigger an appropriate response (stimulus-outcome associations).	Pavlovian system
Instrumental system Form of control learning to learn action-outcome associations. It includes:	Instrumental system
Habitual system Learn to repeat previously successful actions.	Habitual system
Goal-directed system Evaluate actions based on the prior knowledge of their consequences.	Goal-directed system

Remark. Pavlovian and instrumental systems are not independent. By predicting which situations are positive, one can act to reach them through its actions.

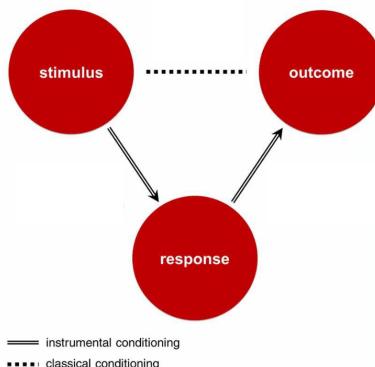


Figure 3.1: Learning systems relationship

3.2 Learning at the neuronal level

Hebbian plasticity Learning and experience change the connections of a neural system.	Hebbian plasticity
Short-term change Functional physiological change that modifies the effectiveness of existing synaptic connections (i.e. amount of neurotransmitters). Lasts from seconds up to hours.	Short-term neuronal change
Long-term change Structural change that leads to anatomical alterations such as pruning or growth of synapses. Lasts days and can cause further short-term changes.	Long-term neuronal change
Remark. Neuronal changes follow a "use it or lose it" policy: only useful changes will last.	
Case study (Phantom limb pain). In amputees, the area of the brain responsible for the missing part of the body is overrun by the neighboring sections. In the case of an arm, the area responsible for the face might "conquer" what once was the area of the arm.	

3.3 Dopamine

Synaptic plasticity Change the synaptic efficacy by changing the amount of:

Neurotransmitters Directly provoke excitatory or inhibitory effects at postsynaptic neurons.

Neuromodulators Neurotransmitters with additional effects.

Dopamine Neuromodulator responsible for processes such as motivation, learning, decision-making, addiction, Parkinson's disease, Huntington's disease,

Dopaminergic pathways

Dopamine
Dopaminergic pathways

Nigrostriatal pathway Originates in the substantia nigra pars compacta (SNc) and primarily projects to the caudate-putamen.

Basal ganglia motor loop Collection of subcortical nuclei responsible for motor control and reinforcement learning.

The direct pathway initiates movement while the indirect pathway inhibits it.

The SNc projects into the striatum and is responsible for releasing dopamine that activates the direct pathway. The striatum can be seen as the component that uses the reward to influence an action.

Meso-limbic pathway Originates in the VTA and projects to the nucleus accumbens, septum, amygdala and hippocampus.

Meso-cortical pathway Originates in the VTA and projects to the medial prefrontal, cingulate, orbitofrontal and perirhinal cortex.

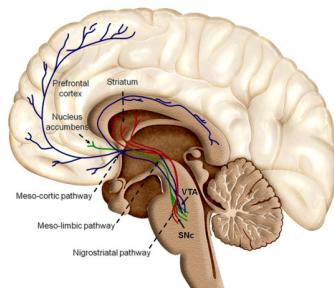
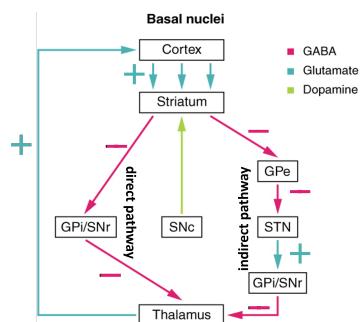


Figure 3.2: Dopaminergic pathways

4 Pavlovian/classical learning

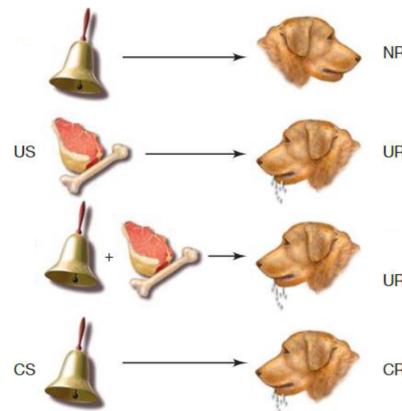
Form of prediction learning that aims to learn stimulus-outcome associations:

- When a reinforcer is likely to occur.
- Which stimuli tend to precede a reinforcer.

This allows the animal to emit a response in anticipation of a reinforcer.

Pavlovian learning works as follows:

- A stimulus that has no meaning to the animal will result in no response (NR).
- An unconditioned stimulus (US) (i.e. a reinforcer) generates an unconditioned response (UR).
- Learning happens when a reinforcer is paired with a non-relevant stimulus.
- The learned conditioned stimulus (CS) generates a conditioned response (CR).



An outcome can be:

Appetitive Something considered positive.

Aversive Something considered negative.

The learned conditioned response can be:

Behavioral Associated to the startle response (i.e. reflex in response to a sudden stimulus).

Physiological Associated to the autonomic system.

Change in subjective response

Remark. Pavlovian learning has its foundations in behaviorism: the brain starts as a blank slate and only observable behaviors can be studied.

4.1 Types of reinforcement

There are two types of learning:

Continuous reinforcement The conditioned stimulus is reinforced every time the unconditioned stimulus occurs.

Continuous reinforcement

Remark. More effective to teach a new association.

Partial reinforcement The conditioned stimulus is not always reinforced.

Partial reinforcement

Remark. Learning is slower but the conditioned response is more resistant to extinction.

4.2 Learning flexibility

Acquisition The probability of occurrence of a conditioned response increases if the conditioned stimulus is presented with the unconditioned stimulus.

Acquisition

Extinction The probability of occurrence of a conditioned response decreases if the conditioned stimulus is presented alone.

Extinction

Remark. Extinction does not imply forgetting. After an association between CS and US is made, extinction consists of creating a second association with inhibitory effects that overrides the existing association.

The extinct association can return in the future (this is more evident when the context is the same as the acquisition phase).

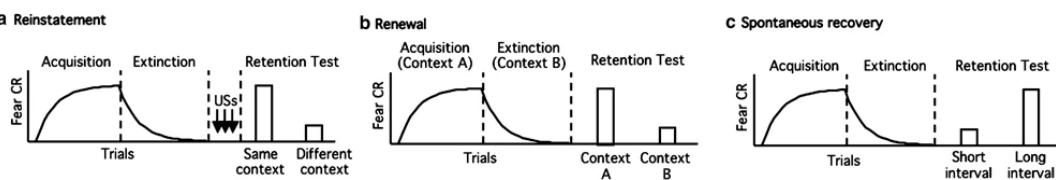


Figure 4.1: Example of acquisition, extinction, and CR return

Generalization A new stimulus that is similar to a learned conditioned stimulus can elicit a conditioned response.

Generalization

Case study (Aplysia Californica).

1. Before conditioning, a stimulus to the siphon of an aplysia californica results in a weak withdrawal of the gill.
2. During conditioning, a stimulus to the siphon is paired with a shock to the tail which results in a large withdrawal of the gill.
3. After conditioning, a stimulus to the siphon alone results in a large withdrawal response.

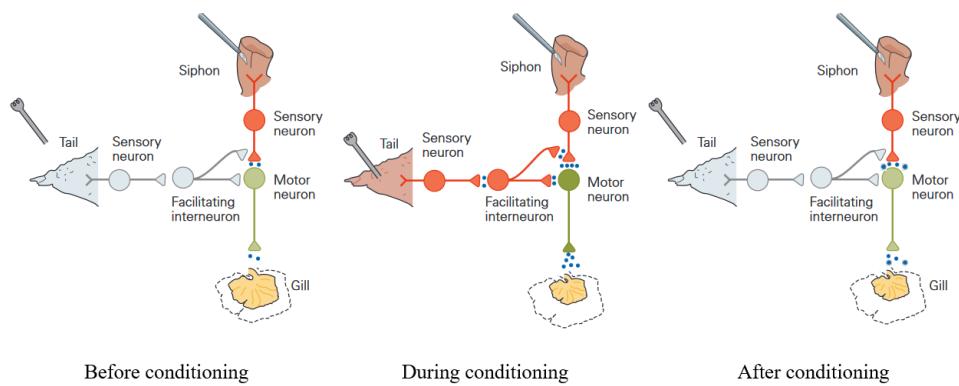
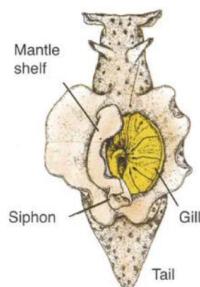


Figure 4.2: Conditioning process

The learned response lasts for days. It can be observed that without training, the response disappears faster.

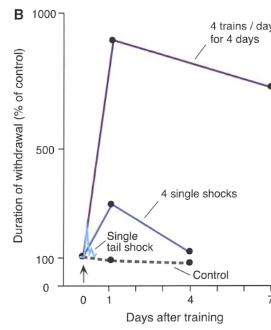


Figure 4.3: Withdrawal response decay

Remark. In mammals, aversive Pavlovian conditioning involves the amygdala. The CS and US are relayed from the thalamus and the cerebral cortex to the amygdala, which in turn connects to various motor responses such as:

Amygdala in Pavlovian learning

Central gray region (CG) Controls the freezing behavior.

Lateral hypothalamus (LH) Controls autonomic responses.

Paraventricular hypothalamus (PVN) Controls stress hormones.

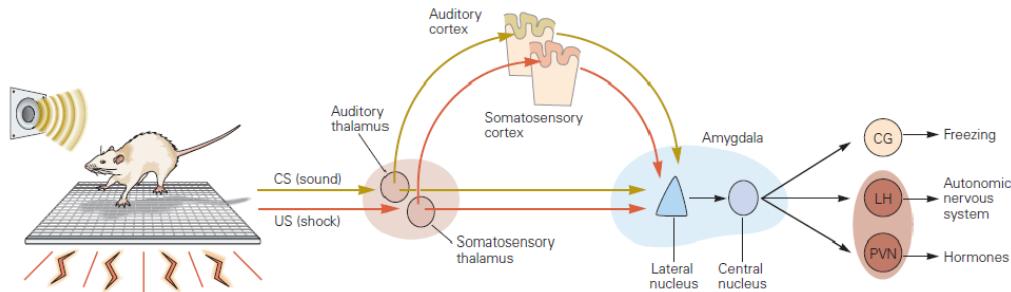


Figure 4.4: Neural circuits during aversive conditioning

4.3 Memory

Memory is vulnerable to alteration. Once reactivated, the subsequent reconsolidation phase might store a modified version of the memory.

Memory

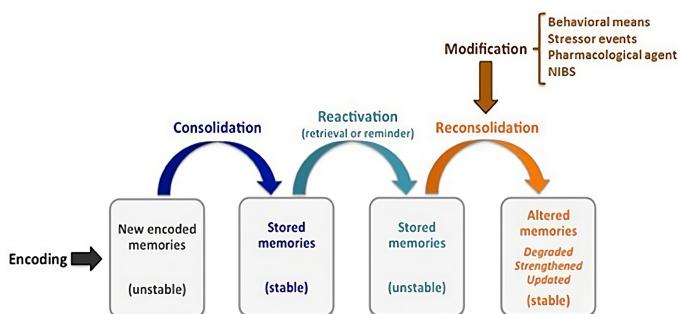


Figure 4.5: Memory flow

Remark. This mechanism is useful against traumatic memories.

Remark. The amygdala is responsible for storing conditioned responses while the hippocampus recognizes conditioned stimuli.

Patients with a damaged amygdala only recognize CS but do not act with any CR. On the other hand, a damaged hippocampus results in patients that present a CR without recognizing the CS.

Case study (Reconsolidation disruption). Propranolol is a drug that disrupts amygdala-specific memory reconsolidation (i.e. the physiological response). A possible therapy to suppress a phobia is to trigger the fear memory and then administer propranolol to prevent its reconsolidation.

4.4 Learning preconditions

4.4.1 Contiguity

Closeness between the conditioned stimulus and the unconditioned stimulus.

Contiguity

Remark. The closer in time the stimuli are presented, the more likely the association will be created.

Depending on when the CS and US are presented, conditioning can be:

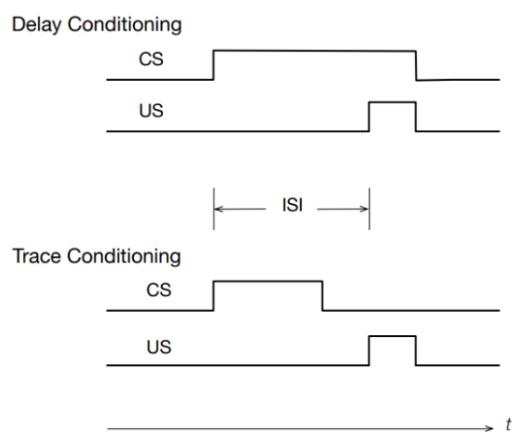
Delay conditioning The CS is extended through the interstimulus interval (ISI) (i.e. time between the start of the CS and the US).

Delay conditioning

Trace conditioning There is a delay (trace interval) between the CS end and the US start.

Trace conditioning

Learning requires more trials and might be impossible if the trace interval is too long as the mental representation of the CS decays.



Case study. Two groups of rats were exposed to a 6 seconds tone (CS) followed by food delivery (US) with a delay of:

- 6 seconds (red).
- 18 seconds (purple).

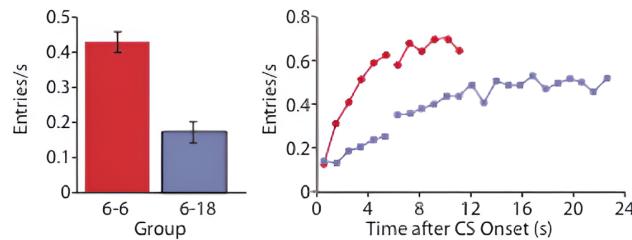


Figure 4.6: Number of entries (i.e. the rat checks the food tray) per second

4.4.2 Contingency

Causal relationship between the conditioned stimulus and the unconditioned stimulus.

Contingency

Remark. Learning happens when:

$$\mathcal{P}(\text{US with CS}) > \mathcal{P}(\text{US with no CS})$$

In other words, the CS should provide information regarding the US.

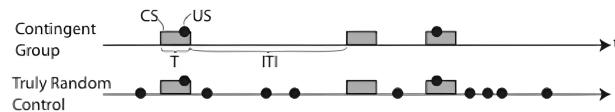


Figure 4.7: Example of contingent and random group

Case study. Two groups of rats are exposed to a shock paired with a bell ring. Contiguity is the same but contingency differs.

Only the group where the shock with the bell is more likely learned the association.

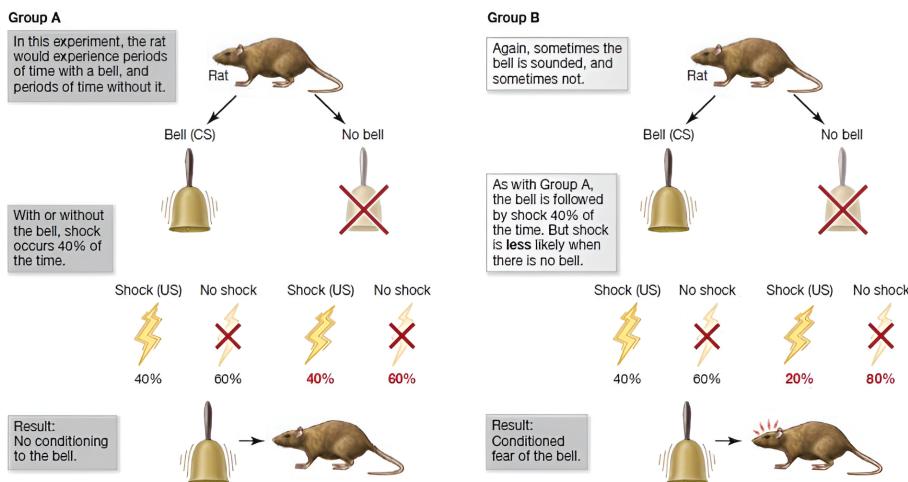


Figure 4.8: Representation of the experiment

4.4.3 Surprise

Prediction error Quantitative discrepancy between the expected and experienced outcome.

Prediction error

Remark. Learning happens when the outcome is different from what was expected.

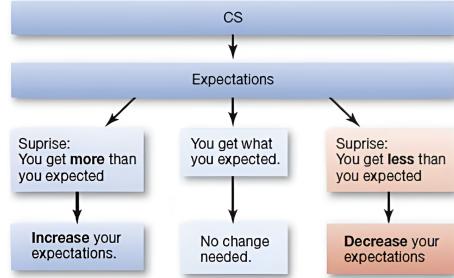
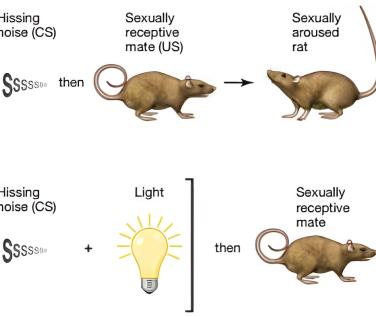


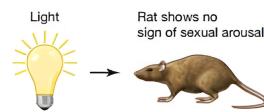
Figure 4.9: Learning outcome due to surprise

Case study (Blocking effect).

1. A rat is taught that a hissing sound (CS) is paired with a sexually receptive mate (US).
2. A light is added together with the hissing sound.
3. When only the light is presented, the rat does not provide a response.



The light is not learned as a CS as it does not provide any new information on the US.



4.5 Computational model

4.5.1 Rescorla-Wagner model

Error-driven learning model where the change expectancy is proportional to the difference between predicted and actual outcome:

$$\delta_{tr} = R_{tr} - V_{tr}$$

where:

- δ_{tr} is the prediction error.
- $R_{tr} = \begin{cases} 1 & \text{if the US is delivered at trial } tr \\ 0 & \text{if the US is omitted at trial } tr \end{cases}$
- V_{tr} is the association strength (i.e. expectancy of the US or the expected value resulting from a given CS) at trial tr .

Then, the expected value V_{tr+1} is obtained as:

$$V_{tr+1} = V_{tr} + \alpha \delta_{tr}$$

where $\alpha \in [0, 1]$ is the learning rate.

| **Remark.** A lower α is more suited for volatile environments.

Rescorla-Wagner
model

Remark. The prediction error δ is:

- Positive during acquisition.
- Negative during extinction.

Moreover, the error is larger at the start of acquisition/extinction.

Remark. The Rescorla-Wagner model is able to capture the blocking effect (see Blocking example) as the animal computes a single prediction error obtained as the combination of multiple stimuli.

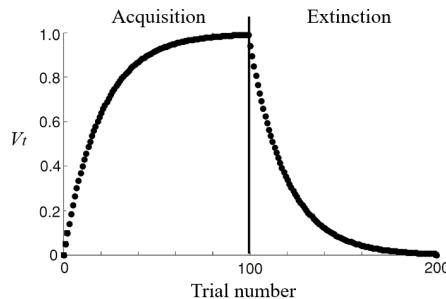


Figure 4.10: Acquisition and extinction in Pavlovian learning according to the Rescorla-Wagner model

Remark. The Rescorla-Wagner model is a trial-level model that only considers the change from trial to trial without considering what happens within and between trials.

4.5.2 Temporal difference model

Real-time model based on time steps within a trial instead of monolithic trials. At each time t of a trial during which a CS is presented, the model computes a prediction of the total future reward that will be gained from time t to the end of the trial.

The prediction error is computed as follows:

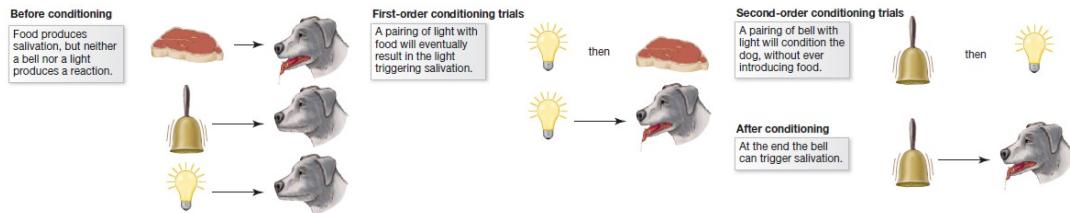
$$\begin{aligned}\delta_t &= R_t + V_t - V_{t-1} \\ V_{t+1} &= V_t + \alpha\delta_t\end{aligned}$$

Temporal difference model

- At the beginning of learning, the CS is presented at time t_{CS} and $V_t = 0$ until the US is delivered at time $t_{US} > t_{CS}$.
- On the next trial, $V_{t_{US}} - V_{t_{US}-1}$ now generates a positive prediction error that updates $V_{t_{US}-1}$.
- On subsequent trials, V_t is updated for each t in between t_{US} back to t_{CS} .

In other words, the value signal produced by the reward (US) is transferred back to an event (CS) that predicts the reward.

Case study (Second-order conditioning). Pairing a new CS to an existing CS.



Remark. The Rescorla-Wagner model is not capable of modeling second-order conditioning while the temporal difference model is.

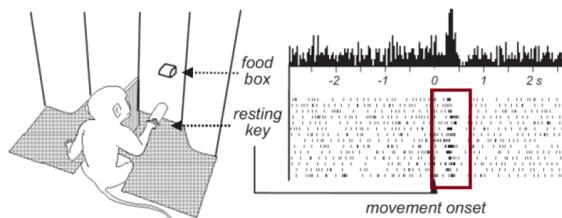
4.6 Reward prediction error hypothesis of dopamine

There is strong evidence that the dopaminergic system is the major neural mechanism of reward and reinforcement.

Response to unexpected rewards Dopaminergic neurons exhibit a strong phasic response in the presence of an unexpected reward.

Dopamine response to unexpected rewards

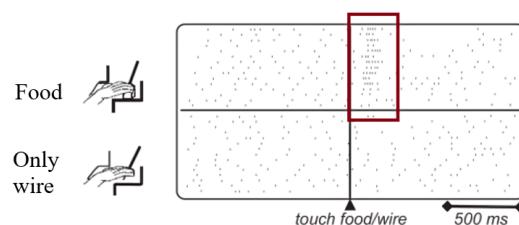
Case study (Monkey that touches food). Some food is put in a box with a hole to reach its content. In the absence of any other stimuli predicting the reward, a monkey presents a high dopaminergic response when it touches the food.



Reward discrimination Dopamine neurons respond differently depending on the actual presence of a reward.

Dopamine reward discrimination

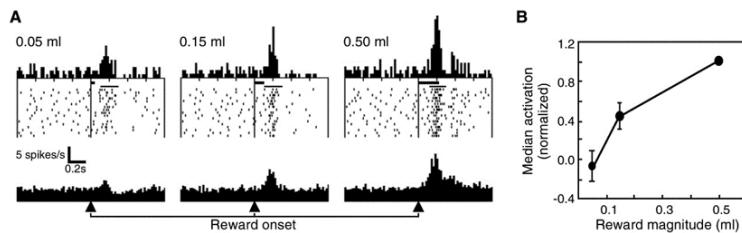
Case study (Monkey that touches food). The dopaminergic response of a monkey that touches an apple attached to a wire in a box is different from the response of touching the wire only.



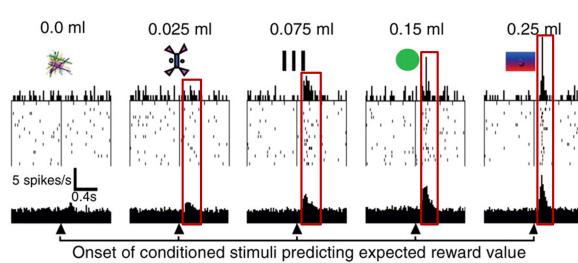
Magnitude discrimination Dopamine neurons respond differently depending on the amount of reward received.

Dopamine magnitude discrimination

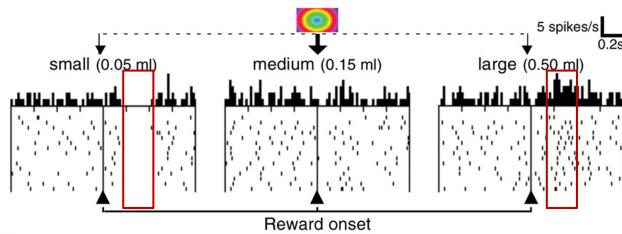
Case study (Monkey that drinks). By giving a monkey different amounts of fruit juice in a pseudorandom order, its dopaminergic response is stronger for the highest volume and weaker for the lowest volume.



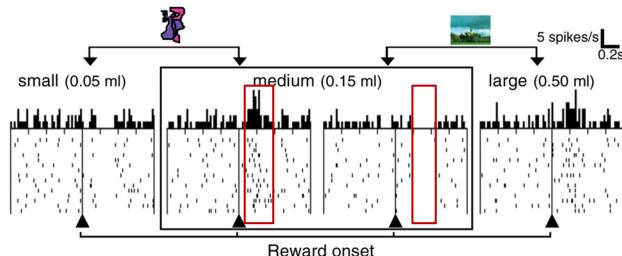
Case study (Monkey with juice and images). Using different CSs, it can be seen that the dopaminergic response differs based on the amount of reward.



Case study (Monkey with juice and images). After learning the association between a CS and US (middle graph), a change in the amount of the reward changes the dopaminergic response.



This behavior also involves the context (i.e. the CS image that is shown).



Remark. With the previous observations, it can be concluded that:

- Dopamine neurons increase their firing rate when the reward is unexpectedly delivered or better than expected.
- Dopamine neurons decrease their firing rate when the reward is unexpectedly omitted or worse than expected.

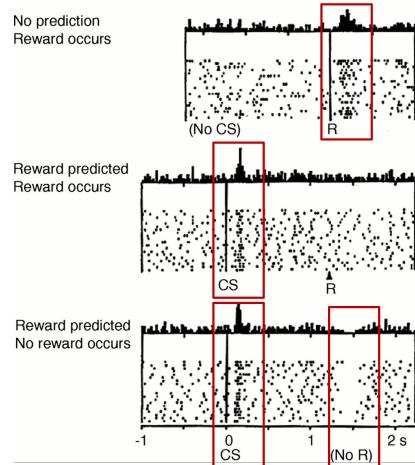
Transfer to CS

- Before training, an unexpected reward (US) causes the dopamine neurons to increase firing (positive prediction error).
- After training, dopamine neurons firing is increased after the CS but not following the reward (no prediction error).

Dopamine transfer to CS

- After training, dopamine neurons firing is increased after the CS but is decreased if the reward is omitted (negative prediction error).

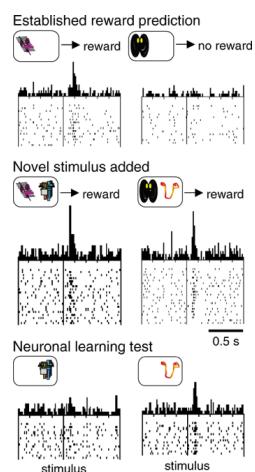
Case study.



Response to blocking Dopaminergic response is in line with the blocking effect.

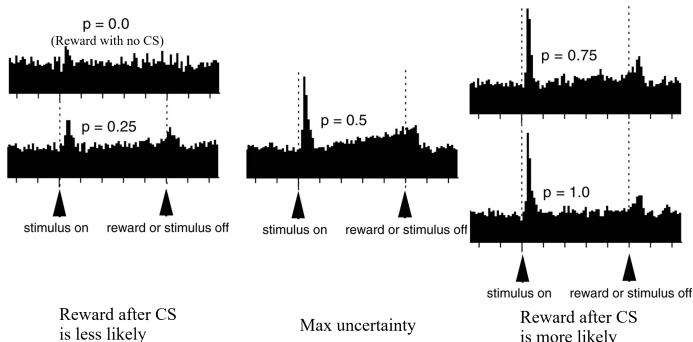
Case study (Monkey with food and images).

A monkey is taught to associate images with food. A new CS alongside an existing CS will not be learned.



Probability encoding The phasic activation of dopamine neurons varies monotonically with the reward probability.

Case study.



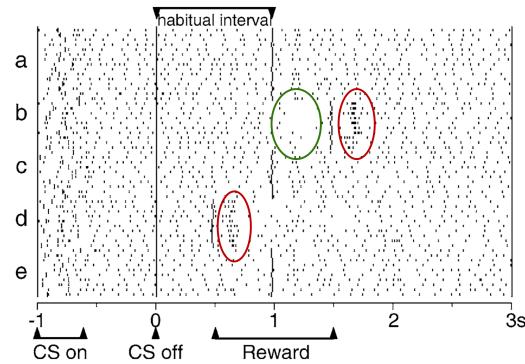
Timing encoding Dopamine response to unexpectedness also involves timing. A dopaminergic response occurs when a reward is given earlier or later than expected.

Dopamine response to blocking

Dopamine probability encoding

Dopamine timing encoding

Case study. After learning that a reward occurs 1 second after the end of the CS, dopamine neurons fire if the timing changes.



Remark. Dopamine is therefore a signal for the predicted error and not strictly for the reward.

5 Instrumental/operant learning

Form of control learning that aims to learn action-outcome associations:

- When a reinforcer is likely to occur.
- Which actions bring to those reinforcers.

This allows the animal to act in anticipation of a reinforcer.

Instrumental learning includes:

Habitual system Learn to repeat previously successful actions.

Habitual system

Goal-directed system Evaluate actions based on their anticipated consequences.

Goal-directed system

Depending on the outcome, the effect varies:

Positive reinforcement Delivering an appetitive outcome to an action increases the probability of emitting it.

Positive reinforcement

Positive punishment Delivering an aversive outcome to an action decreases the probability of emitting it.

Positive punishment

Negative reinforcement Omitting an aversive outcome to an action increases the probability of emitting it.

Negative reinforcement

Negative punishment Omitting an appetitive outcome to an action decreases the probability of emitting it.

Negative punishment

	Delivery	Omission
Appetitive	Positive reinforcement (+prob)	Negative punishment (-prob)
Aversive	Positive punishment (-prob)	Negative reinforcement (+prob)

Table 5.1: Summary of the possible effects

5.1 Types of schedule

There are two types of learning:

Continuous schedule The desired action is followed by the outcome every time.

Continuous schedule

| **Remark.** More effective to teach a new association.

Partial schedule The desired action is not always followed by the outcome.

Partial schedule

| **Remark.** Learning is slower but the response is more resistant to extinction.

There are four types of partial schedules:

Fixed-ratio Outcome available after a specific number of responses.

This results in a high and steady rate of response, with a brief pause after the outcome is delivered.

Variable-ratio Outcome available after an unpredictable number of responses.

This results in a high and steady rate of response.

Fixed-interval Outcome available after a specific interval of time.

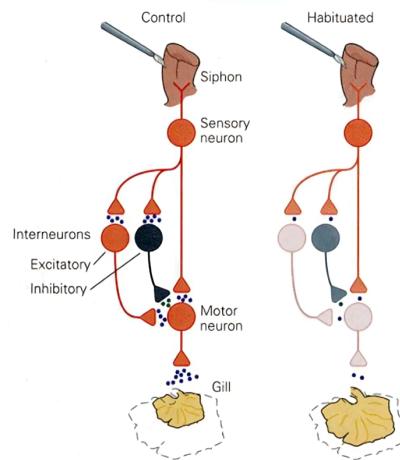
This results in a high rate of response near the end of the interval and a slowdown after the outcome is delivered.

Variable-interval Outcome available after an unpredictable interval of time.

This results in a slow and steady rate of response.

Case study (Aplysia Californica). An Aplysia Californica will withdraw its gill upon stimulating the siphon.

- Repeated mild stimulations will induce a habituation of the reflex.
- Repeated intense stimulations will induce a sensitization of the reflex.



5.2 Dopamine

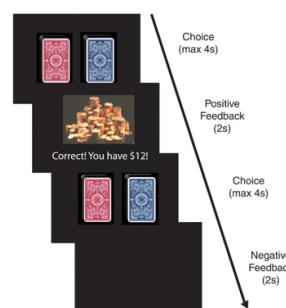
There is evidence that dopamine is involved in learning action-outcome associations.

Striatal activity on unexpected events When an unexpected event happens, there is a change in the activity of the striatum. There is an increase in response when the feedback is positive and a decrease when negative.

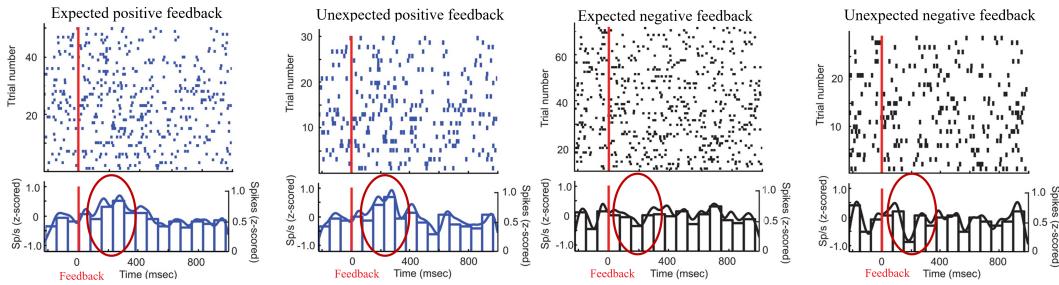
Striatal activity on unexpected events

Case study (Microelectrodes in substantia nigra).

The activity of the substantia nigra of patients with Parkinson's disease is measured during a probabilistic instrumental learning task. The task consists of repeatedly drawing a card from two decks, followed by positive or negative feedback depending on the deck.



The increase and decrease in striatal activity can be clearly seen when the feedback is unexpected.



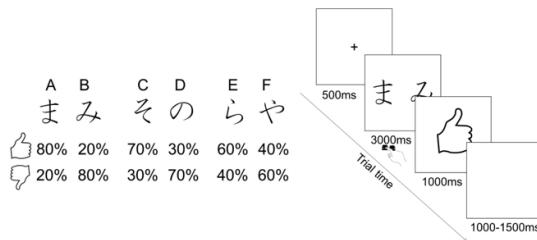
Dopamine effect on behavior The amount of dopamine changes the learning behavior:

- Low levels of dopamine cause an impairment in learning from positive feedback. This happens because positive prediction errors cannot occur.
- High levels of dopamine cause an impairment in learning from negative feedback. This happens because negative prediction errors cannot occur.

Dopamine effect on behavior

Case study (Probabilistic selection task). This instrumental learning task has two phases:

Learning There are three pairs of stimuli (symbols) and, at each trial, a pair is presented to the participant who selects one. For each pair, a symbol has a higher probability of providing positive feedback while the other is more likely to be negative. Moreover, the probabilities are different among the three pairs.



Participants are required to learn by trial and error the stimulus in each pair that leads to a positive reward. Note that learning could be accomplished by:

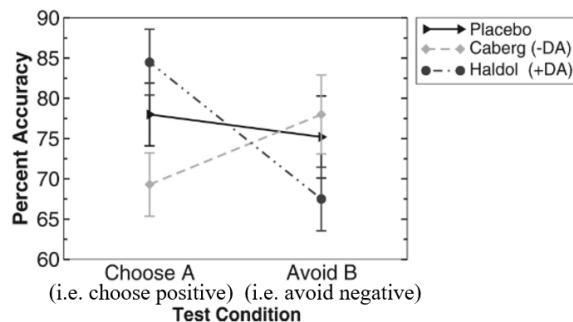
- Recognizing the more rewarding stimulus.
- Recognizing the less rewarding stimulus.
- Both.

Testing Aims to assess if participants learned to select positive feedback or avoid negative feedback.

The same task as above is repeated but all combinations of the stimuli among the three pairs are possible.

Three groups of participants are considered for this experiment:

1. Those who took the cabergoline drug (dopamine antagonist).
2. Those who took the haloperidol drug (dopamine agonist).
3. Those who took a drug without effects (placebo).

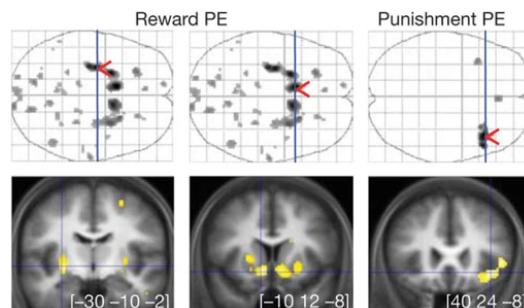


Results show that:

1. Cabergoline inhibited positive feedback learning.
2. Haloperidol enhanced positive feedback learning.
3. Placebo learned positive and negative feedback equally.

Case study. It has been observed that:

- Reward prediction errors are correlated with activity in the left posterior putamen and left ventral striatum.
- Punishment prediction errors are correlated with activity in the right anterior insula.



Actor-critic model Model to correlate Pavlovian and instrumental learning. It is composed by:

- The cortex is responsible for representing the current state.
- The basal ganglia implement two computational models:

Critic Learns stimulus-outcome associations and is active in both Pavlovian and instrumental learning. It might be implemented in the ventral striatum, the amygdala and the orbitofrontal cortex.

Actor Learns stimulus-action associations and is only active during instrumental learning. It might be implemented in the dorsal striatum.

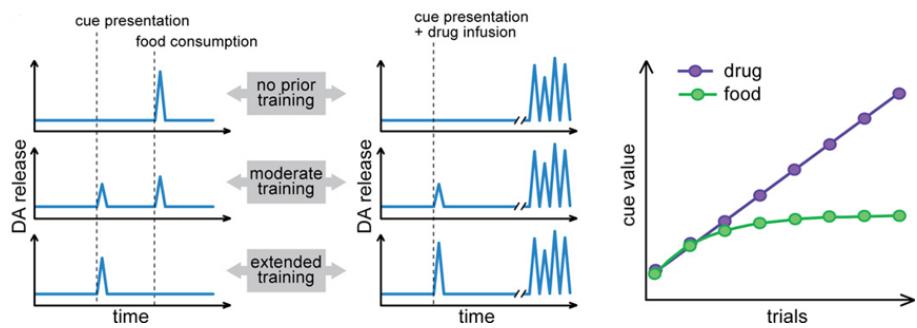
Actor-critic model

Critic

Actor

Case study (Food and cocaine).

- Food-induced dopamine response is modulated by the reward expectations that promote learning until the prediction matches the actual outcome.
- Cocaine-induced dopamine response causes a continuous increase in the predicted reward that will eventually surpass all other cues and bias decision-making towards cocaine.



5.3 Learning strategies historical evolution

5.3.1 Generation 0

There were two possible learning strategies:

Stimulus-response theory Learning happens by creating stimulus-response associations.

Learning does not happen if there is no reward.

Stimulus-response theory

Cognitive map / Field theory A mental map is created and used to find the best action in a given state based on the expected reward.

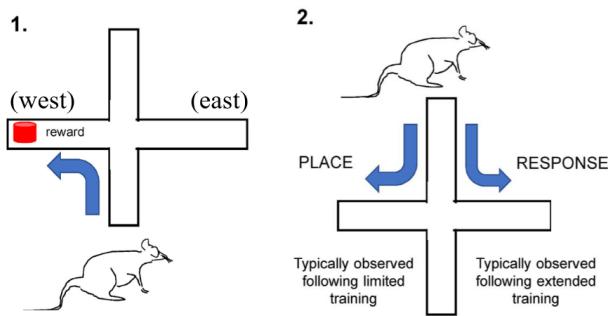
Cognitive map / Field theory

Latent learning Learning that is not shown behaviorally unless there is enough motivation.

Latent learning

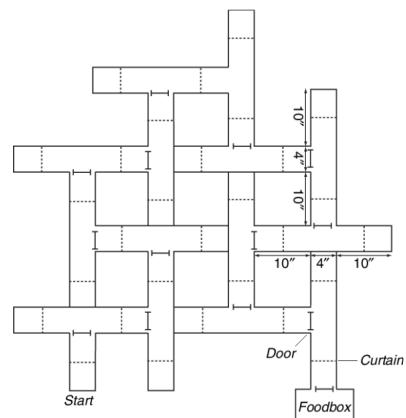
Case study (Maze). An animal is put at the start of a maze where a reward is located in the west arm. After some training iterations, the animal is put at the other entrance:

- If it goes to the west arm, it learned to solve the maze using a cognitive map/place strategy.
- If it goes to the east arm, it learned to solve the maze using a stimulus-response strategy.



It has been observed that rats start by learning a cognitive map (i.e. the environment is unknown). After enough training, they start relying on a response strategy (i.e. the environment is stable).

Case study (Tolman's maze). Consider a maze with curtains and doors to prevent a long-distance perspective.

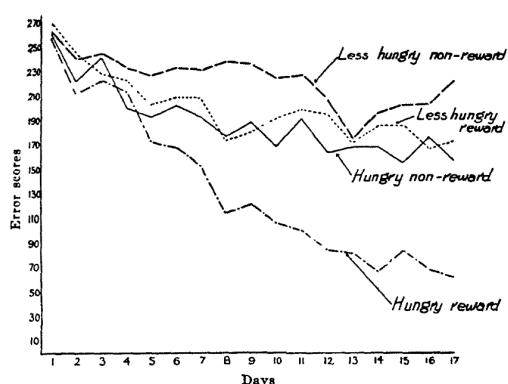


Two groups of hungry rats have been considered to solve the maze:

Group 1 No reward for solving the maze.

Group 2 Reward for solving the maze.

It has been shown that the second group completes the maze faster.



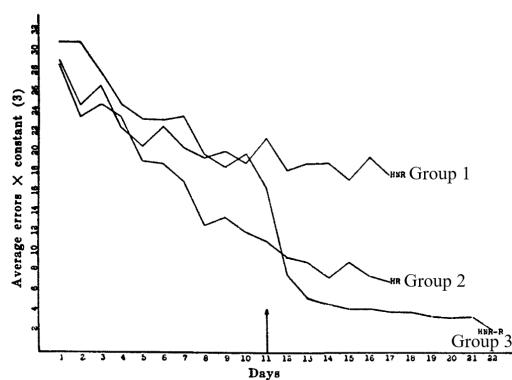
To show latent learning, three groups of hungry rats have been considered:

Group 1 No reward for solving the maze.

Group 2 Reward for solving the maze.

Group 3 Reward for solving the maze starting from day 11.

It has been shown that rats of the third group complete the maze faster as soon as they receive food.



5.3.2 Generation 1

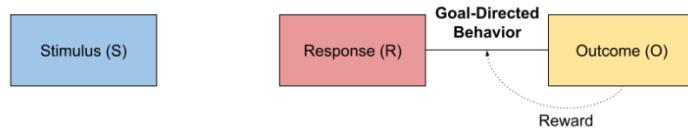
Shifted from studying the spatial domain to a more general domain. Based on two types of actions:

Goal-directed action Actions made because a desired outcome is expected. An action is goal-directed if:

- There is knowledge of the relationship between action and consequences (response-outcome).
- The outcome is motivationally relevant.

Goal-directed behavior has the following properties:

- Involves active deliberation.
- Has a high computational cost.
- It is flexible to changes of the environmental contingency (i.e. stops if no reward occurs)

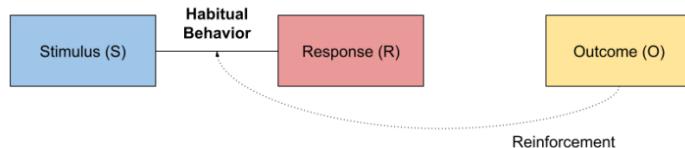


Habitual action Actions made automatically just because they were rewarded in the past. They are not influenced by the current outcome even if it is undesired.

Habitual action

Habitual behavior has the following properties:

- Does not require active deliberation.
- Has a low computational cost.
- It is inflexible to changes of the environmental contingency.



Case study (Goal-directed vs habitual behavior). The experiment is done in three steps:

Training The animal undergoes instrumental learning (e.g. associate that by pressing a lever some food will be dropped).

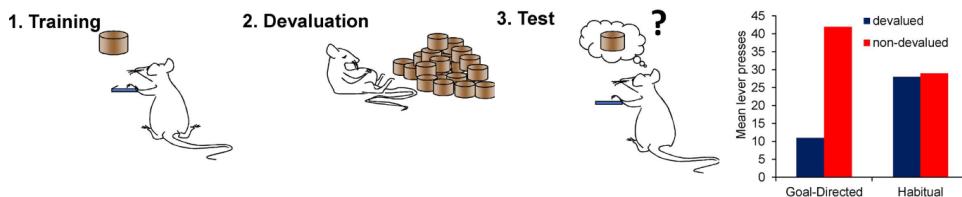
Devaluation Manipulate the learned behavior by either:

- Devalue the reinforcer.
- Degrade the contingency.

Testing Repeat the training scenario without reward:

- If the action associated with a devalued reinforcer is performed less, the behavior is goal-directed.
- If the frequency of the action is the same, the behavior is habitual.

Remark. The training phase aims to instill a goal-directed behavior. On the other hand, if the animal is overtrained, it will learn a habitual behavior. The experiment can be done both ways.



It has been hypothesized that the striatum might be the interface where rewards influence actions as:

Striatum

- The basal ganglia are involved in the selection of actions.
- The SNC affects the plasticity of the striatum through the release of dopamine.

Moreover, different sections of the striatum are responsible for different types of behavior:

Dorsomedial striatum Supports goal-directed behavior.

Dorsolateral striatum Supports habitual behavior.

This also hints that goal-directed and habitual behaviors act simultaneously and competitively (see Generation 3).

5.3.3 Generation 2

Studied goal-directed and habitual behavior in humans.

Functional magnetic resonance imaging (fMRI) Measures the ratio of oxygenated to deoxygenated hemoglobin molecules in the brain. It allows to indirectly measure the neuronal activity of the brain on a high spatial resolution (i.e. allows to see where things happen but not when).

Functional magnetic resonance imaging (fMRI)

Case study (Goal-directed behavior in humans). Candidates are trained to select between two fractals of which one leads to a reward with a high probability and the other with a low probability. The possible rewards are chocolate, tomato juice and orange juice (used as a control outcome).

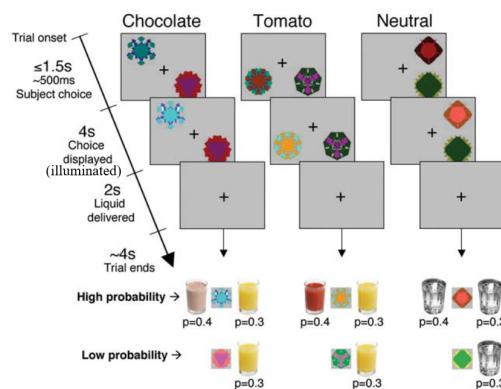


Figure 5.1: Structure of the task. The high probability choice leads to the primary reward (chocolate or tomato juice) with probability 0.4, to the control reward (orange juice) with probability 0.3 and to nothing with probability 0.3. The low probability choice leads to the control reward with probability 0.3 and nothing in the other cases. The neutral case leads to an empty glass or nothing.

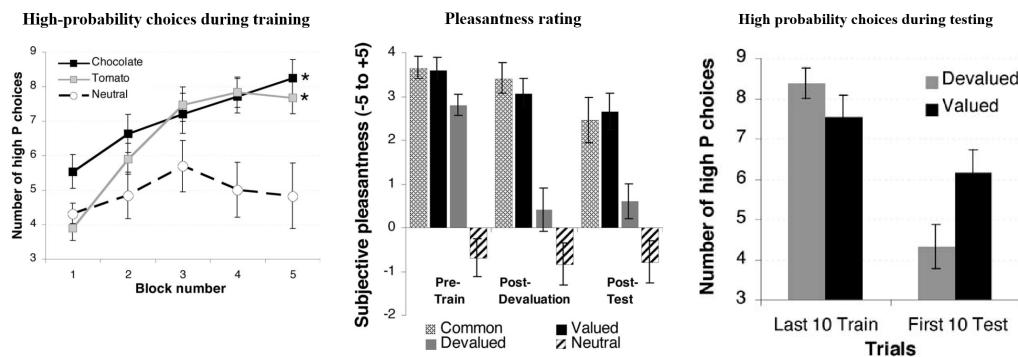
After training, one of the primary rewards is devalued through selective satiation and the other is labeled as the valued outcome. Then, the training task is repeated.



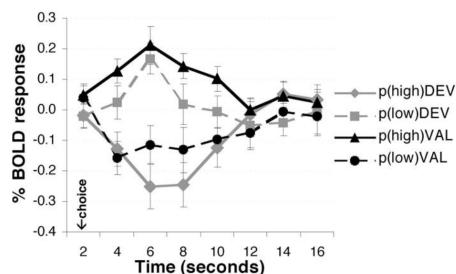
Figure 5.2: Steps of the experiment. In this figure, the devalued reward is the tomato juice.

Behavioral results show that:

- During training, candidates favored the high-probability actions associated with chocolate and tomato juice. On the other hand, choices for the neutral condition were evenly distributed.
- The pleasantness rating for the devalued reward lowered after devaluation while the valued reward remained higher.
- During testing, candidates reduced their choice of the high-probability action associated with the devalued reward.



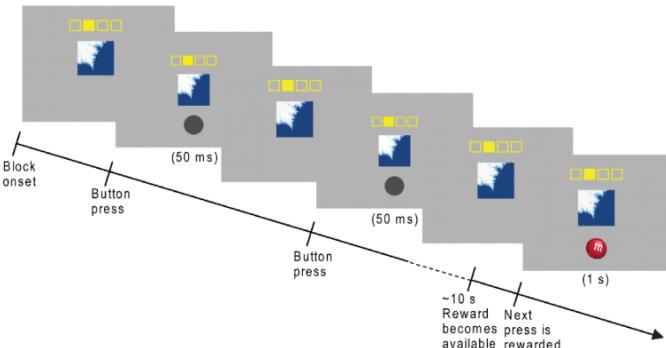
During both training and testing, the fRMIs of the candidates were taken. Neural results show that the **medial orbitofrontal cortex** has a significant modulation in its activity during instrumental action selection depending on the value of the associated outcome.



Case study (Habitual behavior in humans). Candidates are presented, at each round of the trial, with a fractal image and a schematic indicating which button to press. The button can be pressed an arbitrary number of times and, at each press, on the screen appears:

- A gray circle (no reward).

- The image of an M&M's © or Frito © (reward) with probability 0.1. To each fractal, only a type of reward can appear.



After training, one of the food rewards is devalued through selective satiation. Then, during testing, the same training task with the same stimulus-response-outcome is repeated without a reward.

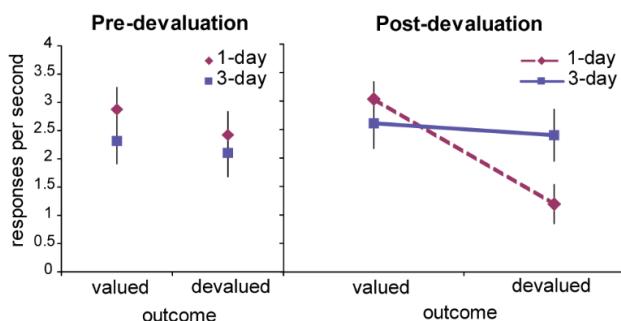
Two groups have been considered:

1-day group with little training.

3-day group with extensive training.

Behavioral results show that:

- Before devaluation, there were no significant differences between the responses of the two groups independently of the type of food.
- During testing, the 1-day group showed a goal-directed behavior while the 3-day group showed a habitual behavior.



During both training and testing, the fRMIs of the candidates were taken. Neural results show that, in the 3-day group, the **dorsolateral striatum** had significant activity.

5.3.4 Generation 3

Formalized goal-directed and habitual actions:

Model-based (Goal-directed) Use a model to predict the consequences of actions in terms of future states and expected rewards from future states.

Model-based

When the environment changes, the agent can update its policy of future states

without the need to actually be in those states.

Model-free (Habitual) Select actions based on the stored state-action pairs learned over many trials.

Model-free

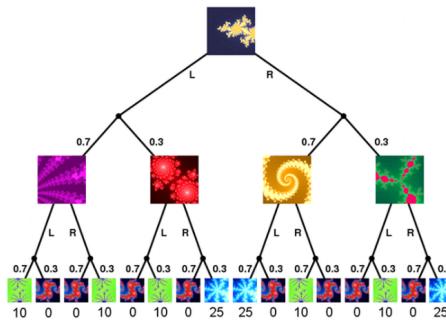
When the environment changes, the agent has to move into the new states and experience them.

Hybrid model Integrated computational and neural architecture where model-based and model-free systems act simultaneously and competitively.

Hybrid model

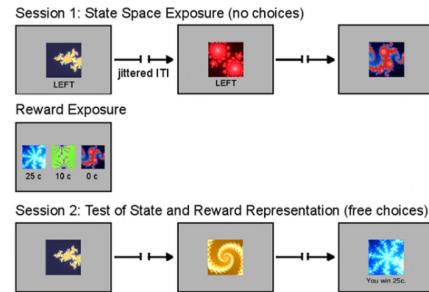
This is the currently favored model for behavior.

Case study (Latent learning in humans). The experiment consists of a sequential two-choice Markov decision task in which candidates navigate a binary decision tree. Each state contains a fractal image and candidates can choose to move to the left or right branch, each of which will lead with probability 0.7/0.3 to one of the two subsequent states. When a leaf is reached, a monetary reward (0€, 10€ or 25€) is delivered.



The experiment is divided into two sessions:

First session Candidates choices are fixed but they can learn the transition probabilities.

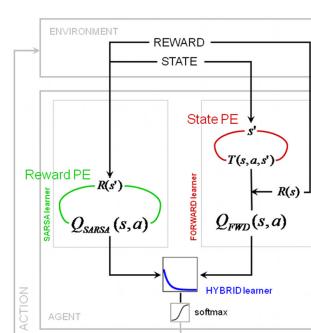


Before second session Candidates are presented with the association between fractal and reward.

Second session Candidates are free to choose their actions at each state.

Behavioral results show that the majority of the candidates are able to make the optimal choice. This indicates that their behavior cannot be explained using a model-free learning theory (as learning only happens with a reward). A hybrid model has been proposed to model the candidates' behavior. It includes:

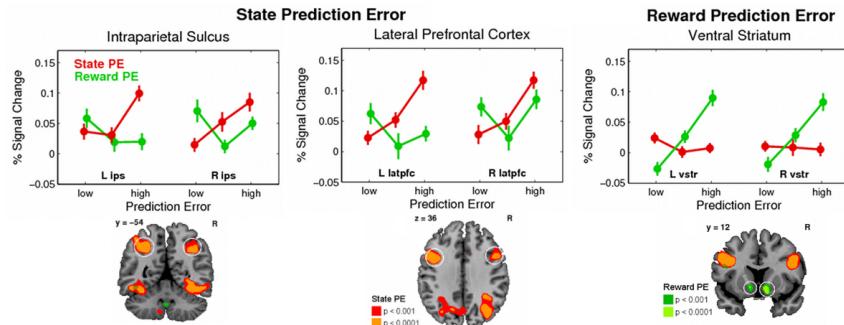
Reward prediction error Associated to model-free learning.



State prediction error Associated to model-based learning.

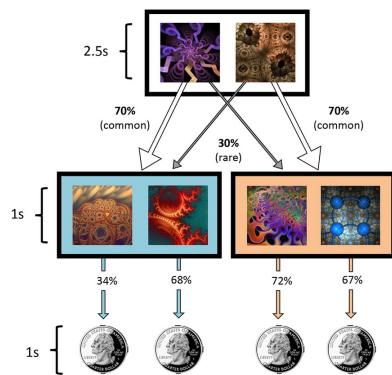
On a neuronal level, fMRI show that:

- State prediction error activates the **intraparietal sulcus** and the **lateral prefrontal cortex**.
- Reward prediction error activates the **ventral striatum**.



Case study (Model-free vs model-based in humans). Consider a Markov decision task that works as follows:

- In the first stage, candidates have to choose between two fractal images, each leading to one of the two subsequent states with probability 0.7 (common) and 0.3 (rare).
- In the second stage, candidates have to choose between two fractal images, each of which will lead to a monetary reward with a certain independent probability.
- The probability of receiving the reward changes stochastically during the trials.



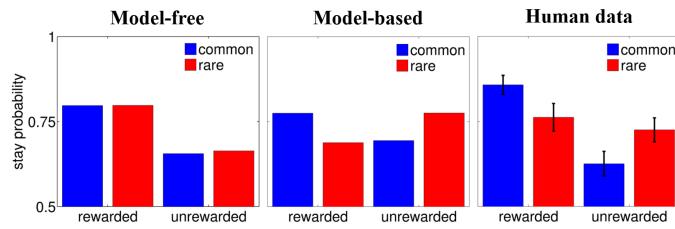
It is expected that:

Model-free agents Ignore the transition structure and prefer to repeat actions that lead to a reward in the past.

Model-based agents Respect the transition structure and modify their policies depending on the outcome.

They are more likely to repeat an action following a rewarding trial only if that transition is common.

Despite that, the actual results on human candidates show that a hybrid model is more suited to explain human behavior.



Neural results from fRMIs also show that the activity in the **striatum** increases for both model-based and model-free prediction errors.

<end of course>