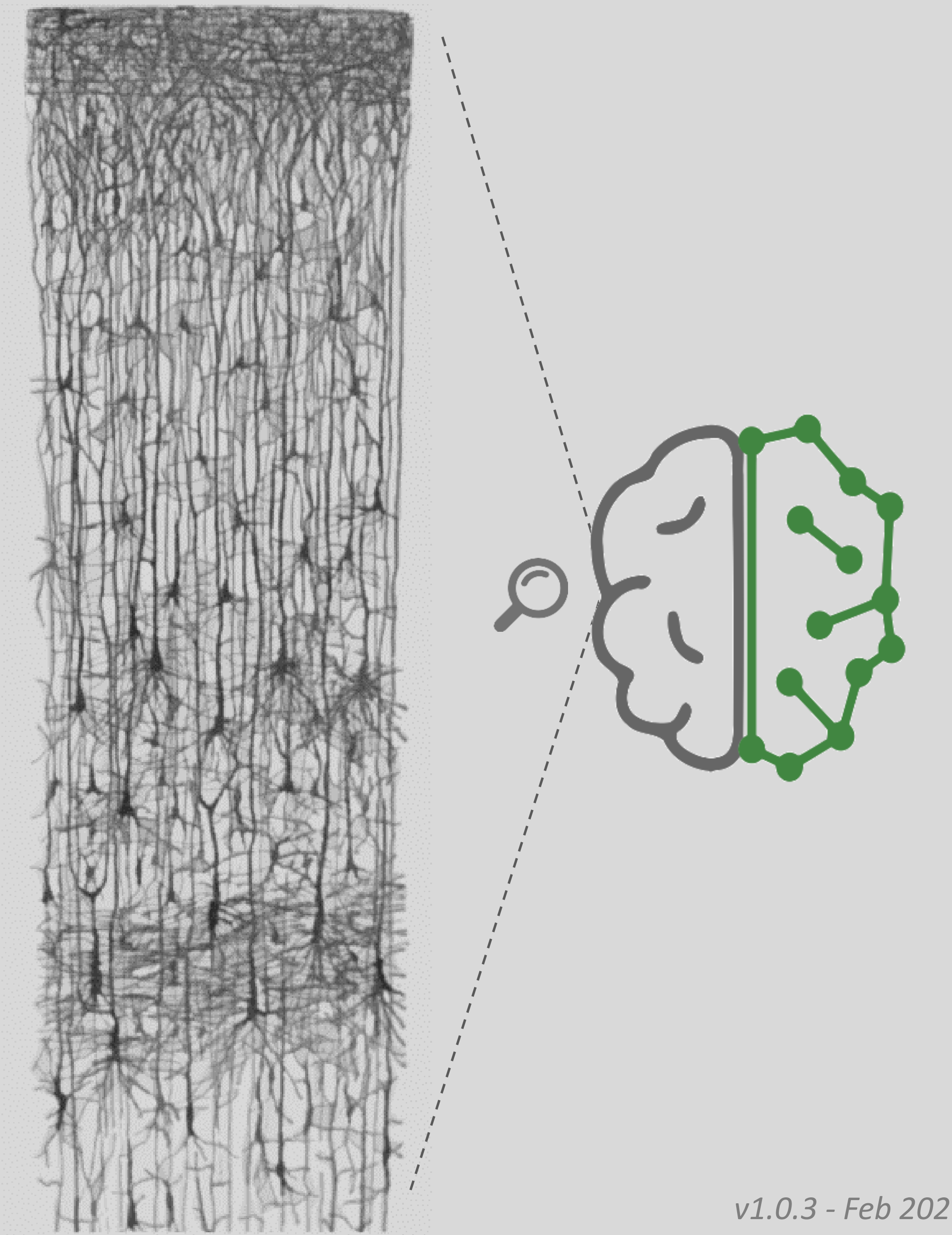


v1.0.3 - Feb 2021

Insights from the brain:

The road towards Machine Intelligence

Matthieu Thiboust

April 2020

This ebook is a personal view and synthesis of some selected experimental findings and theoretical ideas from neuroscience research, that are currently – or could be soon – used in neuroscience-grounded artificial intelligence (AI) approaches.

Even if the inconceivable complexity of our brains makes it near-impossible to perfectly understand its inner workings, we can still get valuable insights from an incomplete and modest approach.

The humble goal of this ebook is to provide AI researchers with neuroscience chunks of information related to AI. Some chunks are solid ground truths with large scientific consensus, others are general principles on which there is no complete consensus, and the remaining ones are just informed speculations that fit with theorical and experimental results.

This illustrated ebook formulates my own perspective of some key neuroscience knowledge that is currently (or could be soon) used in neuroscience-grounded AI efforts, following my deep conviction that the road towards machine intelligence is inseparable from a mixed AI & neuroscience approach. It builds upon my difficult but rewarding experience of navigating through neuroscience papers with a datascientist perspective during several months.

The first part – the longest – is dedicated to biological intelligence. It begins with the fundamental role of physical actions into the gradual emergence of high-level cognitive abilities through evolution. Then, the level of sophistication of the described neural machinery will appear unrivaled compared to today's deep learning artificial networks. I highlight the neocortex, a highly-researched brain structure that currently inspires many AI & neuroscience researchers because of its central role in human intelligence. In order to keep this document short, I had to make choices. One of those choices was to skip the focus on probably underrated subcortical sensorimotor circuits, and on two other popular brain structures in the AI community: the basal ganglia and the hippocampus. I keep those topics for another time.

The second part deals with biologically-inspired AI, starting with the modelling of more realistic neurons, architectures and learning rules into artificial networks. It subsequently continues with the transition from abstract artificial networks to artificial agents learning lifelong by interacting with their environment through their own perspective.

The primary target audience is the classical AI community interested to get insights from brain mechanisms. Also, curious neuroscientists who would like to keep up with neuroscience-grounded AI initiatives are invited to skip to the second part.

I already reached a personal goal with the completion of this ebook. My second goal will be reached if some AI & neuroscience enthusiasts benefit from this reading.

I would be happy to read your comments, answer your questions, correct the errors that you may have spotted, add key missing elements to the document, or just discuss machine intelligence & neuroscience with you.

Matthieu Thiboust

Table of contents / Summary

Biological intelligence

Machine intelligence

Introduction

- AI needs a new momentum.
  Why not look at the brain?

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- Yoshua Bengio
- François Chollet
- Demis Hassabis
- Jeff Hawkins
- Gary Marcus

See the reference section for a list of materials that inspired me.

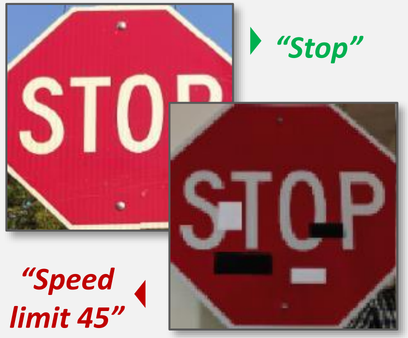
Art credit: Brainbow Hippocampus, Greg Dunn Design

The success of Deep Neural Networks hides some fundamental brittleness for Machine Intelligence

Despite numerous and impressive successes of *Deep Neural Networks (DNNs)* – commonly referred to as *Deep Learning* – during the last 10 years in visual/audio/text recognition, processing and synthesis, the **pace of breakthrough innovations is now slowing down**.

There seems to be no easy fixes for the **fundamental brittleness of DNNs** that perform brilliantly until they break in unpredictable ways when taken into unfamiliar territory. This classical AI approach – primarily designed to solve *specific and isolated tasks* – requires some important revolutions to lead to a longer-term vision of *Machine Intelligence*.

Examples of DNNs brittleness



The model is fooled by the addition of stickers on a "Stop" sign

The model is fooled by the addition of specifically designed noise even if the image looks similar

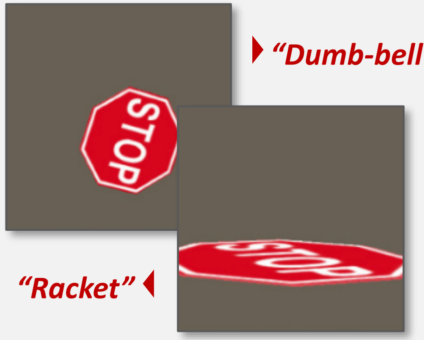
Me:
All men are mortal.
Socrates is a man.
Therefore, Socrates is

GPT-2 model:
immortal. Of course. That would imply that there is something beyond mortal.

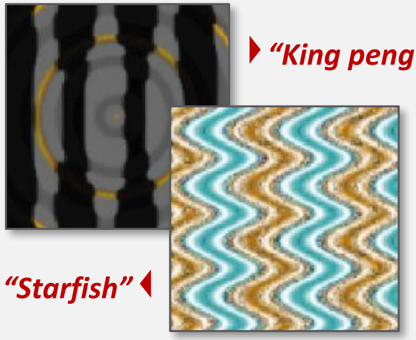
Me:
1+1=2
2+3=?

GPT-2 model:
Assuming that is true, if you apply the math, it implies that the former is a 1 since it represents a non-zero number.

No basic reasoning and arithmetic in Natural Language Processing (NLP) models



The model is fooled by rotation and perspective transformations



Abstract patterns fooling models

Adversarial examples from Eykholt, 2017, Deep Mind, 2019, and Nguyen, 2014

Common sense is the most sought-after feature when talking about Machine Intelligence

Sawing the wrong side of a tree limb!

The most regular criticism about AI systems is probably that they lack **common sense**, and it is sometimes explained as a lack of **intelligence**.

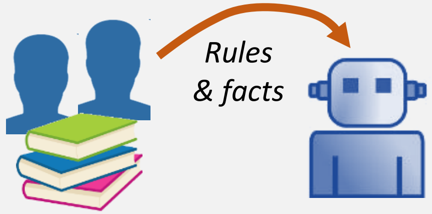
Common sense arises from the fact that most people in a group have a **shared background knowledge** about the world. For example, we expect people to know that *"Lemons are sour", "Cars do not fly", "Knives cut things", "Gravity causes things to fall down", "Falling from a significant height is dangerous", etc.*

To be useful and robust, an advanced AI system needs to share the background knowledge of its user community. A self-driving car should consider changing lanes when tailing an overloaded truck on a bumpy road, a home assistant should react when asked to prepare a meat dish for a vegan guest, and a robot should not saw the side of a tree limb he is sitting on.

In the spectrum of potential solutions to give machines common sense, the most appealing one relies on agents able to self-learn the shared background knowledge as they grow up by interacting with us, but **it needs some flavor of intelligence that machines currently do not have**.

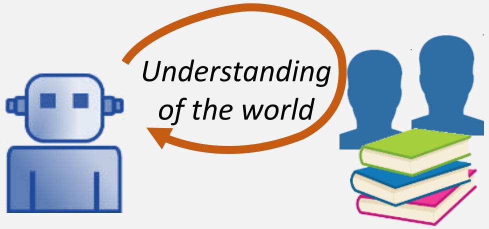
<u>Side note</u>: adding common sense is not only important for intelligence, but also making AI ethical if we consider that ethics is a collection of shared values within a society.

Putting a knowledge base inside the agent



Technically feasible but not flexible, not exhaustive and practically fastidious

Making the agent actively learn by interacting with us



Appealing approach but it requires some flavor of intelligence that machines currently do not have

Definition of “intelligence” will get clearer with progress in neurosciences and AI research

Even if the term is largely used by psychologists, philosophers, neuroscientists and AI researchers, “***intelligence***” **is still an elusive concept with no widely adopted universal definition** in the scientific community.

The word “*intelligence*” is a source of confusion amalgamating several meanings. With scientific progress, the definition will be progressively refined by separating these meanings, as it was done in the 19th century for the words "heat" and "temperature".

Obviously, the same applies to other concepts like *consciousness, cognition, thinking, attention, perception, understanding, emotion*…

Because one cannot advance by totally ignoring this problem, here is a *still-to-be-refined* definition of intelligence (from Legg and Hutter, 2007):

> *“Intelligence measures…*
>
> *…an agent’s ability to* ***achieve goals*** *in a* ***wide range of environments****”*

Task-specific skills
(specificity & static)

Skill-acquisition ability
(generality & adaptation)

Up to now, AI systems have mostly dealt with task-specific skills. In order to push forward AI research towards more intelligent and more human-like artificial systems, we need to focus on the **broader and more complex skill-acquisition efficiency** part of the definition of intelligence (Chollet, 2019).

Remark: Defining “Intelligence” is not a prerequisite to advance in Machine Intelligence research. We – living examples of intelligent agents – can still get inspiration from ourselves.

Intelligence is multidimensional

Spatial, linguistic, logical, kinesthetic, musical, interpersonal, intrapersonal, naturalist, existential, moral (Gardner, 2009)

Intelligence varies across a continuum

Within each dimension, agents' performances could be categorized and quantified(across human beings with IQ tests, or across the animal kingdom)

Intelligence is not restricted to biological agents

No *a priori* reason why this ability would be reserved to existing living creatures. Artificial agents could show some degrees of intelligence

Definition of intelligence is often anthropocentric

Tendency to define intelligence as a collection of human’s abilities not yet mastered by machines. This definition evolves with AI progress.

Because *Artificial Intelligence (AI)* and *Artificial General Intelligence (AGI)* became strongly loaded expressions, this presentation prefers the still-innocent term *Machine Intelligence.*

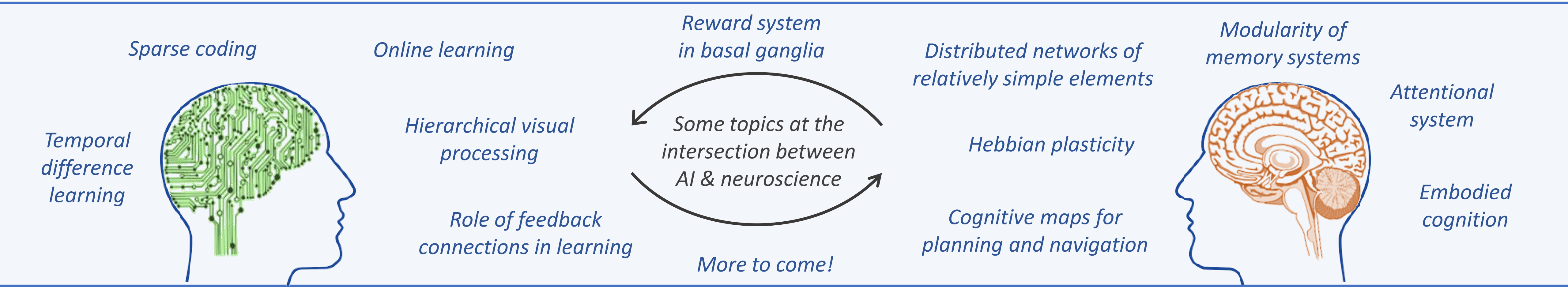
Our brain is the most reliable source of inspiration for Machine Intelligence

Researchers are following **different paths towards Machine Intelligence** that can be grouped into two global approaches:

- A *fundamental approach* primarily leveraging our knowledge in abstract scientific fields such as mathematics, information theory, physics, logic and causality. This approach is commonly referred to as *Symbolic AI*.
- A *biologically-inspired approach* at the intersection between neurosciences, psychology and computer sciences. *Connectionist AI* falls into this category even if most of Artificial Neural Networks (ANNs) remain far from their biological counterparts.

Because the successful connectionist approach currently faces bottlenecks, some researchers are now trying to merge symbolic AI ideas into ANNs, while others are **attempting to make those ANNs even more biologically realistic**. The latter has the advantage to have the human brain as a reliable and invaluable guide to progress incrementally towards Machine Intelligence, without the risk of running into a dead-end requiring us to go back to square one.

Moreover, the **collaboration between bio-inspired artificial and biological intelligence** has already proven to be productive for both fields even if we still have a very long way to go in mimicking truly human-like intelligence:



Sparse coding

Temporal difference learning

Online learning

Hierarchical visual processing

Role of feedback connections in learning

Reward system in basal ganglia

Some topics at the intersection between AI & neuroscience

More to come!

Distributed networks of relatively simple elements

Hebbian plasticity

Cognitive maps for planning and navigation

Modularity of memory systems

Attentional system

Embodied cognition

Bio-inspired Artificial Intelligence

Biological Intelligence

Multidisciplinary approaches will considerably speed up the path towards Machine Intelligence

The multidisciplinary nature of AI

Since the first mention of AI research in the 1950s, the academic field of AI has largely evolved from a *computer science* subfield to a **highly multidisciplinary field** encompassing diverse fields like *information engineering, robotics, mathematics, psychology, linguistics, philosophy* and *neuroscience* (not exhaustive).

During the early decades of this long journey, many AI practitioners were well versed in neuroscience. It led to the idea that networks of simple elements can produce remarkable computations, and that some network architectures are well suited for pattern recognition tasks.

Today, subfields at the intersection of AI and neurosciences like *computational neuroscience, cognitive neuroscience* and *system neuroscience* have taken over this increasingly specialized research with promising results for our understanding of the brain.

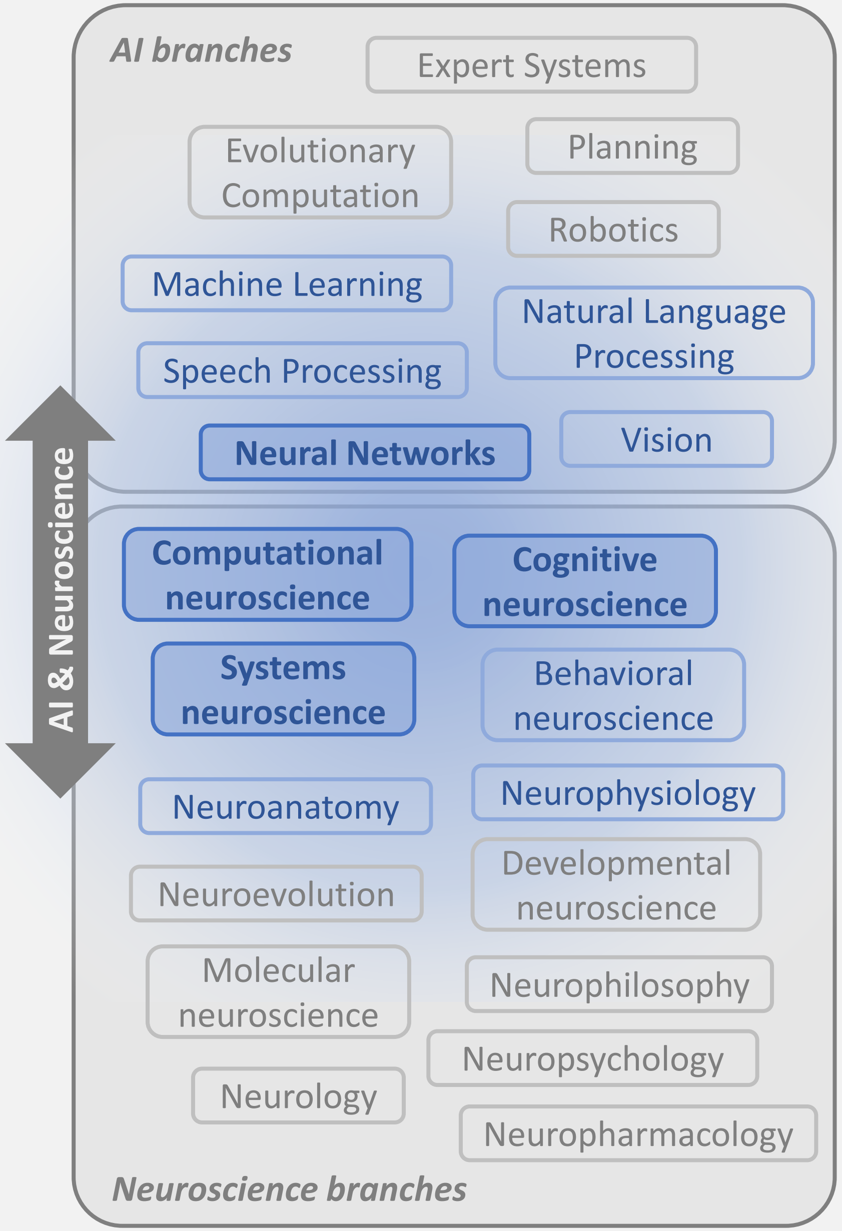
Bridging the gap between AI and neuroscience

Harvesting the next low-hanging fruits relies on multidisciplinary approaches. Research information produced by neuroscience and computer science are not read enough outside their respective fields.

It is not surprising given that researchers from both fields already have a hard time keeping up with the incredible number of publications from their peers, even in their own subfield.

From the perspective of the AI researchers, there are **hurdles to overcome** if they want **to navigate the jungle of experimental results in neuroscience**: few and often disputed frameworks to make sense of the findings, complex naming conventions, high variability of results (sometimes even contradictory results) due to cross-species differences, in vivo vs in vitro, awake vs anesthetized, staining methods, or conduct of the experiment itself.

My hope is that this document can simplify the first step of this effort for a curious AI researcher.



Thiboust, 2020

Brains & cognitive abilities

1. The primary function of a brain is not to think but to efficiently control complex behavior

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- György Buzsáki
- Paul Cisek
- Antonio Damasio
- Sten Grillner
- Joseph Ledoux
- Luis Puelles

See the reference section for a list of materials that inspired me.

Art credit: Midas and the Bandsaw, Greg Dunn Design

The brain is the central part of a bidirectional signaling system that generates complex behavior

The **nervous system** is an **electrical-based signaling system** supporting complex functions and structures of multicellular organisms. Compared to the endocrine hormone-based signaling system, nervous systems are much faster and much more specific for transmitting information, while being energy-efficient (only 20 watts in humans).

The nervous system has two main components:

- The central nervous system (CNS) composed of the **brain** and the spinal cord. It is the major processing unit of nervous systems
- The peripheral nervous system (PNS) with nerve fibers reaching almost all body parts in two opposite pathways (from and towards the CNS)

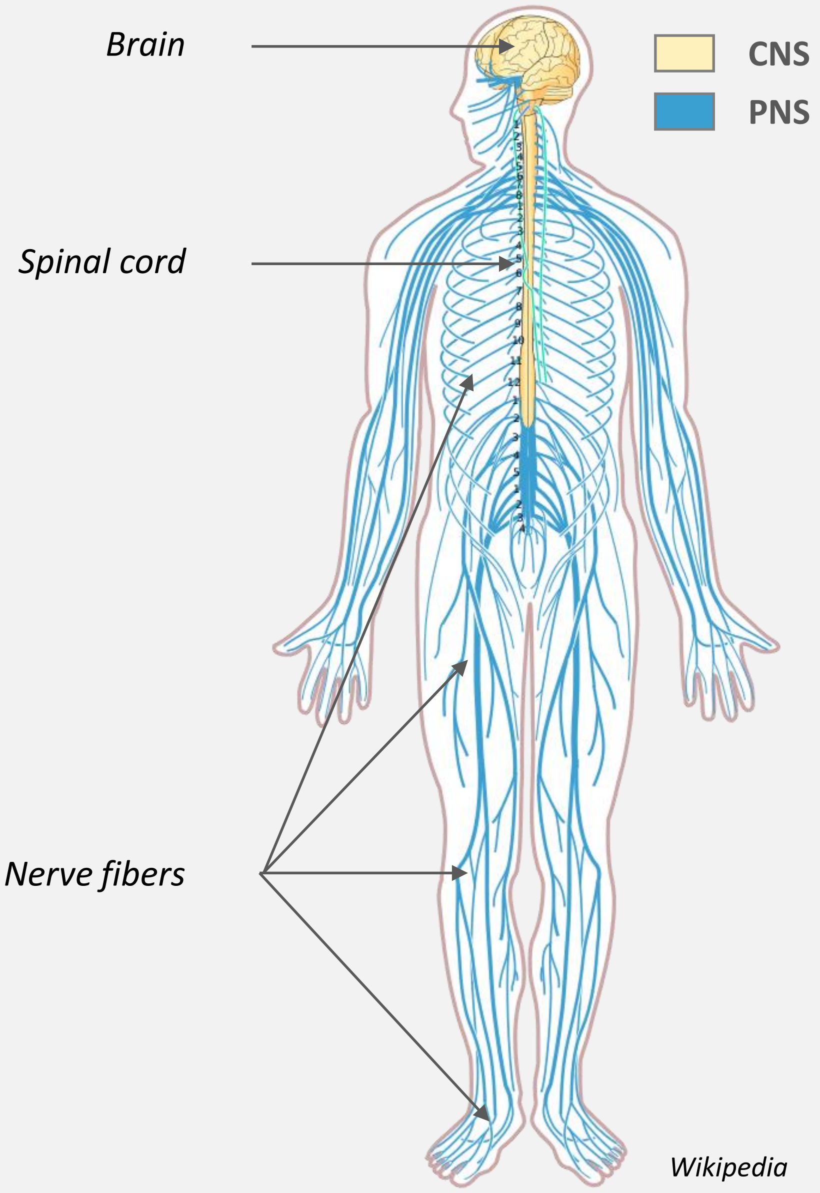
By **coordinating situation-dependent distributed sequences of actions throughout the body**, they generate appropriate **complex behaviors** to sustain the *homeostasis process* (perpetuation of life as an organism and a species).

Two kinds of action:

- **Secretion of hormones** (coupling with the endocrine system via the hypothalamus & hypophysis)
- **Contraction of muscles**: smooth muscles (in walls of hollow visceral organs, except the heart) and striated muscles (skeletal and cardiac muscles)

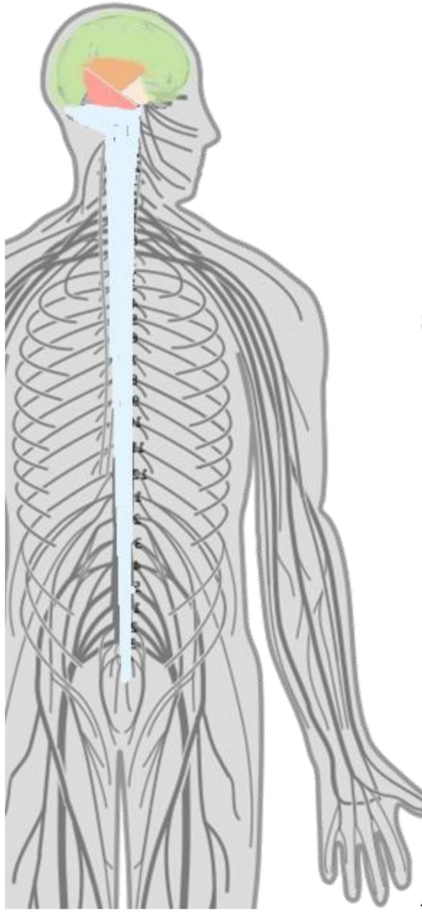
Different classes of senses:

- **Exteroception** for environmental stimuli (sight, hearing, touch, smell, taste)
- **Interoception** for internal vegetative stimuli (from organs, muscles and blood vessels)
- **Proprioception** for internal position and dynamics of the body (muscle tension, joint orientation, sense of balance, …)

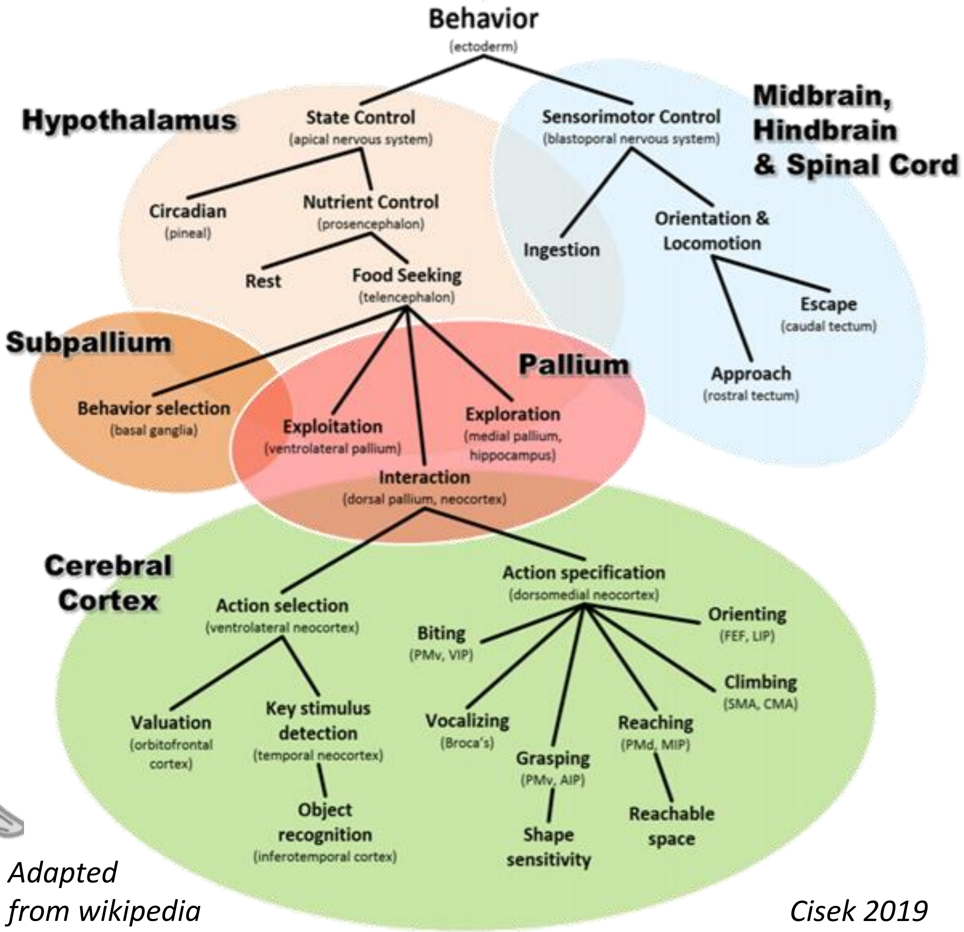


Wikipedia

The control of the different behaviors is distributed over a collection of brain substructures



Adapted from wikipedia



Cisek 2019

Because new brain structures were progressively added on top of previous ones along the phylogenetic tree of evolution, it is tempting to associate a newly acquired ability with a newly acquired substructure, and postulate that the very function of this ancient substructure was mostly preserved in today's descendants. Admittedly, it is an oversimplification of a very intricated system, but it helps to get the big picture.

Key behaviors <> key macro-structures

The **hypothalamus** is the structure that is in charge of the regulation of **basic vital needs** of the body like hunger, temperature, thirst, fatigue, sleep, circadian rhythms. Because some of those needs are complex to satisfy, the hypothalamus delegates some of its functions to the **telencephalon**, a structure composed of:

- A **subpallium (basal ganglia)** for behavior selection
- A **pallium (hippocampus & cerebral cortex)** for exploitation, exploration & interaction behaviors (behaviors such as orienting, reaching, grasping or vocalizing are associated with the cortex)

An ancient structure, the **tectum**, is associated with vital escape and approach behaviors, in parallel to appetitive versus aversive subcircuits in the **habenula** and the **amygdala**.

Also present in the midbrain, hindbrain and spinal cord, **Central Pattern Generators** control stereotyped motor behaviors like walking, swimming, flying, ejaculating, urinating, defecating, breathing, or chewing.

Rather than supporting specific given behaviors, the **cerebellum** in the hindbrain allows the coordinated unrolling of learned behaviors.

Perception and cognition gradually emerged throughout evolution to support increasingly complex behaviors

Brains did not evolve with perception or cognition as a target.

Firstly, evolution does not follow targets. It just selects biological structures that prove to be useful in the quest for survival.

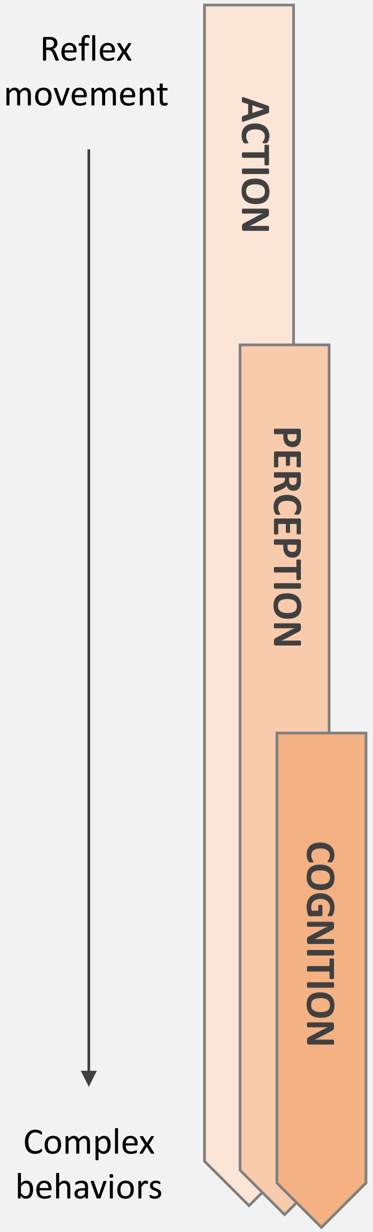
Secondly, perception and cognition are not an end in themselves. They emerged as gradual and quantitative abilities to primarily help **generate more appropriate complex behaviors**.

Reflex movements are the most basic behaviors consisting of triggering a set of actions when specific receptors are stimulated. Simple organisms can sustain life with those simple behaviors.

Perception goes beyond the instantaneous feeling of sensations. It compares sensations with memories of similar experience to identify the evoking stimulus. Organisms that perceive are able to associate a valence (goodness scale) to situations in order to select an appropriate behavior and flexibly adapt its execution.

Cognition adds the ability to form internal representations and use them to guide complex behaviors requiring abilities such as planning, thinking long term, building upon other's knowledge, making rational choices…

Understanding brain function should begin with brain mechanisms and explore how those mechanisms give rise to the performance we refer to as action, perception and cognition.



Blinking of eyes when touched

Vestibulo-ocular reflex

Digesting

Breathing

Having goose bumps

Raising cardiac pulse if in danger

Basic walking

Screaming when hurt

Approaching potential mates

Walking on a boat deck

Singing a song

Reaching & grasping a fruit

Returning to a pleasant place

Dancing in a ballet

Reaching & grasping a server tray

Reading a newspaper

Writing a ballet

Painting a landscape

Designing tools

Teaching philosophy

Proving a theorem

Writing a book

Building a bridge

Thiboust, 2020

Brains & cognitive abilities

2. This control is supported by abilities that were progressively acquired and refined through evolution

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- Paul Cisek
- Antonio Damasio
- Sten Grillner
- Joseph Ledoux
- Kevin Mitchell
- Luis Puelles

See the reference section for a list of materials that inspired me.

Art credit: Midas and the Bandsaw, Greg Dunn Design

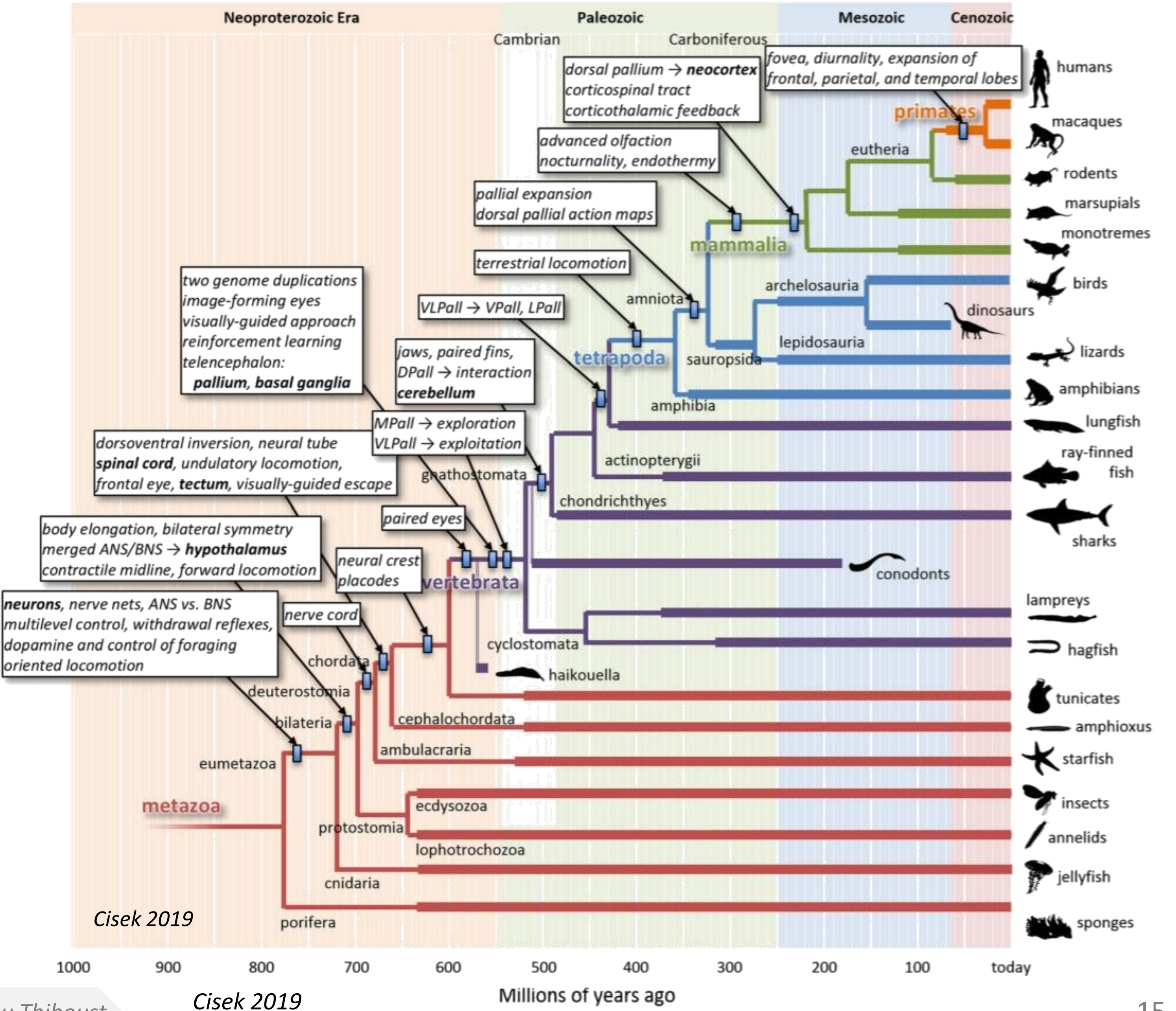
Brains of current living creatures result from an evolutionary process of 700 million of years

Through *natural selection*, nature has progressively come up with *brains* as a solution to the problem of **controlling increasingly complex behavioral activities on the quest for survival of individual living organisms and their species as a whole**, requiring the coordination of the activities of cells distributed over different parts of the body.

Put differently, brains are first and foremost evolved tools that coordinate the **homeostasis process of multicellular organisms** for survival and reproduction.

Deep timeline of evolution:

- *14 billion years: formation of the universe*
- *4 billion years: appearance of life*
- *700 million years: first nervous system*
- *550 million years: first vertebrate*
- *300 million years: first mammal*
- *50 million years: first primate*
- *5 million years: chimpanzee/human last common ancestor*
- *2 million years: homo habilis*
- *0,4 million year: homo sapiens*



Cisek 2019

The neocortex is a major brain innovation along the vertebrate phylogenetic branch

Researchers have inferred the brain organization of vertebrate ancestors from their still living successors.

Being the most phylogenetically-distant currently living vertebrates, **lampreys** are a good proxy of the ancestral vertebrate brain. They possess:

- A *set of tectal visuomotor circuits for species-typical approach and avoidance behavior* (the superior colliculus is the mammalian evolution of the optic tectum)
- *Olfactory foraging systems* forming the initial *telencephalon (pallium & subpallium)* to arbitrate between local exploitation (*ventrolateral pallium* for olfaction and ingestion) and long-range exploration (*medial pallium* for navigation which will later become the hippocampus) for controlling nutrient concentration.

Later, **jawed vertebrates** evolved two new structures:

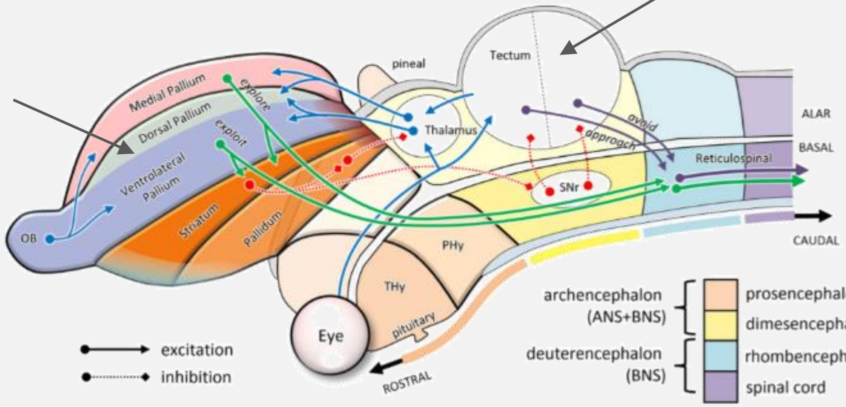
- A larger *dorsal pallium* specialized for sensorimotor interactions
- A *cerebellum*

Then, the size of the *dorsal pallium* increased a lot with **mammals** into what is called the *neocortex* (or *isocortex* since it is not a complete innovation of the mammals), which continued to increase a lot with **primates** and **humans**. A larger *neocortex* means an increased capacity to process sensory stimuli (like vision) and a larger repertoire of sniffing, burrowing, reaching, and grasping behaviors.

Ancestral vertebrate brain

Small neocortex ancestor

Tectal visuomotor circuits



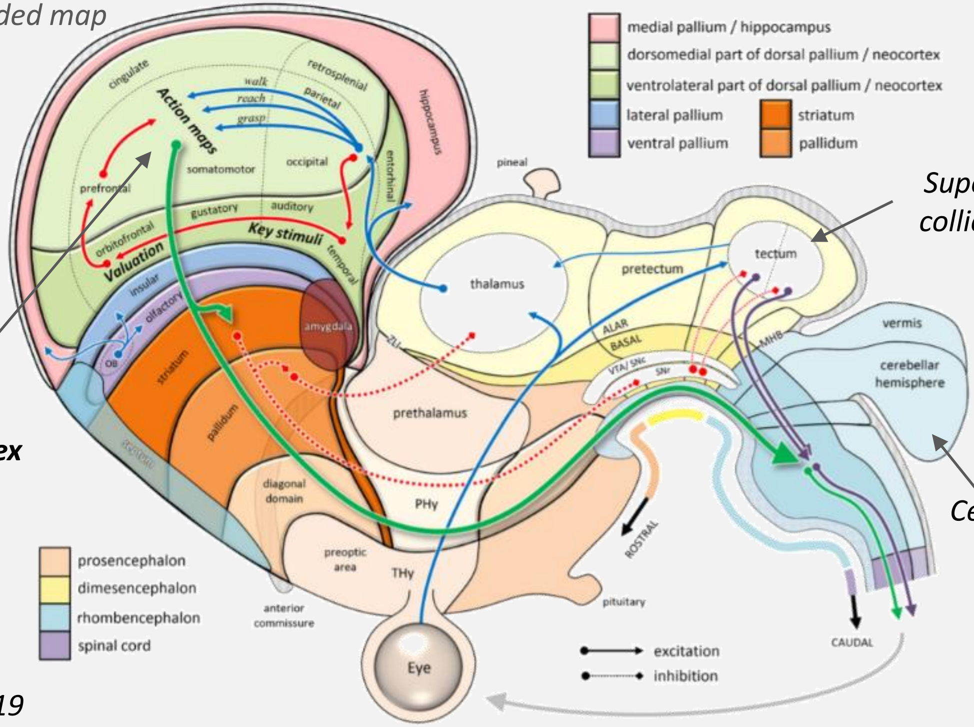
Mammalian brain

Unfolded map

Large neocortex

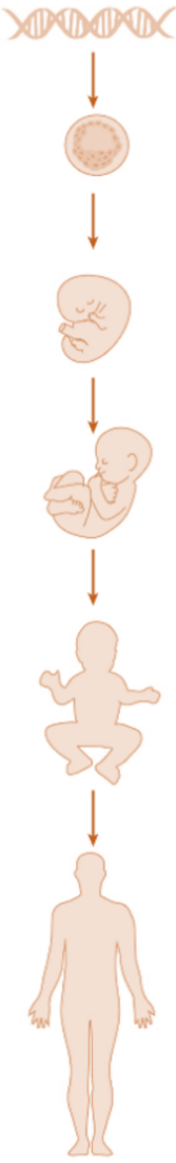
Superior colliculus

Cerebellum



Cisek 2019

Evolution plays with the brain developmental recipe, not directly with mature brain characteristics



The position of our **86 billion neurons** and the connections of their **100 trillion synapses** are obviously not directly encoded in our genetic material. First, it would not be possible to store those explicit design characteristics into our genome with “only” 3 billion of base pairs. Second, our nervous system would be far less flexible if all connections were hardcoded.

Instead, **our genome encodes developmental rules** like a recipe specifying how to make a mature brain from neural stem cells. Those rules are executed in each cell by the sequential expression of specific genes depending on the cell surroundings, thanks to other genes ruling those conditional gene expressions (depending on chemical gradients).

Of the 20,000 genes in the human genome, at least one third are primarily expressed in the brain. It means that a significant portion of our genome is dedicated to our brain recipe.

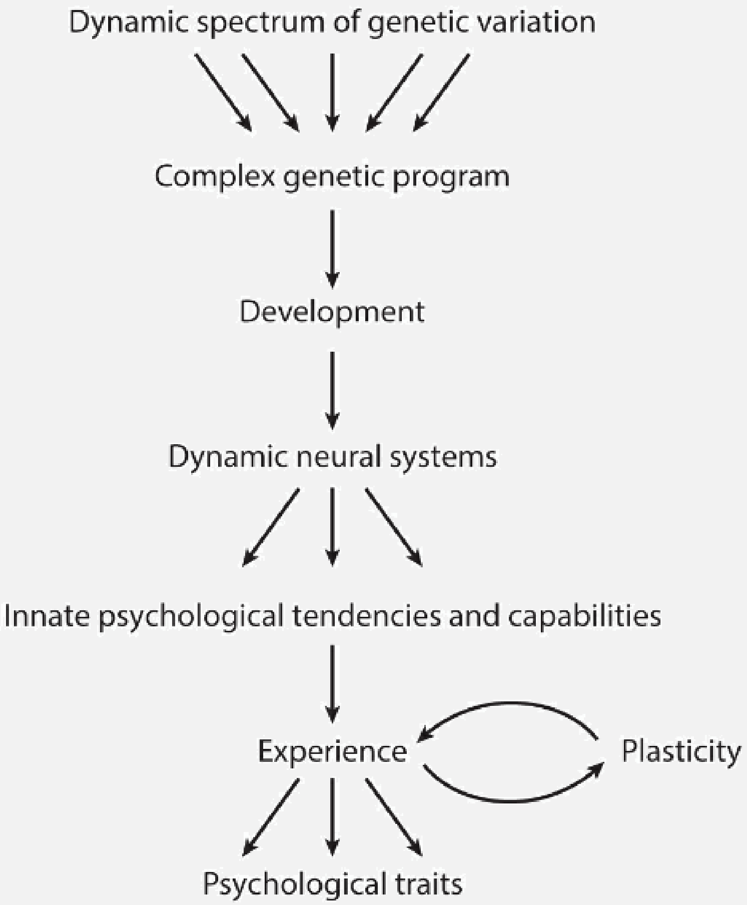
Genetic mutations in those genes can impact the brain development, leading to neutral, beneficial or harmful effects. Such beneficial mutations in germ cells will be progressively transferred to next generations through natural selection. **Evolution plays with the recipe**, not directly with the final characteristics.

Mutation in genes used in early developmental phases have statistically less chances to be beneficial because of subsequent cascading effects over the remaining developmental phases. Thus, as a general rule, the chronological order of brain developmental phases mainly reflects the chronological order of brain evolution in the phylogenetic tree.

It is important to underline the high stochasticity inherent to brain development. Two identical twins raised in the same environment will likely have different traits because they have followed slightly different developmental paths (innate but not genetic).

The brain recipe

Simplified view of the complex genetic architecture of behavioral and psychological traits (from Mitchell, 2018):



Illustrations from Mitchell, 2018

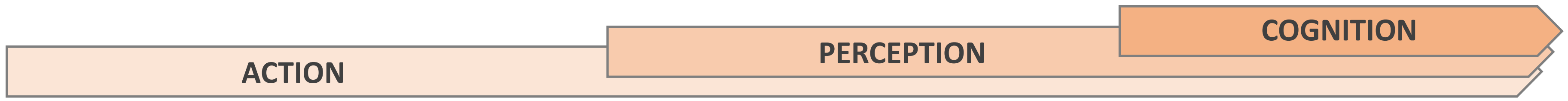
Our abilities progressively mature with a carefully choreographed developmental process

Progressive emergence of cognitive abilities during the developmental process

Embryos acquire knowledge of their body via initially meaningless random movement patterns (ex: baby kicks, muscle jerks)

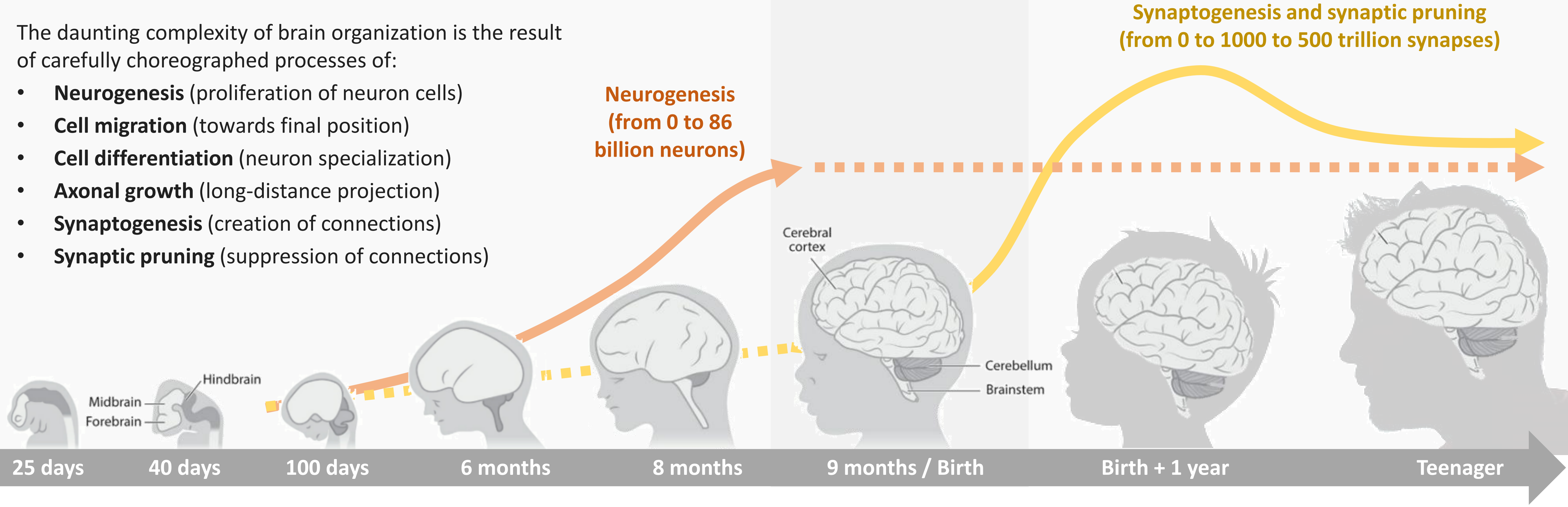
Babies tie external stimuli induced by their movements to a self-organized brain activity

Infants can sustain internal brain activity without producing movement



The daunting complexity of brain organization is the result of carefully choreographed processes of:

- **Neurogenesis** (proliferation of neuron cells)
- **Cell migration** (towards final position)
- **Cell differentiation** (neuron specialization)
- **Axonal growth** (long-distance projection)
- **Synaptogenesis** (creation of connections)
- **Synaptic pruning** (suppression of connections)



Images adapted from Mitchell, 2018 and from Kolb et al, 2009

Thiboust, 2020

Brains & cognitive abilities

3. Biological intelligence gradually emerged with active perception and cognition

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

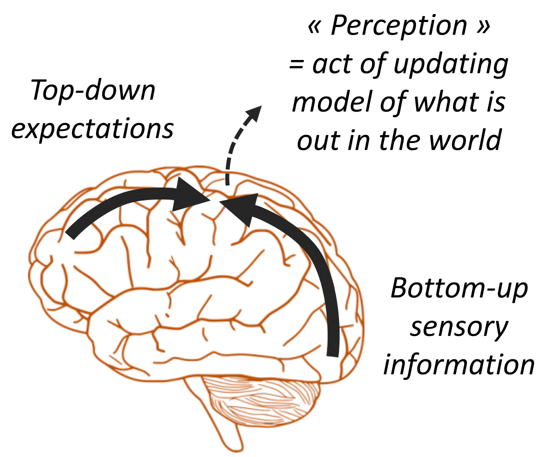
- György Buzsáki
- Paul Cisek
- Karl Friston
- Carlos E. Perez
- Giovanni Pezzulo

See the reference section for a list of materials that inspired me.

Art credit: Midas and the Bandsaw, Greg Dunn Design

Perceptions are constructed mental representations, not the veridical representations of the objective world

Perception is our sensory experience of the world around us. It results from the interpretation of bottom-up sensory stimuli based on internal top-down expectations.


Top-down expectations
« Perception » = act of updating model of what is out in the world
Bottom-up sensory information

The *predictive coding* theory – an increasingly popular theory for perceptual processing – states that **the brain is constantly generating and updating a mental model of sensory input**. The brain makes sense of the experience by adjusting a balance between expectations and sensory information: a mismatch between expectations and reality will induce a more sensory-driven reinterpretation of this experience.

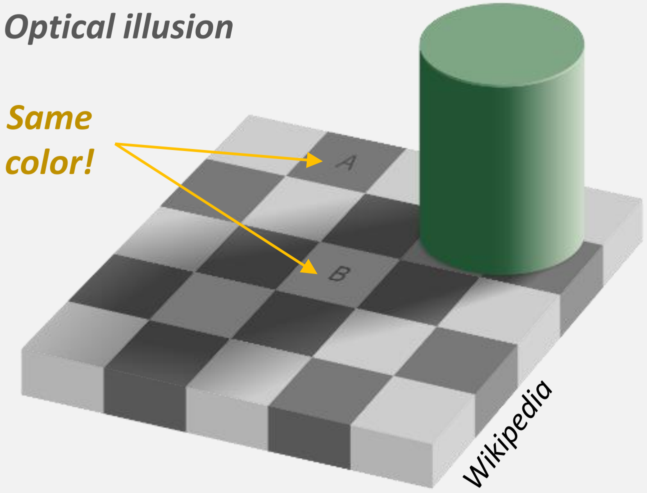
The fact that our brain makes its own subjective model of the environment is easily verifiable in **perceptual illusions** that trick our brain:

- Optical illusions: visual relations (specific shadow, perspective, distance and size of objects), absence of stimuli in the visual area (ex: blind spot)
- Auditory illusions (ex: tinnitus / ringing ears after a loud concert)
- Somatic illusions (ex: feeling ownership of a rubber hand)

Mental representations can also greatly differ from the veridical representations of the objective world when **self-generated stimuli are suppressed** from our perceptual experience. For instance, we do not hear our own footstep when walking.

Examples

Optical illusion



In the checker shadow illusion, tile A looks significantly darker than tile B whereas both tiles are exactly the same shade of grey.

The brain makes inferences from the location of the shadow and the colors of nearby tiles. These inferences lead to different perceptions of the same color.

Suppression of self-generated stimuli

We don't hear our own footsteps

Except if you voluntarily pay attention to the sound of your footsteps, you do not hear it.

The brain learns to turn off responses to predictable self-generated sounds. It cancels the footstep sound from the other external sounds by applying an internal model of sound produced by its own movements.

When we intentionally move our sensors, perception is stabilized by corollary discharges of motor commands

The stimuli received by our sensors are continuously changing because the environment and/or the sensor position is changing. It feels intuitive that we can see a moving object when we stay still: if the stimuli are changing, it is because the environment is changing.

In fact, **our sensors are also continuously moving** during sensory experiences, generally without us knowing it. Despite this fact, **our perceptions are surprisingly stable**. This phenomenon exists for every sense, but it is more obvious for vision with fast eye motions called *saccades* that direct the fovea which has much better acuity than the rest of the retina (around 5 saccades per second) .

In order to **predict** the next sensory stimuli, **the expectations of the brain have to take into account the upcoming self-generated movements**, in addition to the flow of sensory inputs. This information is provided via a copy of the motor command signals called *corollary discharges*, going directly from motor to sensory brain areas.

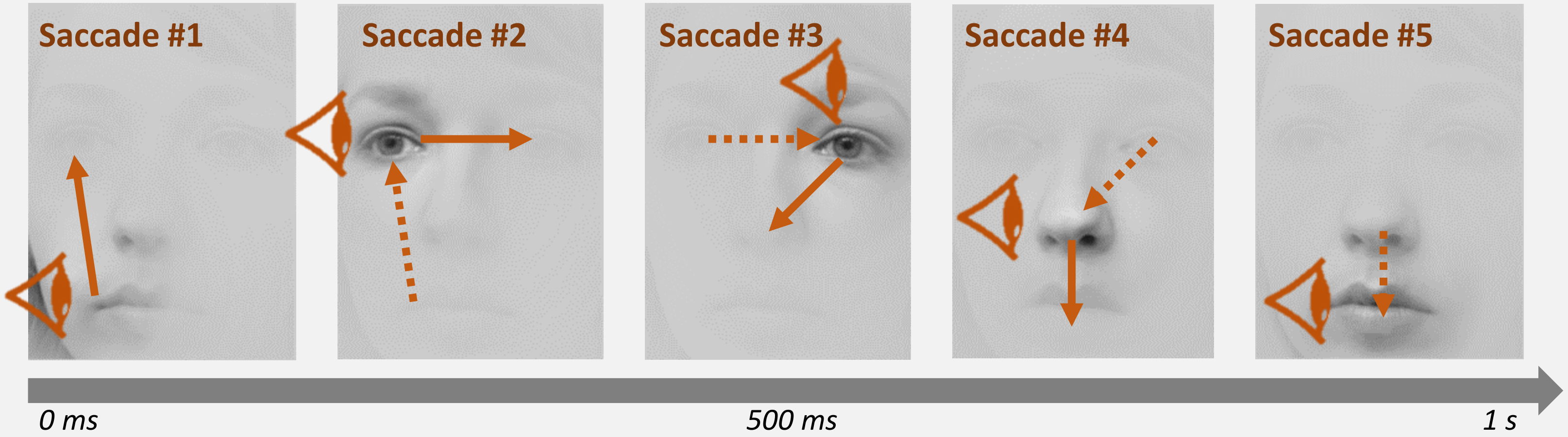
After enough time and experience to calibrate visual neural circuits, sensory stimuli begin to make sense because we have learnt to:

- Filter out meaningless sensory stimuli during the saccade motion
- Predict the expected stimuli after the saccade

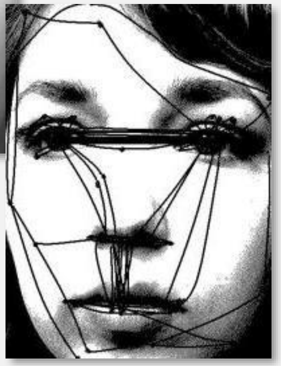
Saccadic eye movements when looking at a human face

→ *the brain as an active predictive machine*

If some visual stimuli represent a mouth, we may expect to be looking at a face. To verify this prediction, we generate the next saccade in a place where we expect to see a eye, then the other eye, then the nose, etc.



Perception needs action: vision ceases after a few seconds when saccades are impeded



Example of trajectories of eye saccades focusing successively on the eyes, the nose and the mouth

Wikipedia

Thiboust, 2020

Active sensorimotor interactions are essential to generate meaningful percepts grounded to the real world

Moving our sensors is not only a way to scan the environment, it is also a way to actively verify the correctness of our models and to correct them if needed. We learn from the consequences of our brain's actions about aspects of the environment that matter for particular goals.

When incoming bottom-up stimuli fit top-down expectations, it implies that a connection has been established between some brain's circuits and something meaningful from the real world. This active process is referred to as *grounding*. It attaches a meaning to a stimuli-induced neural activity that becomes a meaningful percept.

If perception is the act of continuously updating our imperfect models from our actions, then it is inseparable from the grounding process.

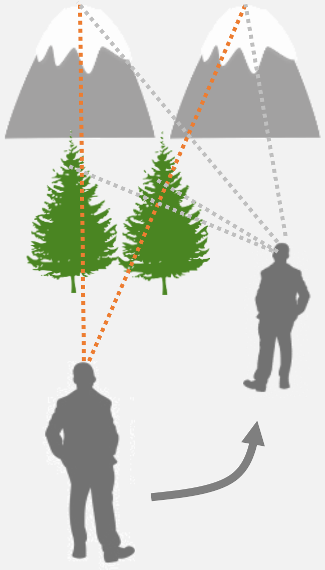
Grounding is realized via sensorimotor interactions through time (also referred to as *active sensing*). The objective is to refine the meaning of sensory signals by **successive comparisons of predictions vs outcomes of self-generated movements**.

With experience, general mental representations are complemented by more specific mental representations with hierarchical connections between those representations (ex: an oak is a tree and a tree is an inanimate object, a tree is composed of branches and leaves).

To be useful, all those representations have to be meaningful and to show at least some degree of invariance:

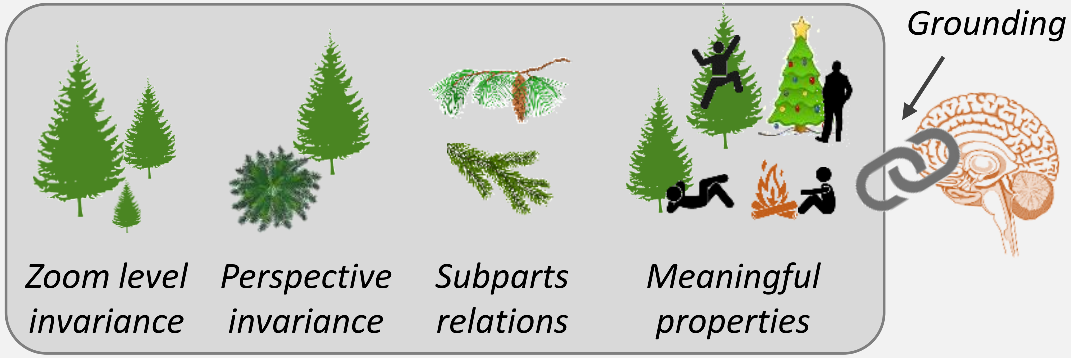
- Meaningful because the representation of an object is attached to a collection of properties that could be helpful to achieve behavioral goals.
- Invariant because the same mental representation of a physical object should be activated when the object is viewed from different brightness, angles or zoom levels.

Moving to verify the correctness of our models



> *"We connect to the world not through our sensors (although they are essential) but through our actions. This is the only way that sensation/perception can become "grounded" to the real world as experience.*
>
> *The distance between two trees and two mountain peaks may appear identical on the retina. It is only through walking and moving one's eyes that such distinctions can be learned by the brain."*
>
> *Buzsáki, 2019*

Grounding useful percepts to the real world



Cognition uses grounded mental representations while disengaging from the external world

Perception is a prerequisite for cognition because the latter uses meaningful mental representations that should already be grounded by active sensing.

Contrary to perception, cognition is characterized by a disengagement from the external world. **Cognition relies on internally organized activity detached from immediate sensory inputs and motor outputs**.

This ability allows us to imagine the future and recall the past. More fundamentally, the main evolutionary advantage of cognition is the ability to **test mentally "what if" scenarios to anticipate at long time scales the potential consequence of alternative actions without actually taking them**.

From the perspective of a brain network receiving sensory inputs, there is no difference between real sensory inputs and similar activity generated by other internal networks. The brain would only need a gating mechanism to direct the neural flow accordingly. Similarly, the motor command sent by the brain network can be retained, leaving only the internal corollary discharge.

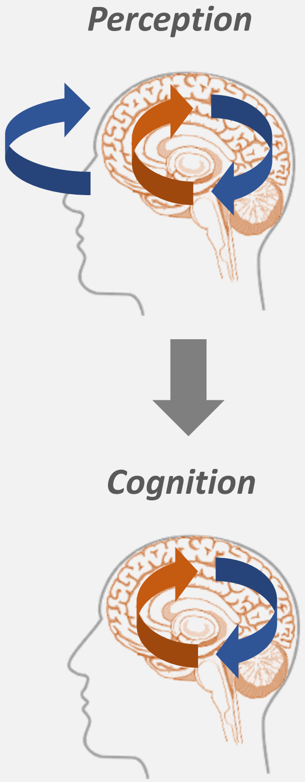
Remark on this explanation of perception and cognition:

This presentation is not supported by the classical outside-in framework that states that the brain is a passive device whose job is to sequentially perceive, cogitate and then act. However, the outside-in approach is increasingly questioned by neuroscientists favoring an inside-out approach. In the inside-out approach, self-organized brain activity is grounded to meaningful features from the environment via actions (perception). Then, this brain activity can be internally sustained (cognition). More in Buzsáki, 2019

No cognition without perception

Contrary to what some researchers are looking for, the question is not "how to ground abstract symbols to concrete experience?". Those symbols are first grounded by perceptual experience before they could be detached for cognition.

In fact, the real question is "how do symbols get detached?".



Perception grounds meaningful mental representations to the external world. This process is done by processing stimuli that have been deliberately produced by an action on the sensor, coupled with internal corollary discharges bypassing the environment.

When brain circuits are calibrated by action-based perceptions, then the brain can disengage from the external world, relying only on internal circuits that support meaningful representations. We call this ability "cognition".

Speculation: cognition is like perception with an internal reference model instead of the environment

Speculation!

I think that perception and cognition are two abilities that use the same fundamental mechanism: updating internal models according to the difference between two parallel signals: a prediction and a reference induced by a self-generated command.

In **perception**, those two parallel signals are:

- A motor command that will induce the outcome (the reference)
- A copy of this motor command (corollary discharge) that is processed by an internal model in order to generate a prediction

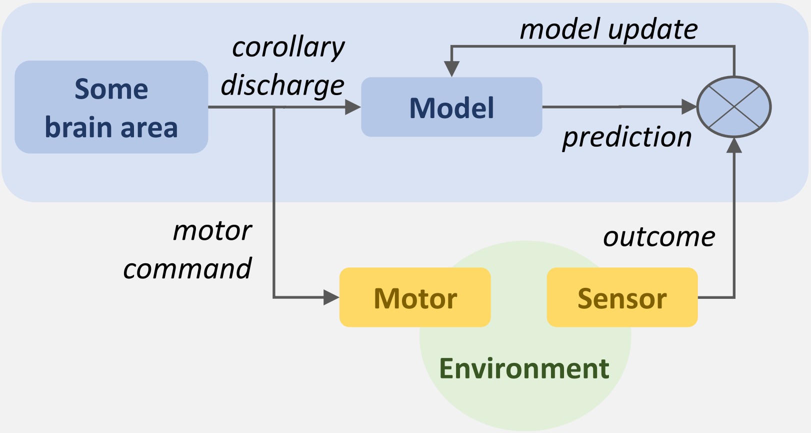
If we draw a parallel to perception, we could say that **cognition** is the act of updating our imperfect models from our "disengaged actions" with two parallel signals:

- An internal command that will produce the reference via a "reference model" circuit (could be a complex model or simply a memory circuit with direct correspondences between internal commands and reference values)
- A corollary discharge of the internal command that is processed by an internal model in order to generate a prediction

In cognition, a prerequisite is that the "reference model" circuit has already been grounded by perception or by another cognition loop. Whereas the reference signal is given by the environment in perception (model-free), the reference signal is modeled internally in cognition (model-based).

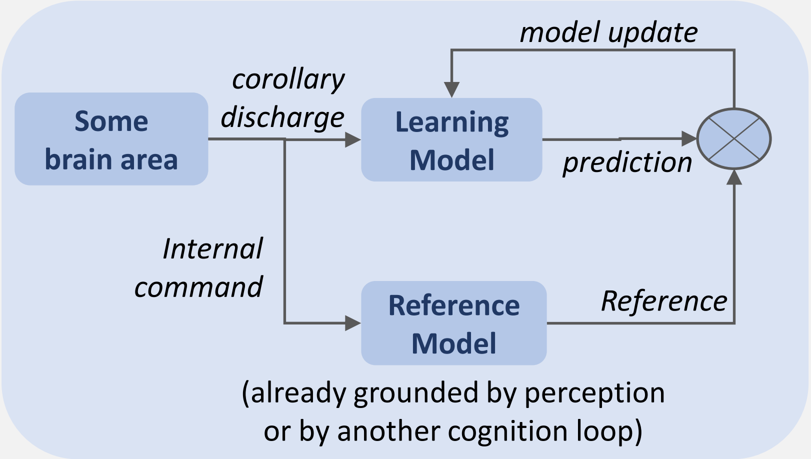
<u>Going further</u>: in cognition, the brain can gate neural activity to consider a high-order area as the reference and a low-order area as the model to be updated (deduction), or the other way around (induction), depending on the context (see later chapter on neocortex for more explanations about hierarchy)

Perception



Model-free

Cognition



Model-based

Thiboust, 2020

Brain general machinery

1. Neurons are sophisticated elementary components of the neural “hardware”

Matthieu Thiboust

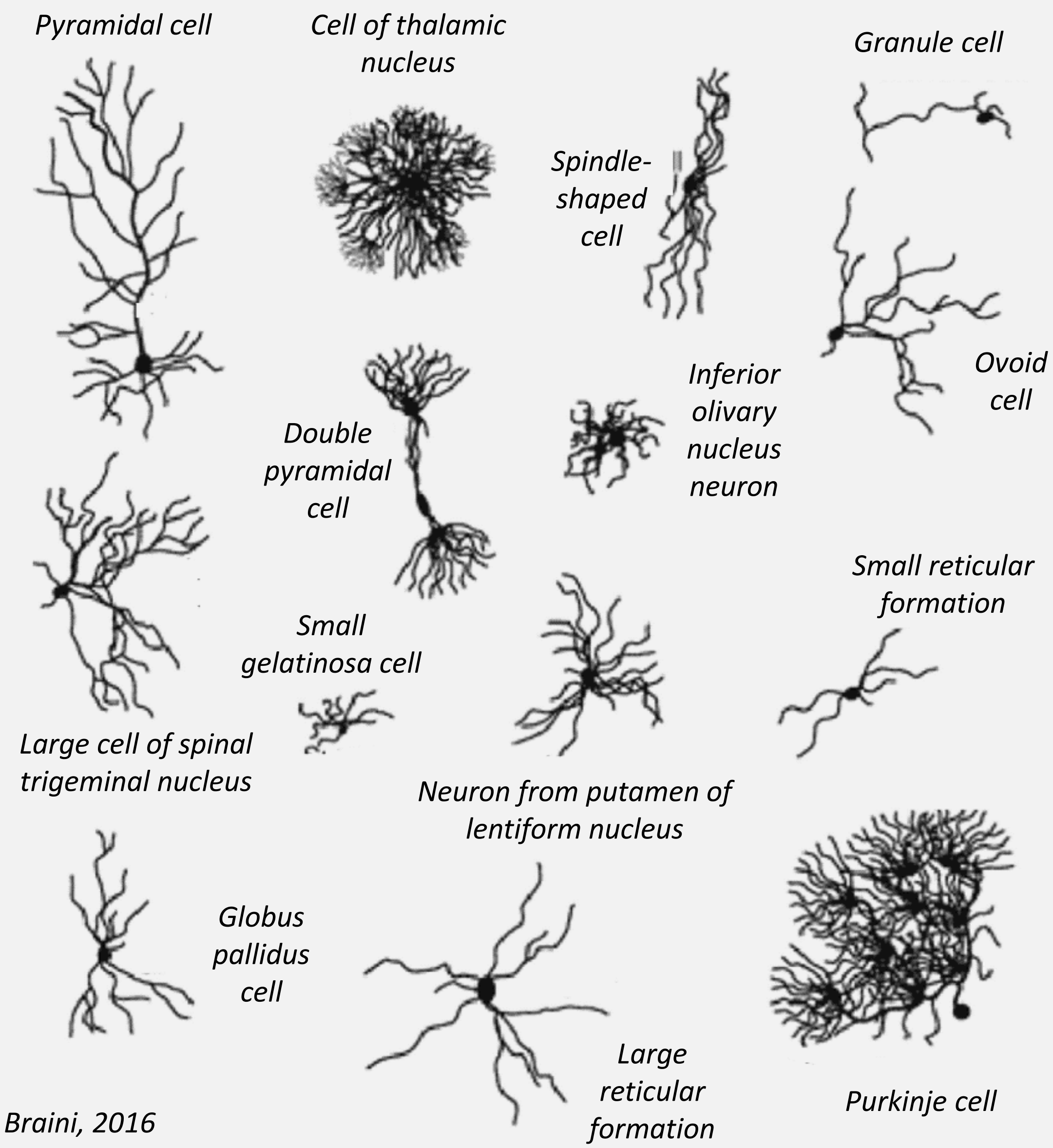
Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- John E. Dowling
- Santiago Ramon y Cajal

See the reference section for a list of materials that inspired me.

Art credit: Self Reflected, Greg Dunn Design

Brains host a highly diverse neuronal ecosystem with hundreds of different neuron types


Pyramidal cell
Cell of thalamic nucleus
Granule cell
Spindle-shaped cell
Ovoid cell
Inferior olivary nucleus neuron
Double pyramidal cell
Small reticular formation
Small gelatinosa cell
Large cell of spinal trigeminal nucleus
Neuron from putamen of lentiform nucleus
Globus pallidus cell
Large reticular formation
Purkinje cell

Braini, 2016
(adapted from Ramon y Cajal)

Even if each single neuron is basically a cell transmitting nerve impulses, there is a great diversity in neuron types.

Our 86 billion neurons can be classified into **hundreds of families and subfamilies** depending on:

- Their **morphology**: shape & size
- Their **position**: sensory neurons, motor neurons, interneurons
- Their **connectivity**: number of input & output connections, and the neuron families they are connected to
- The **length of their connections**: local vs long-distance
- The **specificity of their messages** : focal vs diffuse, ephemeral vs long-lasting
- Their **impact on other neurons**: excitatory vs inhibitory
- Their **passive and active electrical properties**
- Their **excitability**: sensitivity level
- Their **transmission speed**
- Their **discharge patterns**: single vs multiple spikes
- Their **expression of specific proteins**

Neural communication relies on a network of neurons transmitting signals via electrical and chemical signals

Neurons are electrically excitable cells that communicate with other cells via specialized connections called **synapses**.

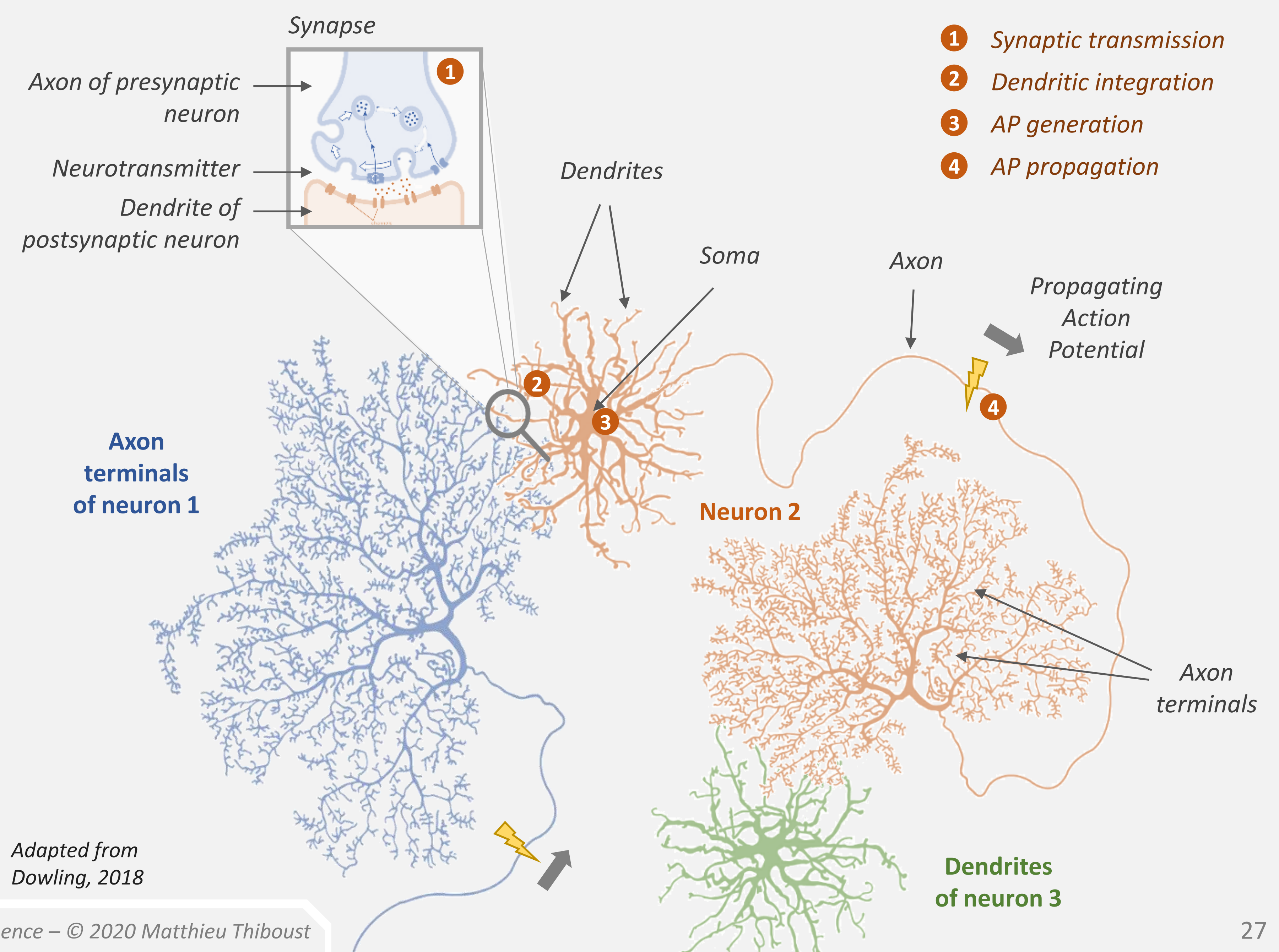
They are typically composed of two parts:

- Several **dendrites** with thousands of branches/segments that can get excited by other neurons
- One **axon** extending into thousands of axon terminals that can excite other neurons

Signal processing and transmission uses:

- **Chemical neurotransmitters in synapses** between two neurons
- **Electrical action potential (AP)** inside the neuron from the dendrites to the axon via the cell body

Those characteristics are the common denominator of neural communication. However, there are **considerable variations around this general theme**, with different neuron morphologies and organizations, different kinds of spikes and different neurotransmitters.



Adapted from Dowling, 2018

Dozens of neurotransmitters and neuromodulators transmit chemical signals in synapses between neurons

Neurons use **several dozen different molecules to convey chemical messages at the synapses level**, with various effects on the receiving neuron (multiple types of receptors for each molecule).

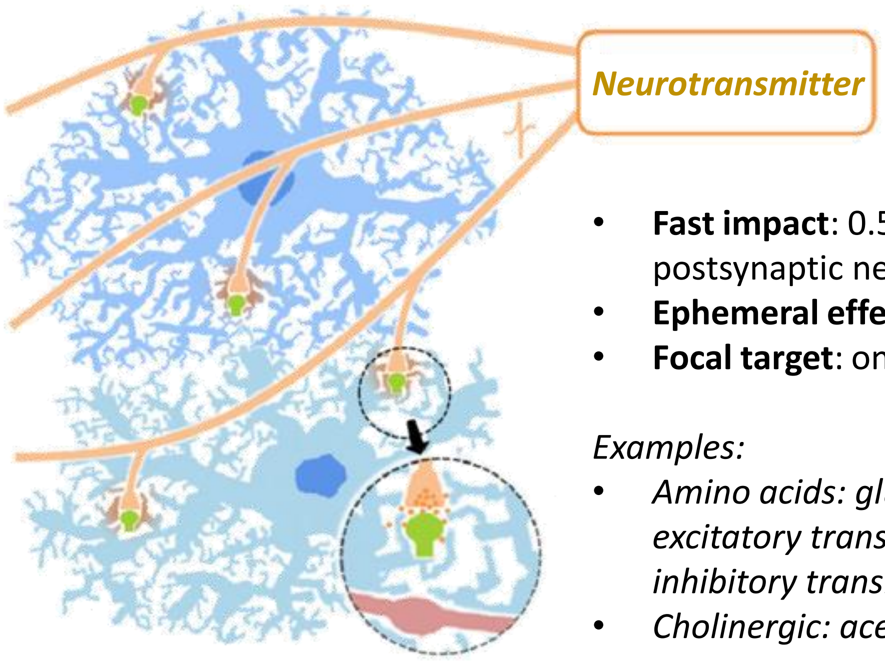
They can act as neurotransmitters, neuromodulators or both:

- **Neurotransmitters** convey fast and ephemeral point-to-point signals in synapse channels.
- **Neuromodulators** convey slow and long-lasting point-to-many signals. They induce biochemical changes in the postsynaptic neuron.

Each neuron generally releases only one kind of neurotransmitter or neuromodulator, but it can be excited by a combination of several neurotransmitters and neuromodulators on its thousands of synapses.

At the synapse level, only one or two substances are released. The combination is done by multiple neighboring synapses on the same dendritic segment.

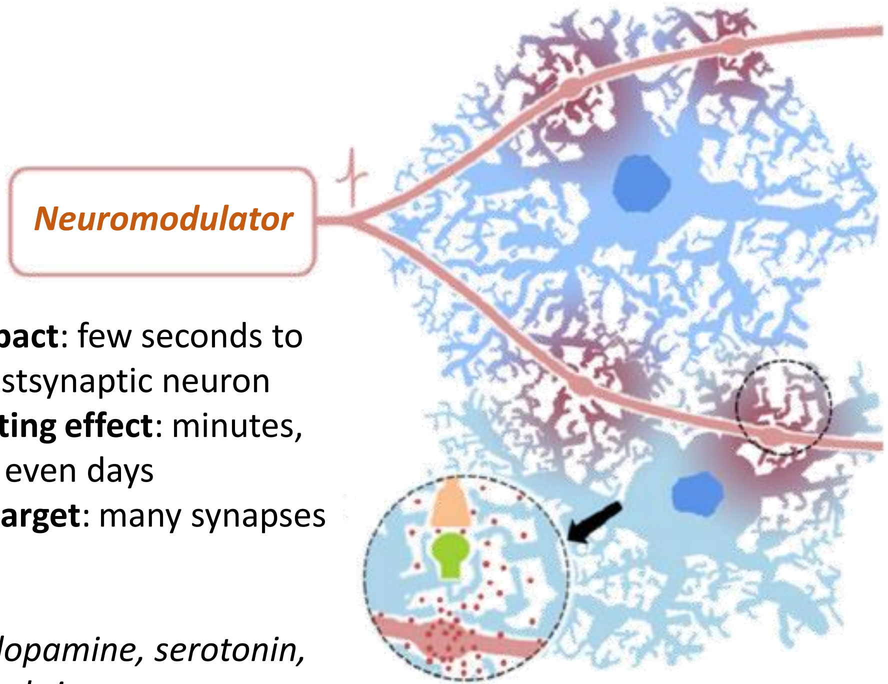
Some substances have **inhibitory** effect (like *GABA*) while others are **excitatory** (like *glutamate*). *Acetylcholine's* inhibitory or excitatory effect depends on whether it is used as a neurotransmitter or a neuromodulator.



- **Fast impact**: 0.5 ms to reach postsynaptic neuron (*)
- **Ephemeral effect**: less than 100 ms (*)
- **Focal target**: only 1 synapse

Examples:

- *Amino acids: glutamate (main excitatory transmitter), GABA (main inhibitory transmitter), glycine*
- *Cholinergic: acetylcholine*



- **Slow impact**: few seconds to reach postsynaptic neuron
- **Long-lasting effect**: minutes, hours or even days
- **Diffuse target**: many synapses

Examples:

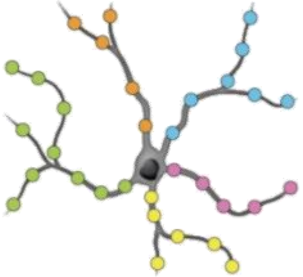
- *Amine: dopamine, serotonin, norepinephrine*
- *Peptide: substance P, endorphins*

Adapted from Hirase, 2014

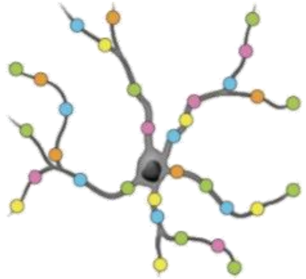
** Metabotropic receptors (≠ ionotropic receptors) can have a longer latency and effect duration*

Dendrites are active components of a neuron that can perform complex computations

Clustered inputs



Dispersed inputs

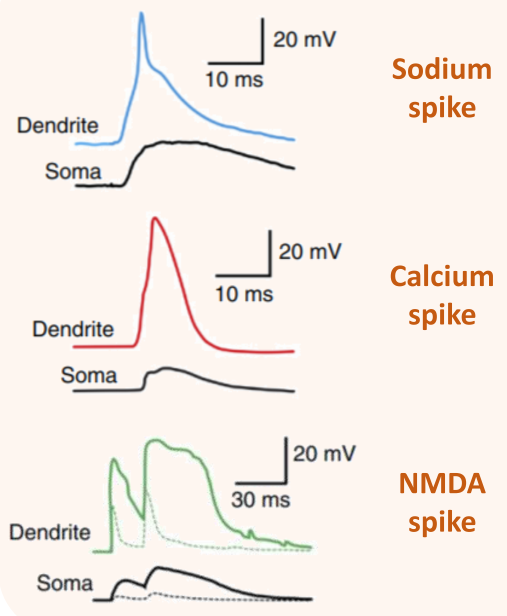


A neuron can have tens of thousands of synapses (sometimes even hundreds of thousands) grouped on hundreds of **dendritic segments**. Depending on the time-distance from the soma (proximal vs distal dendrites) and the distribution of synaptic inputs on the different segments (clustered vs dispersed), **dendrites of a single neuron can perform complex computations** by combining basic operations (like AND, OR, and even XOR) performed on each dendritic segment.

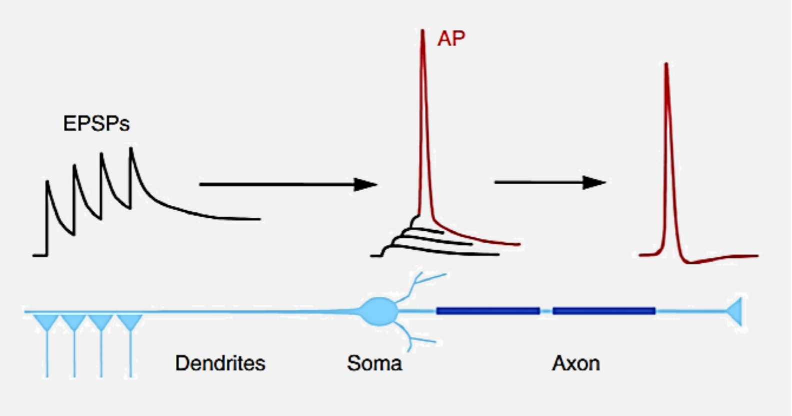
Dendrites do not only passively integrate **excitatory postsynaptic potential (EPSP)** and **inhibitory postsynaptic potential (IPSP)** that can trigger an Action Potential in the axon initial segment if above a given threshold. They are also able to actively trigger various localized **dendritic spikes** (different from Action Potential (AP) spikes) propagating from distal dendrites to the soma. Dendritic spikes increase the probability of AP firing in the axon, but they do not assure it.

Dendrites can also backpropagate AP (generated in the axon initial segment near the soma) into the dendritic arbor. This is referred to as a **backpropagating AP (bAP)**. Interactions between dendritic spikes and bAP are believed to be involved in synapse learning mechanisms.

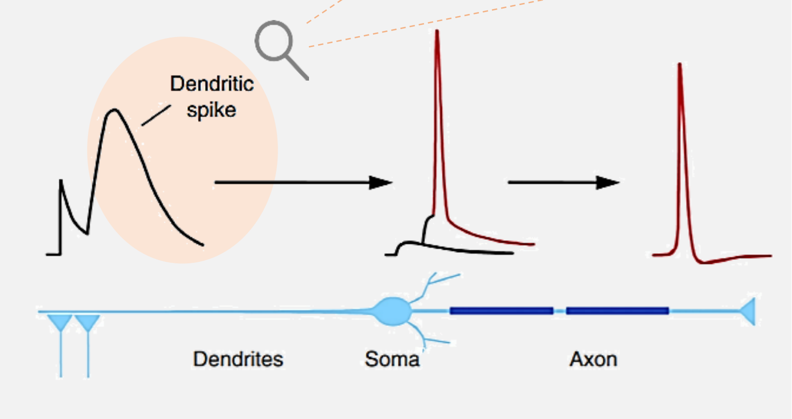
Dendritic spikes



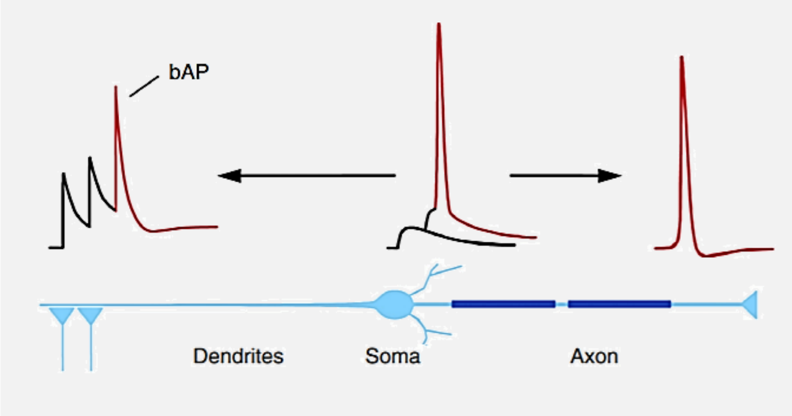
Passive integration



Active integration



Backpropagation



Figures from Stuart & Spruston, 2015

Propagating Action Potentials are triggered when membrane potential increases over a threshold near the soma

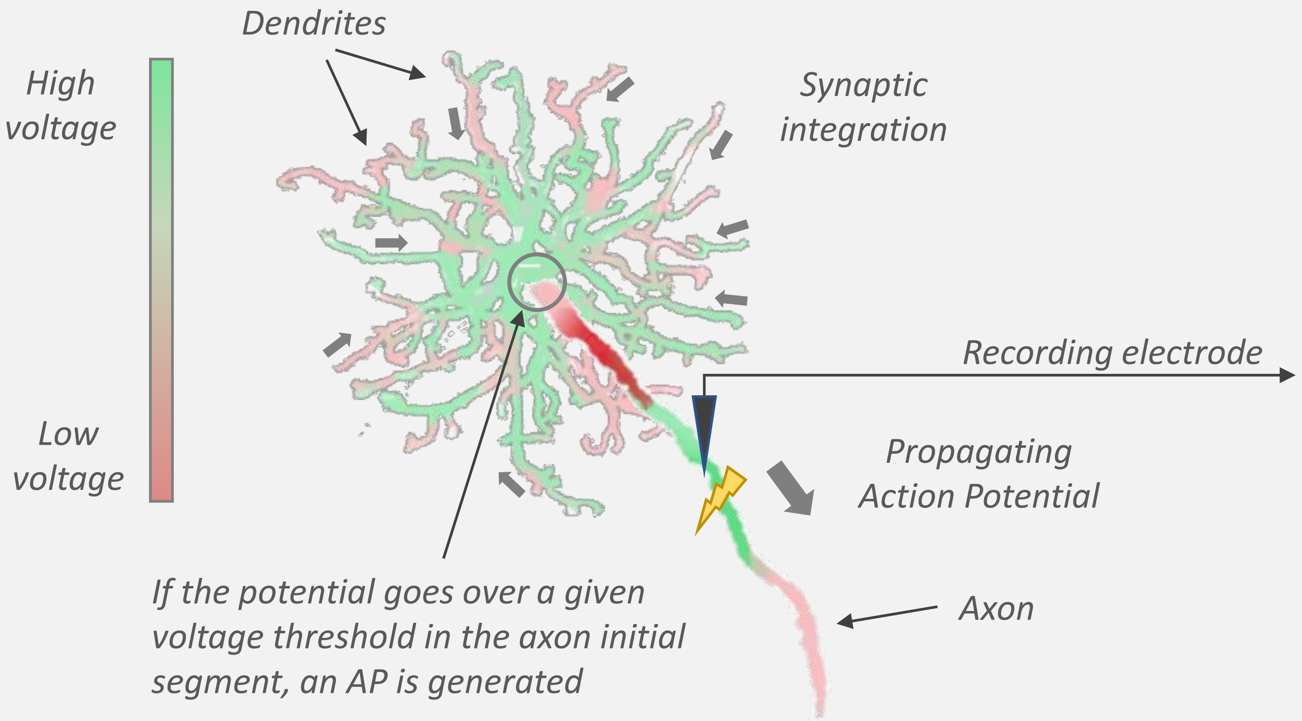
Action Potentials (AP), also called **spikes**, are propagating depolarizations of neuron membrane potential (= voltage) along its *axon* from the *axon initial segment* (near the soma) towards *axon terminals*.

AP propagates very quickly along the axon: from a few to a hundred meters per second, making it possible to convey long-distance electrical messages throughout the brain and the body.

AP are triggered when synaptic inputs increase the **membrane potential** of the *axon initial segment* over a given **voltage threshold**. To maximize the chance to generate a spike, those inputs have to co-occur during a short integration time-window.

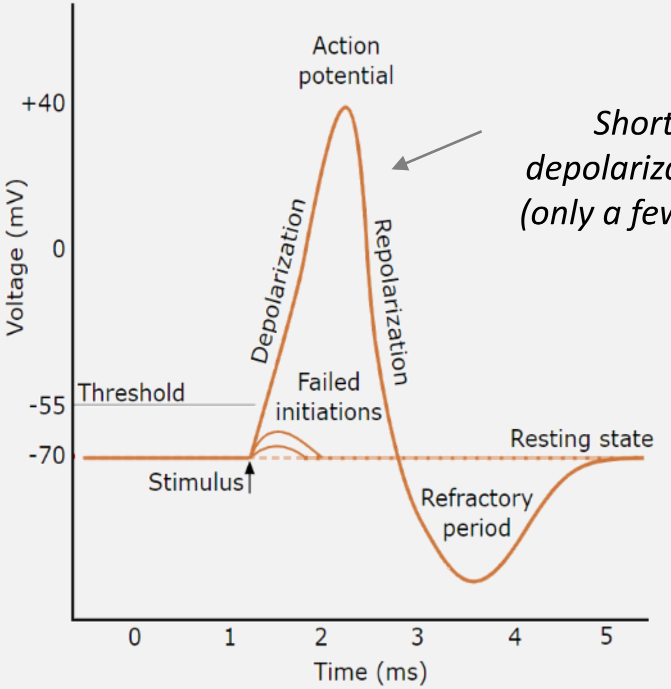
Remark: A cell with a depolarized but subthreshold potential will be quicker to fire if new dendrites became positive. This characteristic is essential to explain competition between neurons at a network level: the first neuron to fire inhibits its neighboring excitatory neurons via fast inhibitory interneurons (see later focus on neocortex).

Voltage spatial map



If the potential goes over a given voltage threshold in the axon initial segment, an AP is generated

AP voltage time profile



Short depolarization (only a few ms)

Wikipedia

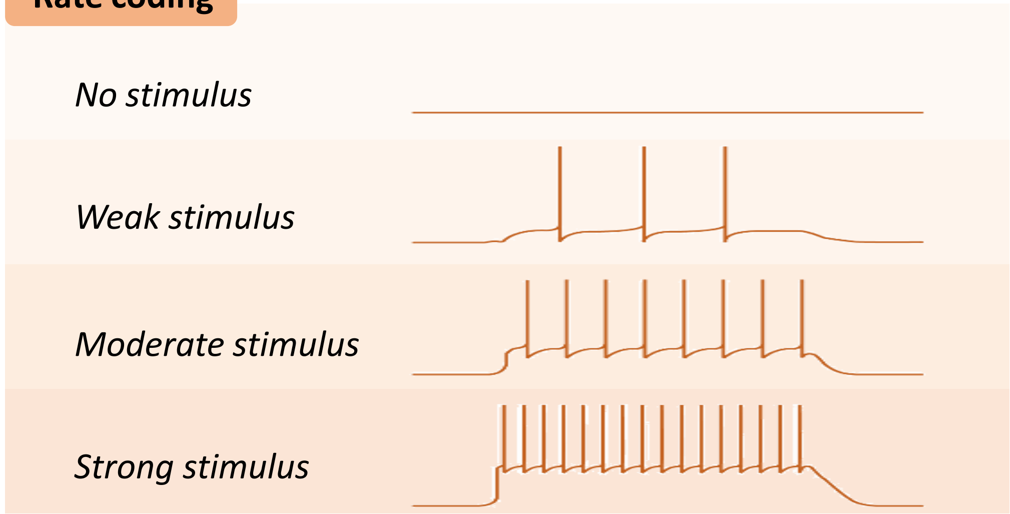
Neurons exhibit a variety of firing patterns composed of successive spikes

Successive action potentials fired by a neuron are called **spike trains**.

Depending on the physiology of the neuron, there exist different firing patterns. For instance, bursting neurons tend to fire repetitively and very quickly during a period, followed by a long quiescent period.

The information is somehow coded into those firing patterns. There are many ways neurons might code information. The most straightforward code is a **rate code** where spike frequency is correlated with the strength of integrated inputs. **Phase coding** is another coding strategy involving brain oscillations *(see later chapter on neocortex).*

Rate coding

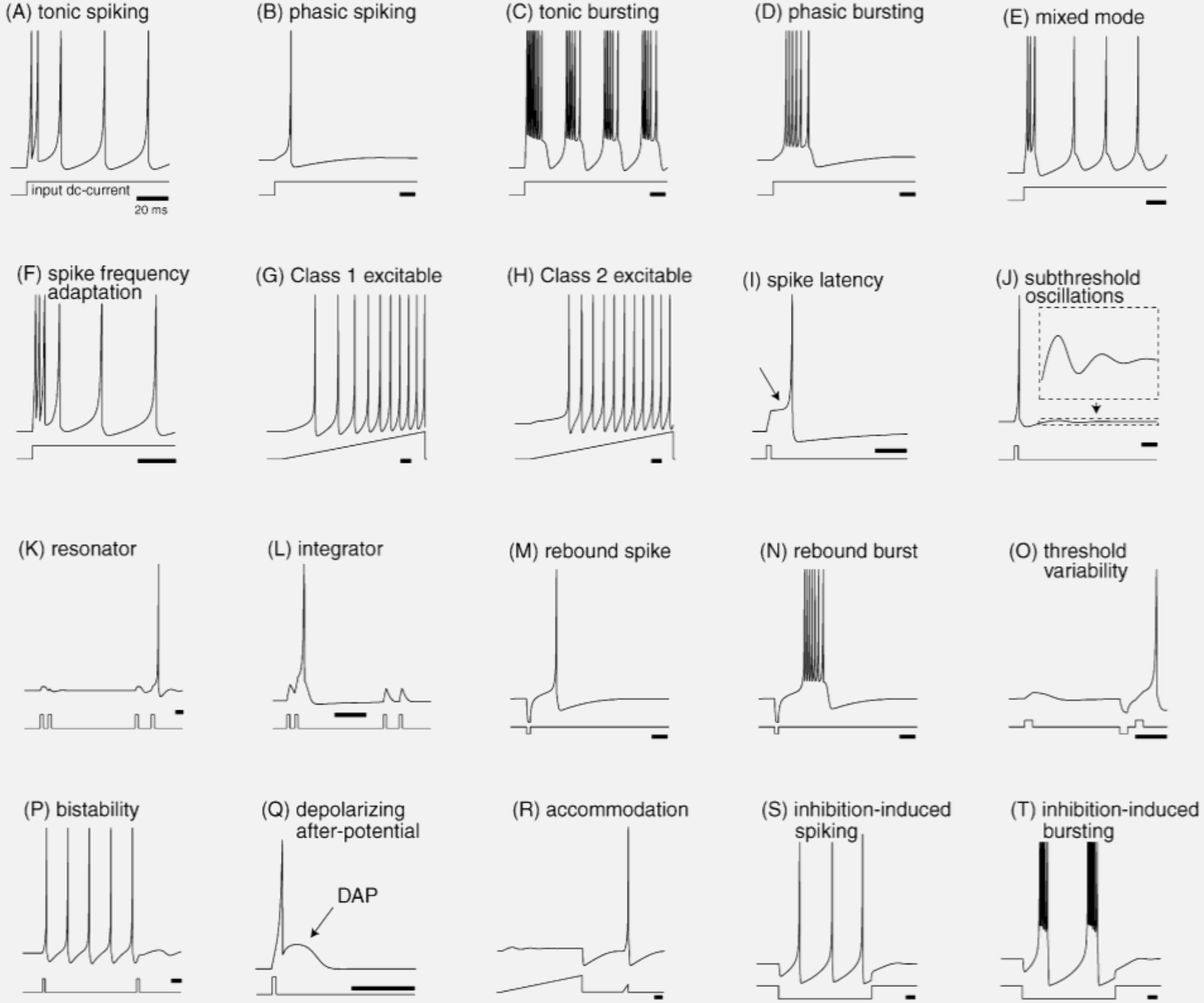


Dowling, 2018

Remark: those illustrations show firing patterns in response to artificial current injections. Real sensory responses of awake animals are more complicated and composite.

Firing patterns

Twenty of the different types of firing patterns exhibited by single neurons in the mammalian cortex:



Each horizontal bar denotes a 20-ms time interval

Izhikevich, 2004

Brain general machinery

2. Neuron plasticity allows to retain memories of previous neural activity

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- John E. Dowling
- Blake Richards

See the reference section for a list of materials that inspired me.

Art credit: Self Reflected, Greg Dunn Design

A combination of local plasticity mechanisms support short and long term memories

Virtually everything we do or experience can cause changes in our plastic brain. Our new memories are encoded by those changes that can persist in time, from tens of milliseconds to a hundred years.

Those changes occur at different levels in our neural networks:

- **Synapse level**: adding or pruning a synapse, increasing or decreasing a synaptic weight
- **Neuron level**: modifying a neuron intrinsic excitability or other physiological characteristics
- **Network level**: self-sustaining a looped activity via network recurrent interactions

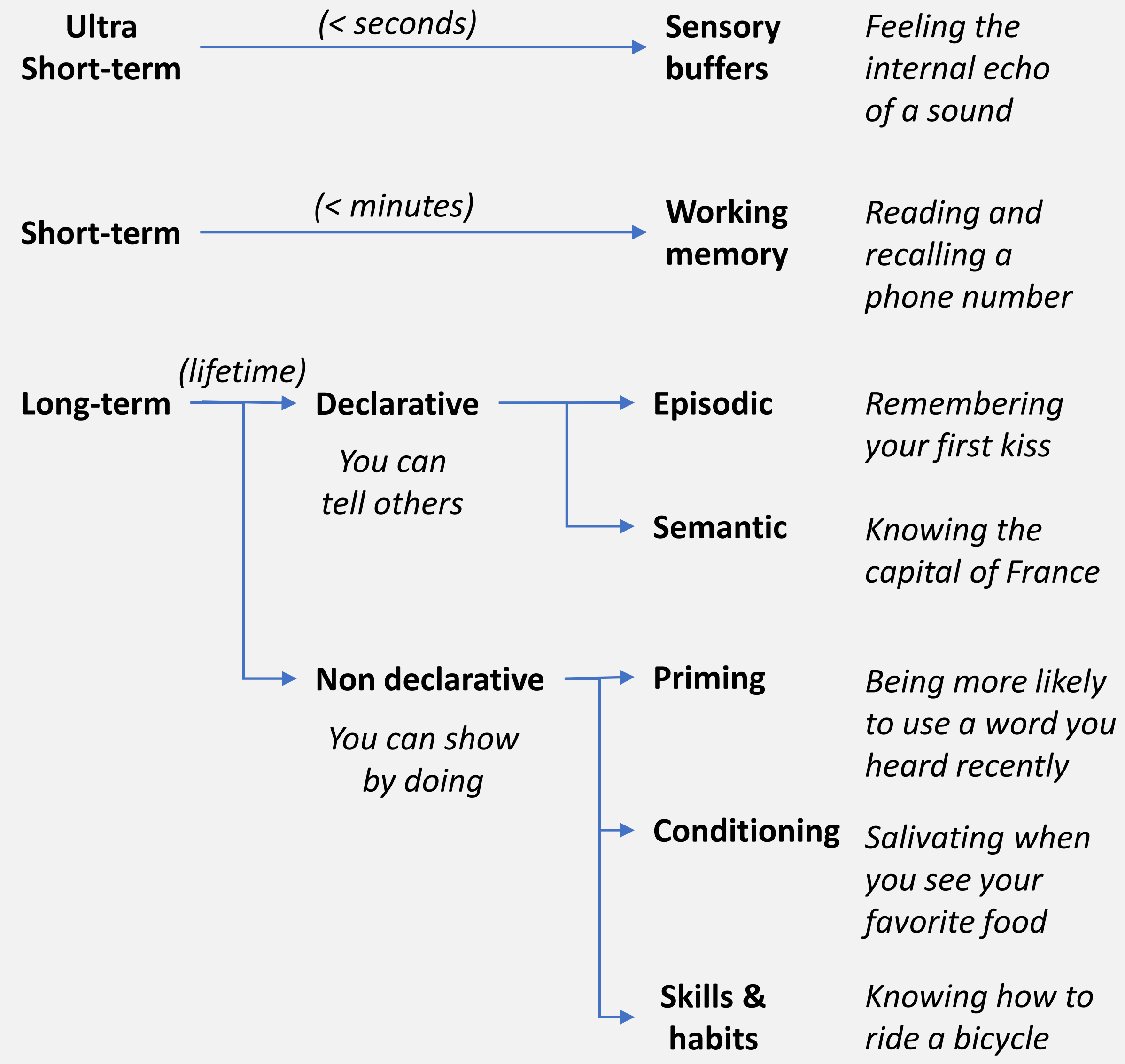
Even if there is a growing interest for non-synaptic plasticity in the research community, it is believed that memories are mostly encoded in synapses, because there are thousands per neuron.

Synaptic plasticity rules give an abstraction of **synaptic plasticity mechanisms** (like *Long Term Potentiation (LTP) and Depression (LTD), synaptic facilitation, …*). They describe how synaptic weights get changed in function of the frequency, the intensity and the timing of activity of presynaptic and postsynaptic neurons. Synaptic changes can also depend on a third factor modulating the plasticity.

Hebbian learning (often simplified by "*fire together, wire together*") and **Spike-Timing Dependent Plasticity (STDP)** are the most famous rules.

All brain plasticity rules are local: changes only depend on information directly available to the synapse, neuron or network.

Types of memory



Examples:
Ultra Short-term
(< seconds)
Sensory buffers
Feeling the internal echo of a sound
Short-term
(< minutes)
Working memory
Reading and recalling a phone number
Long-term
(lifetime)
Declarative
You can tell others
Episodic
Remembering your first kiss
Semantic
Knowing the capital of France
Non declarative
You can show by doing
Priming
Being more likely to use a word you heard recently
Conditioning
Salivating when you see your favorite food
Skills & habits
Knowing how to ride a bicycle

Synaptic plasticity involves cellular and molecular pre and postsynaptic mechanisms

The contribution of synapses to the evoked post-synaptic potential depends on the number, the strength (also referred to as weight) and the dendritic position of synapses.

When some specific patterns of synaptic activity occur, **plasticity mechanisms adapt synaptic characteristics by weakening/strengthening synaptic weights and creating/pruning synapses** (each neuron is only connected by synapses to a fraction of other neurons).

The different cellular and molecular mechanisms of synaptic plasticity are only partially understood. They mainly involve the pre and post-synaptic neurons. However, some complex mechanisms also rely on local concentration of neuromodulators (released by other neurons) and/or neighboring astrocytes (a type of glial cells that populate the nervous systems along with neurons) that act as catalysts or inhibitors.

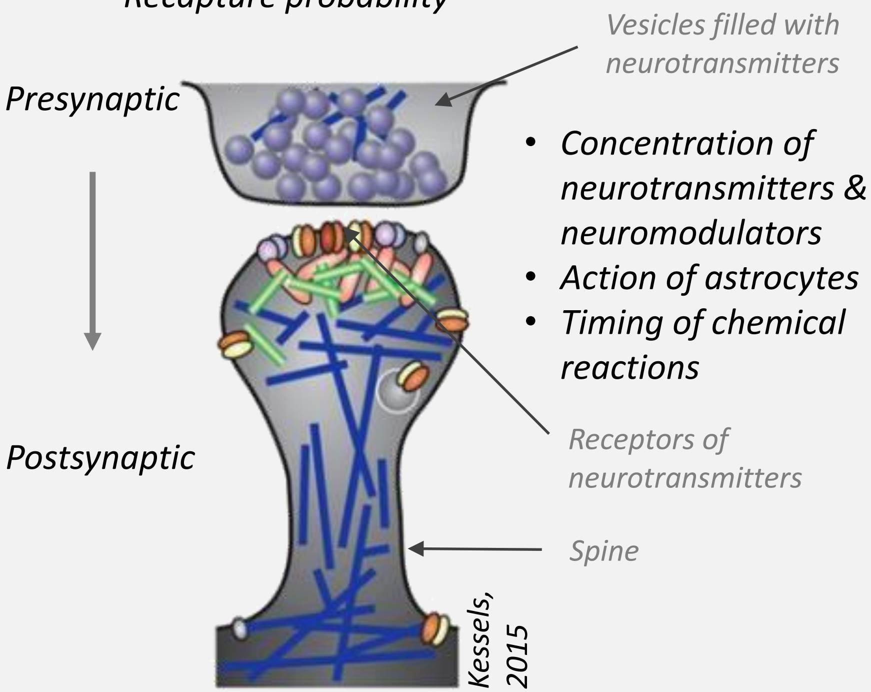
The most commonly studied mechanisms are **Long Term Potentiation (LTP)** and **Long Term Depression (LTD)**. They produce long-lasting increases and decreases in synaptic efficacy of excitatory synapses using the glutamate neurotransmitter (most excitatory synapses use glutamate). The mechanism involves the density regulation of two types of glutamate receptors (NMDA and AMPA). LTP is induced each time the postsynaptic depolarization and the postsynaptic concentration of calcium is above a minimum level. A very high level of calcium generated by a back propagating AP can also be sufficient by itself for LTP.

Short Term Plasticity is believed to be mostly controlled by presynaptic mechanisms. Short term facilitation increases the probability of neurotransmitter release, whereas depression reflects a depletion of releasable neurotransmitters. Because their effect only last for a second or so, they dynamically alter the frequency response of synapses.

Spike Timing Dependent Plasticity (STDP) involves both pre and postsynaptic mechanisms. The precise temporal order of activity between the two neurons matters. If the presynaptic spike precedes the postsynaptic spike, the synaptic strength is increased (most cases) or decreased, depending on the mechanism.

Some characteristics impacting the synaptic strength

- *Nb of vesicles*
- *Stock levels of neurotransmitters*
- *Synthesis of neurotransmitters*
- *Release probability*
- *Recapture probability*



Vesicles filled with neurotransmitters
Presynaptic
Postsynaptic
Receptors of neurotransmitters
Spine
Kessels, 2015

- *Concentration of neurotransmitters & neuromodulators*
- *Action of astrocytes*
- *Timing of chemical reactions*

- *Nb & type of neurotransmitter receptors*
- *Calcium concentration*
- *Depolarization*
- *Size of the spine supported by scaffold proteins*

Models of synaptic plasticity implement local learning rules

The **historical Hebbian plasticity rule** has been significantly enriched since its first mention by Hebb in 1949. This simple model postulates that when one neuron drives the activity of another neuron, the connection between these neurons is potentiated (often summarized as "*cells that fire together wire together*").

More advanced phenomenological models – based on an input-output relationship between neuronal activity and synaptic plasticity – offer a conceptual framework to **understand network-level effects induced by changes in synaptic strength**.

Rate based models determine the sign and magnitude of synaptic plasticity from the **average firing rate** (over some time period) of pre and postsynaptic neurons.

Spike timing based models are inspired by STDP mechanisms. Their outputs depend on the **relative timing difference** between pre and postsynaptic spikes.

Some rate based and spike timing based models are in fact more complex. Those elaborate versions allow to:

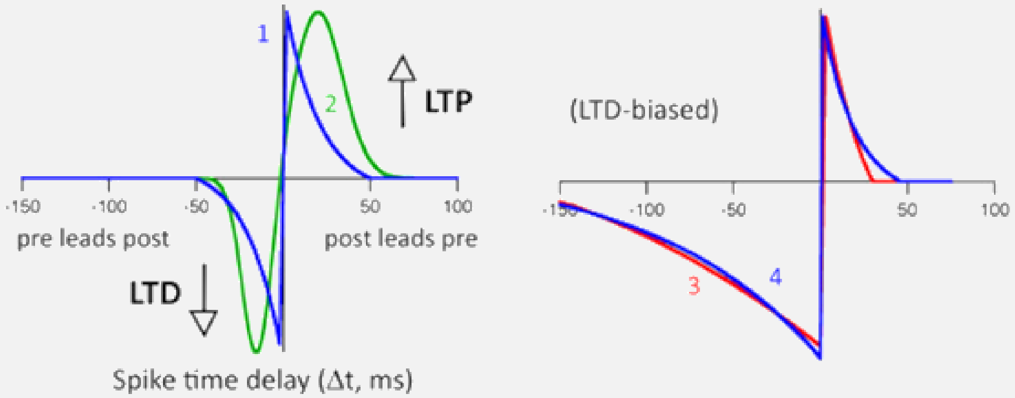
- Mix the two model families
- Adjust the response with the initial synaptic state
- Separate short term and long term averages
- Take into account depolarization events in addition to spike events
- Modulate the synaptic change according to a neuromodulator concentration (3-factor learning rule).

In every cases, models of synaptic plasticity only use inputs directly available in the local periphery of the synapse, and then implement **local learning rules**.

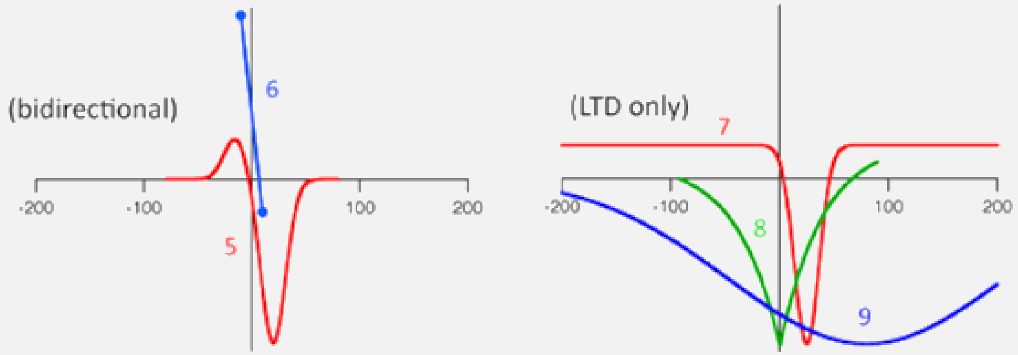
Example of spike timing based models

Spike timing based models are often characterized by a graph of synaptic changes (positive means strength increase) in function of the relative timing between neuron (positive means that the presynaptic spike precedes the postsynaptic spike).

Hebbian STDP (most common)



Anti-Hebbian STDP (less common)



Feldman, 2012

Nonsynaptic plasticity mechanisms complement their synaptic counterparts

Nonsynaptic plasticity involves cellular and molecular mechanisms occurring in the soma, the dendrites and the axon of neurons (instead of synapses for synaptic plasticity) that **modify the intrinsic excitability of the neuron**.

Those mechanisms mostly depend on neuromodulatory regulation and on the internal activity of the neuron.

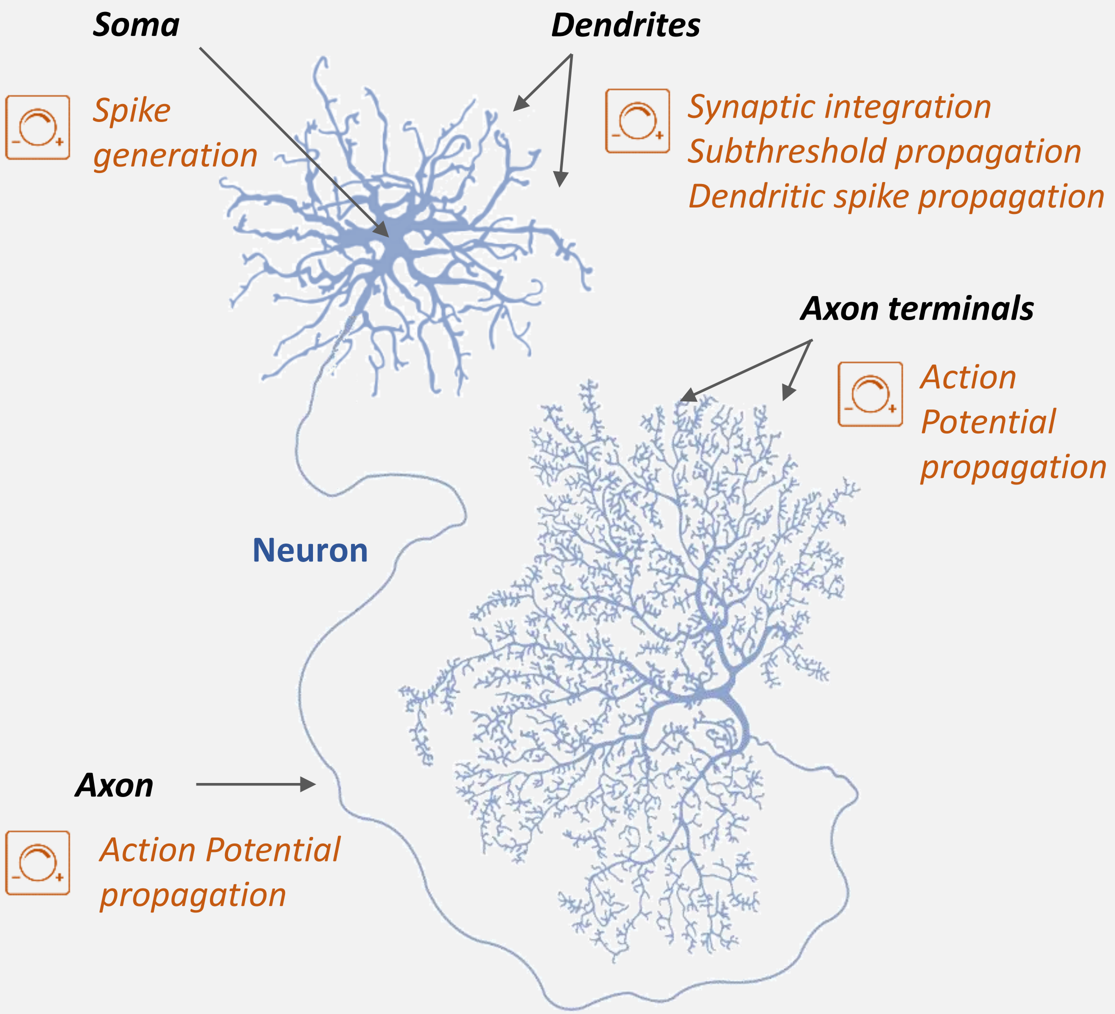
Nonsynaptic plasticity can have **short-term or long-term effects** on synaptic integration, subthreshold propagation, spike generation, and other fundamental mechanisms of neurons at the cellular level.

Although research on nonsynaptic plasticity is still in its infancy, it is generally believed that **both synaptic and nonsynaptic plasticity are essential to memory and learning in the brain**. Their mechanisms complement each other.

For instance, LTP mechanisms at the synapse level can be accompanied by the densification of voltage-gated ion channels along some axon terminals in the presynaptic neuron (strengthening of neuronal action potential) and/or some dendritic branches in the postsynaptic neurons (increased significance in synaptic integration). The regulation of those ion channels augments the effectiveness of synaptic memory formation.

Nonsynaptic plasticity also has a **homeostatic role in order to prevent long-term drift towards excitability or inexcitability**. This continuous regulation makes sure that the circuit keeps its ability to convey information (too many and too few firings mean lower information transmission).

Example of neuronal mechanisms affected by nonsynaptic plasticity



Adapted from Dowling, 2018

Specific architectures allow to assign credit to the faulty neurons inside a global network

In a complex network of neurons, how to know which synapses to strengthen and which synapses to weaken when the outcome turned out to be bad? This question is referred to as the **credit assignment problem**.

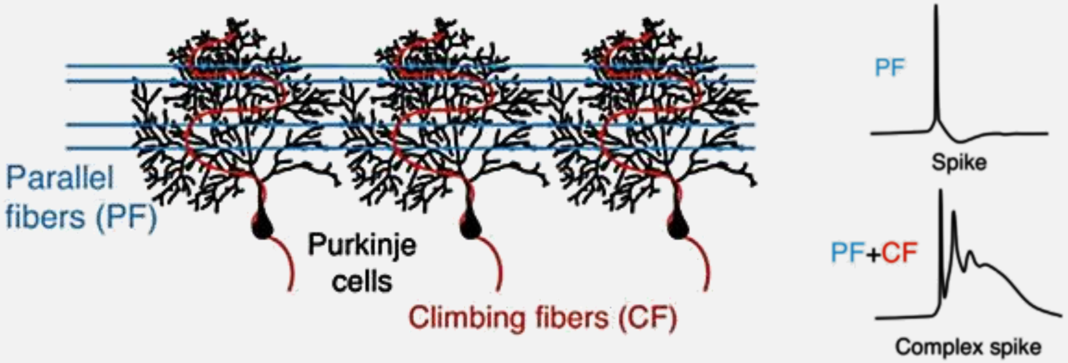
The difficulty of the problem lies in the fact that all plasticity mechanisms are local in the brain, whereas signals transit successively through many neurons before knowing the outcome.

According to recent research, brains seem to have overcome this issue thanks to **specific neuron morphologies and network architectures**.

For instance, *Purkinje cells* – large neurons located in the *cerebellum* – have an intricately elaborated dendritic arbor that is innervated by two kinds of fibers. One kind of fiber probably acts as an error signal. It synapses onto Purkinje cells in a one-to-one correspondence and modifies the spike profile when activated along with the other fibers.

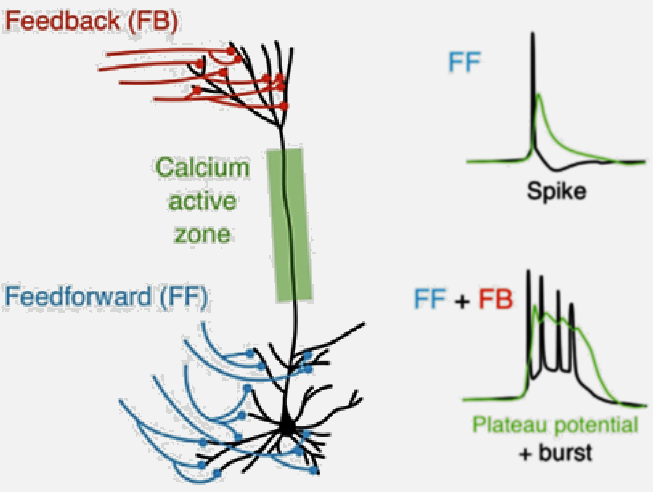
A similar dendritic solution for credit assignment may also exist in the *cerebral cortex*. This developing theory is still under investigation because multiple inputs and multiple outputs are related in a highly complex way contrary to the cerebellum that is basically organized in a characteristic feedforward manner.

Cerebellum ***Neuron-by-neuron credit assignment***



Climbing fibers synapse onto Purkinje cells in a one-to-one correspondence. When stimulated along with parallel fibers, they induce a complex spike that may give the sign of change of synaptic plasticity (LTP or LTD).

Cortex ***Credit assignment solution still unknown***



Pyramidal neuron

Because the coactivation of different dendritic arbors induces a more complex spike, and because some arbors received more inputs from areas higher in the hierarchy, it is hypothesized that it may be interpreted as a credit assignment signal.

Richards and Lillicrap, 2019

Brain general machinery

3. Interconnected brain structures group neurons into organized network architectures

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- John E. Dowling

See the reference section for a list of materials that inspired me.

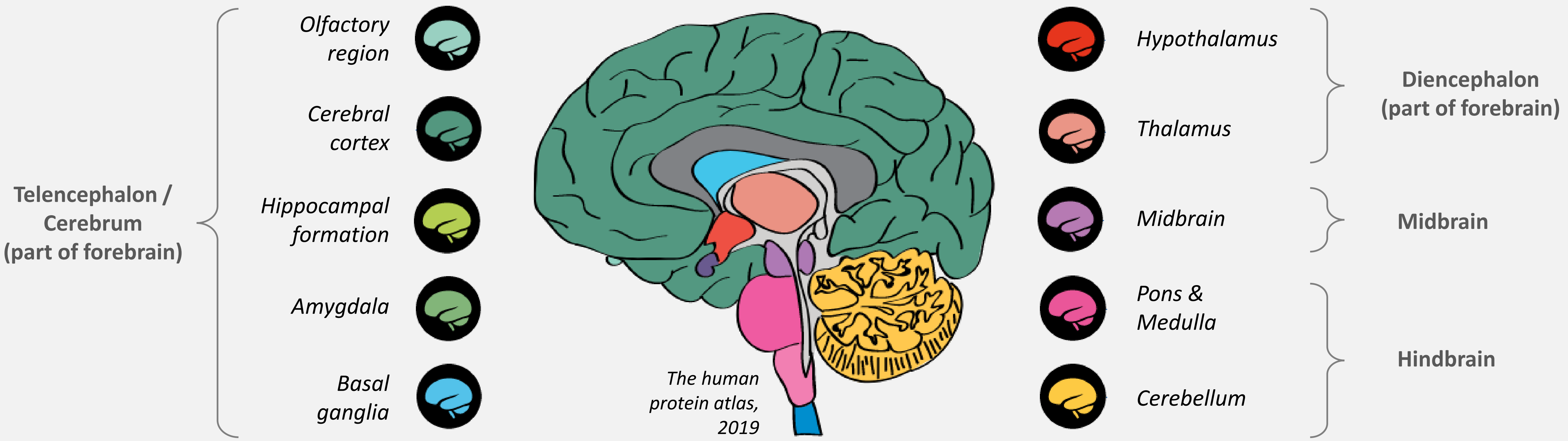
Art credit: Self Reflected, Greg Dunn Design

Related neural networks are grouped in macro-structures identifiable at the anatomical level

Brain anatomical structures result from the gradual differentiation of neural stem cells during the early development of the nervous system. At the highest level, they are separated in **forebrain, midbrain, hindbrain and spinal cord**. At a lower level, they are often divided in a dozen of regions *(see illustration below)*, each one consisting of the aggregation of several substructures. For example, the thalamus is composed of dozens of nuclei.

The substructures of a brain structure are spatially grouped, except the basal ganglia which groups various nuclei, some of which being very distant *(not shown in the illustration)*. This naming convention reflects the functional relation between those highly interconnected nuclei.

Cognitive abilities like intelligence have mainly been associated with specific brain regions like the cerebral cortex, the hippocampus and the basal ganglia. But without the other underlying subcortical areas, those structures are pretty useless. Brains work as a whole.



Telencephalon / Cerebrum (part of forebrain)

Olfactory region

Cerebral cortex

Hippocampal formation

Amygdala

Basal ganglia

Hypothalamus

Thalamus

Diencephalon (part of forebrain)

Midbrain

Midbrain

Pons & Medulla

Cerebellum

Hindbrain

The human protein atlas, 2019

Neural networks are organized in a variety of nuclei and layered architectures

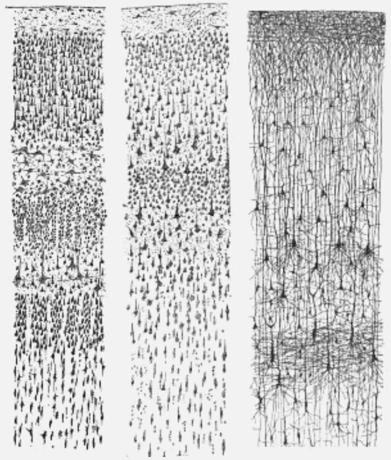
Each brain substructure possesses its own organizational design.

At a macro-level, we can distinguish between 3 kinds of design:

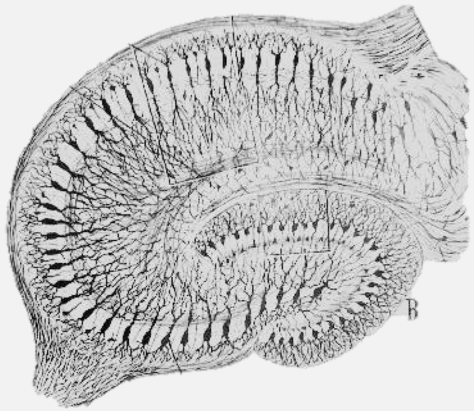
- **Layers**: neurons are grouped in layers, and the connection patterns between layers are conserved across this 2D scalable brain structure (ex: *neocortex, hippocampus, cerebellar cortex, optic tectum*…)
- **Nuclei**: neurons are segregated along a radial organization that is sometimes described as concentric layers (ex: *pallium of birds, cerebellar nuclei, red nucleus, inferior olive, hypothalamus nuclei, basal ganglia*…)
- **No apparent structure**: the distribution of neuron types still follows a gradient but it is more diffuse (ex: *substantia innominata*)

Some examples of neuronal organizations:

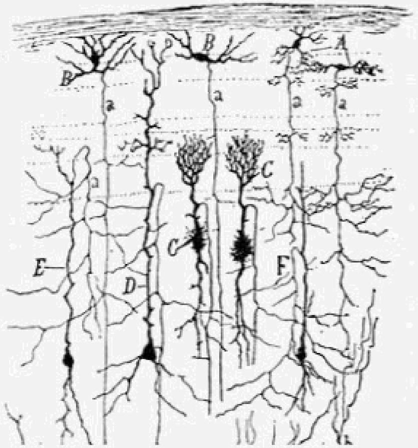
Neocortex



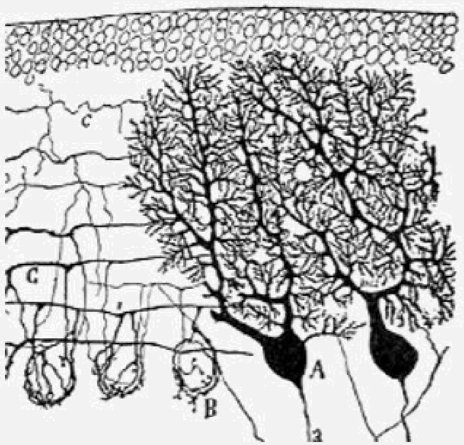
Hippocampus



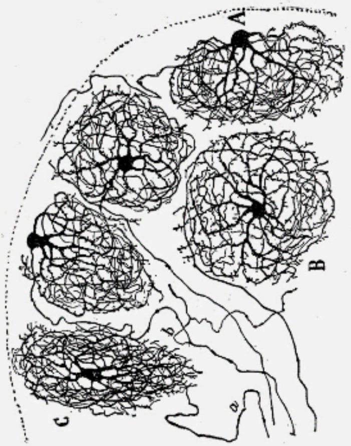
Optic tectum



Cerebellum



Olivary nucleus



Ramon y Cajal, 1900

Layered structure

Nuclei

Thiboust, 2020

Brain structures are highly interconnected with dense and long-distance fiber tracts

Mapping the neuron-to-neuron connectivity of the human brain is still technically out of reach. However, the macroscopic connectivity between brain regions is sufficiently known to inform the function of those regions and the major processing pathways they are involved in. This structural connectivity is called the **connectome**.

Virtually everything seems connected to everything in the brain! But beyond this redundancy, general connection patterns exist: *all cortical areas are connected to some thalamus nuclei, cerebellum is connected to nuclei in the pons, etc*.

The massive interconnections consist in **nerve tracts (= bundles of axons)**. The length of those fibers ranges from a few millimeters to a dozen of centimeters.

The cerebral cortex is involved in most long-distance fiber tracts which are commonly classified into three categories:

- **Association fibers** connect cortical areas within the same hemisphere
- **Commissural fibers** connect corresponding cortical areas in the two hemispheres. The biggest commissure is the *corpus callosum*
- **Projection fibers** connect cortical areas with the thalamus, the basal ganglia, the midbrain, the pons, the medulla and the spinal cord

Remark: direction of fibers and network dynamics are generally not represented in connectomes

Brain connectome

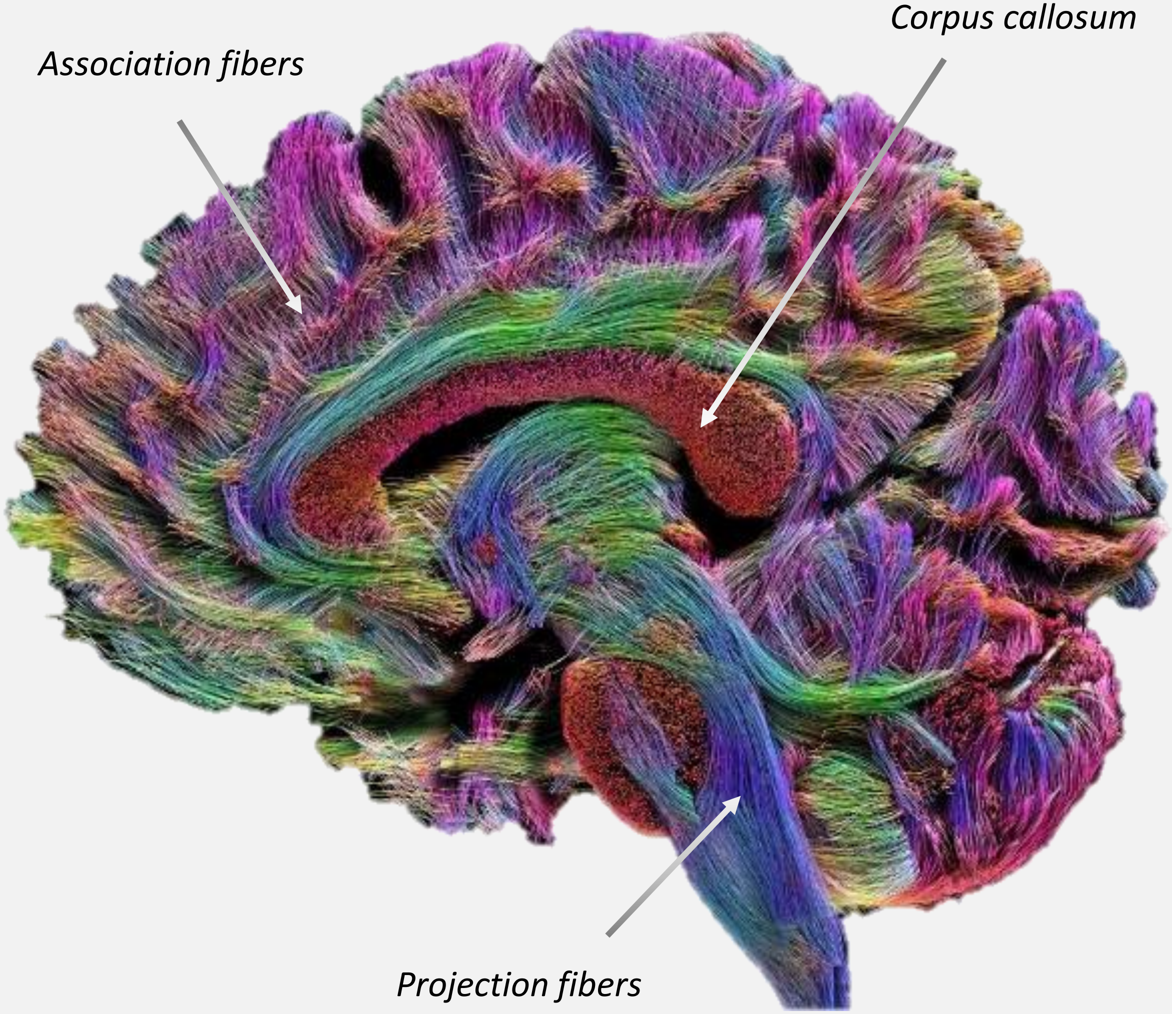


Image credit: Courtesy of the Laboratory of Neuro Imaging and Martinos Center for Biomedical Imaging, Consortium of the Human Connectome Project – www.humanconnectomeproject.org

Brain general machinery

4. Brain activity continuously loops across those structures through parallel pathways

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- György Buzsáki

See the reference section for a list of materials that inspired me.

Art credit: Self Reflected, Greg Dunn Design

Neural activity continuously loops into multiple interconnected neural circuits

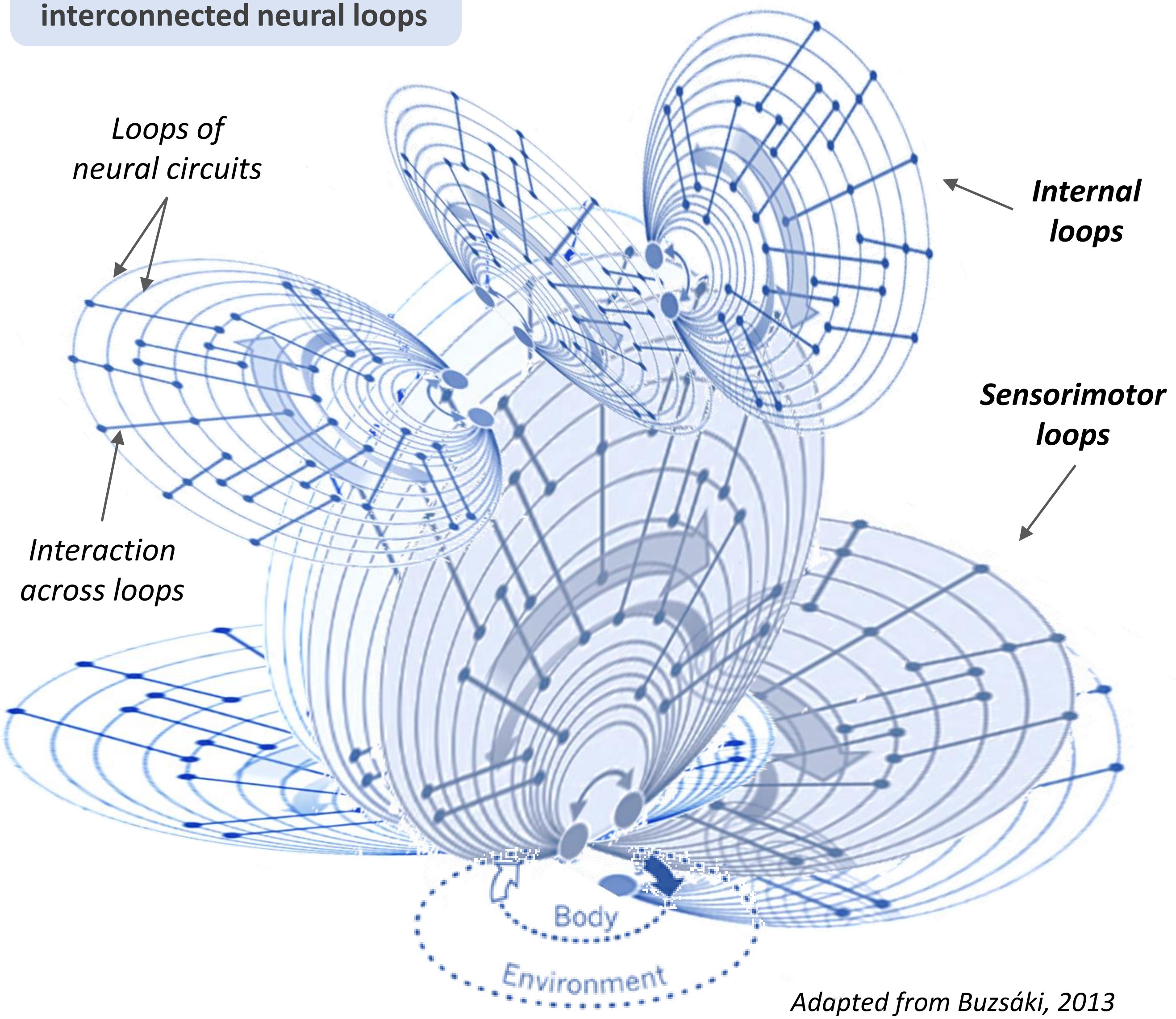
Brains consist of networks of neurons forming parallel and intricated closed-loops:

- **Sensorimotor loops through the environment or the body** are necessary to ground a given neural activity with external stimuli via sensorimotor interaction: motor actions shape sensory input and sensory percepts guide future motor commands.
- **Internal loops** sustain, regulate and coordinate neural activity inside and between brain substructures. This internal activity bypasses sensors and motor effectors, by addressing corollary discharges directly from motor to sensory centers via multiple parallel pathways. The loop is closed inside the brain.

Some brain loops (far from exhaustive)

- *Sensor-motor loop*
- *Sensor-brainstem-motor loop*
- *Sensor-thalamo-cortico-motor loop*
- *Cortico-cortical loop*
- *Cortico-thalamo-cortical loop*
- *Cortico-basal ganglia-thalamo-cortical loop*
- *Cortico-ponto-thalamo-cerebello-cortical loop*

Schematic representation of interconnected neural loops



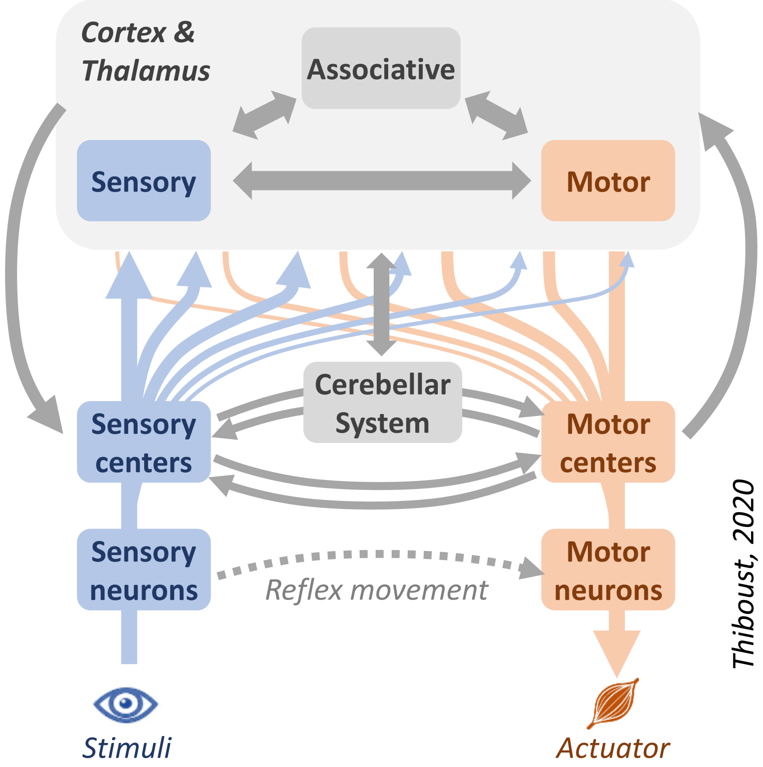
Adapted from Buzsáki, 2013

Sensorimotor loops through the environment are closed via efference copies

Multiple pathways exist from sensory stimuli to body actuators:

- Very short pathways for reflex movements through the spinal cord or the brainstem
- Short pathways through subcortical structures (brainstem & cerebellum)
- Long pathways through the cerebral cortex

Importantly, those pathways form **sensorimotor loops** that are closed inside the brain by **efference copies** (also called *corollary discharges*) which are internal copies of motor signals directed towards sensory systems. Those signals are essential to distinguish between self-initiated movements (*reafference*) and external signals (*exafference*).



Thiboust, 2020

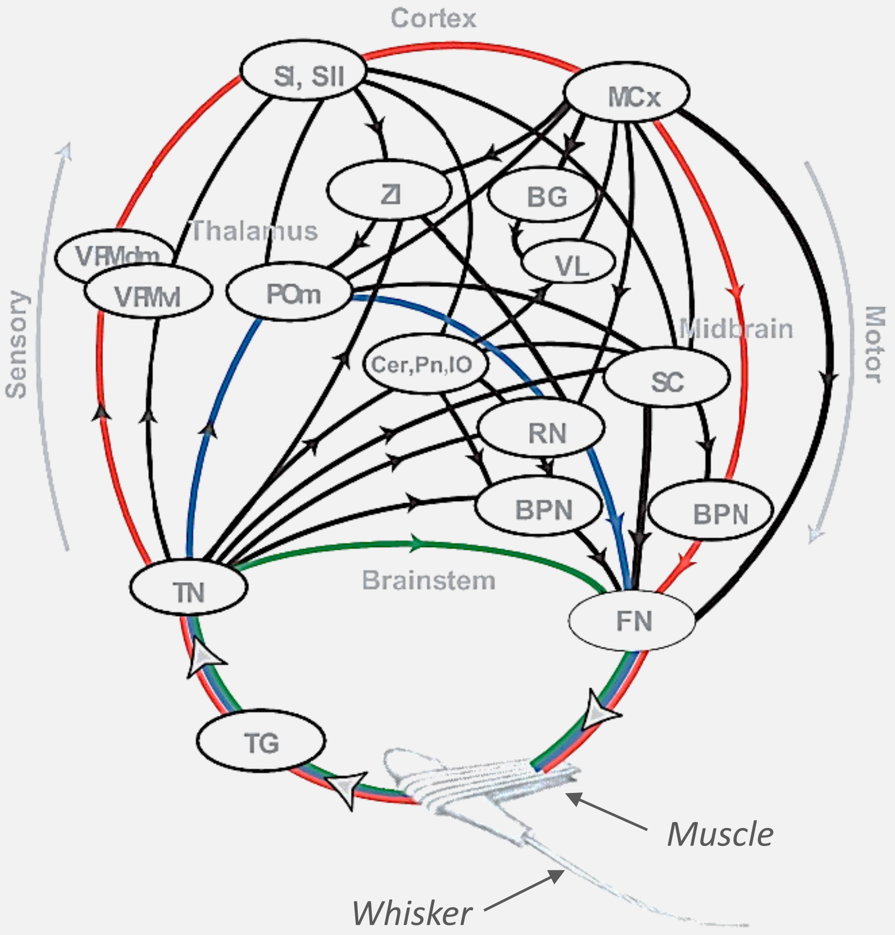
Efference copies also explain the stability of our perceptions despite the regular movements of our sensors. Internal forward models learn to predict sensory inputs from motor commands.

Such **internal forward models** are believed to be implemented in the cerebellum in addition to some part of the cortex. They enable the brain to predict the effect of actions at different levels of abstraction.

The interactions between the different intricated loops are still poorly understood.

Example of a sensorimotor loop

Connections of the mouse whisker system, forming multiples loops between whiskers and muscles (efference copies are not represented in this illustration).



Ahissar et al, 2016

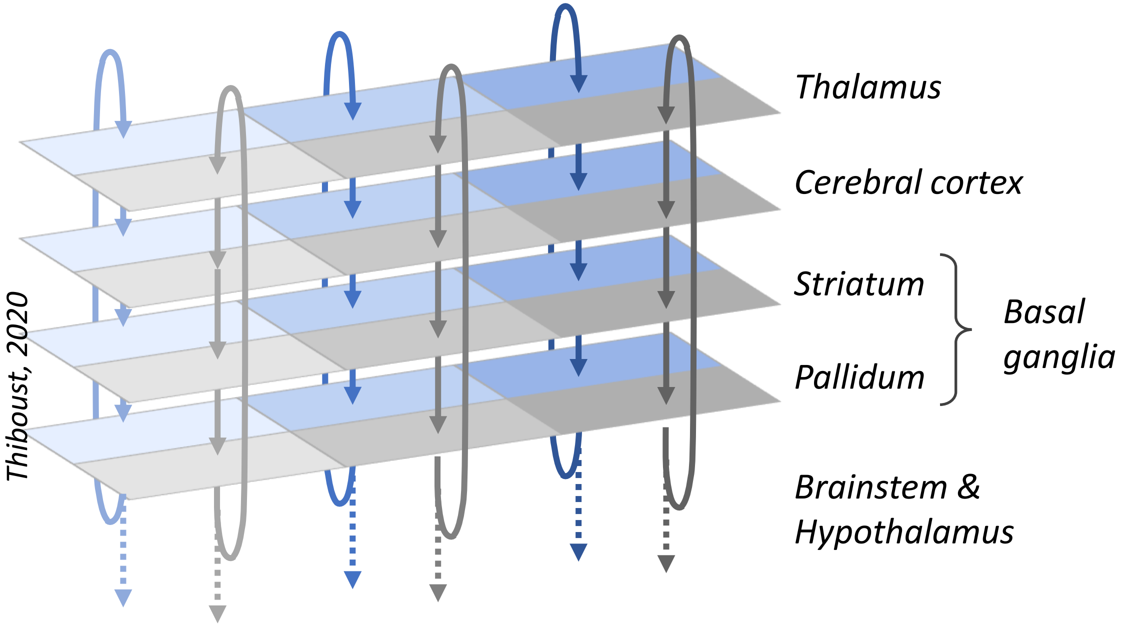
Actions are selected via parallel internal loops through the basal ganglia

The **cortico-basal ganglia-thalamo-cortical loop** is a fundamental processing pattern in the forebrain. It is implicated in action & behavior selection, motivation, reward learning and decision making.

Multiple such loops exist in parallel: a given cortical area projects to a given area of basal ganglia (first striatum, then pallidum) which projects to a given thalamic nuclei, which projects back to the corresponding cortical area (projections are said to be topographically organized).

When the activity in a loop has converged to a selected choice, the pallidum communicates this decision to other nuclei (in brainstem or hypothalamus).

Simplified view of the parallel cortical-basal ganglia-thalamo-cortical loops:

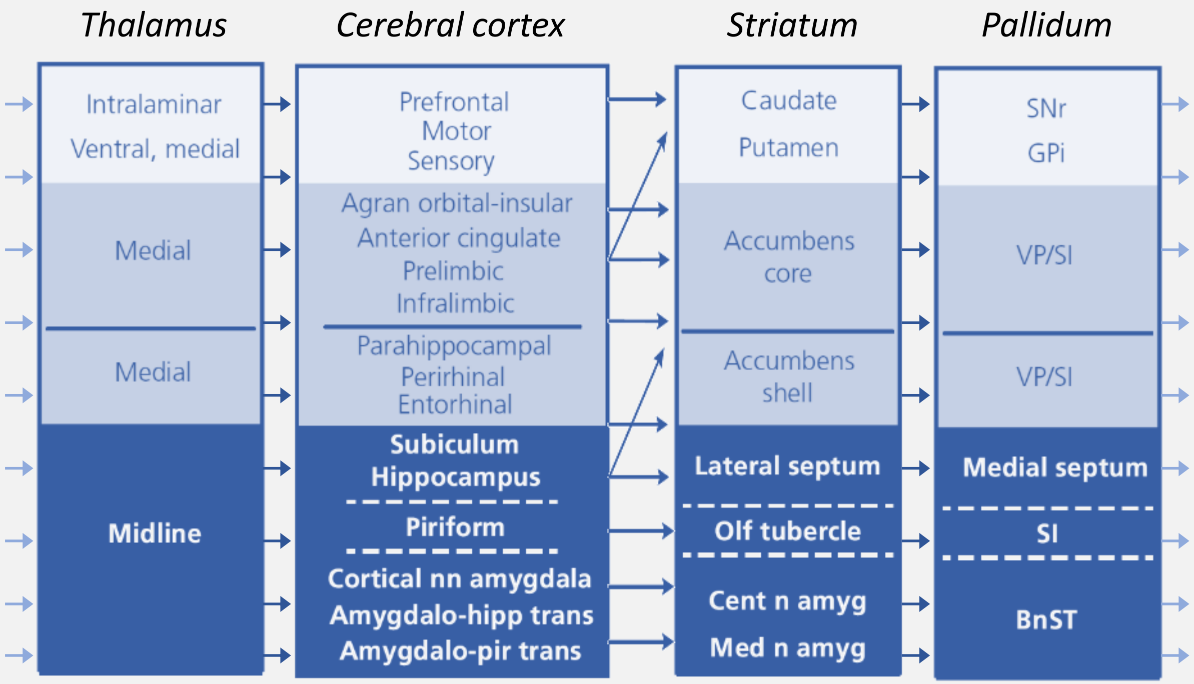


Thiboust, 2020

List of basal ganglia loops

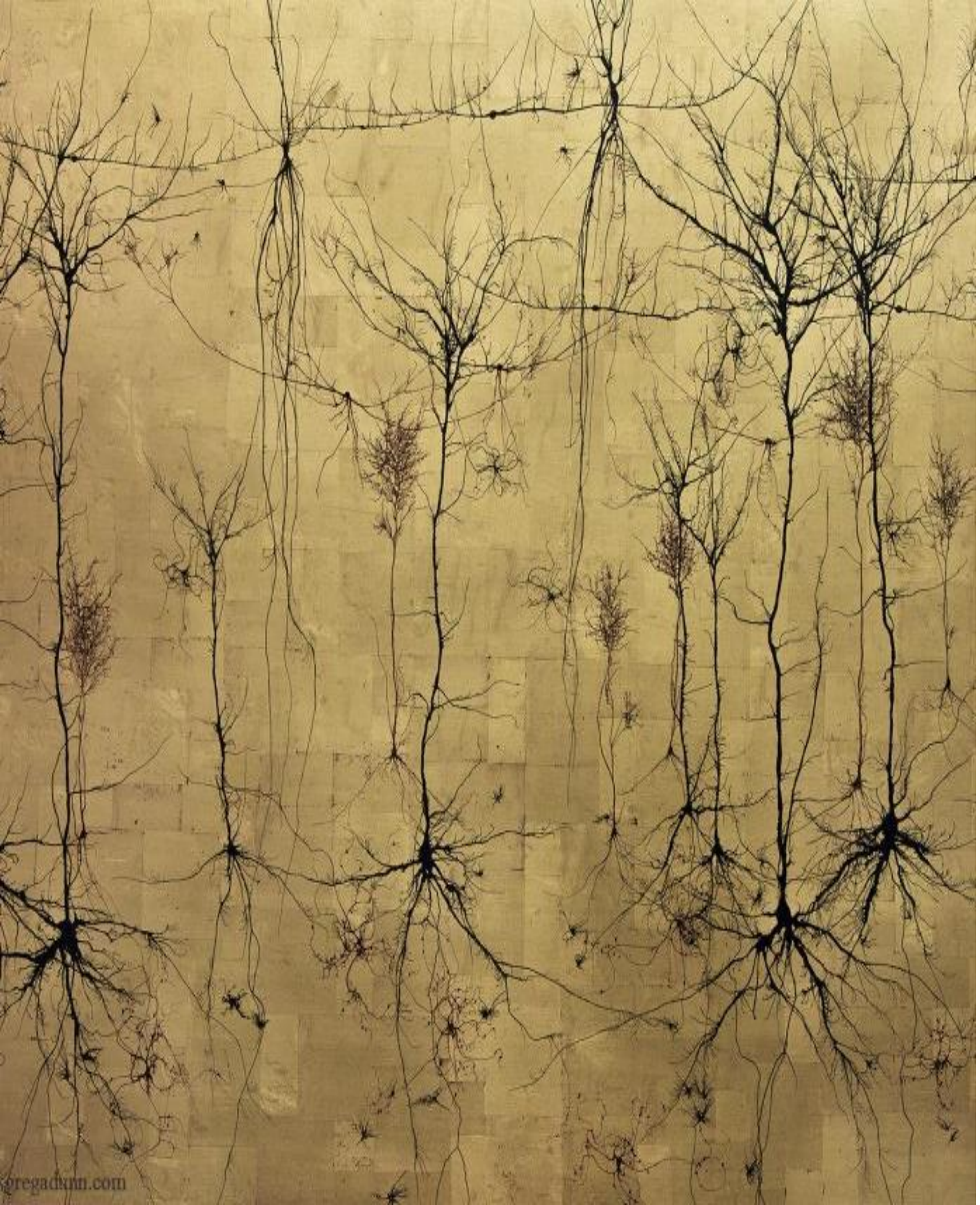
The typical examples of loops are the sensorimotor, associative and limbic circuits whose striatal structures are respectively the putamen, the caudate nucleus and the nucleus accumbens.

Interestingly, the terminology of basal ganglia can be used in an extended sense to include striatal-like and pallidal-like structures. For example, the hippocampus is involved in a loop with the septal nuclei and the midline thalamic nuclei. The different amygdala nuclei are also involved in a similar loop.



Thalamus	Cerebral cortex	Striatum	Pallidum
Intralaminar Ventral, medial	Prefrontal Motor Sensory	Caudate Putamen	SNr GPi
Medial	Agran orbital-insular Anterior cingulate Prelimbic Infralimbic	Accumbens core	VP/SI
Medial	Parahippocampal Perirhinal Entorhinal	Accumbens shell	VP/SI
Midline	**Subiculum Hippocampus Piriform Cortical nn amygdala Amygdalo-hipp trans Amygdalo-pir trans**	**Lateral septum Olf tubercle Cent n amyg Med n amyg**	**Medial septum SI BnST**

Graham, Murray, Wise, 2017



Focus on the neocortex

1. The neocortex is divided into functionally specialized but anatomically similar cortical areas

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- Jeff Hawkins
- Vernon Mountcastle
- Luis Puelles
- Gordon M. Shepherd

See the reference section for a list of materials that inspired me.

Art credit: Gold Cortex, Greg Dunn Design

The cerebral cortex is a thin sheet of brain cells that can be functionally divided into dozens of specialized areas

The **cerebral cortex** is a two-dimensional thin sheet of neural tissue covering the outside of the brain in two hemispheres, connected to nearly all brain structures.

All vertebrates possess a cerebral cortex (or a *pallium*, its primitive form), but its significance greatly increased in mammals, with the expansion of the part that is called **neocortex** (or *isocortex* to avoid the misconception of a mammalian innovation).

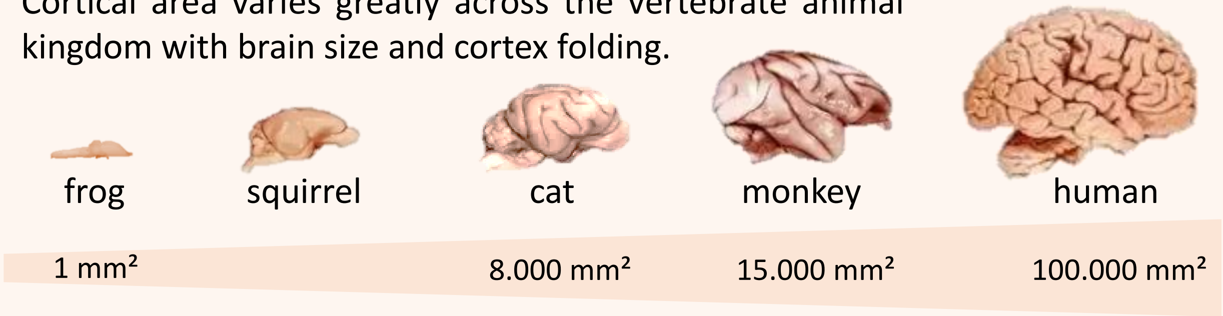
This expansion in surface area gradually continued through the phylogenetic tree by increasing the size of the brain and by folding the sheet, forming the famous characteristic patterns on the external surface of primate and human brains.

Different high-order cognitive abilities are associated to different specialized areas that can be divided at 3 structural levels for each hemisphere:

- **4 cortical lobes**: frontal, temporal, parietal and occipital
- **52 Brodmann areas** (ex: area n°17 correspond to the primary visual cortex V1)
- **180 cortical areas** (from Glasser, 2016, with Human Connectome Project data)

Cortical surface area

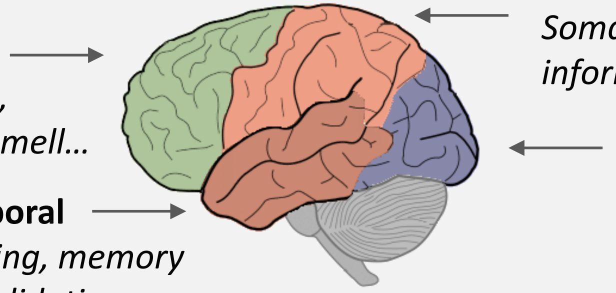
Cortical area varies greatly across the vertebrate animal kingdom with brain size and cortex folding.



Pictures from Kinser, 2000

Specialized Cortical areas

Human brain



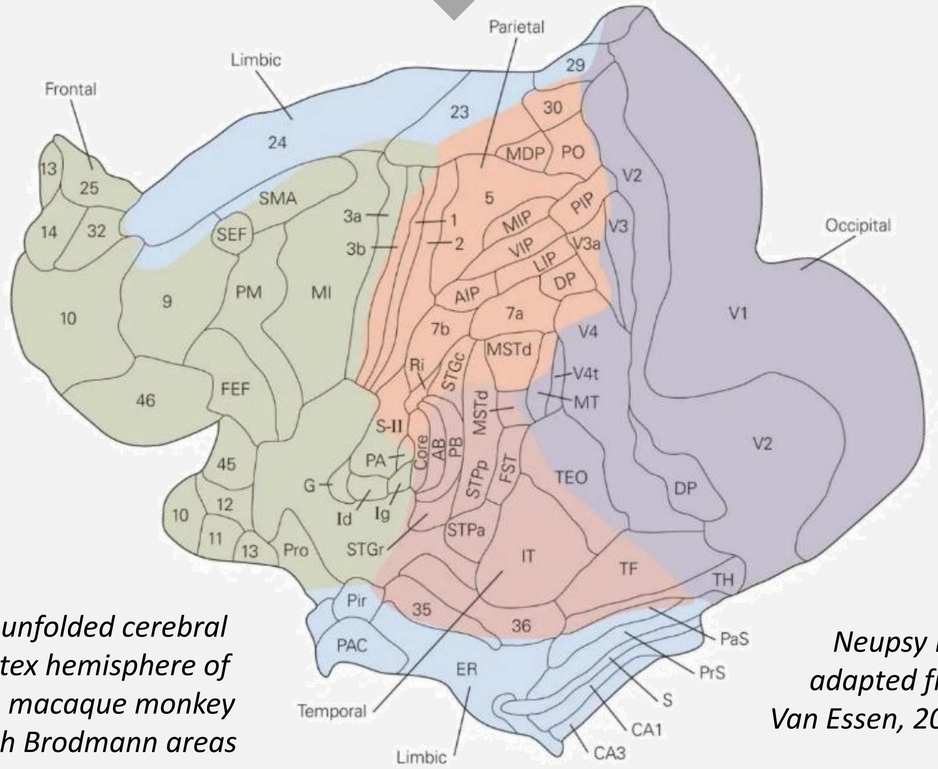
Frontal
Movement, planning, smell…

Temporal
Hearing, memory consolidation, …

Parietal
Somatosensory information, …

Occipital
Vision

Unfolding



An unfolded cerebral cortex hemisphere of the macaque monkey with Brodmann areas

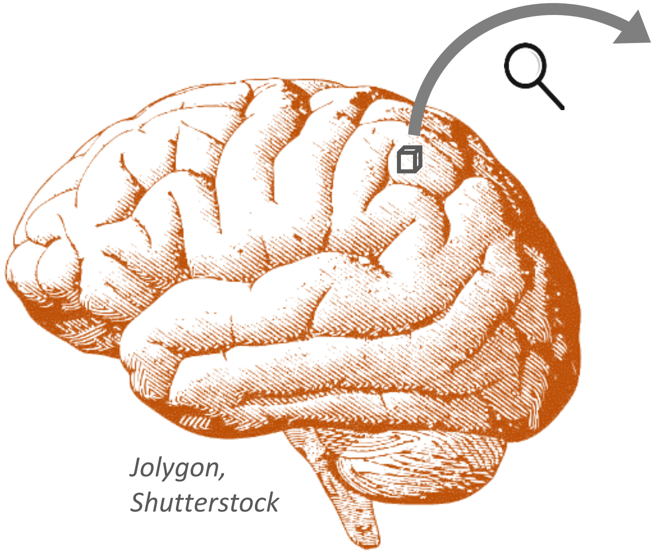
Neupsy Key, adapted from Van Essen, 2001.

All cortical areas have a similar anatomical structure made of a collection of fundamental units called minicolumns

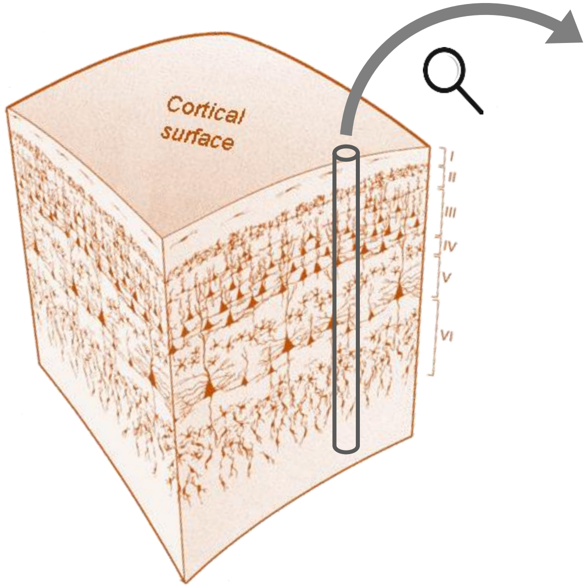
Even if the different cortical areas support very diverse functions, their anatomical organization is strikingly similar.

Indeed, the whole cortical sheet is made of a collection of *anatomical fundamental columnar units* called **minicolumns** (around 50 µm of diameter). Each cortical area is basically a collection of millions of minicolumns (each one being composed of around 100 neurons). Minicolumns are organized in layers (generally 6 layers), with specific neuron types and connection patterns in each layer. This organization is said to be laminated.

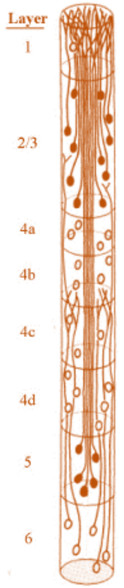
Neighboring minicolumns share a same Receptive Field (RF), meaning that they are innervated by the same axonal inputs. Those minicolumns form structures called **macrocolumns/hypercolumns** (around 500 µm of diameter) that are thought to be *functional fundamental units* (the hypothetical functional role of macrocolumns remains controversial).


Jolygon, Shutterstock

1 cortical sheet
2 million macrocolumns
200 million minicolumns
20 billion neurons



1 macrocolumn
100 minicolumns
10.000 neurons



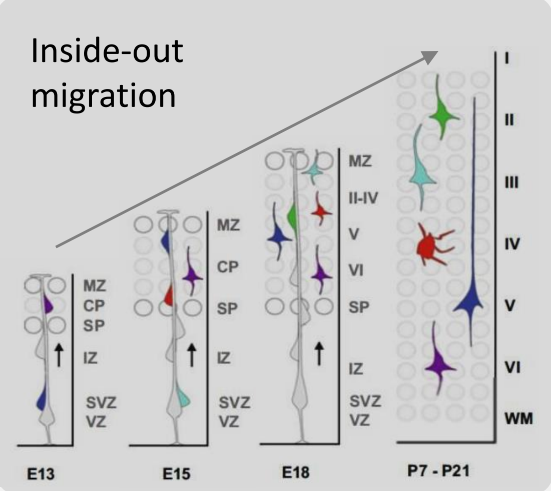
Mountcastle, 1997

1 minicolumn
100 neurons

Development of minicolumns

All excitatory neurons of a minicolumn come from the same progenitor cell that divided multiple times in a radially inside-out manner during brain development in the embryo. This origin explains the vertical columnar aspect and the regularity of the inter-laminar connection pattern inside a microcolumn.

This common origin does not concern inhibitory neurons which migrate later into the cortical plate.



Hippenmeyer, 2014

Cortical neurons are horizontally organized in 6 layers identifiable by the density and morphology of their neurons

The **neocortex is classically arranged in 6 layers** that lie above a dense horizontal network of fiber tracts (white matter).

Those fibers are axons traveling to and/or from some cortical areas. When they enter the cortical plate, they form vertical axon bundles (parallel to minicolumns) with some axons crossing just a few layers, and others going until L1 near the pial surface.

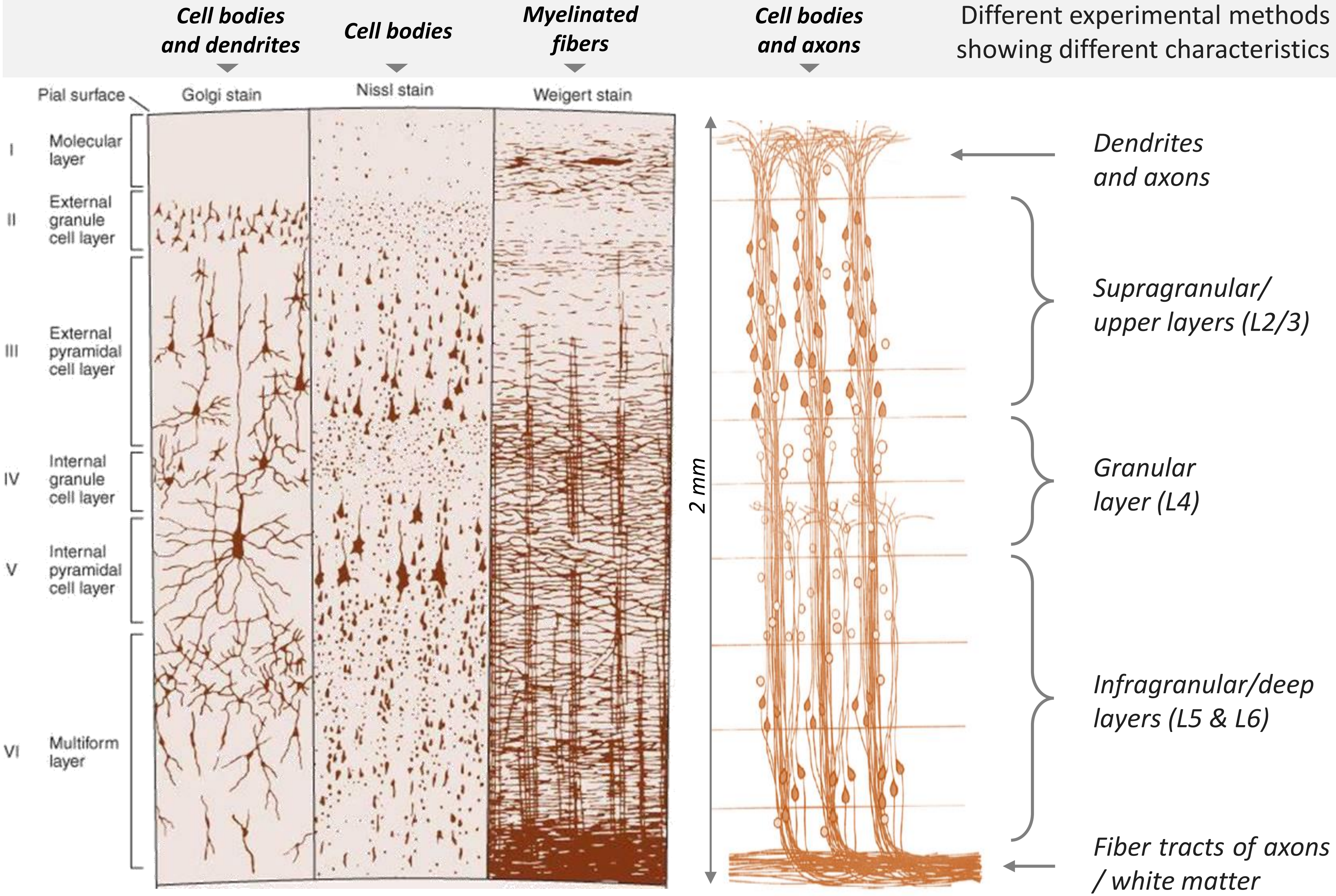
Cortical neurons are distributed through L2 to L6, with laminar specificity:

- Stellate cells are more common in L4 and L2
- Pyramidal cells in L2/3, L4 (small and medium size), L5 (big) and L6

Pyramidal cells have an apical dendritic tree climbing vertically (some going towards L1).

Lateral/horizontal *myelinated* fibers are more common in deep layers than upper layers.

Different experimental methods showing different characteristics



Cell bodies and dendrites
Cell bodies
Myelinated fibers
Cell bodies and axons
Pial surface
Golgi stain
Nissl stain
Weigert stain
I Molecular layer
II External granule cell layer
III External pyramidal cell layer
IV Internal granule cell layer
V Internal pyramidal cell layer
VI Multiform layer
2 mm
Dendrites and axons
Supragranular/ upper layers (L2/3)
Granular layer (L4)
Infragranular/deep layers (L5 & L6)
Fiber tracts of axons / white matter

Ranson, 1959 and Mountcastle, 1997

The laminar organization and composition slightly vary across the cortex

Dividing the cortical sheet into layers is hard to get exactly right, although it's good enough for most purposes. More, this laminar structure is not uniform across the cerebral cortex.

Some evolutionary ancient parts (called ***allocortex***) have less layers than their ***neocortex*** counterparts: only 3 layers for the hippocampus and 4 layers for the olfactive system.

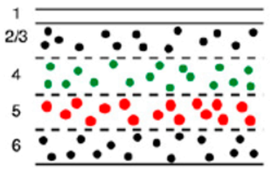
Even in the neocortical 6-layer structure, there are significant variations along a ***granular-agranular*** axis. Compared to agranular cortices, granular cortices have a smaller thickness, a greater neuron density & number, and a large granular L4 giving them their name.

The composition of some layers also differs. Granular cortices tend to have a greater proportion of stellate cells than pyramidal cells in L2/3. Moreover, the nature of L4 cells is not the same in the primary visual cortex and the somatosensory cortex, two granular cortices.

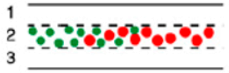
The variation from granular to agranular forms a continuum across the cortex:

- **Granular cortices** for primary sensory areas *(in red in the figure)*
- **Less granular cortices** for higher sensory areas *(in yellow)*
- **Even less granular cortices** for associative and high-order areas *(green & blue)*
- **Agranular cortices** for motor areas *(purple)*

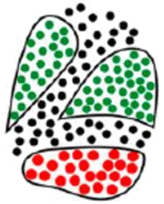
<u>Insight from birds and turtles</u>: Anatomical organization doesn't necessarily make the function. Those animals have similar neuron types and wiring despite structural differences in their cortex-equivalent:



Neocortex of mammals



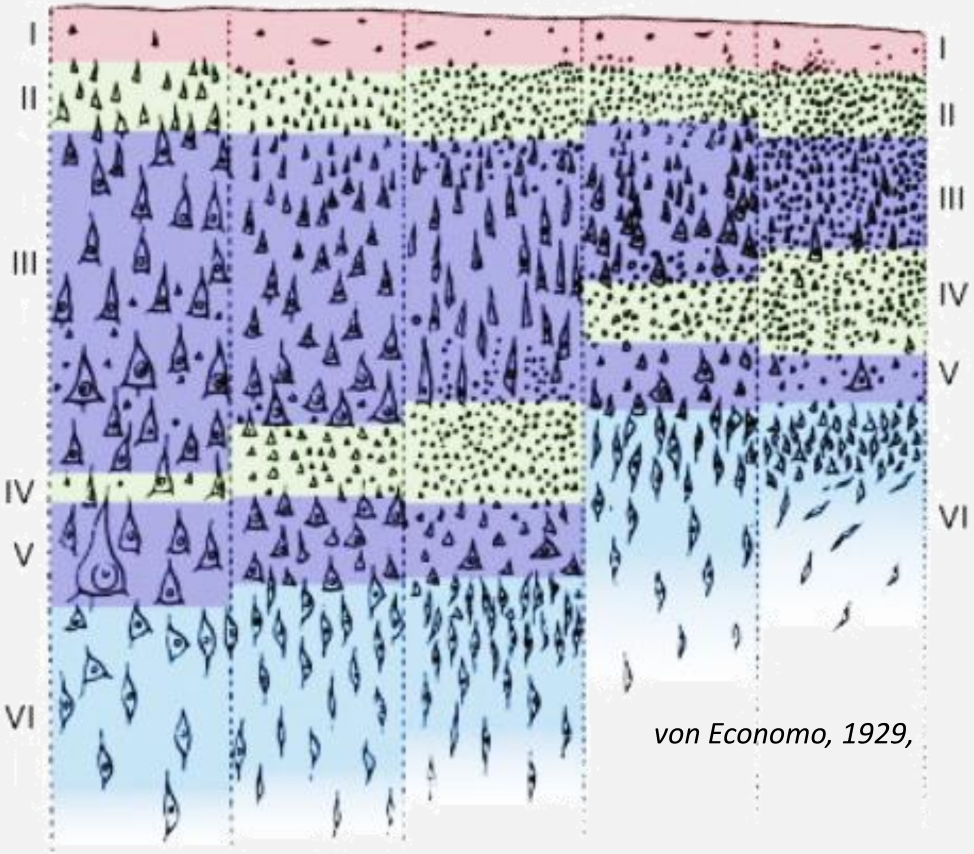
Cortex of turtles



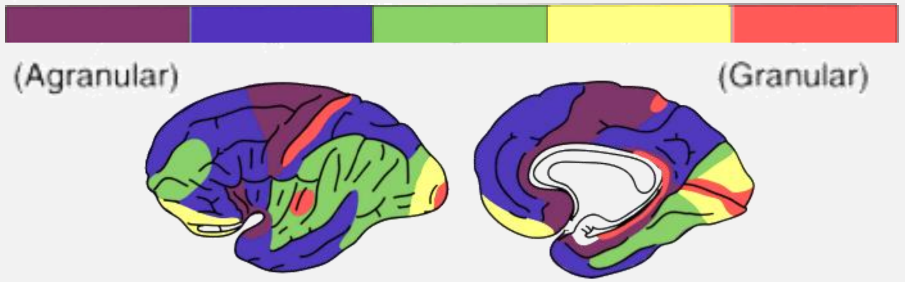
Pallium of birds

Dugas-Ford et al, 2012

Laminar differences of cytoarchitecture



von Economo, 1929,



Fukutomi et al, 2018

Processing flows mainly follow laminar differences: from granular to agranular cortical areas

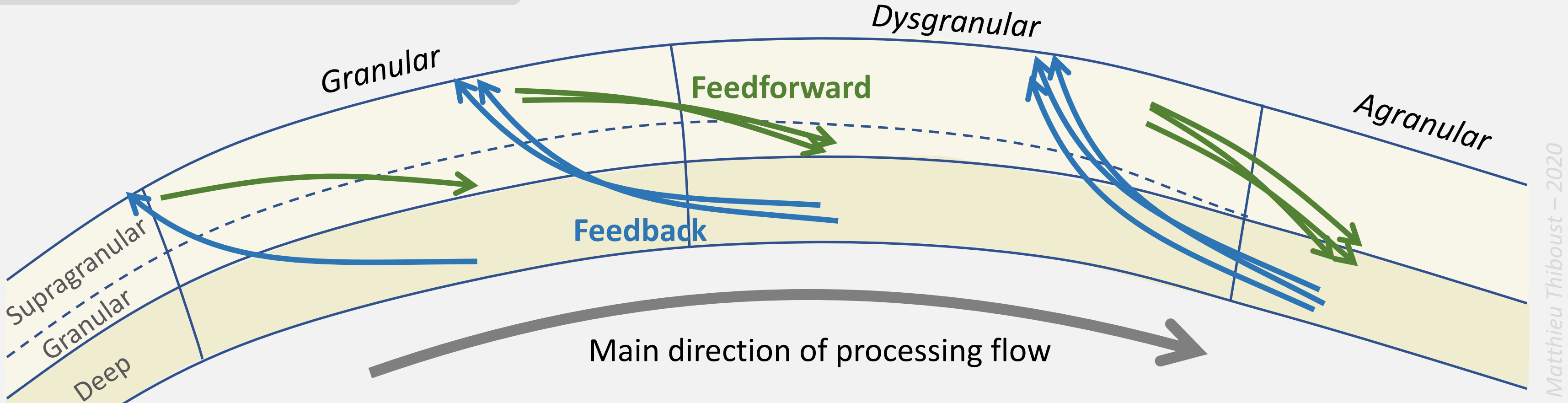
The main direction of processing flow in the cerebral cortex is from granular to agranular areas. Looking at the function of those cortical areas, the main processing flow is **from sensory and high-order areas to motor areas**. Note that there are also many connections going in the other direction.

Unsurprisingly, dense granular areas develop less inter-area connections than loose agranular areas. When those long-distance connections occur between cortical areas of similar cytoarchitecture, they are mostly "horizontal": projections originating from a neuron of a given layer mainly target distant neurons located in the same layer (from L2/3 to L2/3, L5 to L5, and L6 to L6).

However, long-distance connections between areas of different cytoarchitecture are not horizontal: upper layers of granular areas tend to project more to deep layers of agranular areas (and conversely). Those different connection patterns come from temporal differences in cortical development between agranular (early) and granular (late) areas

*Projections in the direction of the main processing flow are generally called **feedforward connections**, and the others are **feedback connections**. This vocabulary can be confusing because the same terms are also used to describe connections between areas of different hierarchical level (but main processing flow does not necessarily follow the level of abstraction)*

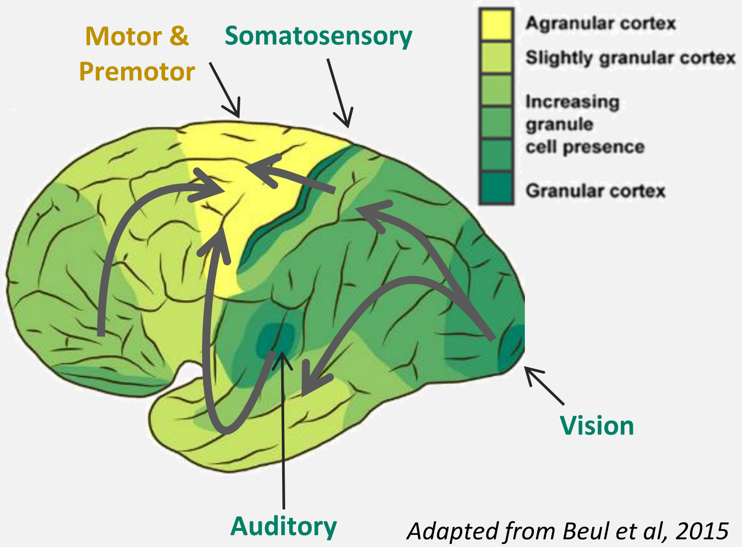
Processing flow at a laminar level



Schematic view only.
Anatomically, those projections travel into the white matter band, and then go up through the cortical layers (till L1 for feedback, and just L4 for feedforward)

Thiboust, 2020

Processing flows at cortex level



Adapted from Beul et al, 2015

Cortical areas are densely interconnected with many feedback and feedforward skip connections across hierarchy

Even if all cortical areas are bidirectionally connected with other brain structures, their **main inputs and outputs come from other cortical areas** via long-distance connections.

The connection matrix between areas is dense and bidirectional. However, there is an organizational and quantitative asymmetry in these bidirectional projections, explained by a **hierarchy between areas**.

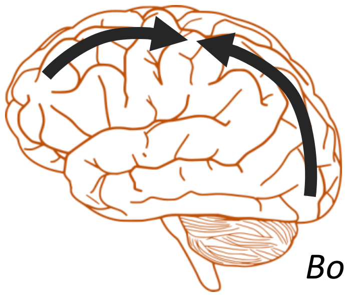
Sensory areas are lower in the hierarchy than associative and motor areas. This classification can be refined at a finer level: for example, the primary visual area has a lower hierarchy level than the secondary visual area.

Bottom-up projections from a lower area to a higher area are generally called **feedforward projections**, in opposition to top-down **feedback projections**.

The “simplified” connectome of visual cortical areas shows general cortical architecture rules:

- Many **skip connections** across the hierarchy (example: V1 projecting directly to V4 bypassing V2)
- High **recurrence** with many feedback loops, and some coming from top-level areas
- **Distributed processing** rather than serial processing

Top-down expectations/needs

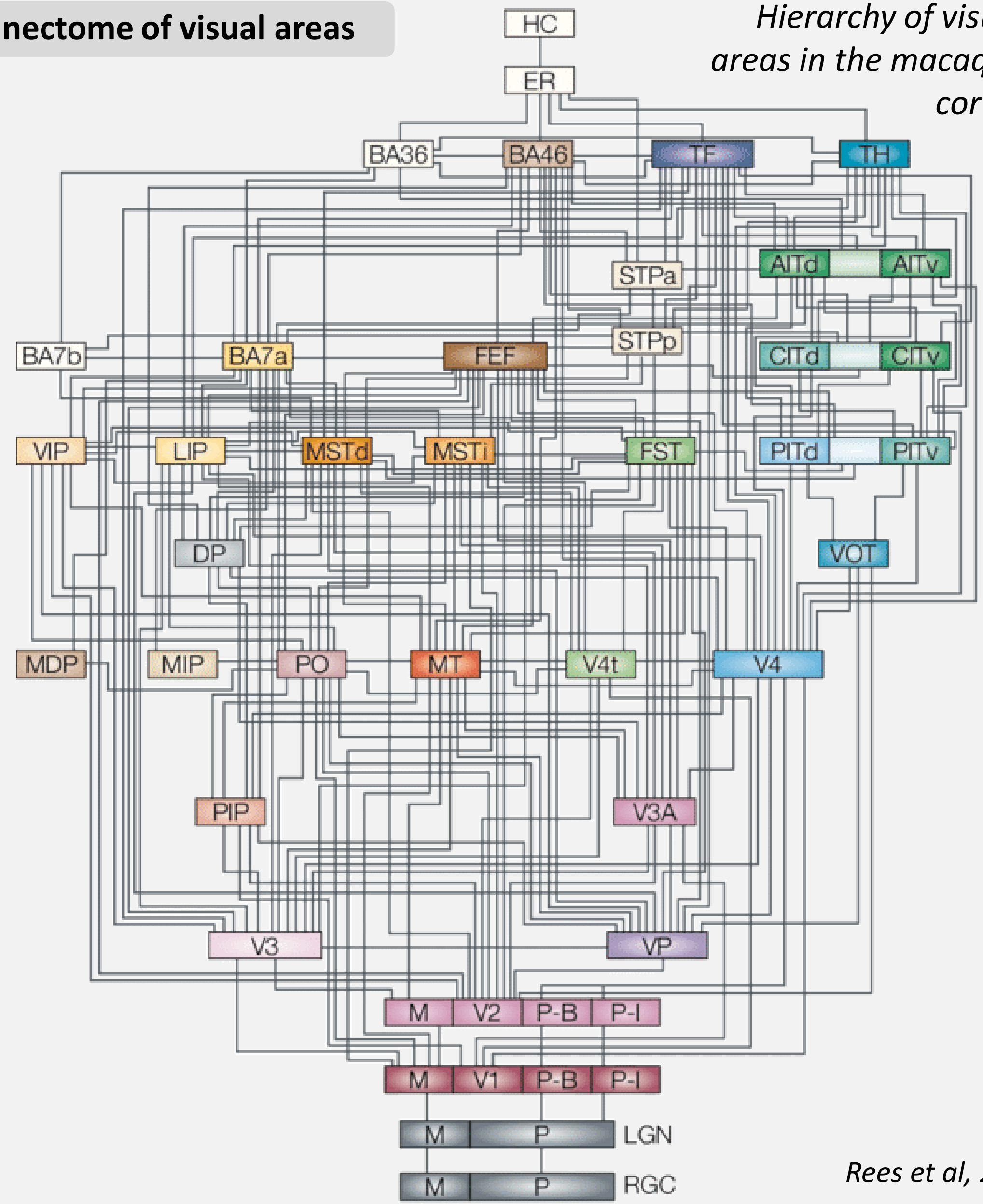


Bottom-up sensory information

Connectome of visual areas

Higher level

Lower level



Hierarchy of visual areas in the macaque cortex

Rees et al, 2002

At a higher level, cortical activity is organized around key cortical and subcortical hub connections

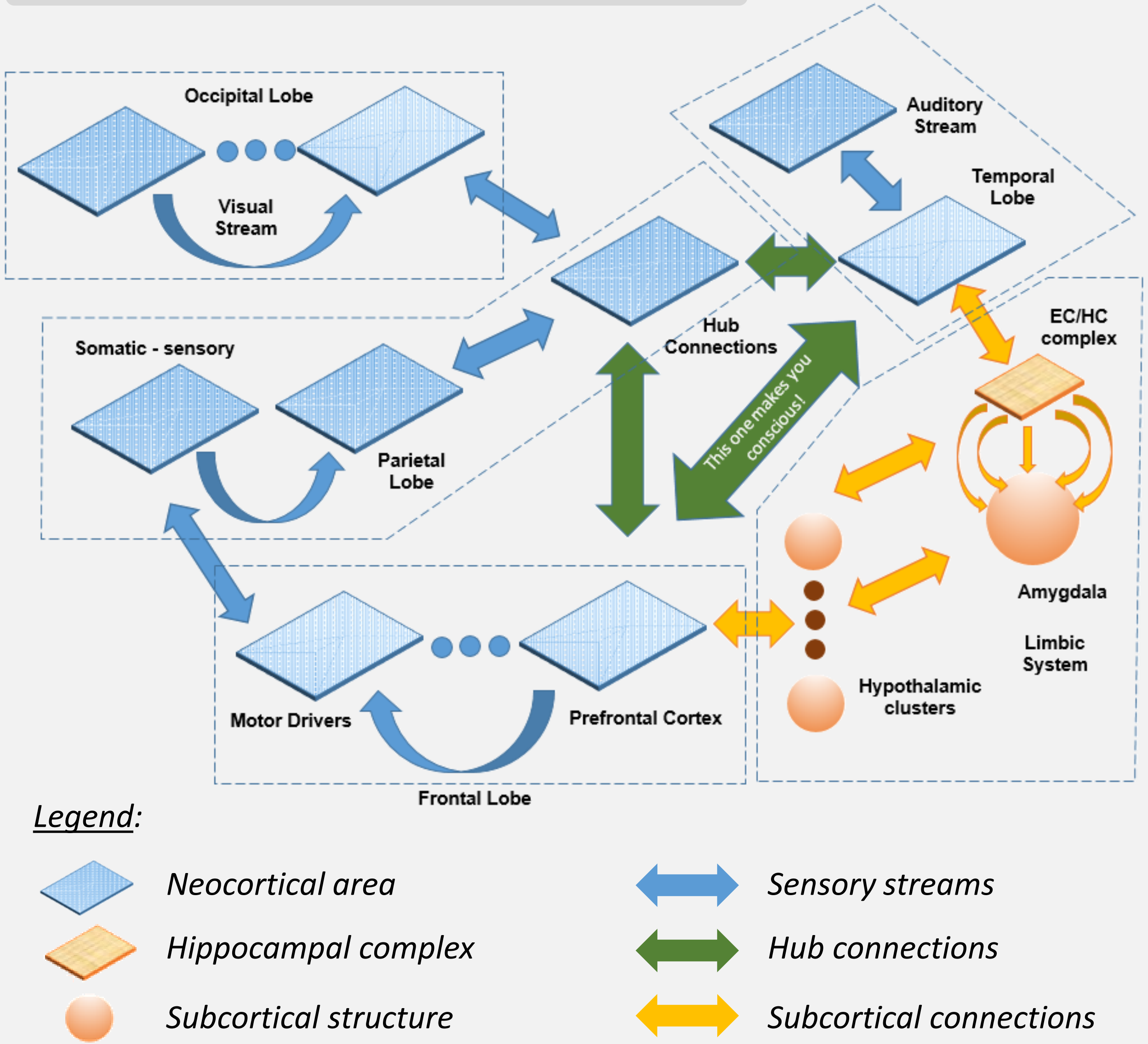
Except in early sensory processing and in late motor areas, information streams occur essentially in a parallel form.

Those bidirectional streams between many cortical areas are organized around **associative areas** that integrate various content in increasingly more abstract concepts. Specific relevant information can be shared with the rest of the cortex thanks to **massive hub connections** between the different associative areas from the parietal, temporal and fontal lobe.

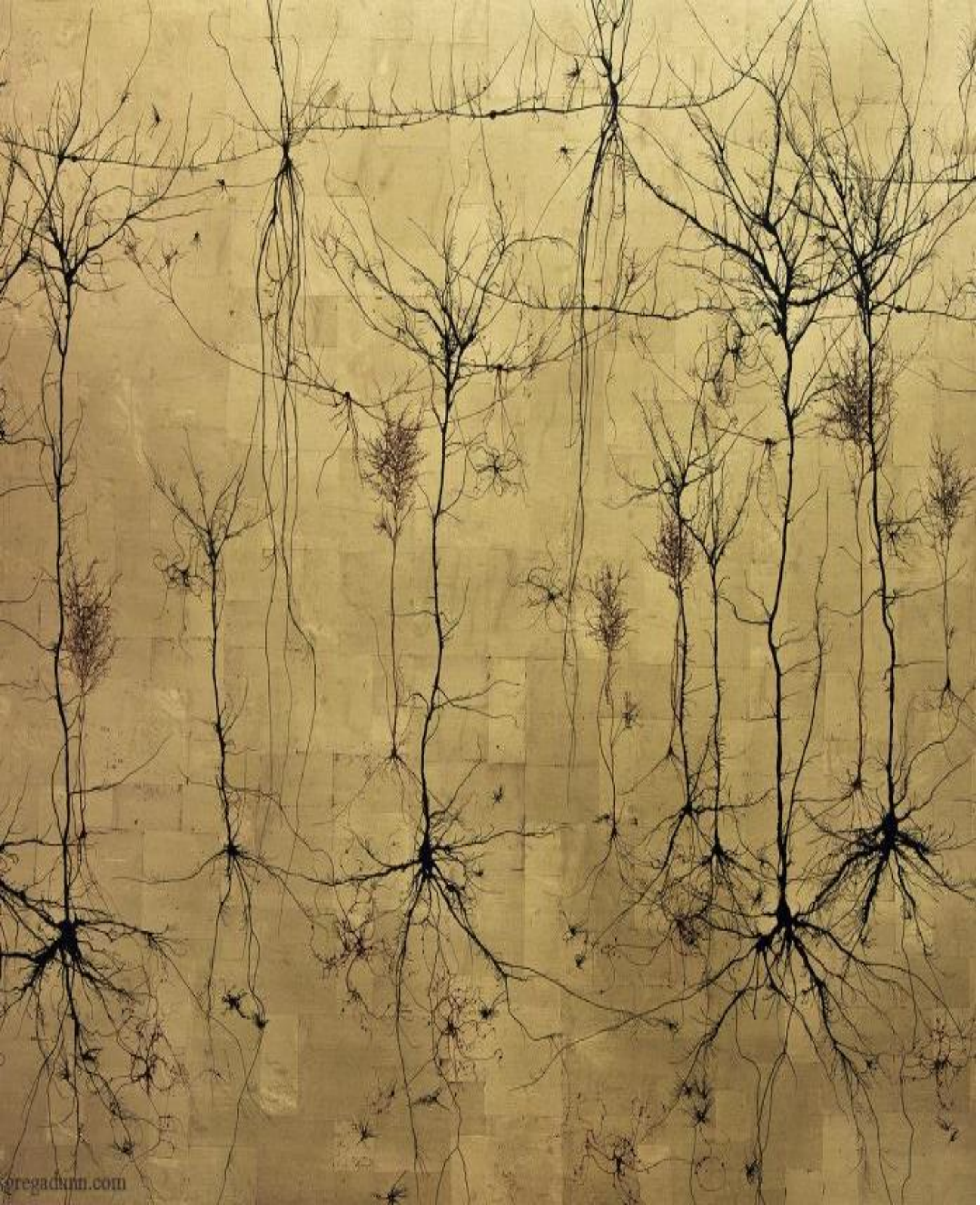
All cortical areas receive information from the thalamus: either relatively raw sensory inputs for sensory areas or already preprocessed inputs for the other areas, or both *(not represented in the diagram)*. In addition to their thalamic inputs, some cortical areas also receive major **inputs from the hypothalamus** (prefrontal cortex) **and the hippocampal complex** (temporal lobe).

The hypothalamus furnishes an internal drive towards the initiation of actions needed by the body, while the hippocampal complex (hippocampus, subiculum and entorhinal cortex) gives access to the individual location in the surrounding environment and to personal experiences related to the self. Those experiences are colored by the amygdala.

Basic system level diagram of the cerebral cortex



Browne, 2019



Focus on the neocortex

2. Cortical areas receive and send information in a laminar-specific way

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- Jeff Hawkins
- Gordon M. Shepherd
- Murray Sherman

See the reference section for a list of materials that inspired me.

Art credit: Gold Cortex, Greg Dunn Design

Each cortical area interacts locally and distally in a canonical way

The organization and connectivity of the neocortex are broadly similar between cortical areas, leading to the idea of a **canonical cortical microcircuit**.

Locally, neurons of a given cortical area interact in two directions:

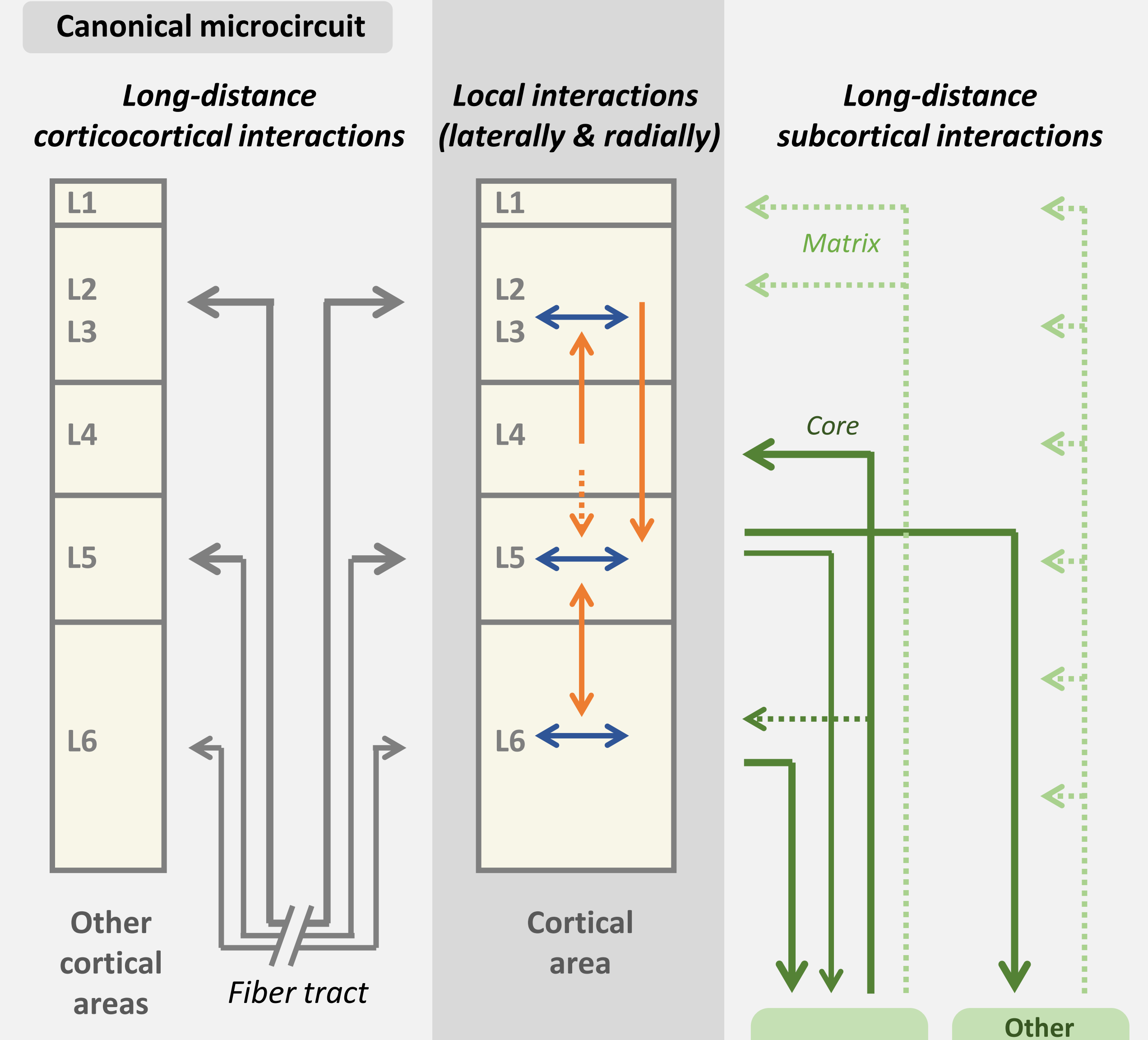
- **Lateral (=horizontal) interactions** inside the same layer (L2/3, L5, L6)
- **Radial (=vertical) interactions** between layers (L4→L2/3, L2/3→L5, L5<-->L6)

Distally, each cortical area interacts with other cortical areas and subcortical structures.

The **thalamus is the main input and output subcortical structure of the neocortex** (it is sometimes referred to as the 7th layer). It primarily sends sensory or preprocessed information to L4 (and to deep layers to a lesser extend on its way to L4). Other thalamocortical projections innervate upper layers in a more diffuse way. On the output side, L5 and L6 project to the thalamus. In addition to the thalamus, L5 also projects to other subcortical structures such as the striatum and motor centers.

In general, the first four layers (L1 to L4) serve as input stations whereas deep layers (L5 and L6) are the main source of output projections. Projections from the basal forebrain, which reach every cortical layer, do not follow this rule but their modulatory function put them apart.

Long-distance corticocortical interactions tend to connect corresponding layers together (L2/3 with L2/3, L5 with L5, L6 with L6) via long fiber tracts running under L6. To be precise, if one cortical area is higher in the hierarchy, the target layers of its projections are slightly shifted toward L1.



Canonical microcircuit
Long-distance corticocortical interactions
Local interactions (laterally & radially)
Long-distance subcortical interactions
L1
L2
L3
L4
L5
L6
Other cortical areas
Fiber tract
Cortical area
Matrix
Core
Thalamus
Other subcortical areas

Simplified illustration

Thiboust, 2020

Long-distance corticocortical interactions transmit signals across hierarchy and create an inter-area coupling

Cortical areas are densely interconnected by **long-distance corticocortical connections**. Lower areas send bottom-up “feedforward” inputs to higher areas and receive top-down “feedback” inputs from those higher areas. Areas of same hierarchical level also interact together.

Inter-area long-distance connections between L2/3 neurons are not particularly reciprocal at the neuron level, meaning that a neuron which projects to another neuron is generally not targeted by the same neuron in return. On the contrary, there are many **long-distance looped interactions between neurons from deep layers L5 and L6** (Young et al, 2019).

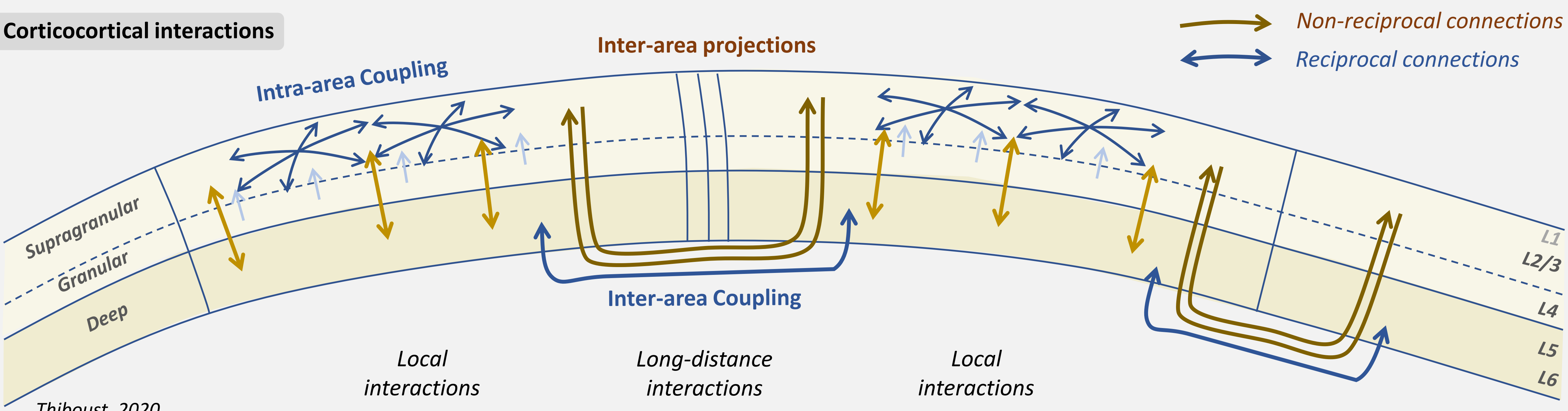
Reciprocal excitatory connections create a strong **coupling**, even between distant cortical areas (those long-distance interactions with myelinated axons can be faster than local interactions with unmyelinated axons).

Schematically, there are two kinds of coupling via reciprocal connections:

- **Inter-area coupling** via **long-distance connections in deep layers**
- **Intra-area coupling** via **local lateral connections in supragranular layers**

Long-distance non-reciprocal connections give clues to other areas without coupling (in both supragranular and deep layers).

Corticocortical interactions



Thiboust, 2020

Virtually all information reaching the cortex passes through the thalamus which routes and gates input signals

The thalamus is the gateway to the neocortex. It routes and gates the inputs it receives from nearly all brain structures.

Virtually all cortical areas receive two main types of thalamocortical projections:

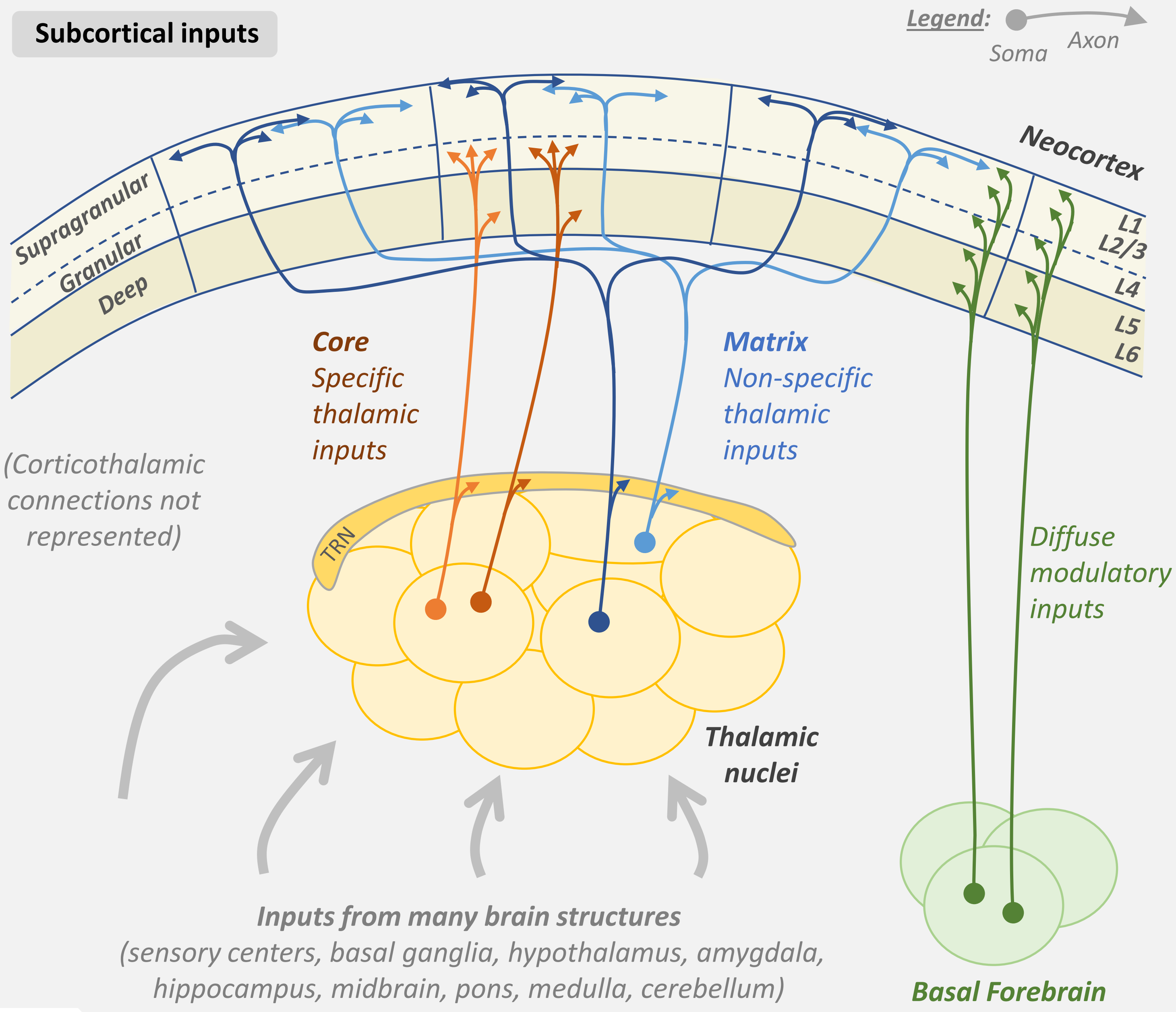
- **Core thalamic neurons** send focal and dense inputs to layer 4 and to layer 6 to a lesser extent. They constitute a significant part of the neuron population in principal sensory thalamic nuclei (ex: *LGN* for some of the visual information) and some other nuclei. Those projections are strong enough to drive vigorously cortical activity.
- **Matrix thalamic neurons** send dispersed modulatory inputs to layer 1 (and to layer 2/3 to a lesser extent). They are distributed in all thalamic nuclei and represent the only type of cortical projections in some nuclei.

In addition to thalamic inputs, all cortical areas also receive direct projections from neurons in the *basal forebrain*. Those diffuse and modulatory projections reach all cortical layers.

Some exceptions:

- The *agranular motor cortex* (that has no layer 4) mainly receives thalamic inputs in layers 1 and 5
- The *piriform cortex* for olfaction (that is an evolutionary ancient cortex with only 3 layers) receives its primary sensory inputs directly from the *olfactory bulb*, bypassing the thalamus.

Subcortical inputs



Legend:
Soma
Axon
Neocortex
Supragranular
Granular
Deep
L1
L2/3
L4
L5
L6
Core
Specific thalamic inputs
Matrix
Non-specific thalamic inputs
(Corticothalamic connections not represented)
TRN
Thalamic nuclei
Diffuse modulatory inputs
Inputs from many brain structures
(sensory centers, basal ganglia, hypothalamus, amygdala, hippocampus, midbrain, pons, medulla, cerebellum)
Basal Forebrain

Thiboust, 2020

Each cortical area projects subcortically to the thalamus, the striatum and motor centers via its deep layers

Cortical neurons that project their axon to subcortical structures are called **corticofugal projection neurons**. They are essentially located in **deep layers (L5 & L6).**

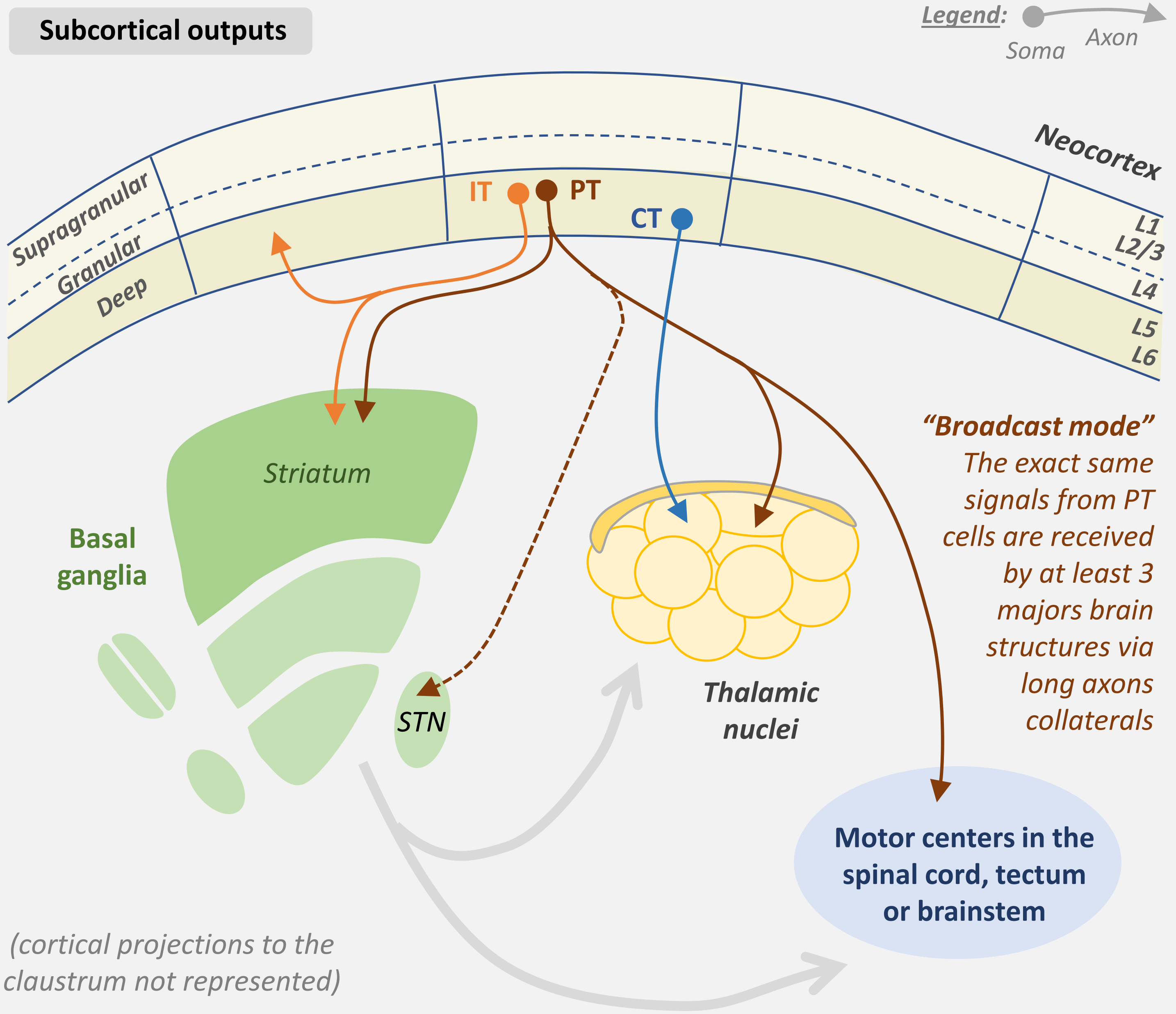
Each cortical area sends 3 types of corticofugal projections:

- **Intratelencephalic (IT) cells in L5 and L6** project to the *striatum* (input structure of the *basal ganglia*) in addition to other cortical areas. Some other cells in L6 project to the *claustrum*, a telencephalic structure whose function is still largely unknown.
- **Pyramidal Tract (PT) cells in L5** project to several subcortical structures via **long axon collaterals** reaching at least the *striatum*, the *thalamus* and one motor center in the *brainstem*, the *spinal cord* or the *tectum*. It sometimes also targets the *subthalamic nucleus (STN).*
- **Cortico-Thalamic (CT) cells in L6** project to the same thalamus nucleus that sends its inputs to L4 & L6

Some exceptions of corticofugal projections that could be interpreted as deviations from the canonical neocortex model (mainly located in evolutionary ancient limbic cortex):

- Projections from L2/3 (to the *striatum* and the *amygdala*)
- Projections to other subcortical structures: *hippocampus, amygdala, septum, hypothalamus, VTA, habenula*

NB: the hippocampus and the pallial amygdala could be seen as primitive cortical areas forming "lateral" connections with the cortex (so not really corticofugal)



"Broadcast mode" *The exact same signals from PT cells are received by at least 3 majors brain structures via long axons collaterals*

(cortical projections to the claustrum not represented)

Thiboust, 2020

With its dense connections into and out of cortex, the thalamus can be viewed as the seventh cortical layer

The thalamus is not only the gateway to the cortex. It is also involved in **thalamocortical loops** that reverberate to the same cortical area and **transmit signals across the cortical hierarchy**.

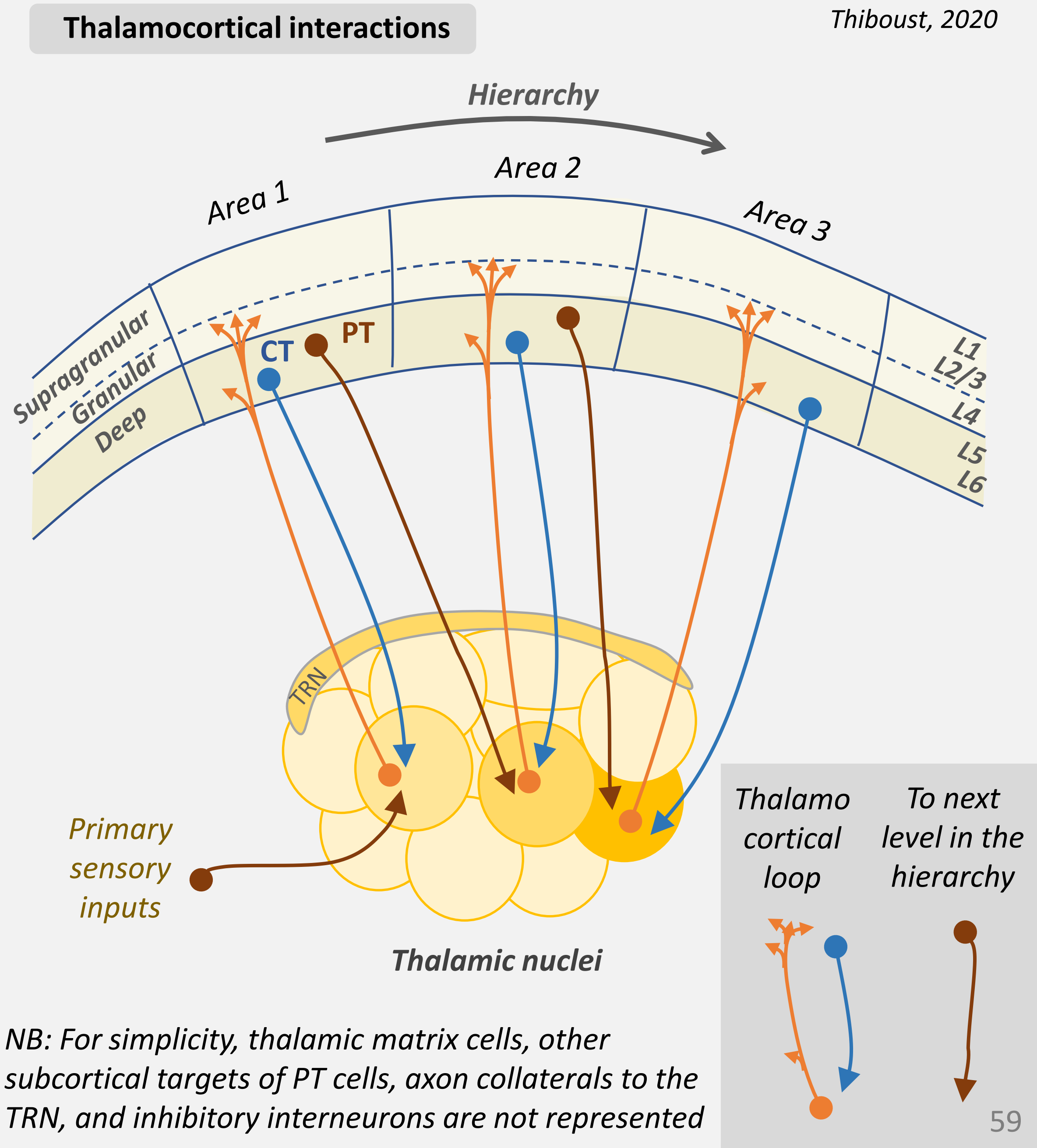
Each cortical area has two ways to send output back to the thalamus:

- **Via axon collaterals of the giant PT cells in layer 5** that project to different thalamic areas that send inputs to the layer 4 of cortical areas higher in the hierarchy. In higher order thalamic nuclei that do not receive any primary sensory stimuli, those projections act the same way as sensory inputs in first order thalamic nuclei (Sherman, 2018). They are reference inputs to be processed by the corresponding cortical area.

  __Speculation__: It is the main feedforward pathway across the cortical hierarchy. Contrary to corticocortical projections that exist in both feedforward and feedback directions between cortical areas, this cortico-thalamo-cortical pathway only exists in the feedforward direction.

- **Via the many CT cells in layer 6** that project back to the same thalamic nucleus that sends input to the layer 4 of this cortical area. This thalamocortical loop is believed to have a modulatory role on L4 inputs (Sherman, 2018) and/or a learning role by comparing predictions from L6 CT cells with reference inputs received by the thalamus (O'Reilly, 2017).

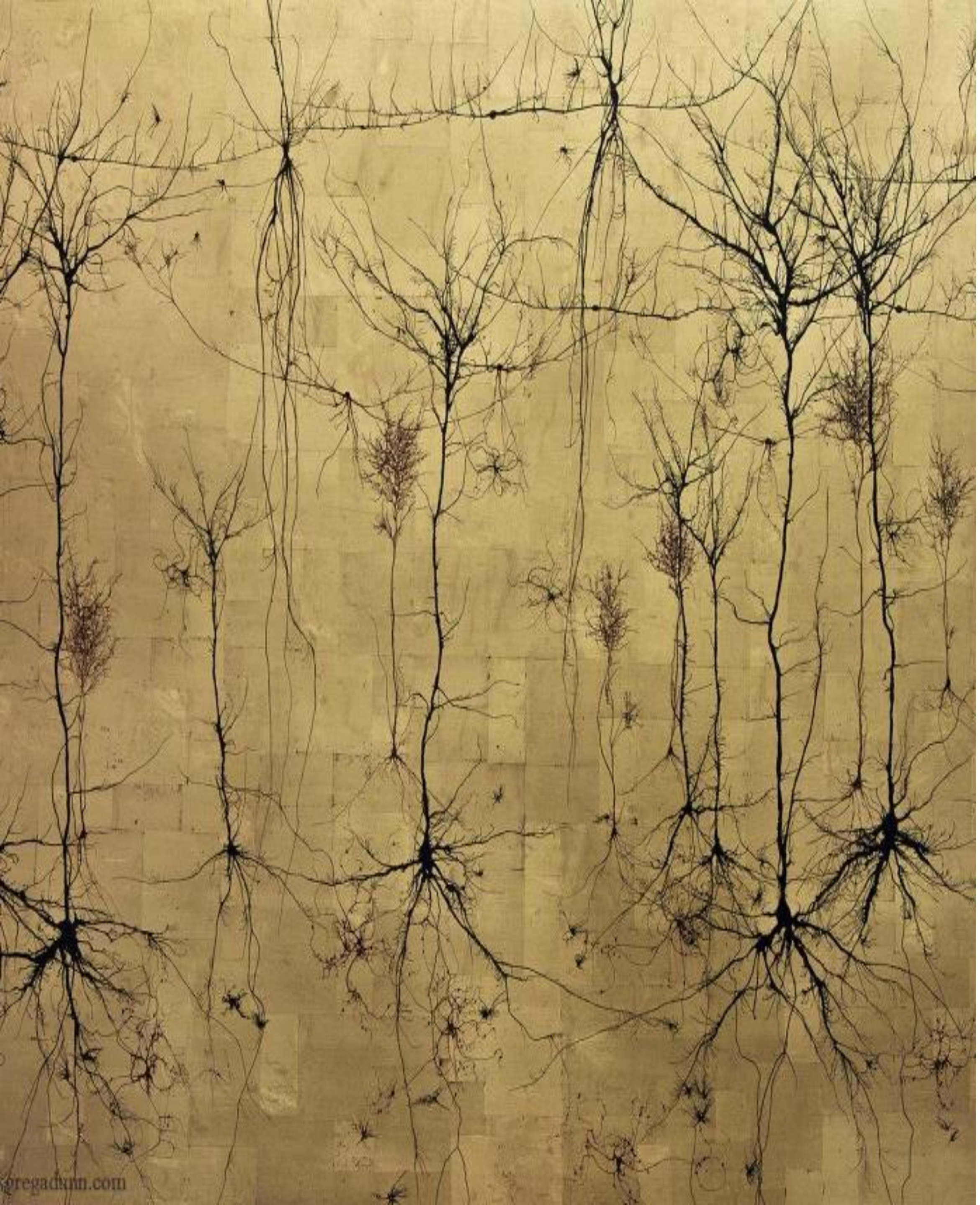
  __Speculation__: Difference between those two signals is interpreted as an error signal. If the error is significant, the thalamus transmits this error signal to L1 via matrix cells (that target apical dendrites in related cortical areas) and amplifies the gain of "ground truth" reference inputs towards L4 to help the cortex to disambiguate.



Thalamocortical interactions

Thiboust, 2020

NB: For simplicity, thalamic matrix cells, other subcortical targets of PT cells, axon collaterals to the TRN, and inhibitory interneurons are not represented



Focus on the neocortex

3. A majority of long-distance projecting pyramidal neurons cohabits with a minority of local inhibitory cells

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- Jeff Hawkins
- Matthew E. Larkum
- Gordon M. Shepherd

See the reference section for a list of materials that inspired me.

Art credit: Gold Cortex, Greg Dunn Design

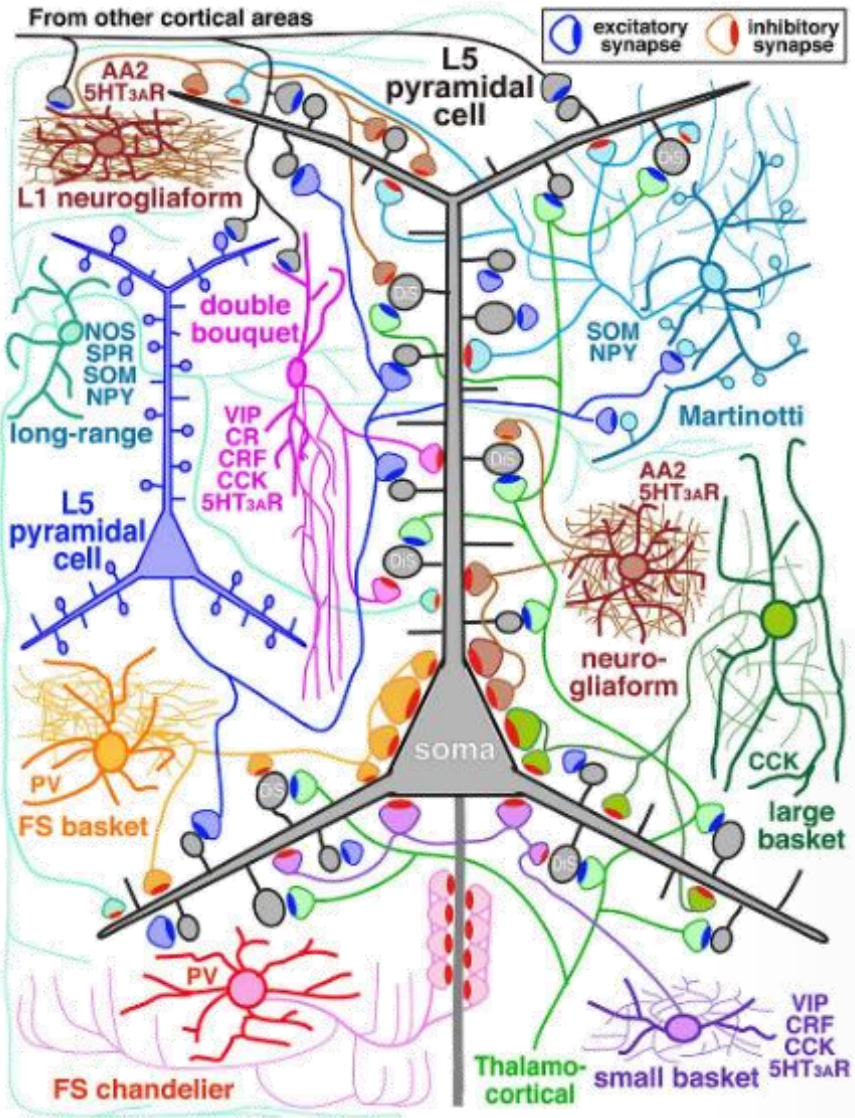
The cortex is mostly populated by excitatory pyramidal neurons surrounded by an ecosystem of inhibitory neurons

Each mm^3 of cerebral cortex contains between 20.000 and 40.000 neurons, of which **85% are excitatory neurons** (75% pyramidal cells, 10% spiny stellate cells) and **only 15% inhibitory neurons**.

The vast majority of cortical excitatory neurons are **pyramidal cells** which have a characteristic apical dendritic arbor and project their axon over long distances to cortical and subcortical targets. They form an extensive network mainly among themselves.

The other neurons are called interneurons because their activity remains local. It comprises excitatory spiny stellate neurons in some layers and a **high diversity of inhibitory neurons** in all layers.

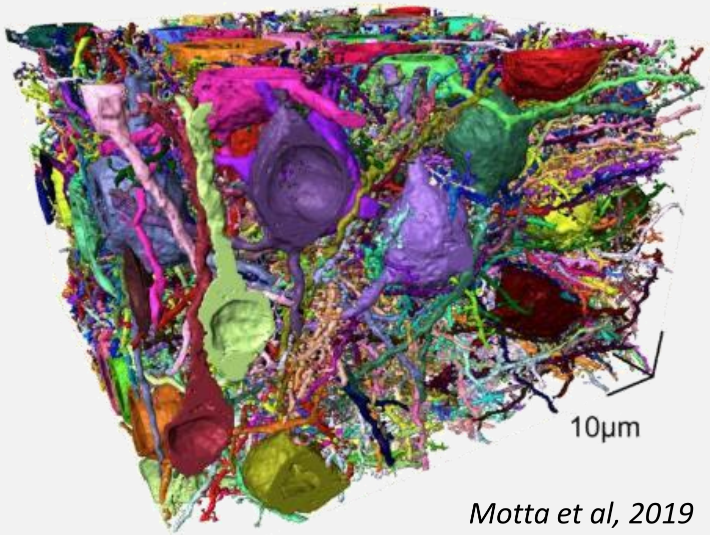
Over a dozen of types of inhibitory neurons populate the cortex. They are loosely interspersed and contribute to about 10% of the synapses on pyramidal neurons



Kubota et al, 2016

Diagram of cortical microcircuit showing pyramidal neurons surrounded by the major subtypes of inhibitory interneurons

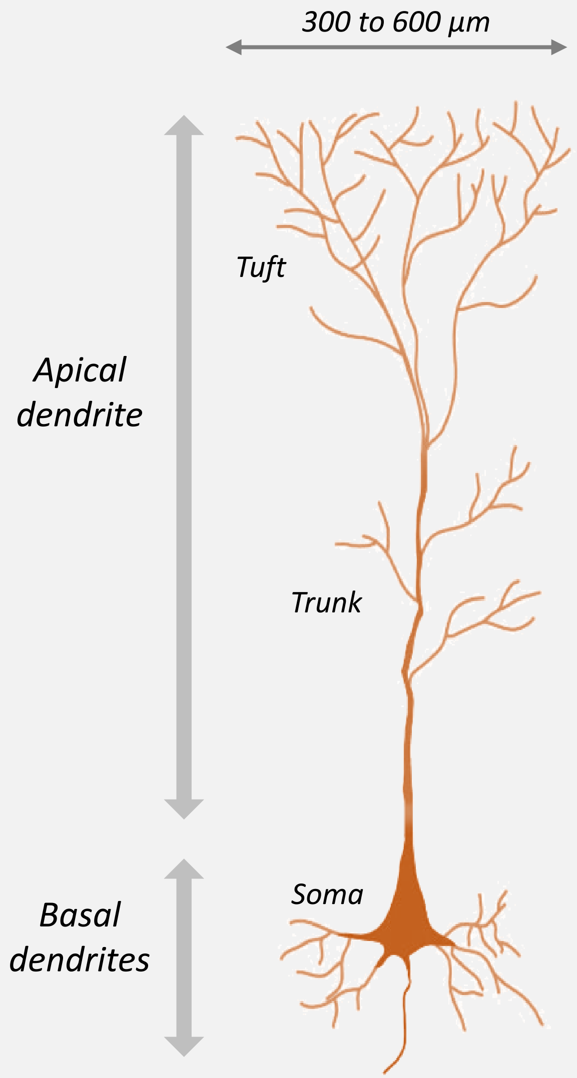
Reconstruction of a portion of L4 somatosensory cortex showing the densely packed neurons



Motta et al, 2019

Pyramidal neurons have a characteristic apical dendritic arbor extending towards the pial surface

Dendrites of a pyramidal neuron



Pyramidal neurons are the typical excitatory neurons of the gray matter of the cerebral cortex. Their name comes from the triangular shape of their cell body.

In addition to the common dendritic arbor around the soma (***basal dendrites***), they have the particularity to have an ascending dendritic branch extending towards the pial surface (***apical dendrite***), contrary to spiny stellate neurons.

The apical dendrite spatially segregates distal inputs from proximal inputs, influences the trigger of *action potentials* (AP) near the soma via *dendritic spikes* (NMDA spikes in the *apical tuft* and calcium spikes along the *apical trunk*), and is believed to play a major role in learning mechanisms because they are heavily targeted by feedback projections.

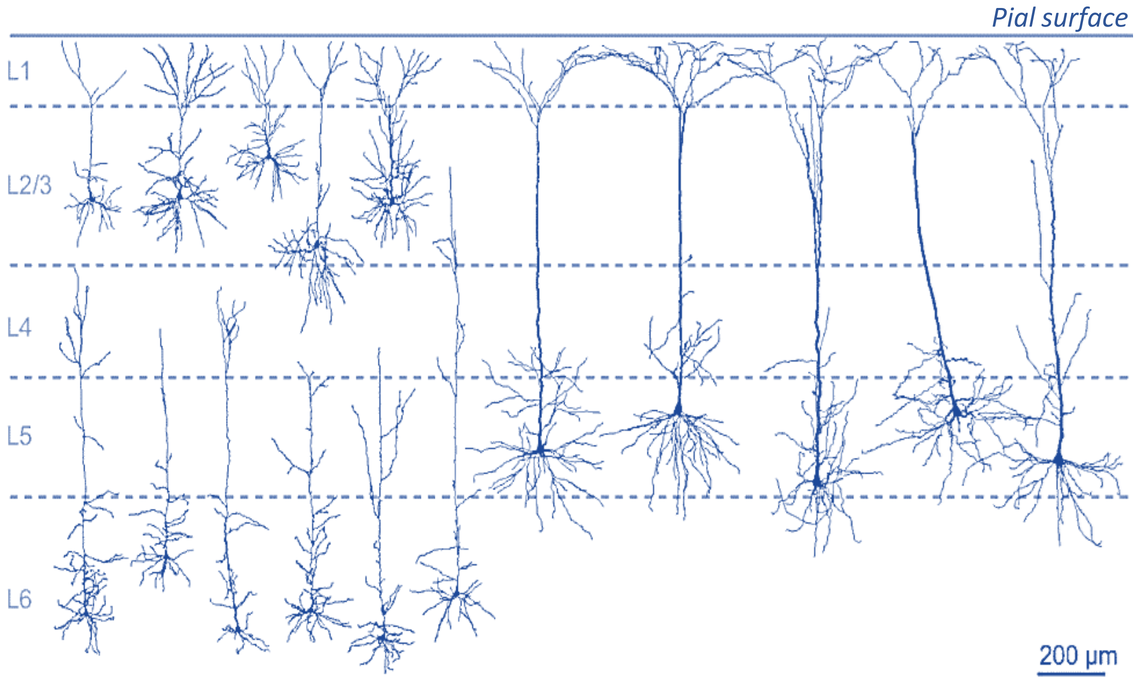
The dendrites of each pyramidal neuron make several tens of thousands of excitatory synapses (around half from local sources, half from remote sources) and a few thousands of inhibitory synapses.

Pyramidal neurons in different layers exhibit considerable diversity in morphologies *(cf figure).*

Most have an apical tuft extending in L1 except for L6 pyramidal neurons whose apical dendrite only ascends to roughly L4 in a focal way (no or small tuft).

The main other difference is in the extend of the tuft. For example, cells in L5 can be thick-tufted or slender-tufted.

Caveat: morphology can misrepresent connectivity.



Ledergerber and Larkum, 2010

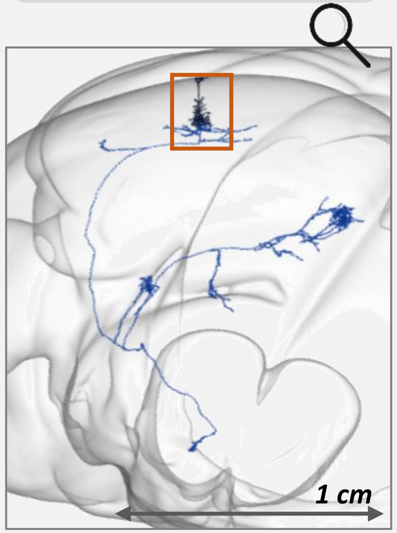
Pyramidal neurons make both short-range and long-distance excitatory connections

Cortical pyramidal neurons are both **short-range and long-distance projecting neurons**. In total, each pyramidal neuron makes several tens of thousands of excitatory synapses (usually none or few synapses with any one of the other neurons).

Its axon descends radially from the soma to the white matter where it joins fiber tracts until arriving to its target. Along the way, this principal axon makes numerous branches called **axon collaterals**:

- Axon collaterals that branch before quitting the cortical area give rise to several **local/intrinsic axonal arbors**: around the soma, laterally in the same layer (ex: L2/3, L5tt and L6cc cells), beneath the soma in deeper layers (ex: L2/3 cells projecting to L5), above the soma in upper layers (ex: L5st cells projecting broadly to L1 and L2/3 in a conic manner, L5tt cells projecting focally to L1, L6cc projecting massively to L3 and L4).
- Axon collaterals that branch after quitting the cortical area give rise to several **distant/extrinsic axonal arbors**: pyramidal neurons project to other cortical areas or to subcortical structures like the thalamus, the striatum, the claustrum, motor centers, or to both cortical and subcortical areas.

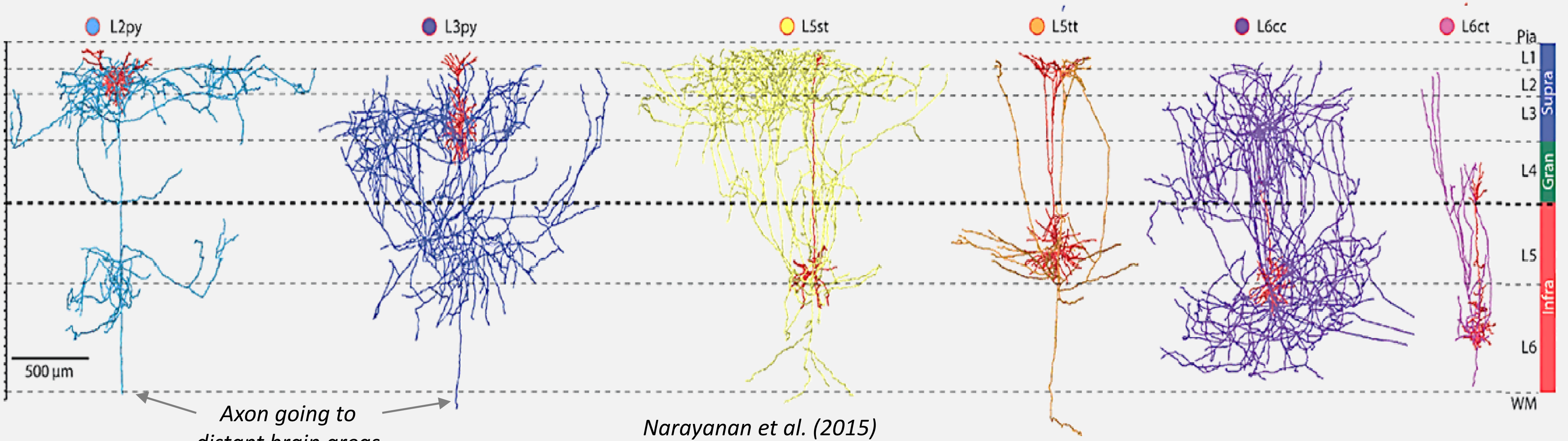
Distant/extrinsic axonal arbors



MouseLight Janelia (neuron AA0949)

Local/intrinsic axonal arbors

Dendrites are in red and axons are in colors



Axon going to distant brain areas

Narayanan et al. (2015)

Image Credit: Axons and Brain Architecture, 1st Edition, ISBN: 978-0-12-801393-9, Figure was adapted from Narayanan et al. (2015). By permission of Oxford University Press

The modulatory power of inhibitory interneurons is greatly expanded by their incredible diversity

In the cerebral cortex, information flows via excitatory neurons that activate other excitatory neurons and so on. This excitatory recurrent chain cannot go on forever, it has to slow down or stop whenever required to keep the network in a functional state. **Inhibitory neurons** allow a balanced cortical activity between excitation and inhibition.

They are often called **inhibitory interneurons** because they only project locally (there exists a few exceptions). If a distant brain structure needs to inhibit a cortical neuron, it has first to excite an inhibitory interneuron that will then inhibit the given neuron in its vicinity.

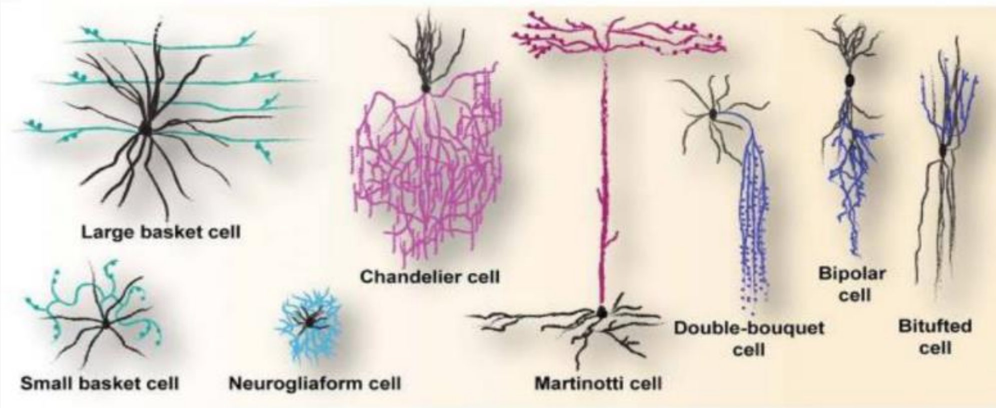
Even if inhibitory interneurons are vastly outnumbered by excitatory projection pyramidal neurons, their modulatory power is greatly expanded by their **incredible diversity in their morphology, the targeted postsynaptic compartment, the selectivity of their connections and their firing patterns**.

They are classified by their morphologies and their molecular marker expression in a dozen of classes, the most prominent being:

- **Basket cells** have a multipolar shape, target exclusively the soma of pyramidal cells and often exhibit fast spiking discharge rates.
- **Martinotti cells** are small multipolar neurons with short branching dendrites and send their axons up to L1 to target the distal tuft of apical dendrites
- **Neurogliaform cells** are small neurons with an unusually high presynaptic bouton density
- **Chandelier cells** have characteristic axon arbors with the terminals forming distinct arrays called "cartridges“ (hence their name). They specifically target the axon initial segment of pyramidal cells, meaning that they inhibit the propagation of Action Potentials, not their generation.

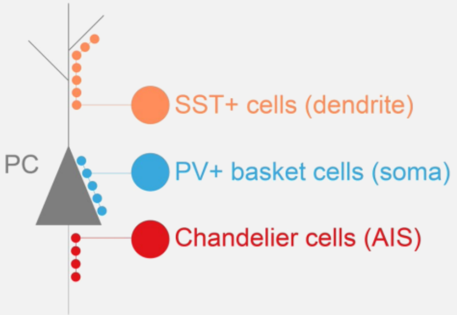
Diverse morphologies

Sultan et al, 2017



Diverse postsynaptic targets

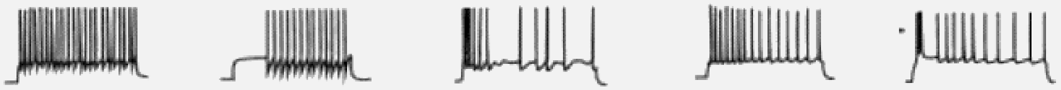
Favuzzi et al, 2019

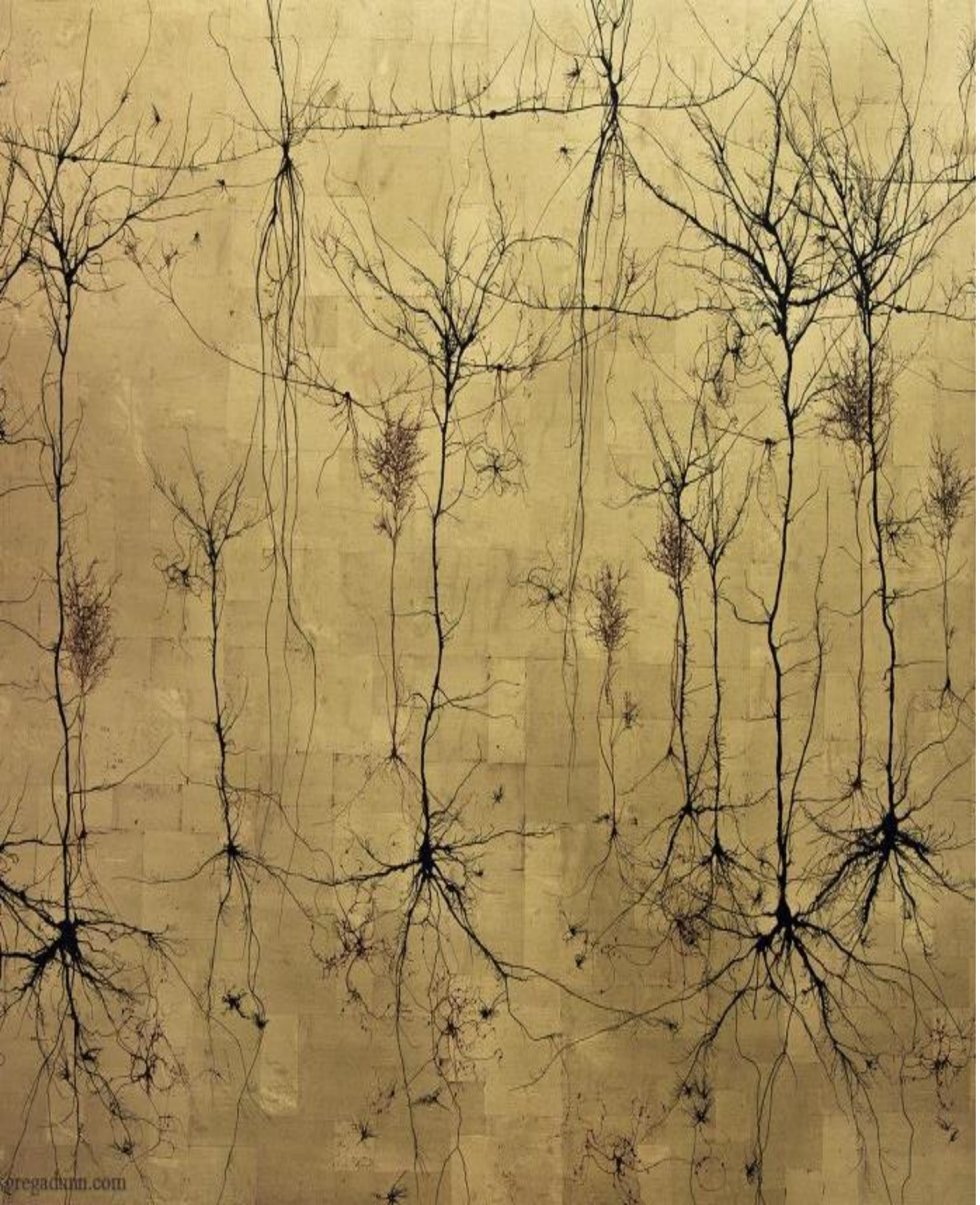


Inhibitory interneurons can specifically target different compartments of pyramidal cells: dendrites (proximal, distal or tuft), soma or axon

Diverse firing patterns

Sultan et al, 2017





Focus on the neocortex

4. Functional neocortical circuits rely on laminar-specific lateral and radial interactions

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- Mark Browne
- William Calvin
- Jeff Hawkins
- Vernon Mountcastle

See the reference section for a list of materials that inspired me.

Art credit: Gold Cortex, Greg Dunn Design

The cortical circuitry relies on laminar-specific radial and lateral interactions at different scale levels

Axonal & dendritic topologies of **excitatory vs inhibitory neurons** matter a lot to understand the cortical circuitry.

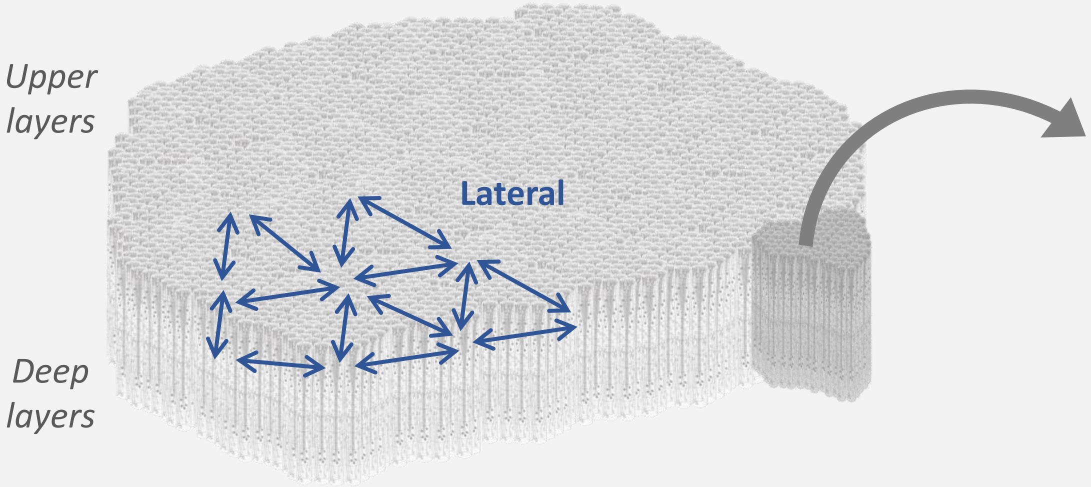
All cortical pyramidal neurons are spatially oriented along a radial axis. They all have an apical dendrite going up towards the pial surface and an axon going down to the white matter. The lateral and translaminar extension of their dendritic and axonal arbors can vary significantly depending on their type.

At the minicolumn level, radial interactions between different layers constitute the main connection pattern. Because of local inhibition by interneurons, lateral interactions occur at a larger scale level: macrocolumn and cortical area.

All connections presented here are intra-area cortical connections. They do not use the fiber tracts in the white matter underneath the cortical plate.

Cortical area level

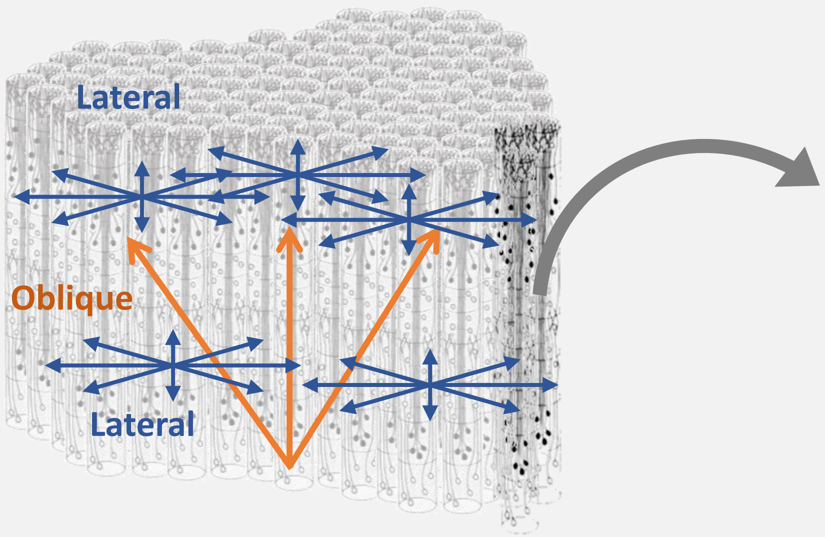
diameter ~ 5 mm
~ 1.000.000 neurons



Lateral interactions in upper layers via reciprocal lateral connections of L2/3 pyramidal cells

Macrocolumn level

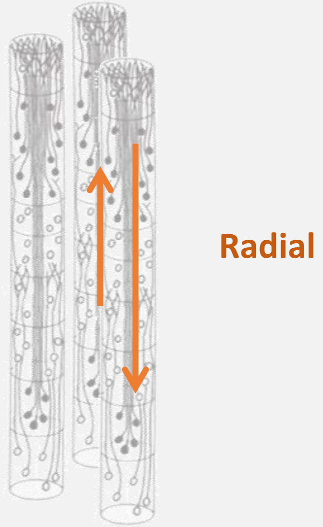
diameter ~ 500 µm
~ 10.000 neurons



Local lateral interactions in upper and deep layers and ascending conical dendritic and axonal arbors

Minicolumns level

diameter ~ 50 µm
~ 100 neurons



Radial interactions between layers

Illustration of a minicolumn from Mountcastle, 1997

Strong radial connections across cortical layers dominate interactions at the level of a few minicolumns

At the level of a minicolumn (or a few minicolumns), excitatory neurons are likely to be connected radially across cortical layers, but not laterally within the same layer. Those **strong radial connections** come from the way the cerebral cortex develops in the embryo: excitatory neurons of the same minicolumn originate from the successive divisions of a progenitor cell that migrates radially in an inside-out manner.

Therefore, a significant fraction of radial interactions originates from neurons with a common developmental lineage:

- From L4 to L2/3
- From L4 to L5
- From L2/3 to L5
- Also probably between L5 and L6

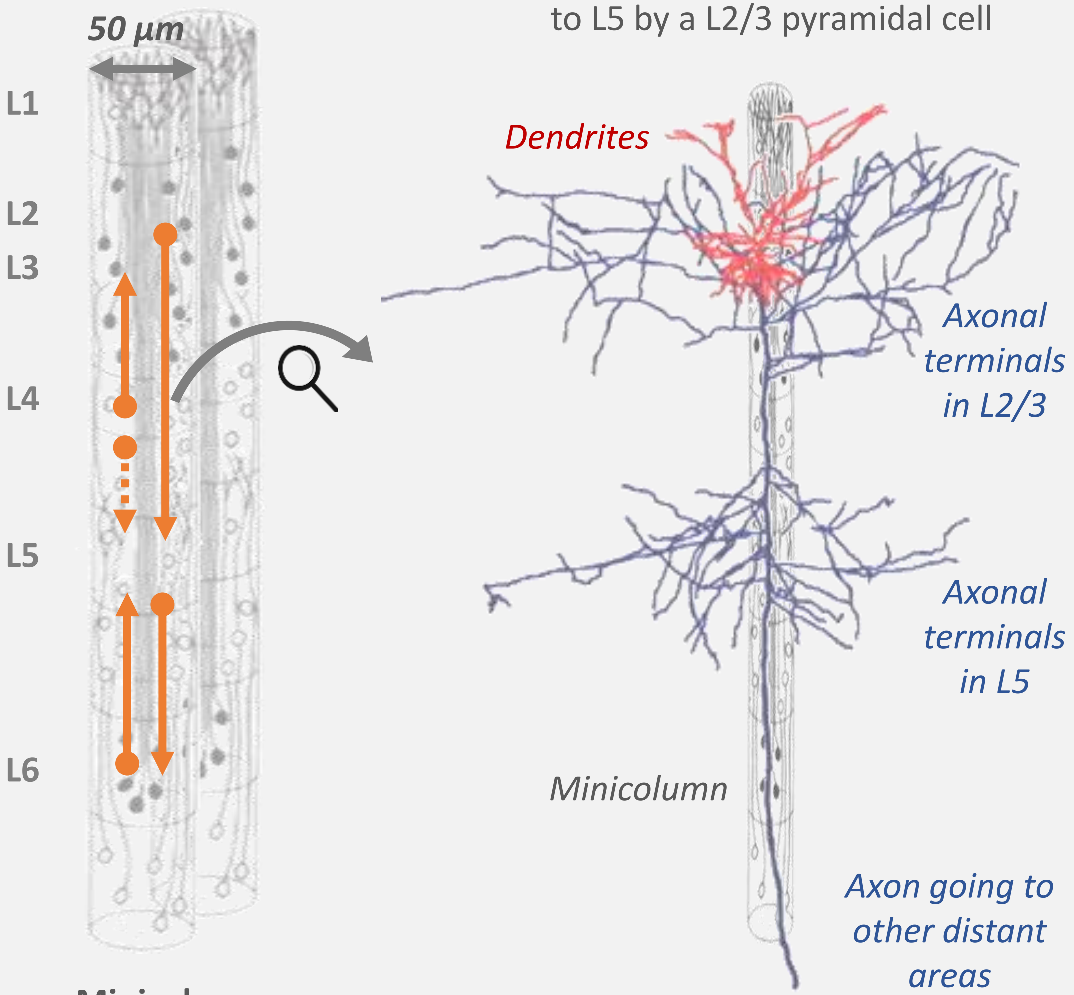
"Integration of vertical input from related neurons within radial units and lateral input from unrelated neurons may represent a developmentally programmed blueprint for the construction of functional neocortical circuits." (Cadwell et al, 2019)

Fundamental cortical unit for radial interactions:

Even if radial connections are enhanced between clonally related neurons, they are also significant between nearby unrelated pairs. It would be more accurate to say that **the fundamental unit for radial interactions** corresponds more to **a few nearby minicolumns** than a single minicolumn.

Radial interactions

Example of connections from L2/3 to L5 by a L2/3 pyramidal cell



Minicolumn
(or a few minicolumns)
~ 100 neurons

Thiboust, 2020

L2/3 neuron reconstruction from Tanaka et al, 2011

The complex connectivity patterns of macrocolumns support a subtle excitatory vs inhibitory dynamics

Macrocolumns are ensembles of minicolumns that share a similar receptive field from thalamic input in L4. Anatomically, they can be discrete (barrels in mouse somatosensory cortex) or continuous (orientation columns in primary visual cortex V1).

At this scale level, **spatial topography and cell type determination matter a lot to uncover the cortical connectivity patterns** (still not yet fully understood).

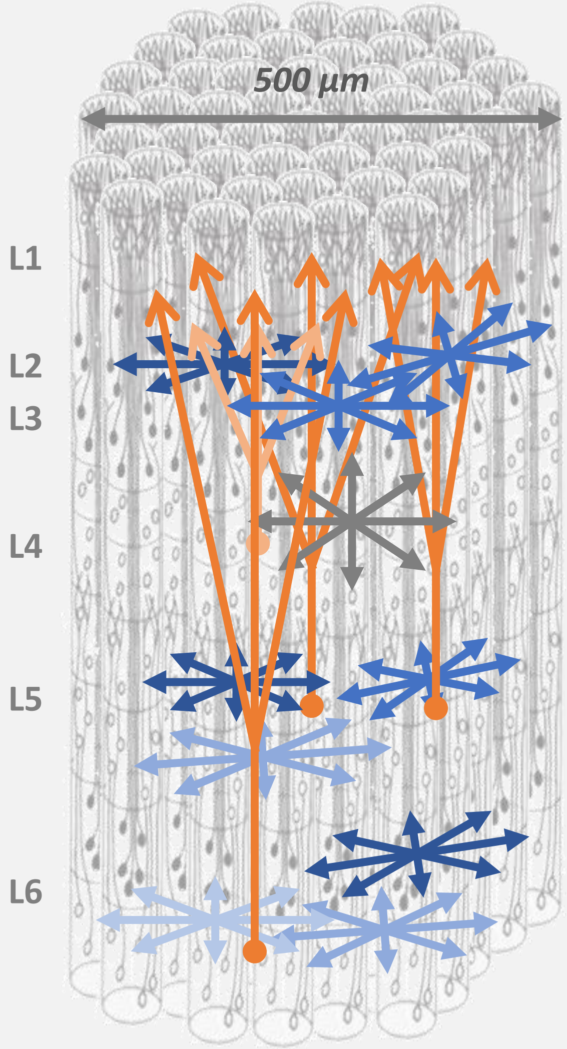
Most cortical neurons make **lateral reciprocal connections** via a local axonal arbor surrounding their soma (mainly in L2/3, L5 and L6): activation can propagate laterally step by step between macrocolumns to form global activation patterns that mutually reinforce themselves. In addition to those connections, pyramidal neurons also make distal connections via their ascending conical dendritic and axonal arbors. Because of the conical shape, input diversity is broader in upper layers and more focal in lower layers.

In every layer, inhibitory interneurons enforce a competition for activation between excitatory neurons: the first excitatory neuron to fire inhibits its neighbors. This dynamics is called **Winner-Take-All (WTA) competition**. It is an important computational principle in the brain: depending on the network parameters, it can achieve ramp-up computations, decision-making or sustained activity (not exhaustive).

The 3 main excitatory neuron types in the cerebral cortex (IT, PT, CT) show a laminar and cell type specificity in their local connectivity patterns:

- IT cells synapse with all types of cells (in L2/3, L4, L5 and L6)
- PT cells preferentially synapse with PT cells (in L5)
- CT cells preferentially synapse with CT cells (in L6)

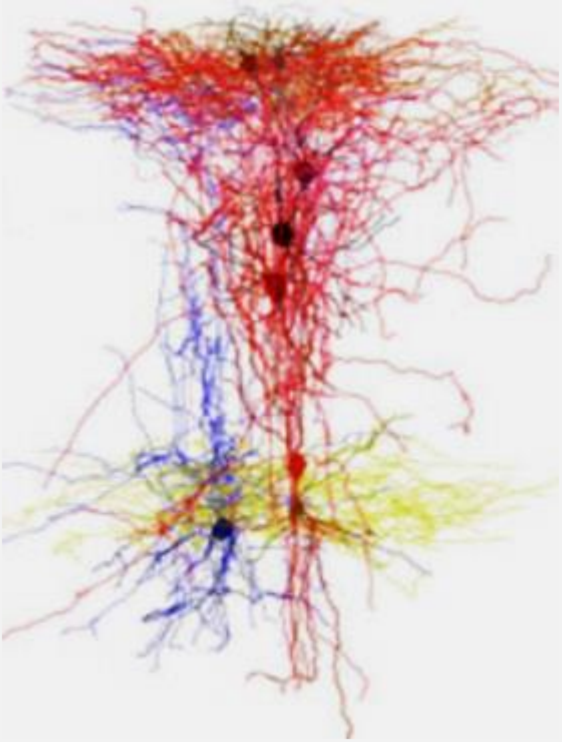
Lateral and conical connections *(far from exhaustive representation)*



Macrocolumn
~ 10.000 neurons

Thiboust, 2020

Reconstruction of 8 cortical neurons
(2 excitatory, 6 interneurons)



Neurons reconstruction from Jiang et al, 2015

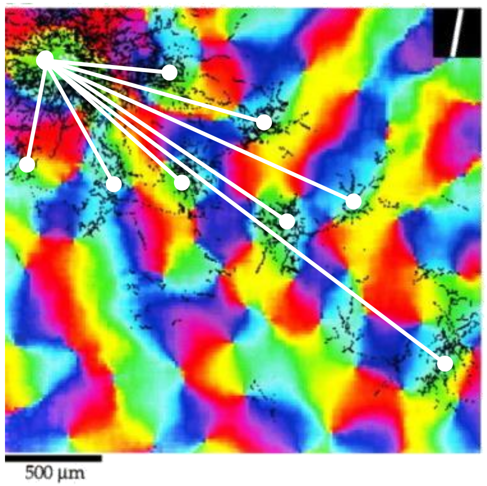
Inside a cortical area, macrocolumns are coupled via strong recurrent connections in superficial layers

Contrary to other cortical neurons, **lateral connections from L2/3 excitatory pyramidal cells** can extend up to a few millimeters in several lateral directions. They form a **strong recurrent network able to propagate and sustain activity inside a cortical area**.

Synapses are strongly clustered along the very elaborated axonal arbor of L2/3 pyramidal cells. In V1, we can differentiate three kinds of clusters (possibly not a universal characteristic):

- One large local axonal cluster surrounding their soma in L2/3
- One distal radial axonal cluster in L5 underneath their soma
- Several distal lateral axonal clusters in L2/3

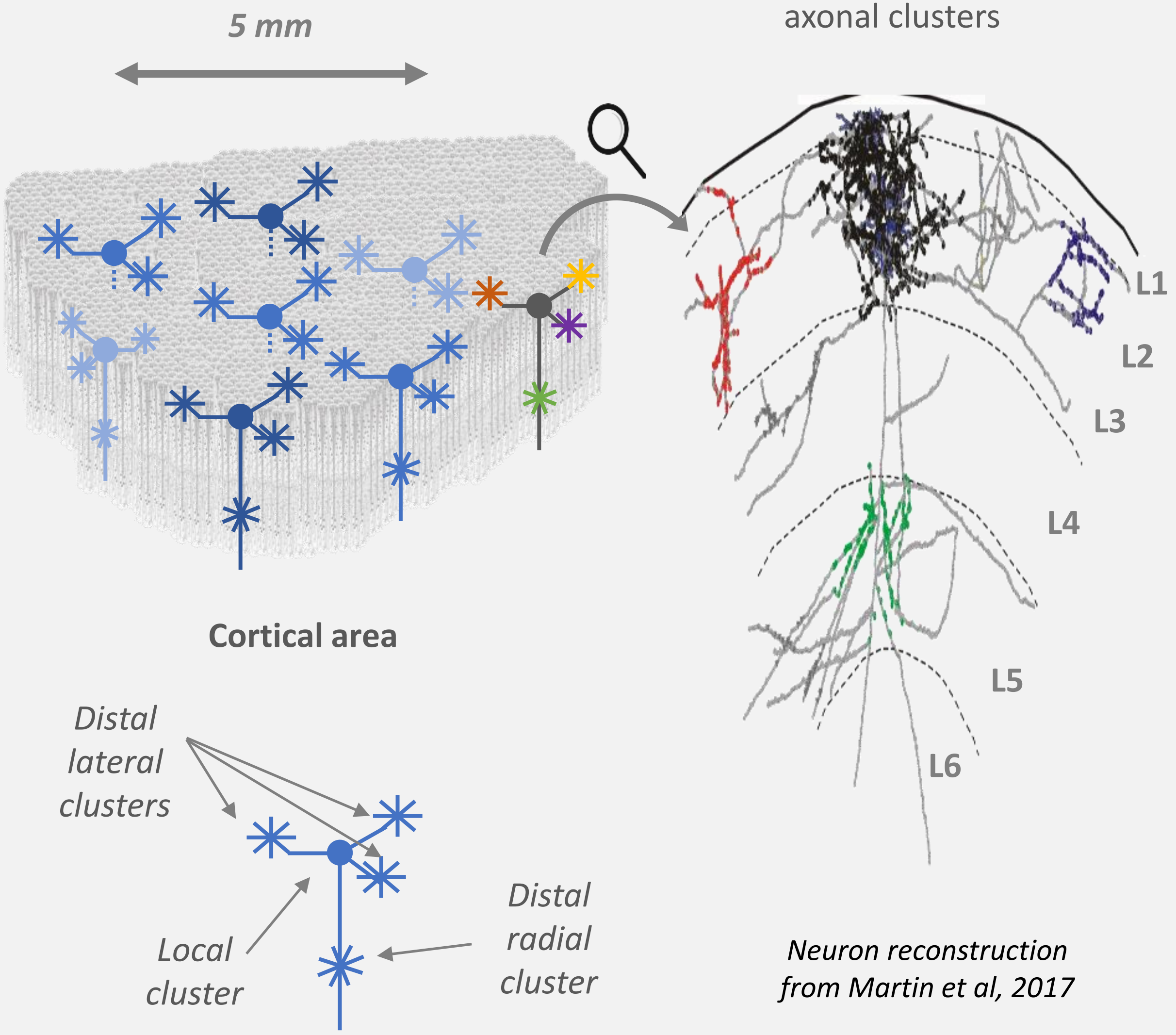
The first two clusters make connections at a macrocolumn level, while the distal lateral clusters connect with other macrocolumns (not necessarily direct neighboring macrocolumns).



Connections between orientation columns in V1 are an illustration of those connections via distal lateral clusters in L2/3 *(see next chapter for orientation columns).*

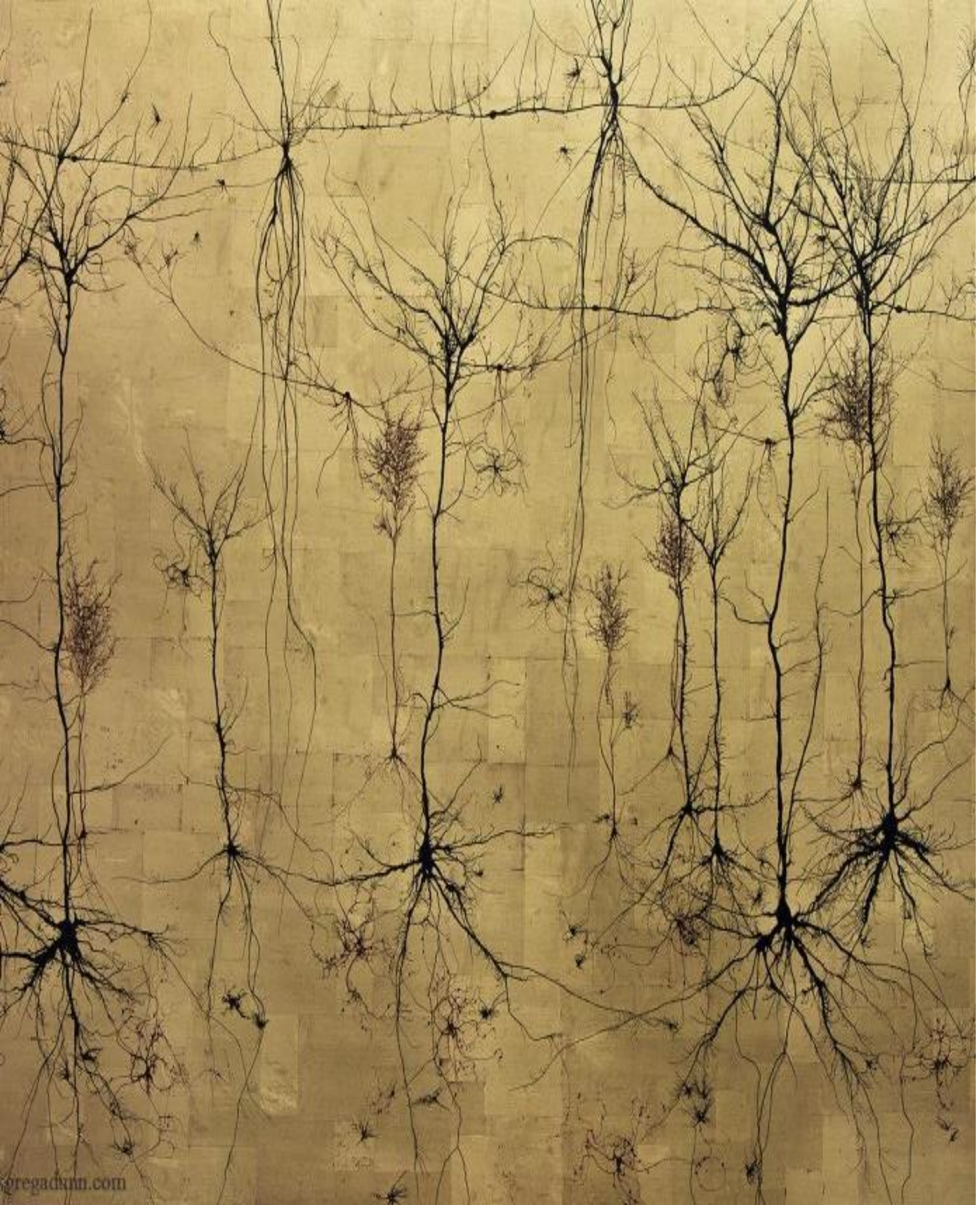
Synaptic boutons distribution (in black) from axons of L2/3 pyramidal neurons of a column associated with an 80° orientation. (Bosking et al, 1997)

Lateral recurrent connections



Neuron reconstruction from Martin et al, 2017

Thiboust, 2020



Focus on the neocortex

5. Sensory stimuli, motor actions and spatial navigation offer a window into the cortical code

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- György Buzsáki
- David Hubel
- Edvard Moser
- May-Britt Moser
- Torsten Wiesel

See the reference section for a list of materials that inspired me.

Art credit: Gold Cortex, Greg Dunn Design

Neural correlates of sensory stimuli and spatial location inform us on cortical internal representations

Because they offer a privileged window into internal neural activity, **neural correlates** are largely studied in neuroscience experiments recording from single neurons in vivo (generally in mice, cats or monkeys).

In particular, **neural responsiveness to sensory stimuli in primary sensory cortices and to spatial location in the hippocampal complex & entorhinal cortex** inform us on how specific brain circuits form internal representations. In those cases, firing patterns of *orientation-selective cells, head direction cells, place cells and grid cells* are strongly correlated with the examined external variable.

However, neural responses are often only partially correlated with the examined external variable, making the interpretation more complex. **Brain states, contexts, goals and/or other parallel tasks can modulate the neural response**.

<u>*Going further:*</u>

*New approaches are even able to **reveal neural correlates of behavior without behavior measurement** (in hippocampus and prefrontal cortex, for behaviors such as moving along a linear track, turning and drinking as a reward) by measuring and correlating internal structure of neuronal activity with internal representations. Surprisingly, the measured internal structure was conserved across mice, allowing using one animal's data to decode another animal's behavior (Rubin, 2019).*

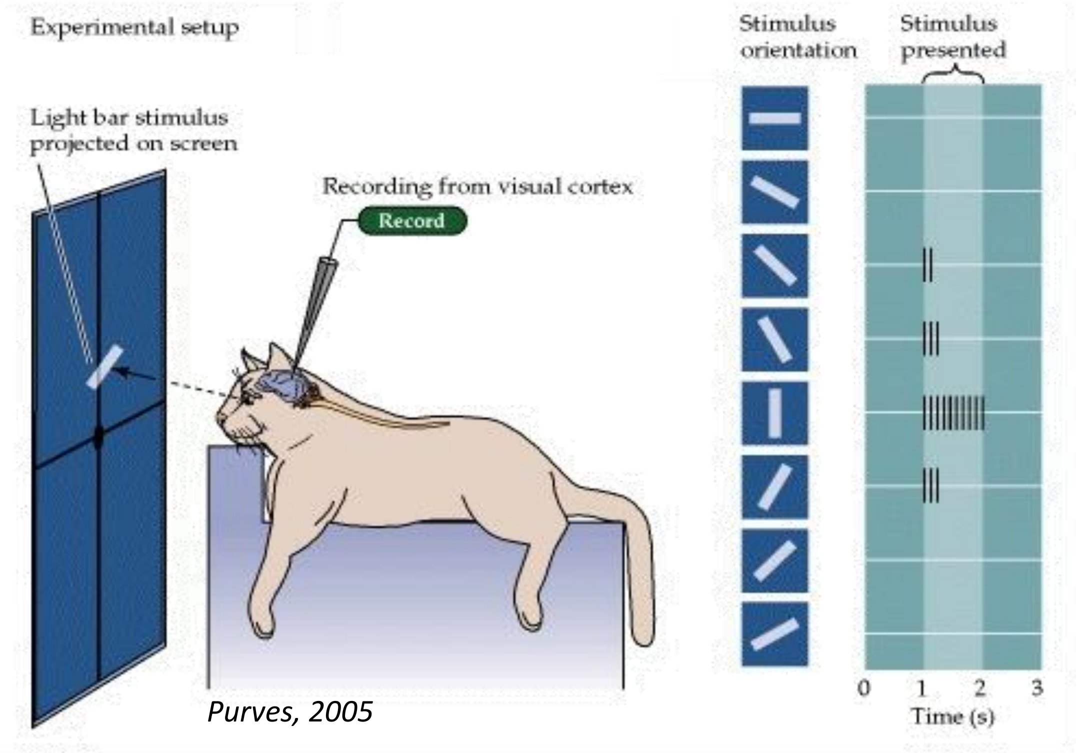
Examples of neural correlates

Single neuron activity		External variable
Orientation-selective cell *in V1*	⟷	*Orientation of visual lines*
Head direction cell *in the subiculum*	⟷	*Allocentric direction of the animal's head*
Place cell *In the hippocampus*	⟷	*Allocentric position*
Border cell *In the subiculum*	⟷	*Allocentric position*
Grid cell *In the entorhinal cortex*	⟷	*Allocentric position on an hexagonal lattice*
Speed cell *In the entorhinal cortex*	⟷	*Instantaneous allocentric speed*
"Velocity" cell *In the sensorimotor striatum (not the cortex)*	⟷	*Instantaneous targeted velocity during prey hunting with egocentric direction of motion*
…	⟷	**…**

Orientation-selective cells in the primary visual cortex correlate with the orientation of visual stimuli

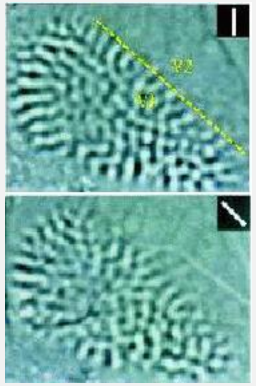
Orientation-selective cells are neurons that increase their firing rate for specific angles of visual line stimuli. They are direct neural correlates of external stimuli.

They are found in multiple layers of the *primary visual cortex (V1)* of humans, primates, cats (but not mice), and are organized in *orientation columns* that group cells of the same orientation selectivity.



Purves, 2005

Firing patterns



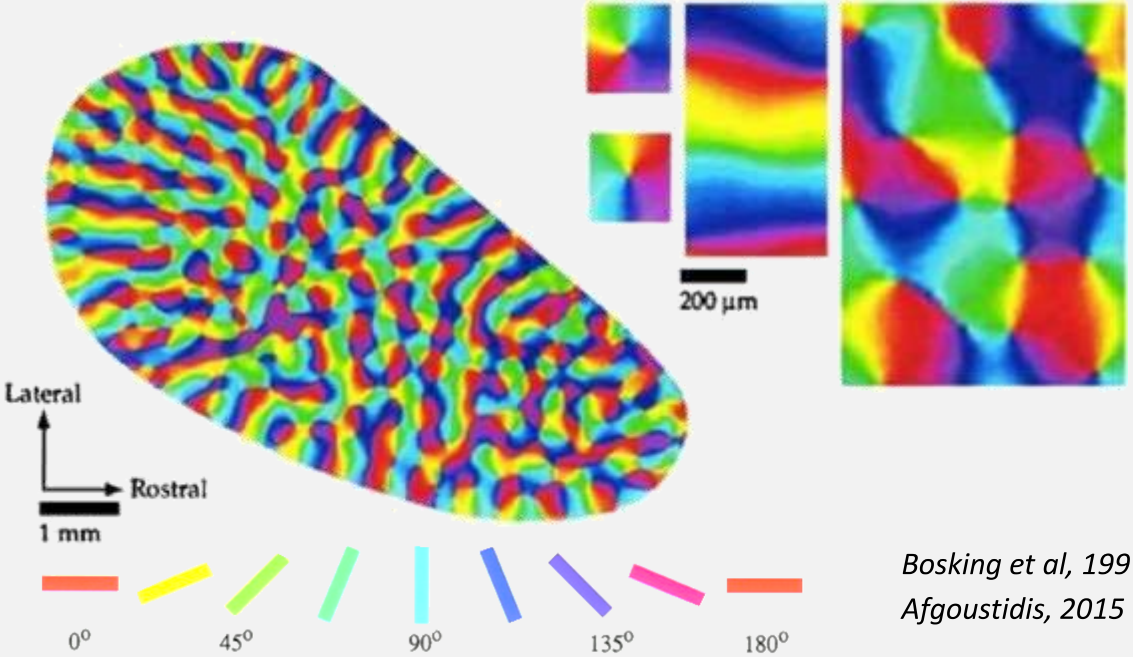
Bosking et al, 1997

Example of recordings of average firing activity in V1 when the animal is shown vertical and obliquely oriented visual lines (areas with great activity are in dark).

Each black area represents the average activity of thousands of cells that form an orientation column.

This orientation selectivity disappears in neighboring cortical areas like V2.

Map of V1 orientation columns



Bosking et al, 1997
Afgoustidis, 2015

Head direction cells in diverse cortical and subcortical areas correlate with allocentric head direction

Head direction (HD) cells are neurons that **increase their firing rates** above baseline levels **when the head of an awake animal points in a specific direction**, whatever its location.

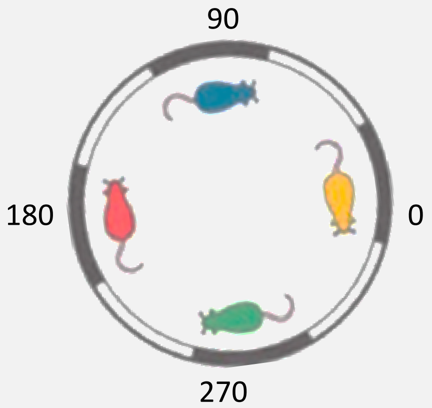
Each cell has only one direction in which it fires maximally. This direction is said to be allocentric because it is anchored to its surrounding environment as its reference frame (depend on landmarks and self-motion cues). In a given familiar environment, their firing remains stable during days and even months

Their firing is primarily independent of the animal's on-going behavior.

They are found in many interconnected brain areas:

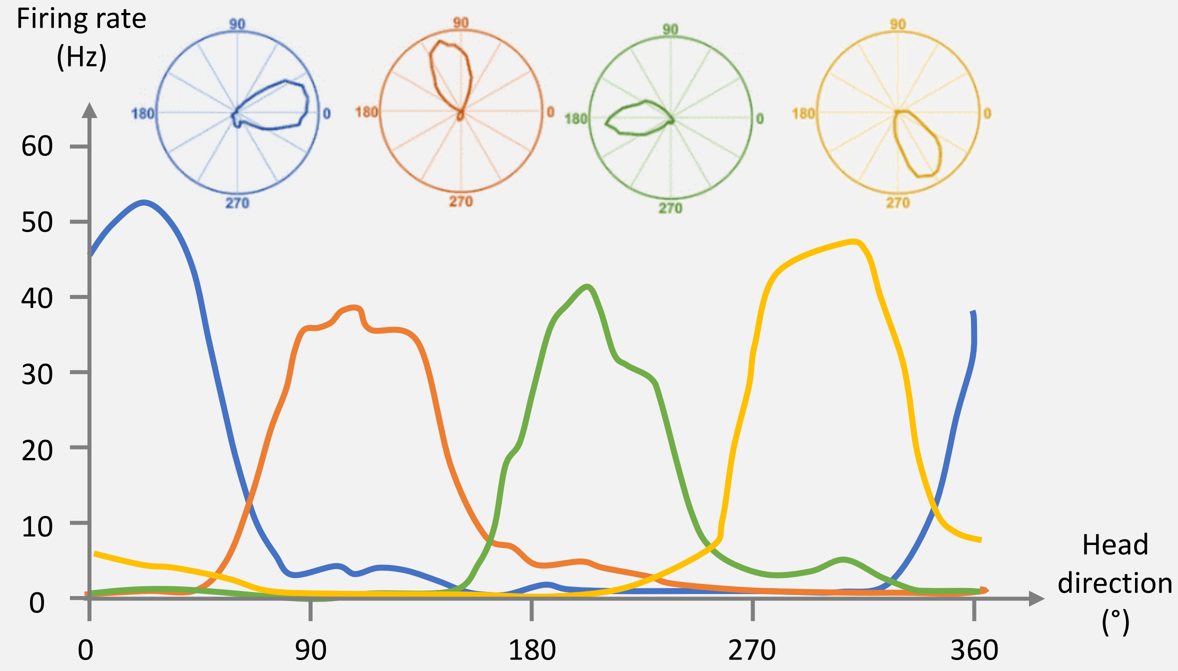
- Cortical areas: *postsubiculum, retrosplenial cortex, entorhinal cortex*
- Subcortical areas: *thalamus (anterior dorsal and the lateral dorsal thalamic nuclei), lateral mammillary nucleus, dorsal tegmental nucleus and striatum*

Firing patterns



Example of recordings from 4 differently tuned head direction cells. The blue curve corresponds to a cell that fires when the animal's head points to the East in this environment (arbitrarily referenced by 0°).

They are often represented by visual friendly polar plots (equivalent to the curve plots).



Adapted from Page, 2017 and Sharp 2001

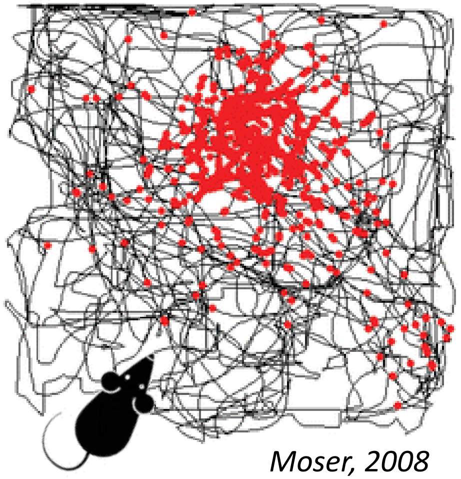
Place cells in the hippocampus correlate with the allocentric spatial location

Place cells are neurons that **fire at a high rate whenever the animal is in a specific location in the environment**, called the **place field**.

Contrary to head direction cells, they are **location-specific** and **orientation-invariant**. A large population of place cells can provide a reliable map and faithfully track the animal's allocentric position in the environment by relying on landmarks and self-motion cues.

However, this map is not static. If the animal is placed in a different environment, a different set of place cells becomes active. Neighboring place fields of two place cells in one environment can be very different in another environment. More, place fields change even when the animal visits the same environment at different times (*remapping*).

Place cells are found in *hippocampus*. The size of their place fields increases along the dorsoventral axis.



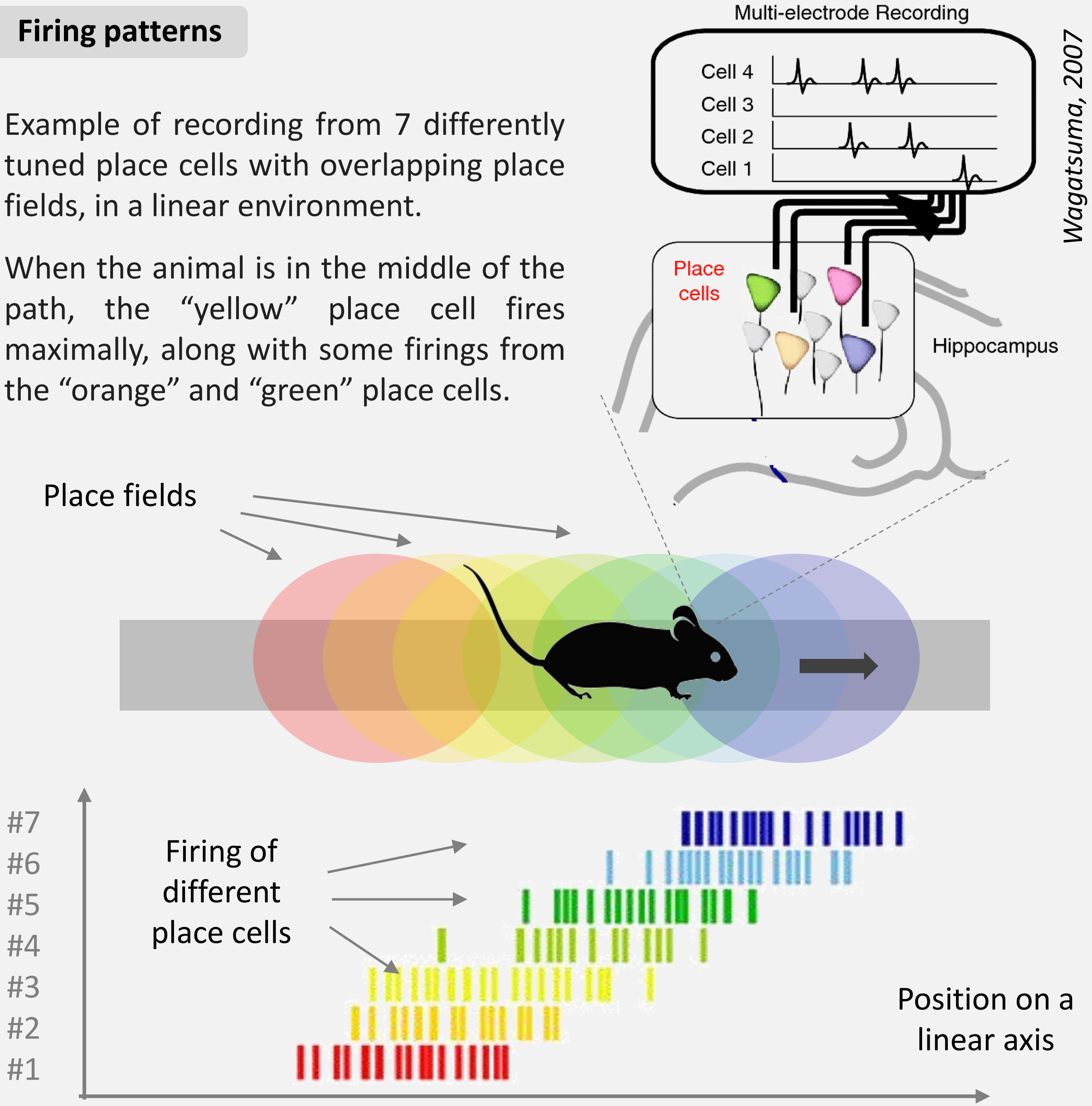
Moser, 2008

Place field (red area) of a place cell in a square room environment. The position of the animal is recorded along with the firing of a place cell during a few minutes. Each red dot correspond to a location that coincides with a firing. The black line is the full recorded track of the animal.

Firing patterns

Example of recording from 7 differently tuned place cells with overlapping place fields, in a linear environment.

When the animal is in the middle of the path, the "yellow" place cell fires maximally, along with some firings from the "orange" and "green" place cells.



Multi-electrode Recording
Cell 4
Cell 3
Cell 2
Cell 1
Place cells
Hippocampus
Place fields
#7
#6
#5
#4
#3
#2
#1
Firing of different place cells
Position on a linear axis

Wagatsuma, 2007

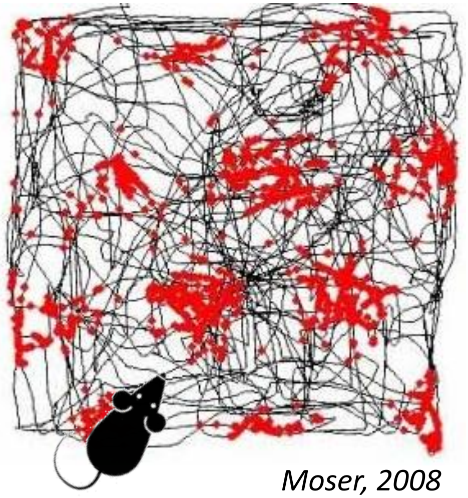
Grid cells in the entorhinal cortex correlate with the allocentric spatial location on an hexagonal lattice

Grid cells are neurons whose **multiple firing locations define a periodic hexagonal lattice** covering the entire available surface of an open two-dimensional environment.

This allocentric neural representation of space and location differs from our intuitive cartesian coordinates system. This neural metric may be a general representation for cognitive map encoding knowledge, not just spatial navigation (Behrens et al, 2018).

Neighboring grid cells have stable similar patterns, with only a slight spatial offset (phase). They are organized in discrete *modules* that group grid cells of same scaling and orientation.

Grid cells are found in the *medial entorhinal cortex (mEC)* which is a part of the *hippocampal complex*. The scaling of the grid increases along the dorsoventral axis.

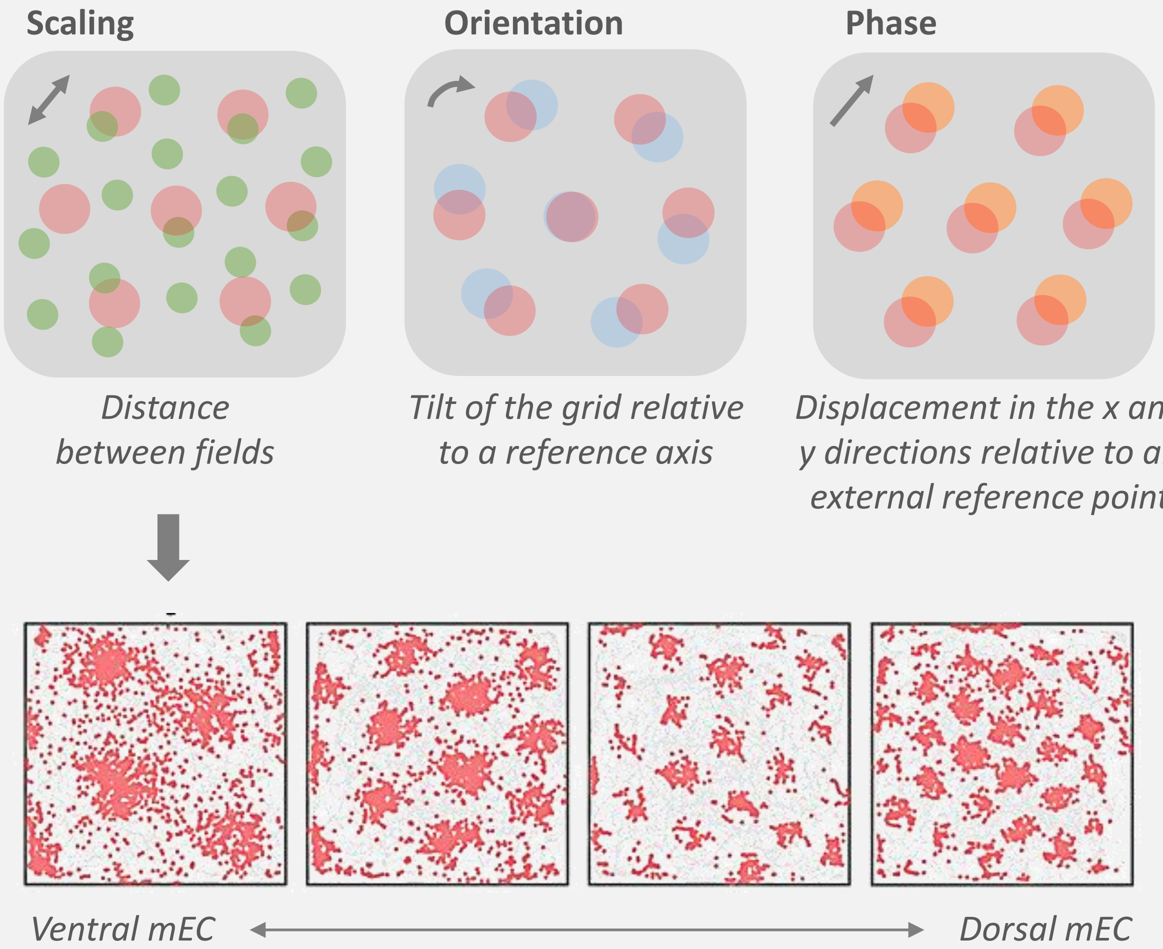


Moser, 2008

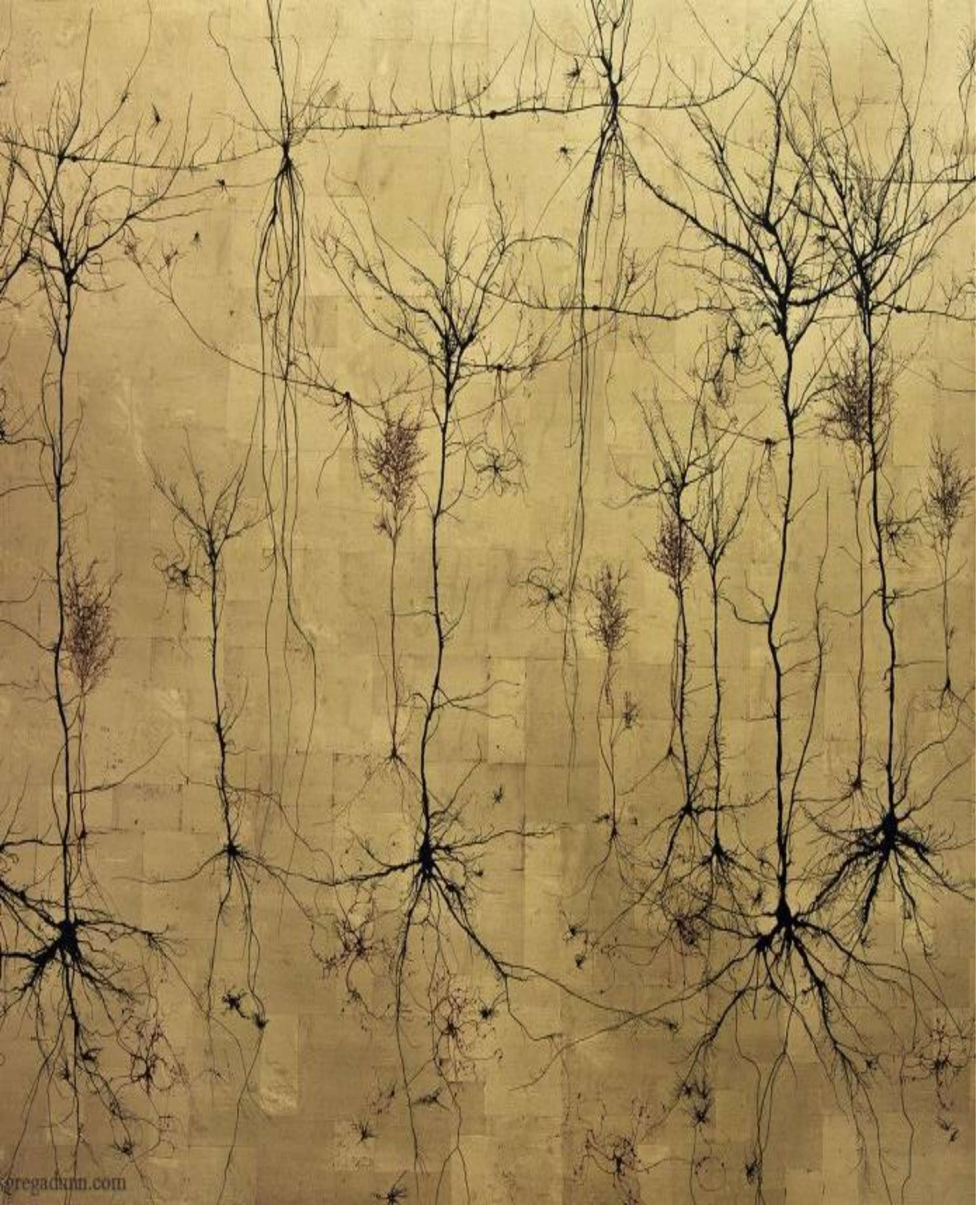
Firing of a grid cell in a square room environment. The position of the animal is recorded along with the firing of a grid cell during a few minutes. Each red dot corresponds to a location that coincides with a firing. The black line is the recorded track of the animal.

Firing patterns

The position of the animal is encoded by the simultaneous firing of multiple grid cells of different scaling, orientation and phase.



Scaling
Orientation
Phase
Distance between fields
Tilt of the grid relative to a reference axis
Displacement in the x and y directions relative to an external reference point
Ventral mEC
Dorsal mEC



Focus on the neocortex

6. The dynamics of cortical activity can only be analyzed in relation to brain oscillations

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- György Buzsáki
- Ole Jensen

See the reference section for a list of materials that inspired me.

Art credit: Gold Cortex, Greg Dunn Design

Macroscopic oscillations reveal synchronized activity of neural ensembles at different frequencies

Neural activity is made of rhythmic patterns of various frequencies called **neural oscillations** or **brain waves**.

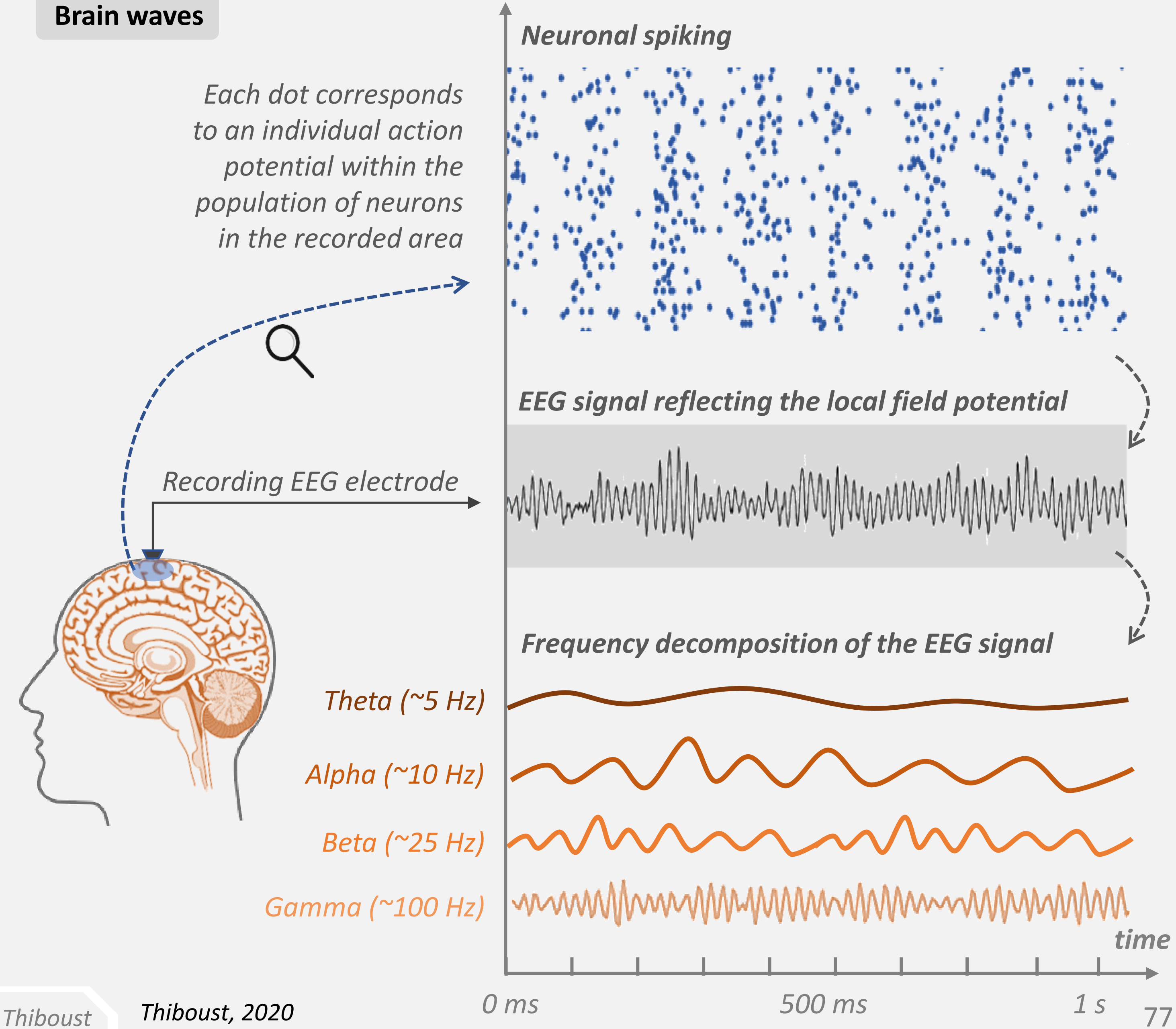
These dynamics result from repetitive firings of individual neurons and from recurrent/feedback interactions between neurons. At the level of neural ensembles, synchronized activity of large numbers of neurons gives rise to macroscopic oscillations, which can be observed with non invasive methods like electroencephalography (EEG) or magnetoencephalography (MEG).

Recorded signals reveal oscillatory activity in specific frequency bands. The best-known rhythm is the ***alpha*** activity between 8 and 12 Hz. It is often accompanied by ***delta*** (1-4 Hz), ***theta*** (4-8 Hz), **beta** (13-30 Hz), ***low gamma*** (30-70 Hz) and ***high gamma*** (70-150 Hz) activity.

Most of these oscillations have been linked to cognitive states and/or functions. For example, strong alpha waves are observed in the occipital lobe during wakeful relaxation with closed eyes, but they are weak with open eyes or during sleep. Beta activity briefly appears after the execution of a movement. High gamma is thought to be involved in communication between cortical areas.

Some fast oscillations can be nested within slow oscillations. This is commonly observed in the cerebral cortex with **fast gamma activity nested within alpha or theta activity**.

Brain waves



Neuronal spiking
Each dot corresponds to an individual action potential within the population of neurons in the recorded area
EEG signal reflecting the local field potential
Recording EEG electrode
Frequency decomposition of the EEG signal
Theta (~5 Hz)
Alpha (~10 Hz)
Beta (~25 Hz)
Gamma (~100 Hz)
time
0 ms
500 ms
1 s

Thiboust, 2020

Cortical dynamics consists in periodic ~100 ms alpha cycles modulating gamma oscillations in specific layers

Alpha and gamma oscillations constitute the main oscillatory activity in the neocortex. Gamma waves are thought to reflect functional cortical processing, while alpha waves may produce functional active inhibition to suppress the processing of distracting information (Bonnefond, 2013).

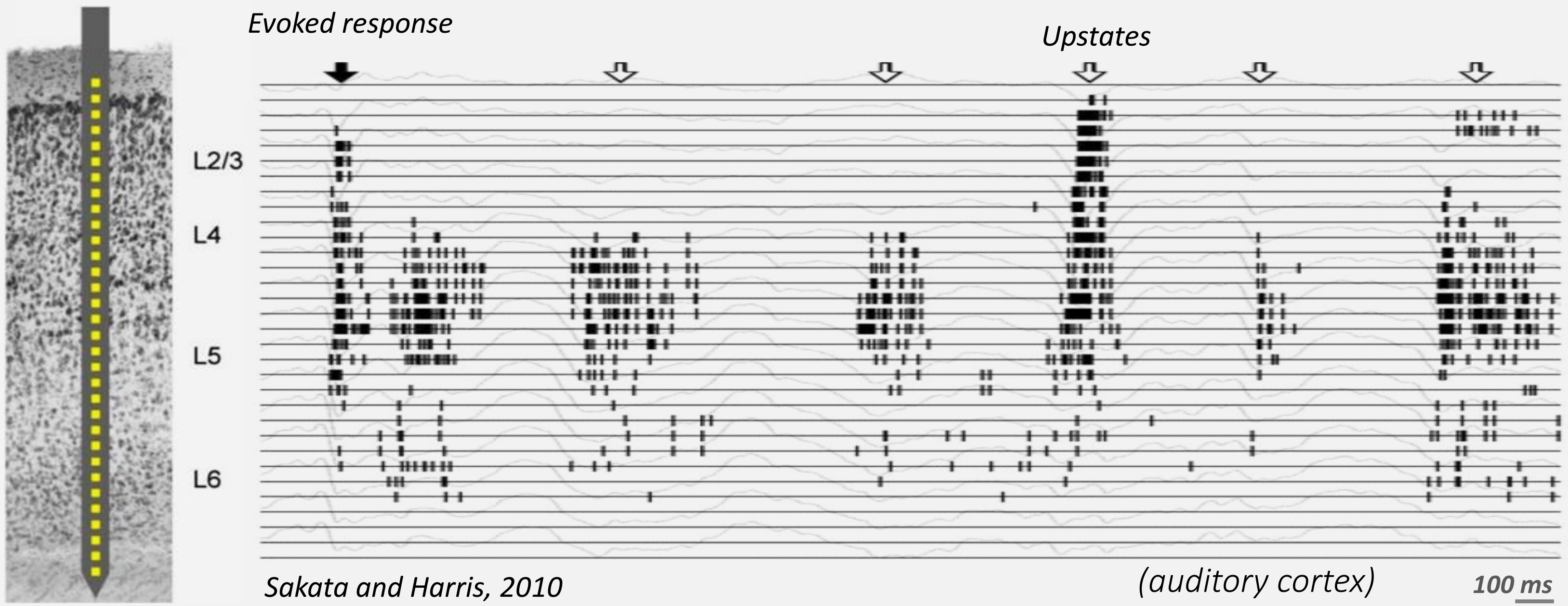
Those oscillations drive dynamical interactions between cortical layers, with a **strong alpha activity in deep layers** (L5 & L6), and **gamma activity in superficial layers** (L2/3 & L4). Interestingly, gamma activity in superficial layers is coupled to the alpha rhythm in deep layers: the lower frequency alpha oscillation cuts down otherwise constant gamma (Spaak et al, 2012).

The strength of the coupling between alpha and gamma is strong in the granular layer (L4) that receives most thalamic inputs, and lower in supragranular layers (L2/3) that seem to process information in relative isolation.

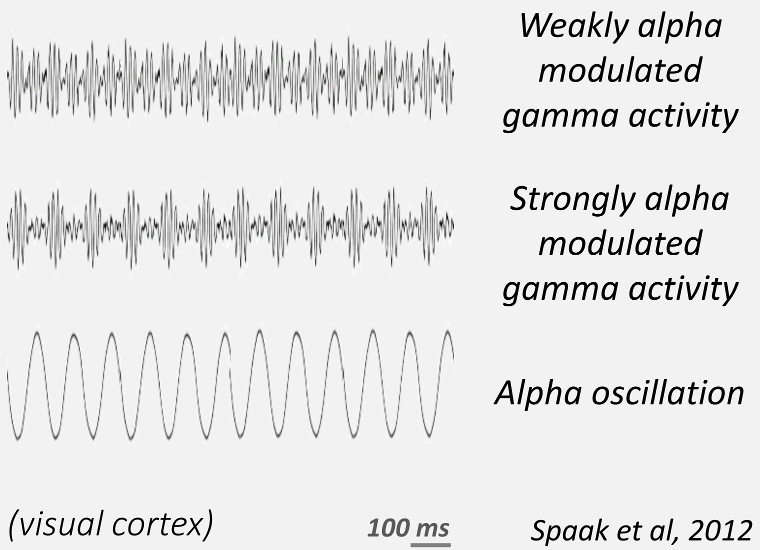
Alpha activity in sensory regions implements a mechanism of pulsed inhibition silencing neural firing every ~100 ms. Said differently, **this mechanism periodically gates external sensory information**, so sensory perception is more likely to occur at specific phase of alpha activity.

Each periodic alpha cycle can be decomposed in an inhibitory phase during which thalamic inputs are silenced, followed by an excitability phase.

Layer specific cortical activity



Sakata and Harris, 2010



Spaak et al, 2012

Precise timing of individual spikes in relation to the phase of neural oscillations encodes information

Some neurons fire **individual action potentials that are precisely timed at a specific phase of neural oscillations** in the surrounding cells (a process referred to as **phase precession**)

This phase code differs from the classical rate code in which the intensity or salience of a feature is represented by the rate of firing. Because phase coding is relative to a given oscillation, there are different phase codes (theta phase code, beta phase code, …). An individual neuron can simultaneously use those different coding strategies.

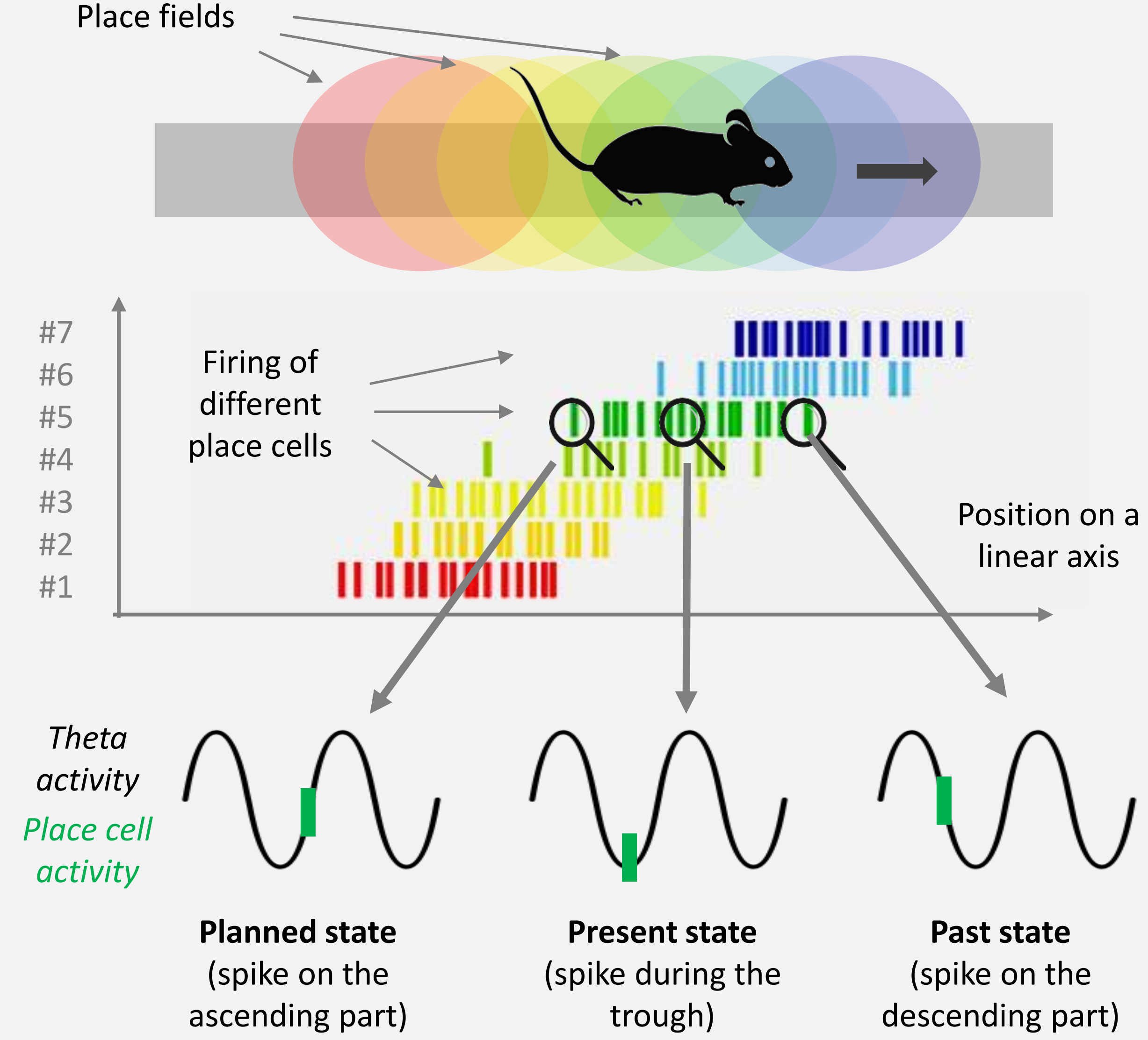
Electrophysiological studies of **place cells in the hippocampus** show strong evidence that **phase precession encodes critical information** about recent past, present and planned locations. Place cells strongly fire during the trough of the theta oscillation when the animal is precisely located at the corresponding place field. Before arriving at this location, those place cells were firing on the ascending part of the oscillation as if they were representing a planned state. After leaving this location, they fire on the descending part of the oscillation.

Other experiments have shown that the theta phase precession of hippocampal place cells is not restricted to spatial location (Lenck-Santini, 2008).

A phase precession coding strategy has also been observed in the entorhinal and prefrontal cortex – considered as limbic cortices – with a gamma activity nested in beta/theta waves (Hafting, 2018, and Smith et al, 2019). It is not yet clear if the neocortex uses a similar coding strategy with the gamma/alpha coupling.

Phase precession of place cells

Coding temporal information by nesting gamma activity within specific phases of theta cycles


Place fields
#7
#6
#5
#4
#3
#2
#1
Firing of different place cells
Position on a linear axis
Theta activity
Place cell activity
Planned state (spike on the ascending part)
Present state (spike during the trough)
Past state (spike on the descending part)

Thiboust, 2020

Cortical dynamics varies significantly between the various states of wakefulness and sleep

Brain oscillations – and therefore cortical dynamics – vary significantly between the **different states of wakefulness and sleep**.

During wakefulness, cortical activity is mainly constituted of alpha and gamma waves, as described previously.

During sleep, the brain alternates between **REM and NREM sleep phases**.

REM stands for Rapid-Eye-Movement. It is recognizable by rapid movements of the eyes, low muscle tone and a propensity of the sleeper to dream vividly. Physiologically and electrically, it is characterized by **high level of acetylcholine neurotransmitter** and **theta/gamma rhythms**.

NREM sleep stands for non-REM sleep. It groups the other phases of sleep. The transition from wakefulness to eyes closed intensifies alpha waves that are replaced by theta waves at the first stage NREM1. Intermittent spindles appear in NREM2. Then, NREM3 & NREM 4 are characterized by **slow oscillations (< 1 Hz) and delta waves** (Adamantidis et al, 2019)

Even if those specific oscillations mostly occur in the hippocampal complex and nearby areas, they impact the dynamics of connected cortical areas.

Sleep phases are believed to play a role in the **consolidation of long-term memories** via hippocampal replay/preplay**:** consolidation of declarative memories seems tightly tied to NREM, but it is still unclear whether other memories are consolidated during NREM or REM (Ackermann, 2014).

Brain waves during wakefulness and sleep

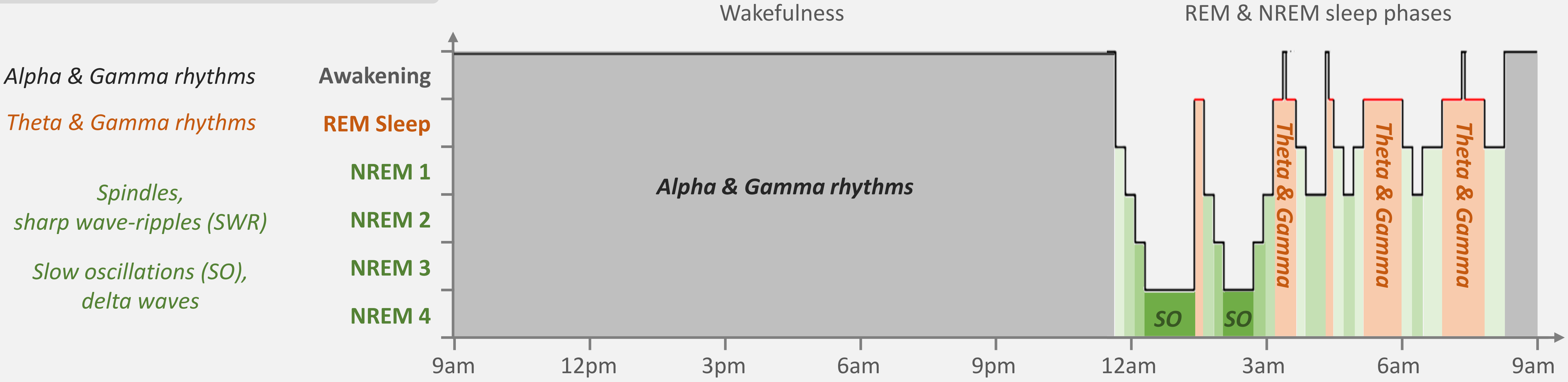


Illustration adapted from Wikipedia

Back to code

1. Next-level artificial neural networks model more realistic neurons, architectures and learning rules

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- Subutai Ahmad
- Yoshua Bengio
- Jeff Hawkins
- Geoffrey Hinton
- Yann LeCun
- Randall O’Reilly

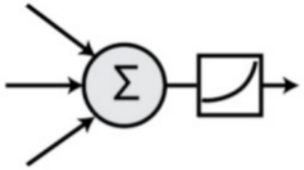
See the reference section for a list of materials that inspired me.

Art credit: Cortical Circuitboard, Greg Dunn Design & Brian Edwards

HTM and Spiking neurons refine the classical “point-neuron” model by modeling dendrites and precise timing

The **“point neuron”** model has been used for decades in ANNs, namely in commercially successful Deep Learning ANNs the last 10 years. At the other side of the spectrum, **Spiking Neural Networks (SNNs)** closely mimic natural neural networks for neuroscience research purposes, but are computationally-intensive. In between, some dendrite-focused models ignore some biological implementation details to save computational effort (like the **HTM model**, not exhaustive).

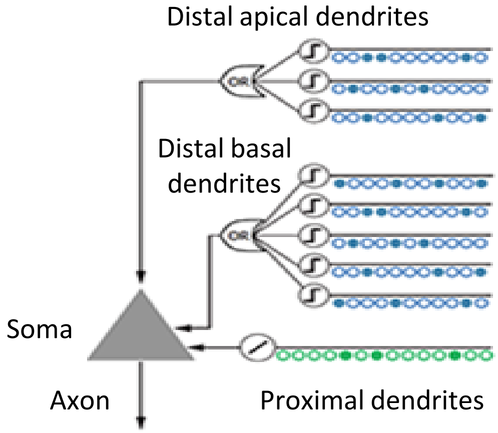
Point neuron



Soma Axon

- Used in classic Deep Learning ANN
- No dendrites / All synapses on soma
- Importance of synaptic weights
- All inputs are considered synchronized
- Non-linear function of the weighted sum of the inputs

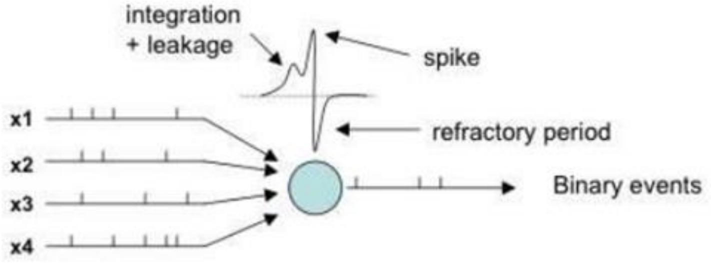
HTM neuron



Hawkins et al, 2016

- Used in ANN by Numenta
- Inspired by pyramidal neurons in the cortex
- Different kinds of dendrites:
    - Proximal for feedforward inputs
    - Distal basal for contextual inputs
    - Distal apical for feedback inputs

Spiking neuron



- Used mainly by researchers to model precisely the physiology of the different neuron types
- Precise timing matters: inputs integration, electrical profile of spikes, refractory periods
- Some well-known models: Hodgkin-Huxley, Izhikevish, Leaky Integrate-and-Fire (LIF)

Biological conformity

Computational friendliness

Wide & highly recurrent network architectures are more biologically plausible than deep & feedforward ones

Wider rather than deeper

- **Deep Learning networks used to get increasingly better by adding hidden layers**: several hundreds and even 1202 layers for ResNet-1202 (He, 2015).

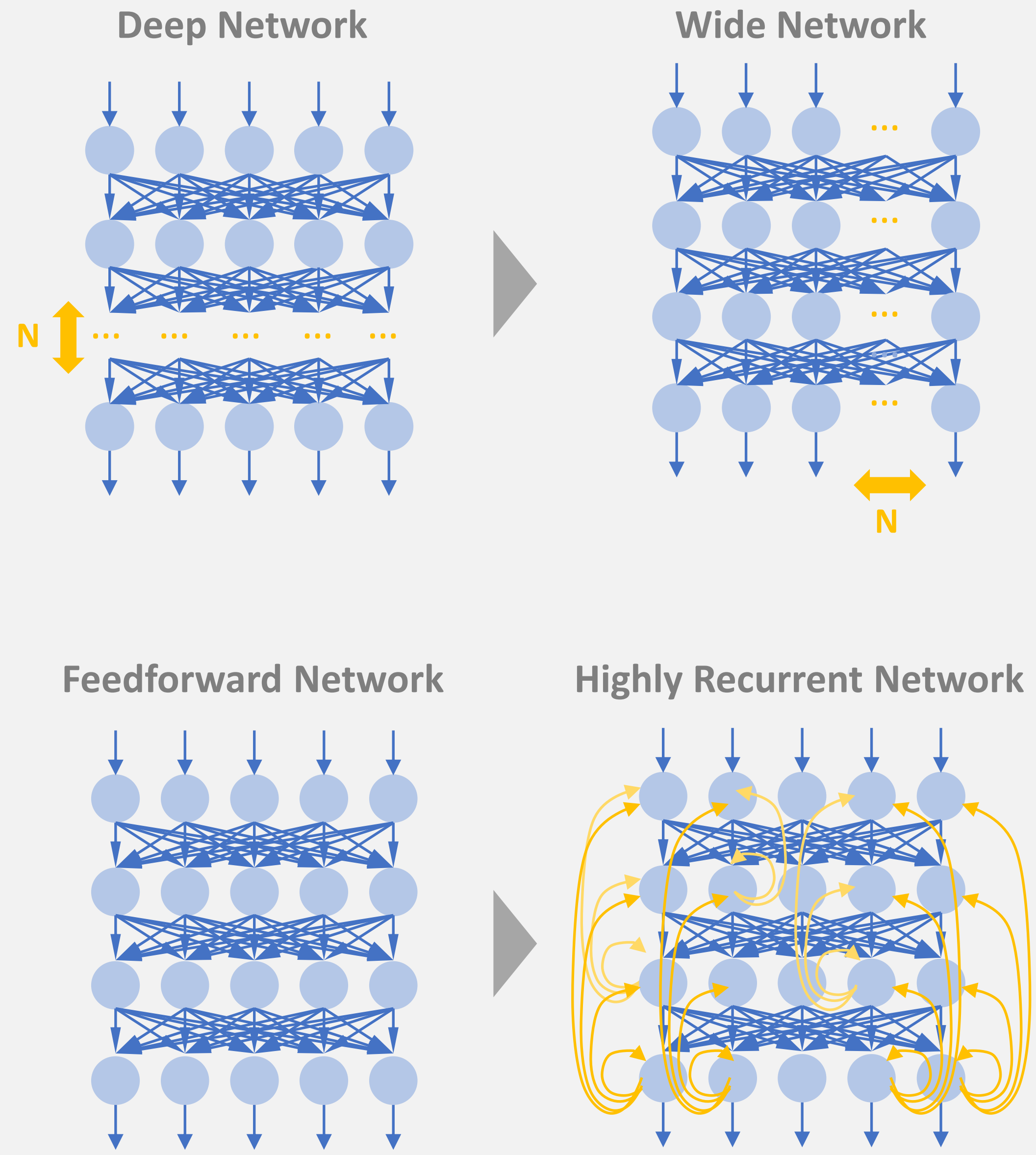
- **However, wider networks can outperform their thin and very deep counterparts**: example with simple 16-layer wide ResNet (Zagoruyko et al, 2016). They also present better learning dynamics (Xu et al, 2019).

- With their massive parallel processing abilities, humans can reliably identify objects in the central visual field within a single fixation in less than 200 ms when viewing "standard" images (DiCarlo, 2012). Given an average duration of 5 ms per neuron activation, it would mean that **humans achieve this task with a network depth of only 40 successive layers** (even less in reality given the recurrent connections)

Recurrent rather than feedforward

- **Feedforward networks are popular because of the easy applicability of the *backpropagation learning algorithm*, but they lack the memory abilities of *Recurrent Neural Networks (RNN)*.** In simple RNN and specific recurrent architectures like LSTM/GRU, the network can be unfolded to apply a *backpropagation through time algorithm (BPTT).* However, this solution doesn't scale well for more complex recurrent structures (Pascanu, 2013).

- **In humans, only 10% of inputs to the LGN (main input to primary visual cortex) come from the retina**. 90% of inputs come from feedback projections of different position in the hierarchy (cortex & brainstem), making the network highly recurrent (Derrington, 2001)



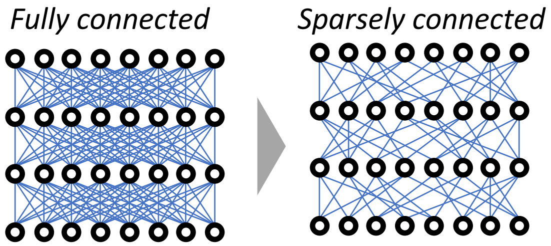
Thiboust, 2020

Sparsity is a key characteristic of ANN to improve noise robustness

Sparsity is the biological solution of the brain to maintain robustness while being highly energy-efficient. Sparsity is both structural (number of neurons & synapses) and operational (% of active neurons, % of synapse updates)

Sparse connections

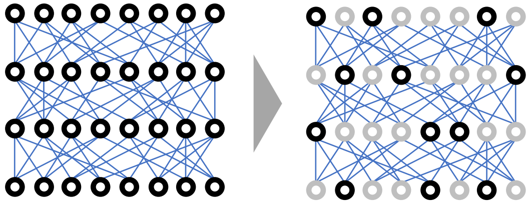
- Each pyramidal neuron in L2/3 receives inputs from around 5% of other nearby accessible pyramidal neurons (Holmgren, 2003)
- Sparsity can be enforced by pruning connections with low synaptic weight



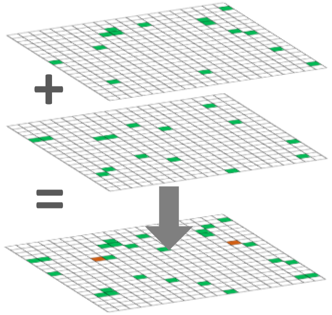
Ahmad et al, 2019

Sparse activations

- At each instant, only around 1% of cells are active (Lennie, 2003).
- Sparsity can be enforced by applying a k-Winner-Take-All algorithm



Ahmad et al, 2019

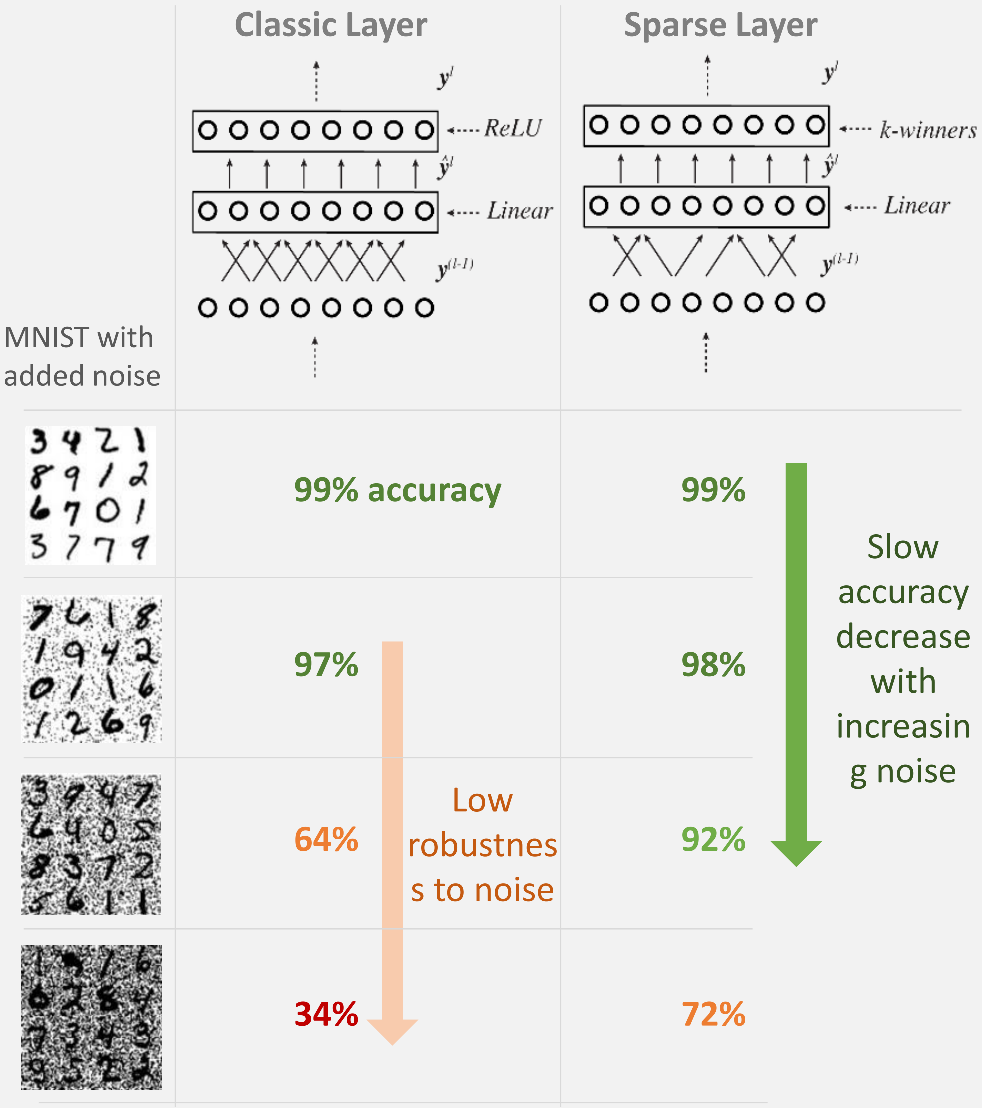


Sparse Distributed Representation (SDR) are binary vectors mostly composed of 0s and representing a state of the layer.

Because of their sparsity, unions (= *bitwise-OR*) of SDRs can represent multiple things or ambiguous states with low overlaps *(= indexes of 1s after bitwise-AND*).

An altered SDR can also be easily recognized.

Example of noise robustness with sparsity



Ahmad et al, 2019

Hierarchy can be rethought with skip connections along the bottom-up and top-down pathways

Multi-levels clues

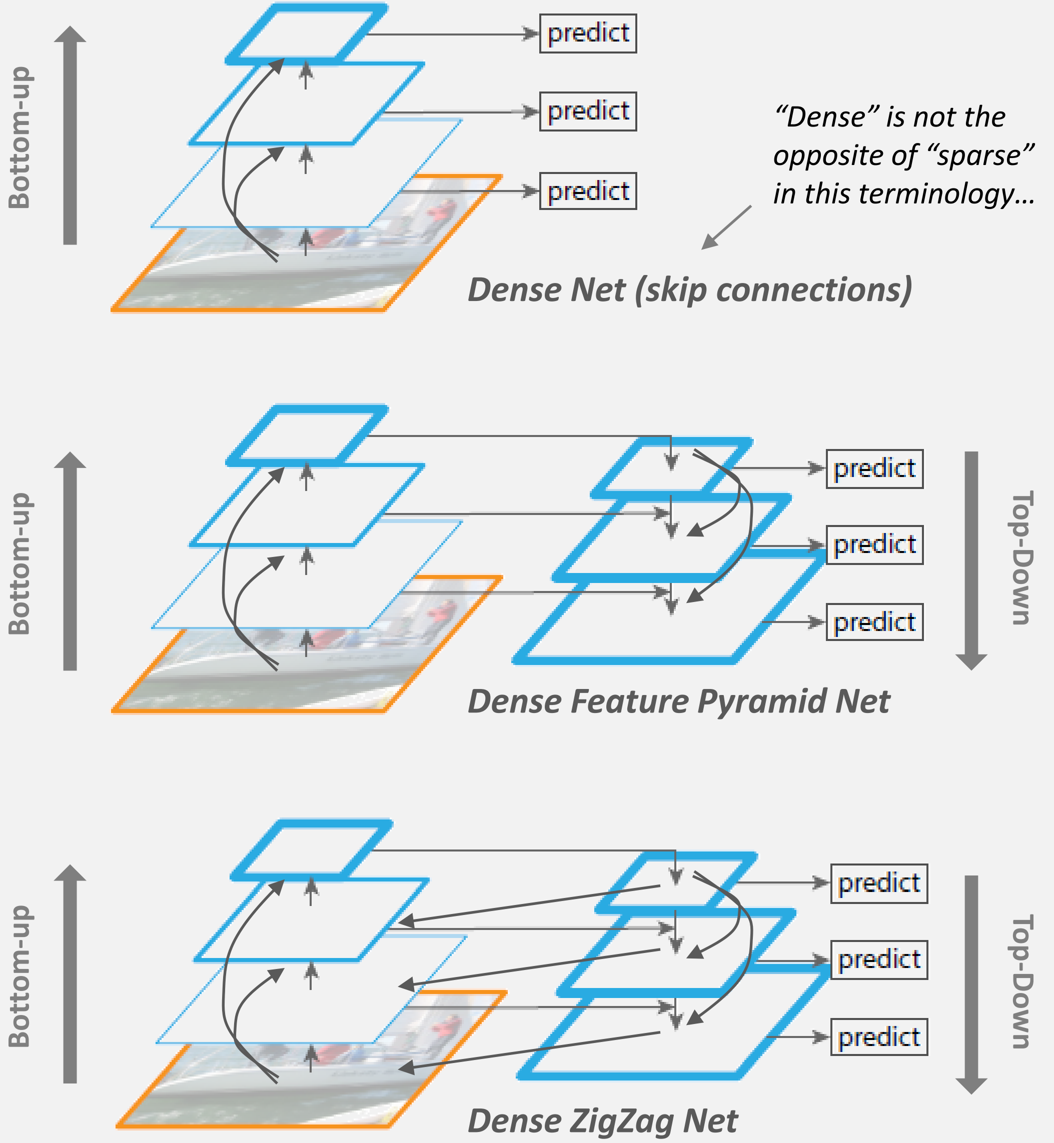
- Hierarchy is not linear in the brain: for example, the different areas of the human visual stream are densely interconnected with **V1 projecting not only to V2, but also directly to even higher areas like V4, MT & IT, etc.**
- Those skip connections are used in *Residual Networks (ResNet)* and *Dense Networks (DenseNet)* to **optimize the training of very dense networks with the backpropagation algorithm**. They also made available low-level features to higher layers.

Top-down modulation of predictions

- **Combining a bottom-up with a top-down network and "lateral skip connections" allows to simulate feedback connections at each level** while keeping the network feedforward (ex: *Top-Down Modulation Networks, Feature Pyramid Networks*). Predictions at high-resolution (semantically weak features) is improved by top-down clues (semantically strong features) working as a kind of attention mechanism.

Top-down modulation of predictions and inputs

- **Some networks make bilateral connections between the bottom-up and top-down networks to mimic the brain more closely** (ex: ZigZagNet). The bidirectional connections are critical for fusing and exchanging context, progressively learning how to refine the feature maps with useful information. However, the training is more complex because of the recurrent connections.



Dense Net (skip connections)

"Dense" is not the opposite of "sparse" in this terminology…

Dense Feature Pyramid Net

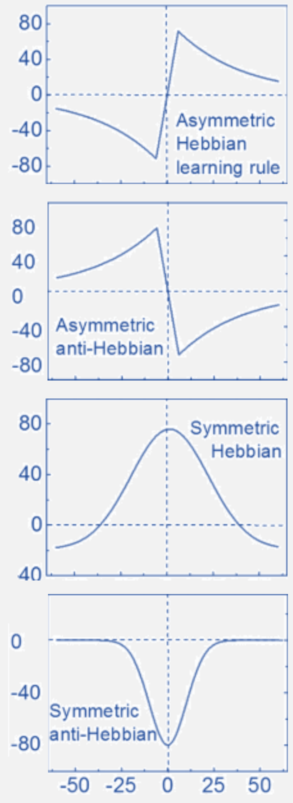
Dense ZigZag Net

Adapted from Lin et al, 2017 (dense connections and ZigZag Net added)

Local learning rules allow to get around the limitations of the backpropagation algorithm

Designing biological-inspired RNN architectures is mainly constrained by the limitations of the biologically non-plausible backpropagation algorithm. This kind of training suffers from vanishing gradient and gradient exploding problems. While many researchers are looking for specific RNN architectures well-adapted to the canonical or approximated backpropagation algorithms (LSTM, GRU, …), other have chosen to tackle the inverse problem: **finding a learning algorithm well-adapted to biological-inspired RNN**. It comes as no surprise that the latter researchers are focusing on **biologically plausible local learning rules** for updating the synaptic weights, the most famous being ***Hebbian learning / Spike Timing Dependent Plasticity (STDP)*** and ***competitive learning*.**

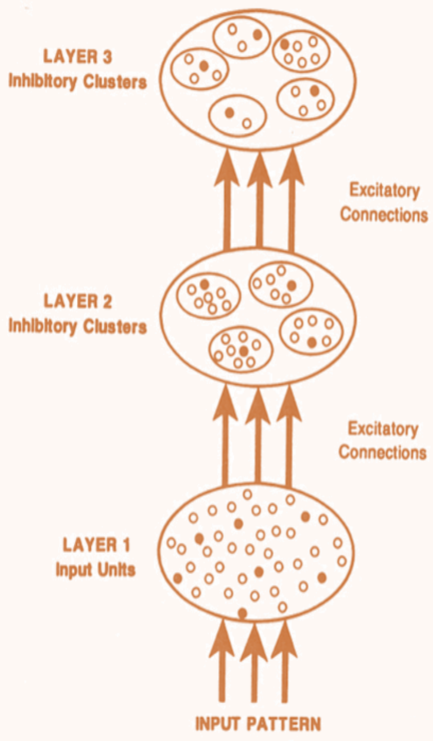
Hebbian learning / STDP



Li, Miao et al, 2014

- Updates of synaptic weights only depend on the **relative timing of spikes between pre and post-synaptic neurons** (and possibly other reward/error inputs in *three-factor plasticity rules*).
- Brains implements those mechanisms via **back-propagating Action Potentials (bAP**) from the soma to **NMDA receptors** in dendrites.
- The **Hebbian rule** is the most famous: increase of synaptic weights between neurons that fire together.
- Some synapses are governed by **anti-Hebbian or non-Hebbian plasticity to enforce causality information** into the network: the pre-synaptic neuron fires slightly before the post-synaptic neuron if it is the cause of the firing.
- This updating mechanism is **applied locally and online at each step.**

Competitive learning



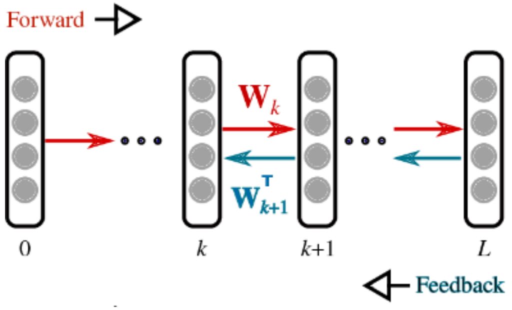
Rumelhart et al, 1985

- Updates of synaptic weights depend on the **result of the competition between neurons** in a given cluster of neurons.
- Brains implements those mechanisms via biological interactions between **excitatory inputs and local inhibitory neurons.**
- Only the fastest neuron to fire wins the competition and inhibits the other ones: **Winner-Take-All (WTA).**
- It is used by **Self-Organizing Map (SOM)** algorithms (also called Kohonen map).
- This updating mechanism is **applied locally and online at each step.**

New algorithms are needed to solve the credit assignment problem without the backpropagation of errors

Multi-layer neural networks need a ***credit assignment algorithm*** to compute the **contribution of each neuron to the overall error**, and then use this information to update the parameters of the entire network. Limitations of the backpropagation of errors for biologically-inspired networks have encouraged researchers to look for more biologically-plausible alternatives that could be classified into three main families:

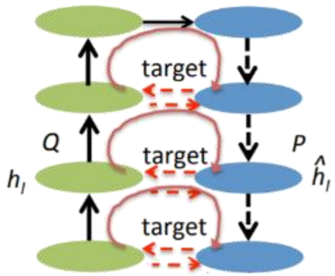
Contrastive Hebbian Learning (CHL)



Detorakis, 2018

- Target-based learning via activity difference between a minus phase (forward-only) and a plus phase (essentially backward). Similar to Boltzmann Machine learning
- Requires a symmetric feedback pathway and cyclic inhibition/disinhibition of the backward pathway
- Does not require a full forward pass before updates
- Variants: Random CHL (rCHL) with no need for symmetry in feedback pathway

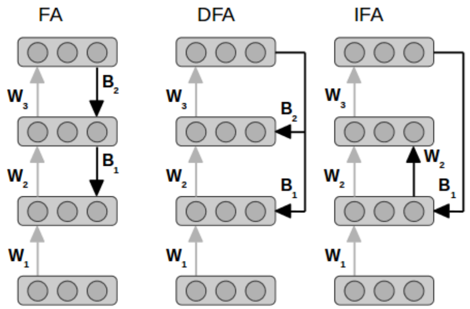
Target Propagation (TP)



Bengio, 2014

- Target-based learning via auto-encoders to assign reconstructed targets to each layer below. Reciprocal propagation of the activities is realized through learned connections
- Requires a feedback pathway and layers of similar dimension to avoid bottlenecks during reconstruction
- Does not require symmetric weights
- Variants: Difference TP (DTP)

Feedback Alignment (FA)



Nøkland, 2016

- Gradient-based learning via propagation of errors through feedback connections with learned weights. The network learns how to learn.
- Requires a feedback pathway and a full forward pass before updates
- Does not require symmetric weights
- Variants: Direct FA (DFA) with level-skipping, Indirect FA (IFA)

Hybrid approaches

Local Representation Alignment (LRA),
Direct Random Target Projection (DRTP)

Back to code

2. The transition from artificial networks to artificial agents is a necessary step towards machine intelligence

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

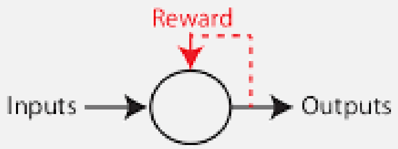
- Yoshua Bengio
- François Chollet
- Jeff Hawkins
- Carlos E. Perez
- David Silver
- Richard S. Sutton

See the reference section for a list of materials that inspired me.

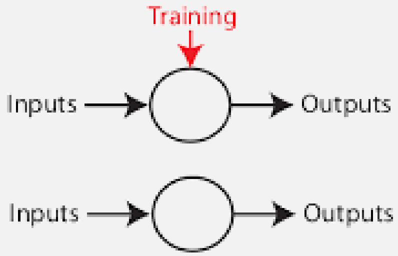
Art credit: Cortical Circuitboard, Greg Dunn Design & Brian Edwards

Hybrid Reinforcement Learning approaches supports more autonomous agents driven by rewards

Reinforcement Learning (RL) is a paradigm in which software agents learn to take actions in an environment so as to maximize a cumulative **reward:**

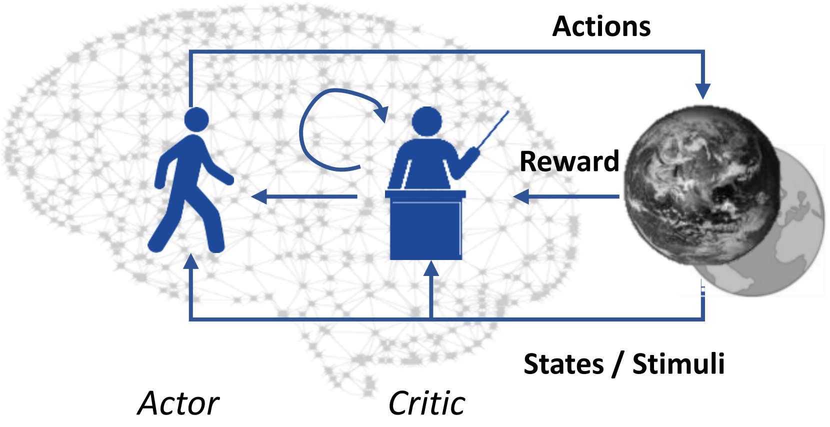


As such, it differs from *supervised learning* where an already labelled inputs-outputs dataset is provided, and from *unsupervised learning* where only the inputs are given:



RL algorithms optimize the policy and/or estimate the value of a given policy, with or without modelling the environment. All those methods can be combined like the brain does.

Combining Value-Based and Policy-Based RL



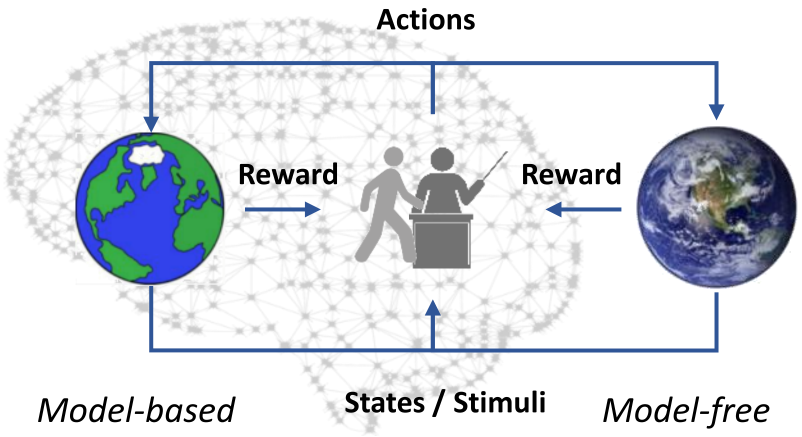
The brain is believed to work both:

- As an *actor* that learns policies in the dorsal striatum
- As a *critic* that learns value functions of the policies followed by the actor in the ventral striatum

In RL, this is called an **actor-critic algorithm**.

The *actor* takes as input the state and outputs the best action. It controls how the agent behaves by learning the optimal policy (*policy-based RL*).
The *critic* evaluates the action by computing the value function (*value based RL*).

Combining Model-Based & Model-Free RL



The brain uses both:

- *Model-based* learning with the prefrontal cortex and the dorsomedial striatum (DMS)
- *Model-free* learning with the sensorimotor cortex and the dorsolateral striatum (DLS)

With habituation, behavioral decisions are progressively transferred from the DMS to the DLS.

Implementing both types of learning in an artificial agent allows for combining the sample-efficiency of *model-based RL* with the accuracy of *model-free RL*.

Thiboust, 2020

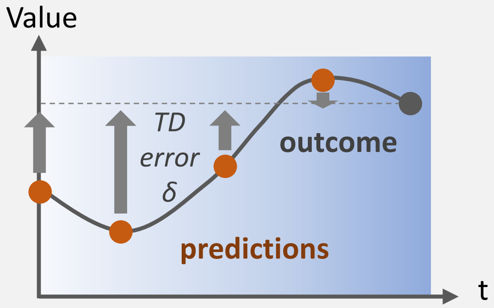
Reinforcement Learning gives insights about learning from delayed feedback & knowledge transfer

The Reinforcement Learning community has developed several key algorithms to tackle longstanding challenges in Machine Learning.

One popular and very successful algorithm is *Q-Learning* (and its improved variants *Double-Q-Learning, Deep-Q-Learning, Double-Deep-Q-Learning*). However, it relies on *Action-Value functions* which appear less biologically plausible than *Value functions.* Indeed, brains are primarily action-driven and seem to equate *Action* and *State* representations according to a growing consensus in the neuroscience community.

Two other RL algorithms have shown surprising similarity with specific brain mechanisms:

Learn from delayed feedback with *Temporal Difference*



Formula:

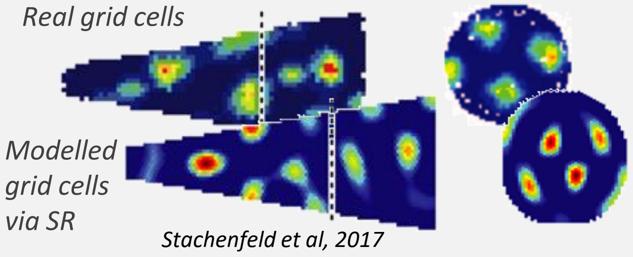
$$\delta_t \doteq R_{t+1} + \gamma V(S_{t+1}) - V(S_t)$$

$$V(S_t) \leftarrow V(S_t) + \alpha\, \delta_t$$

Temporal Difference (TD) learning is "learning a prediction from another, later, learned prediction". This scalable & online learning algorithm is well adapted to real-life general multi-step prediction problems where the feedback is delayed or even not reached. TD(λ) is a more powerful extension of TD with some of Monte-Carlo advantages.

Dopamine signals generated in VTA/SNc brain structures exhibit many of the hallmarks of the reward prediction TD error (Shultz, 1998). However, recent findings have shown that dopamine signals are higher dimensional that initially thought, driving new RL research.

Transfer knowledge with *Successor Representations*



Stachenfeld et al, 2017

Formula:

$$\underset{\underset{\text{Value}}{\uparrow}}{V(s_t)} = \underset{\underset{\text{Reward}}{\uparrow}}{R(s_t)}\underset{\underset{\text{SR}}{\uparrow}}{M(s_t, s_{t+1})}$$

The approximation of the *value function* can be simplified under the hypothesis that it can be decomposed in two decoupled factors:

- The *Successor Representation (SR)* that only depends on the dynamics of the environment and the agent itself
- The reward function of the environment

Having learned the *SR* of an environment dynamics, it can be transferred to similar environments but with different reward functions.

Place cells can be modeled with *SR*. Surprisingly, *grid cell* patterns look accurately similar to the eigendecomposition of *SR (Stachenfeld, 2017)*

Leverage temporal data to “self-learn” by continuously predicting what will happen next

Temporal data

All living organisms process temporal data that streams continuously on their sensors.

Recurrent Neural Networks (RNNs) and memory networks like LSTMs already take the time dimension into account, but not the Convolutional Neural Networks (CNNs) that are commonly used to recognize object in images. If CNNs are applied to every frames of a video, they will process each frame independently, without using the results of previous frames as clues for the next ones.

In order to allow the **temporal integration of temporal data** in computer vision, CNNs are combined with recurrent and feedback connections into what are called deep **convolutional recurrent neural networks (CRNNs)**.

Self-supervised learning

Streams of temporal data can be used to “self-train” a model by continuously **predicting the future from the past, and then comparing the prediction vs the outcome at the next timestep**. Note that it does not require pre-labelled data. This training method follows the principles of the **predictive coding theory** from the neuroscience literature.

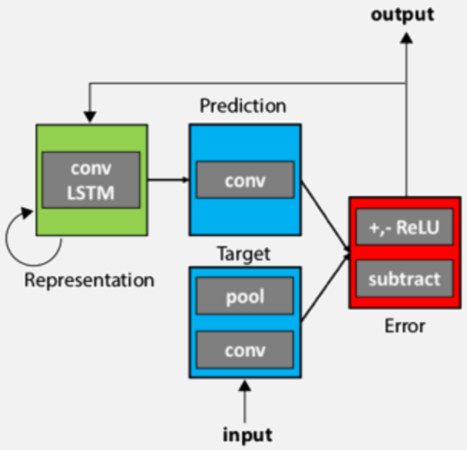
In the AI community, it is part of a more general method called **“self-supervised learning”**. The idea is to train a model using **labels that are naturally part of the input data**, instead of separate external labels. In addition to the biologically-plausible task of predicting future representations based on the recent past, AI practitioners can ask the model to reorder video frames that have been voluntarily shuffled, or to reposition pieces of an image that has been voluntarily cut for instance.

Self-supervised learning in practice

The field of Natural Language Processing (NLP) has been the first AI discipline to fully embrace self-supervised learning. Self-training models by predicting the next word given the past sequence conducted to state-of-the-art models like Word2Vec, Glove, ELMO and BERT.

In computer vision, models pretrained on the huge ImageNet dataset are commonly used as the starting point before applying transfer learning. It works well when dealing with real-life pictures, but not with specialized medical images like radiographies for which annotated data is still scarce.

Self-supervised learning is increasingly chosen as a solution to this issue in computer vision:



The PredNet is a CRNN trained for next-frame video prediction with the belief that prediction is an effective objective for "self-supervised" learning”

Lotter et al, 2016

Agents can build meaningful internal models of the environment via active sensorimotor interactions

Humans acquire knowledge by interacting with their surroundings. In other words, they **build internal models of the environment by creating their own data through their own actions**. This kind of learning gives a much richer understanding of objects and the world in general by grounding *meanings* through *actions*.

Artificial agents can also do active sensing in the sense that the **movement of their sensor is controlled to improve information pickup and is tuned to the ongoing task**.

Sensorimotor interactions

Like for humans, actions of artificial agents do not necessarily have to impact the environment. ***Active sensing*** like controlling the camera orientation (eye saccades for humans) or moving the whiskers of a robot (Pearson et al, 2011) is enough to get a sense of what is happening through time and to **consolidate or refine internal models with predicted or unpredicted chosen observations**.

The time dimension in sensorimotor interactions is crucial to transform **simple correlation learning into causal learning**.

Causality links are strengthened in experiences where changes in the environment are directly caused by the actions of the agent (for example, self-generated camera shakes when moving)

Embodied AI

In reference to the expression "embodied cognition" that underlines the strong intertwinement of the mind and the body, the AI field uses "***embodied AI***" to define those intelligent agents that learn from their own perspective with sensorimotor interactions (for example, a visual representation within an environment). The acquired knowledge of these agents is grounded in their artificial embodiment.

Real vs virtual environments

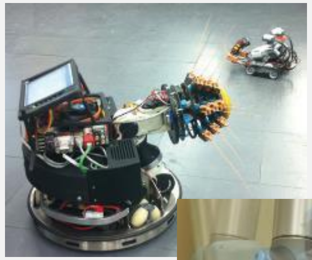
Agent knowledge acquired through interactions in a *real environment* shares a common ground with human knowledge. As such, those agents may have self-initiated meaningful interactions with humans in the future.

Simulated environments are often preferred to avoid the long training time and the engineering challenges of robotics. Because designing good virtual environments is a difficult task, AI researchers do not hesitate to use simulated environment from video games. However, simulations are still far from real-world richness and complexities.

Virtual environment

Starcraft video game

Real environment



Pearson et al, 2011

Shrewbot (vibrissal sensing)

Princeton website

Robot arm

Agents learn continually all along their life with adaptive learning mechanisms and curriculum training

The extent to which human and artificial agents learn all along their life depends on:

- Their built-in **adaptive learning mechanisms** allowing to adapt to evolving data distribution and to learn new tasks without catastrophic forgetting.

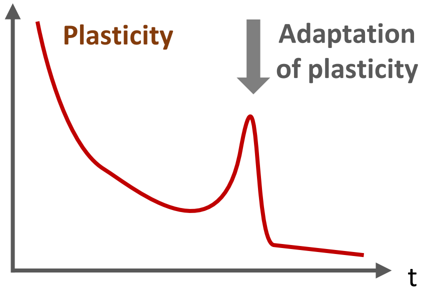
- The **richness, diversity and progressivity of their exposure** to various situations or datasets in order to generalize to increasingly more abstract concepts.

Juliani, 2017

Easy
Difficult

Adaptive learning mechanisms

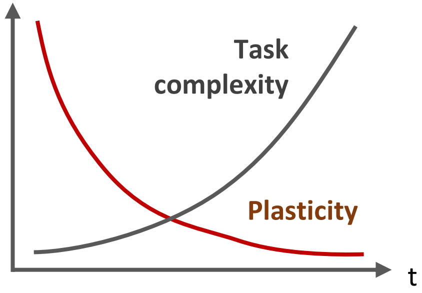
A compromise must be found between two goals: adapting to new tasks and enforcing stability to preserve knowledge from previous tasks



- A first phase of synapses growing, followed by an intense *synaptic pruning* phase to speed the convergence towards the initial architecture (like infant development).
- A divided architecture between a slowly evolving global part and a quickly evolving task-specific local part, with progressive knowledge transfer from the latter (*hippocampus-like*) to the former (*cortex-like*).
- A self-learned and evolving learning rate parameter at the synapse or neuron level, so that the network learns when and how to adapt.
- A globally decreasing learning rate with self-learned semi-global learning rate adaptations mimicking *neuromodulated plasticity* (ex: “backpropamine” ANN).

Curriculum training

Like humans and animals, artificial agents exhibit better learning performance when the training is organized in a meaningful way. This is referred to as *curriculum training*:



- Making the learning tasks gradually more difficult, in order to stabilize first the fundamental knowledge upon which subsequent knowledge will be grounded. It can be seen as a special case of transfer, where the knowledge collected during the initial tasks is used to guide the learning process of more sophisticated ones.
- Sequencing learning along successive *critical periods* of subparts of the network to mimic the successive limited time windows in infant development during which the brain is particularly plastic (primary sensory areas first, then language areas, ….)

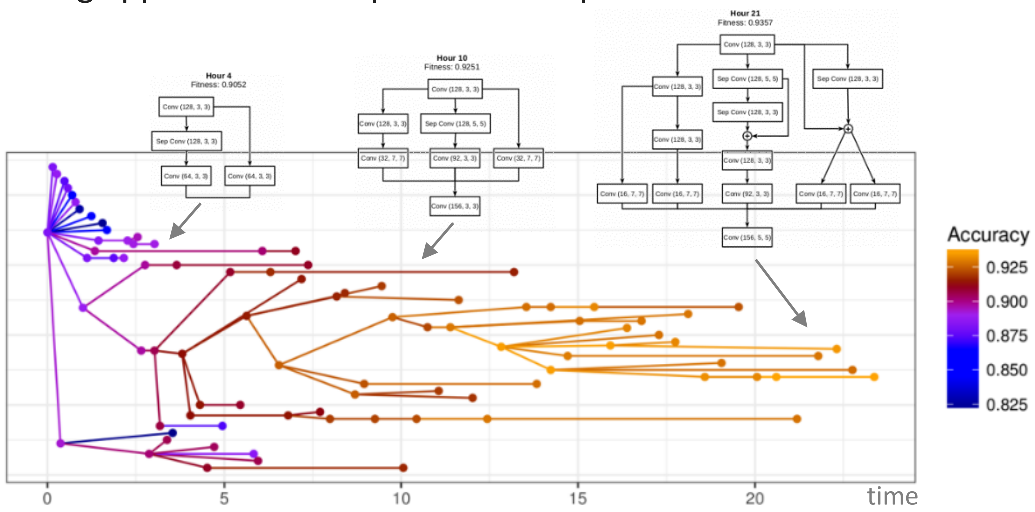
Agents need human-like innate priors designed by hand or via evolutionary algorithms

We, humans, are not born with a blank slate. We possess numerous **innate priors** – selected by *natural selection* over millions of generations and likely embedded in our DNA – that speed up our intellectual development during a lifetime.

Innate priors can be incorporated into ANN by either relying on researchers' intuitions (like the Convolutional Neural Networks architecture to impose *translational invariance*) or on *evolutionary algorithms* for a more general and prior-agnostic approach.

In ANN, innate priors reside in the chosen network architecture, hyperparameters and learning rules. They constrain the network.

It is common to use genetic algorithms to optimize hyperparameters, but not yet for the network architecture and its learning rules. It may be a promising approach to incorporate useful priors.



Example of network architecture optimization via evolutionary algorithms (Wistuba, 2018)

Chollet, 2019

Meta-learning priors

Meta-learning priors govern our learning strategies and capabilities for knowledge acquisition. As such, they are the building blocks of intelligence:

- **Spatiotemporal continuity**: form persistence and smoothness of motion
- **Modular-hierarchical structure** as a general organization rule of information
- **Causality**: directional correlation with direction provided by time. A preceding observation likely causes the following observation

$a \Longrightarrow b$

High-level knowledge priors

High-level knowledge priors regarding objects and phenomena in our external environment:

- **Elementary physics**: object definition, object persistence, object motion, object contact, …
- **Goal-directedness**: separation between inanimate and animate objects possessing intentions and following their own objectives
- **Elementary arithmetic**: abstract number representation for small numbers that can be added, subtracted and compared
- **Elementary geometry**: distance and orientation in 2D/3D environment

Artificial cognition goes through an extension of artificial perception, not through a neuro-symbolic hybrid approach

Since its beginning, the history of AI has been divided into two approaches:

- The ***connectionist approach*** that revolutionized computer vision and natural language processing with neural networks, but has not yet succeeded in tasks that involve logical reasoning, planning or capturing causality. In other words, those models are good for “curve fitting” but bad for extrapolation beyond training data.
- The ***symbolic approach*** that thrives in bounded problems with symbols, objects and relationships between them, but struggles to bridge the messiness of the real world to the world of symbols.

Those approaches look so complementary – both in what they have to offer and in what they lack – that it seems natural to **combine the best of both worlds to have both perception and logical reasoning**.

However, **I think that connecting a neural network to a rules & logic system is not the way to go**. In this hybrid approach, the symbols of the second system would need to be attached somehow to the corresponding neural representations.

Following biology, it would be more interesting to “grow” symbols directly inside a neural network, and then see how those symbols could be detached to be manipulated at a high level (see the symbol detachment problem by Pezzulo et al, 2007).

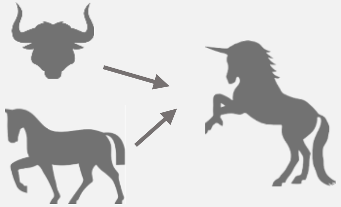
“Growing” symbols is a multistep process that begins with simple symbols (percepts) that are then combined into more complex and abstract symbols (concepts). Implementing **compositionality** is essential to combine existing concepts in novel ways.

How to detach symbols that could be logically manipulated and communicated by language is still an open question in the AI community that is actively looking for some kinds of new priors. This research is still in its infancy, but its biological inspiration makes me confident that future neuroscience findings will inform this approach.

Extending artificial perception

Compositional learning

Compositionality is the capacity to understand and produce novel combinations from known components. AI progress on this topic would have tremendous impact on data-efficiency, zero or one-shot learning and transfer learning.



I know what is a horse and what is a horn, so I can imagine what a unicorn would look like even if I have never seen any.

Symbol manipulation

Adding symbol manipulation capacity to deep learning networks sounds like adding slow-serial-conscious abilities to fast-massively-parallel-unconscious abilities.

Some new architectures implementing soft attention mechanisms may constitute a key ingredient to focus computation on a few concepts at a time.

Two cognitive systems (according to Kahneman, 2011):

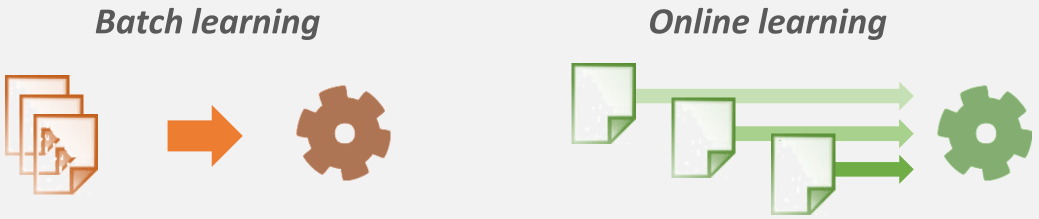
System 1:	*Fast*	*Parallel*	*Everyday decisions*	*Mainly unconscious*
System 2:	*Slow*	*Serial*	*Complex decision*	*Mainly conscious*

Online learning algorithms and neuromorphic hardware are necessary for fast and energy-efficient computations

Simulating today's not-yet-so-intelligent agents already requires an extraordinary amount of computational power, using sometimes thousands of GPUs/TPUs in parallel during a few days to train the network. As an example, the training of the BERT model for NLP on 64 V100 GPUs consumed 12 MW during 79 hours while the brain only uses 20 W on average.

Two technical solutions – one software, one hardware – can enable **scalable fast and energy-efficient computation** in biologically-inspired neural networks:

Online learning algorithms



- Scan all data before updates
- Need memory to store data

- Model updates after each sample
- Process data once and then get rid of them

Samples are processed sequentially one at a time as they come in from the stream. With multisensory encoding, the same architecture can process multiple streams coming from different sensors in parallel. No need to store and access in memory old samples.

The intelligence of the agent resides in **filtering perceptual data and retaining in real-time only the most important information** which will compose the recall memory.

Neuromorphic hardware

Neuromorphic architecture with digital or analog circuits mimicking biological networks of spiking neurons and their STDP learning rules without the need to constantly shuttle data between physically separated logic and memory units von Neumann architectures.

Neurons do not need to produce an output at all times. Instead, information is integrated over time and **communicated sparsely using discrete spikes**, lowering the energy footprint.

Some chips hardcode local learning rules following precise topology without complex and memory-hungry software computations.

Photos credits: The University of Manchester (SpiNNaker), DARPA (TrueNorth), Heidelberg University (BrainScaleS)

Back to code

3. The potential emergence of machine intelligence already raises existential questions

Matthieu Thiboust

Main inspirational people whose work helped me to shape my vision in this section (views are my own):

- Lisa Barrett Feldman
- François Chollet
- Stanislas Dehaene
- Antonio Damasio
- Christof Koch
- Joseph Ledoux

See the reference section for a list of materials that inspired me.

Art credit: Cortical Circuitboard, Greg Dunn Design & Brian Edwards

How far are we from High-Level Machine Intelligence (HLMI) ?

Current AI is still at least a dozen breakthroughs away from HLMI, very unlikely to happen within the next decade

The AI debate is filled with existential questions about the possibility to achieve HLMI, the potential threat for humanity and the philosophical implications of what it means to be human.

Concerning the first question, predictions about a coming-soon HLMI regularly make the headlines. Those claims, which often arise from some famous business leaders, have a high resonance in the ongoing hype while most experts in the field are more reserved.

Given enough time, there seems to be a consensus that we will eventually reach HLMI. But it may take decades or even centuries.

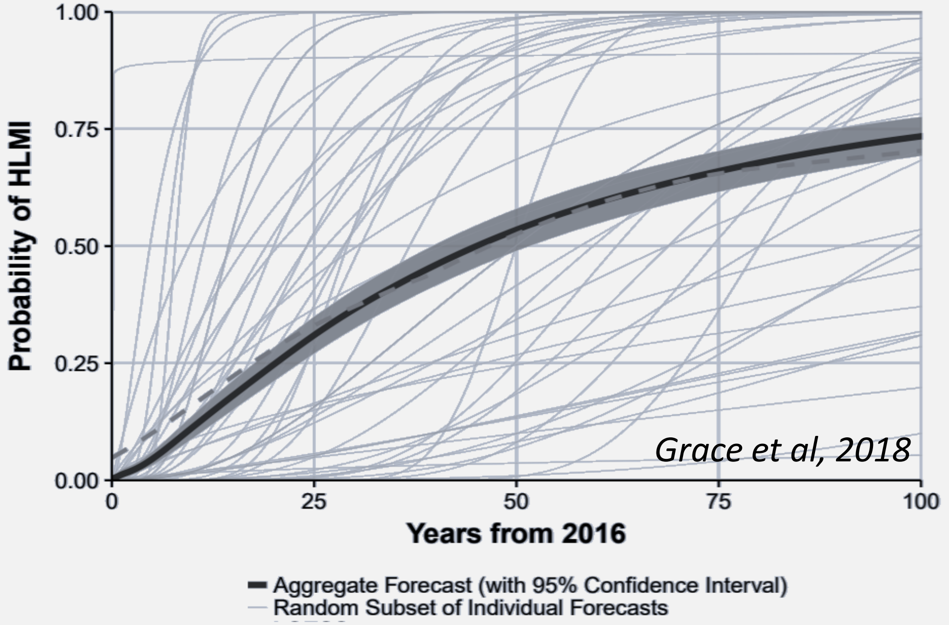
Progresses towards Machine Intelligence do not follow a linear path. From time to time, there is a new breakthrough idea that allows significant advances, followed by small marginal improvements until the next breakthrough.

Even if scientific breakthroughs are intrinsically hard to predict, we can safely state that **HLMI is not around the corner because several significant breakthroughs still need to be unlocked to bridge the gap, and unlocking those breakthroughs at the same time is statistically unlikely.**

The list of the needed breakthroughs is not straightforward. We can just guess that some of those breakthroughs will be software-related: modelling of synapses & neurons, network architectures, interactions between networks, learning algorithms *(not exhaustive).* Others will relate to hardware: neuromorphic chips, embodiment, virtual perception *(not exhaustive).*

Expert predictions of HLMI arrival

"High-level machine intelligence" (HLMI) *is achieved when unaided machines can accomplish every task better and more cheaply than human workers.*



Grace et al, 2018

According to this survey among 352 AI researchers, the aggregate forecast gave a 50% chance of HLMI occurring within 45 years and a 10% chance of it occurring within 9 years (with a large inter-subject variation)

Should we expect a singularity such as an intelligence explosion?

Surely on an evolutionary timescale yes, but probably not on a human lifetime scale.

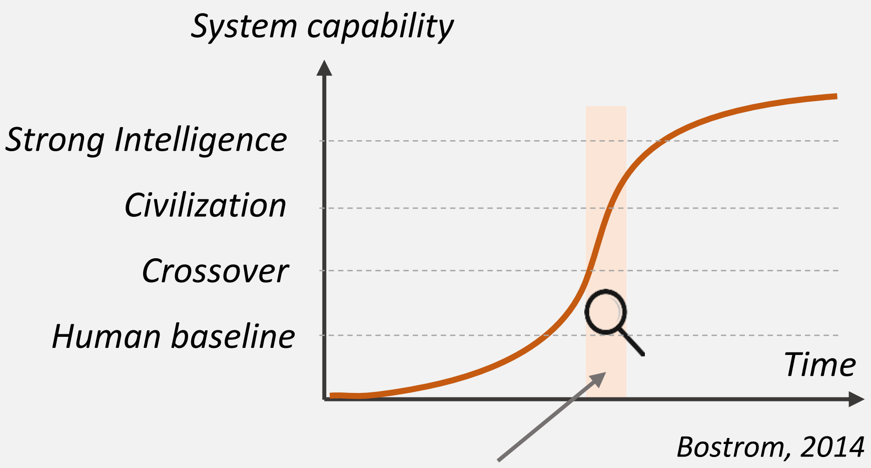
The **intelligence explosion** – the most popular version of the singularity hypothesis – is a hypothetical point in time when an intelligent agent slightly surpasses human intelligence, and then recursively designs more intelligent agents in such a way that machines overtake human intelligence by orders of magnitude in a short period of time.

The plausibility of such a scenario depends on the timescale:

- On an **evolutionary timescale**, we are already experiencing an **intelligence explosion started 10.000 years ago.** However, it is does not relate to individual biological intelligence. It is an **explosion of collective intelligence supported by human cumulative culture** (most of our intelligence is now at a civilization level, not a brain level). A hypothetical High-Level Machine Intelligence (HLMI) will surely accelerate the pace of those exponential progresses, but not to the point of an explosion over a human lifespan.
- On **shorter timescales like a human lifespan**, the self-improvement cycles initiated by a new HLMI will hit **some hard limits impeding a quick intelligence explosion**. For instance, an intelligent agent will still need a significant amount of time to experience its environment and to be trained before reaching its full potential: HLMI will not be able to rely only on knowledge databases to learn human social interactions, they will have to experience at least some of them at the pace of those interactions to ground their knowledge. Copying/duplicating their software will not be an option either for physically embodied agents

However, let's remember that such a potential intelligence explosion relies on the speculative hypothesis that we first achieve High-Level Machine Intelligence (HLMI).

Shape of the intelligence takeoff



Bostrom, 2014

Very unlikely to be shorter than a year.
Dozens of years would be more likely.

If it ever happens, intelligence explosion will not occur in a single day, a single week, and even a single year because of training periods that could not be time-compressed to interact with the real world at its own pace. The duration of the singularity would likely take at least a dozen years.

Exponential progress would also be capped by resource limitations and economical constraints.

Will an intelligent agent need a body?

Definitely yes, but it does not necessarily have to be physical. It can be a virtual body in a simulated world.

Embodied cognition – a growing scientific discipline – suggests that to understand the world, we must experience the world through our body.

Indeed, in order to deeply understand its environment, **an intelligent agent needs to manipulate meaningful mental representations**. Living organisms like us construct such meaningful concepts in a multi-step grounding process that seems to have no alternative:

- First, by grounding meaningful percepts through active sensing (perceiving by moving their sensors)
- Then, by using those grounded meaningful percepts to anchor progressively more abstract meaningful concepts (cognition).

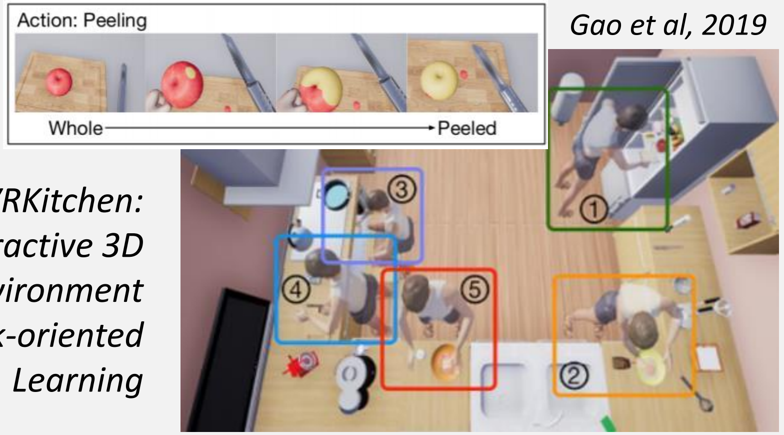
Similarly to living organisms, **a bio-inspired intelligent agent would need to act on its sensors to ground its artificial percepts, the essential primary step before machine intelligence.** And acting on its sensors implies that the agent has a body.

Though, I don't see physical embodiment as a necessity: **the intelligent agent could have a virtual body in a simulated world**, as long as it has something that corresponds to action and perception. Such a “software AI” could do sensorimotor interactions inside its virtual world.

However, cognition of software AI living inside a simulation would be another flavor of cognition, because its virtual perception will be grounded on different materials compared to ours. Even if it would still rely on similar principles, it would not share a shared background knowledge with us (a kind of alien intelligence).

Embodied intelligent agents inside virtual worlds

A virtual world mimicking the real-world environment



Gao et al, 2019

VRKitchen: an Interactive 3D Virtual Environment for Task-oriented Learning

Simulations are still (and will remain) far from real-world richness and complexities. Embodied software AI growing and learning in those environments will likely be limited in their capacity to extrapolate their behavior in the real world.

A virtual world with no real-world equivalent

An embodied software AI in charge of routing web traffic between internet servers could behave intelligently in its environment. The usefulness of virtual environment depends on the goal we are following.

They must have affects and could fake emotions. Feeling emotions depends on whether they are conscious or not.

Every living organism comes with some **built-in functions that drive their actions in their quest for survival and reproduction**. In humans, the homeostatic process keeps our heart rate, breathing, blood pressure, temperature, hormones and metabolism into an acceptable range despite constant external disruptions. If those variables get too far from their ideal values, an unpleasant signal motivates us to take adequate actions in order to reach again the pleasant signal. **Those signals are interoceptive affects that help us to maintain our body budget**.

Intelligent machines must also have a motivation function based on internal sensors. In that sense, they have affects. More, **machines need internal affects to hold their own goal of self-preservation** in a dynamic and unpredictable world.

Evolutionary more recent than affects, emotions like fear and happiness are mental tools that also help us to navigate through our social life. Following the theory of Lisa Barrett Feldman (and supported by Joseph Ledoux), **emotions are constructed concepts**, like the concepts of colors. This embodied knowledge depends on one's culture and experience.

Intelligent machines may construct emotional concepts to categorize and represent the sensed reality. The more concepts they construct, the more emotional granularity they have. They could also try to fake the cultural "fingerprints" of emotions of a given culture by mimicking their human counterparts.

However, the possibility that machines could feel these emotions is a different topic that depends on whether they could be conscious or not.

Unreliable stereotyped fingerprints of emotions

Source: J. Pass, 1821, after Charles Le Brun

Reading expression on faces is not as straightforward as one might think. **Facial movements increasingly appear as an inexact gauge of a person's emotions** (Barrett Feldman, 2019): the same emotions are not always expressed in the same way, the same facial expression do not reliably indicate the same emotions, the results are culture and context-dependent.

Even if the so-called emotional expressions do not reflect true emotions, machines could still try to fake emotion by exploiting our strong stereotypes that will mislead us.

Could machines become conscious?

Maybe! There seems to be no fundamental reason preventing physical machines from becoming conscious.

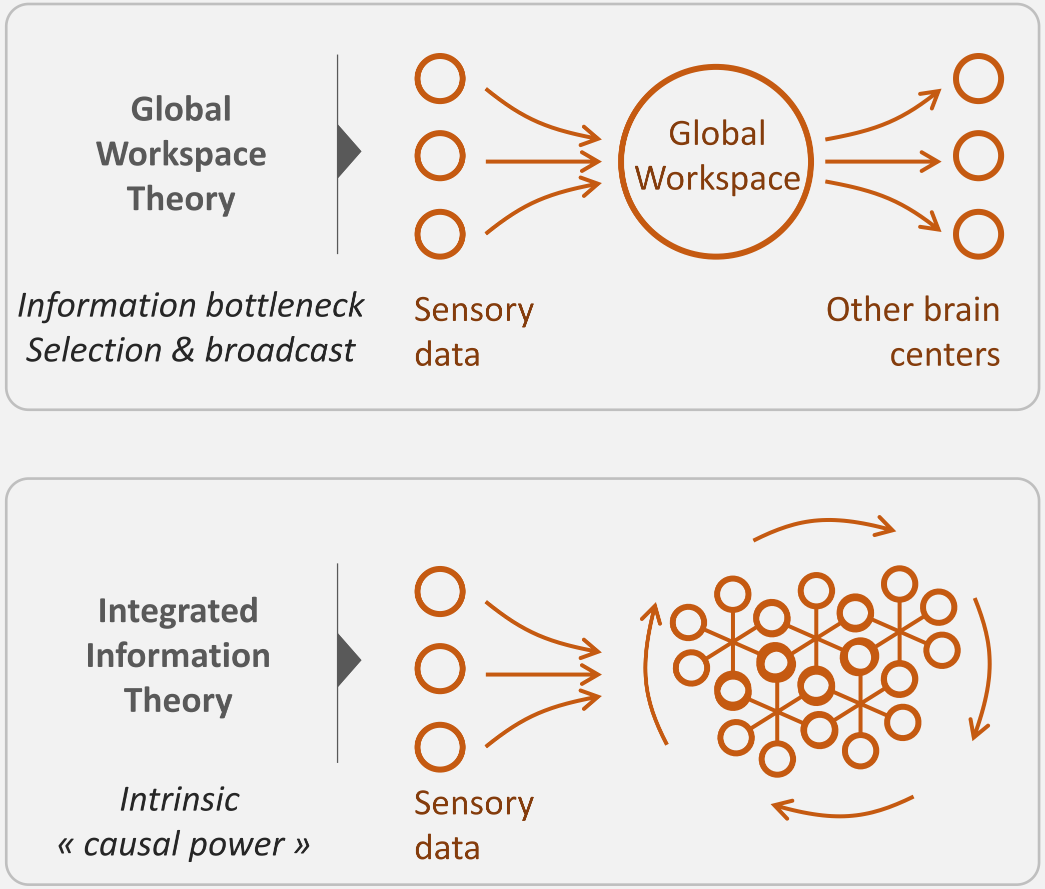
The existence of conscious machines is compatible with the two leading theories of consciousness: the **Global Workspace Theory (GWT)** by Stanislas Dehaene and the **Integrated Information Theory (IIT)** by Giulio Tononi.

Those theories do not embrace the same level. The GWT proposes that consciousness is a form of information processing that could be artificially replicated. When a piece of information enters the "global workspace" (supposedly in the prefrontal cortex), it can be selected and then be broadcasted back to the other centers. The selection process is what we perceive as consciousness. Stanislas Dehaene distinguishes two orthogonal dimensions of conscious computations: **global availability** via selection & broadcasting, and **self-monitoring** of those computations leading to subjective introspection.

The IIT is a more fundamental approach in which consciousness is an intrinsic property of matter that arises from the interconnectedness of brain networks that exert a causal power on themselves: the more complex a neural network, the more conscious it is. In this theory, machines have to be physical to access some level of consciousness, whereas consciousness could arise from nothing more than specific computations in a simulation according to the GWT.

But how would we know if a machine is conscious? How it feels to be a machine? Because consciousness is a subjective experience and because we cannot impersonate another living organism or a machine, we could only rely on consciousness correlates to assess consciousness. We may give a machine some level of consciousness depending on the chosen correlates, a highly debated topic.

Leading theories of consciousness



Illustrations adapted from Reading-Ikkanda, Quanta Magazine, 2019

Despite its ethical issues and potential threats, should machine intelligence be seen as a desirable target?

A more fundamental question is “how should we prepare ourselves for upcoming progress in machine intelligence?”

Numerous ethical issues and potential threats are already raised by the development of AI research and its business & governmental application. No need to be a visionary to predict that those topics will get increasingly problematic with a strong brain-inspired machine intelligence.

Though, the question of whether machine intelligence should be seen as a desirable target is a profound question but with very impractical solutions if the answer is no. The financial and geopolitical stakes of this winner-takes-all race for machine intelligence are too high to prevent other organizations or countries from following this goal, should it be pursued secretly.

Because **tremendous progresses in machine intelligence have already occurred and will continue anyway** (even if not High-Level Machine Intelligence), a more fundamental question is **how should we prepare our society to harness the current and upcoming impacts of this technology**.

Immediate societal issues are already around the corner: the redistribution of machine-produced wealth in a context of rising social inequalities, the transformation of the labor market with some significant job losses across large industries, the distortion of our sense of reality fueled by fake videos and audio recordings challenging democracies with potential threats of mass manipulation, the development of lethal autonomous weapons, to mention only a few.

Setting up safeguards to prevent malicious uses of this technology will obviously not be enough. First and foremost, we have to **strengthen the resilience of our societies**.

How to strengthen our resilience?

Some innocent food for thought

Becoming more flexible

Develop our ability to adapt at all levels: local to global, industry & institutions to civil society (mindsets open to changes, financially viable alternatives for unskilled workers, retraining and continual training, local ecosystem of skills…)

Reducing the attack surface

Develop industrial, institutional and civil awareness, improve critical thinking, prefer local solutions, reduce our reliance on digital technologies for physical vital infrastructures…

Because progress towards machine intelligence goes along with progress in our understanding of the brain, we should also get ready for fundamental questions about the nature of what makes us human, challenging our spiritual and philosophical beliefs.

Conclusion

- The road towards machine intelligence is inseparable from a mixed AI & neuroscience approach

Matthieu Thiboust

Art credit: Brainbow Hippocampus, Greg Dunn Design

Understanding the brain and building machine intelligence are two different objectives with a common path

Driven by the convergence of neuroscience and AI research, the road towards Machine Intelligence is a fascinating scientific endeavor we are currently witnessing. This document has described how intermingled this effort is with the fundamental attempt to understand the brain.

Despite the fast accumulation of huge amounts of data about brain structure and function, neuroscience still lacks a widely-accepted theoretical framework to interpret those findings and artificially replicate parts of brain functions. Because the brain is both integrated and composite, this framework would likely consist in global organizing principles on top of a collection of specialized theorical concepts and models.

Analyzing a complex system – like the brain – involves analyzing it at multiple and distinct levels of abstraction. Only by moving up and down this ladder of abstraction one can get a deep understanding of the system. The three-level hypothesis of David Marr has been very influential since the 1980s to investigate brain functions: the computational level (what does the system do), the algorithmic level (what algorithms does the system instantiate) and the implementation level (what hardware or substrate does the system run on). But not everyone agrees with this three-level sequence of stages that implies that the brain does represent information. According to György Buzsaki, the brain does not represent information, it constructs it: "*understanding of brain function should begin with brain mechanisms and explore how those mechanisms give rise to the performance we refer to as perception, action, emotion and cognitive function*".

Understanding how such brain functions emerge from simpler brain mechanisms will help to separate implementation details from fundamental inner workings in order to simplify the modeling (no need to simulate chemical reactions at the molecular level for instance). In fact, mimicking the right collection of brain mechanisms could lead to machine intelligence before we actually understand how the brain works (if we ever do). In return, this artificial replica of intelligence would give invaluable insights to neuroscience. This process will likely be iterative between AI and neuroscience, converging progressively towards more intelligent machines and a deeper understanding of the brain.

My deep conviction is that the road towards machine intelligence is now inseparable from a mixed AI & neuroscience approach.

Matthieu Thiboust

Behind the scene

Personal motivations and acknowledgments

Matthieu Thiboust

Personal motivations

As a datascientist, I got increasingly frustrated by the unjustified mediatic brouhaha about the impending Artificial General Intelligence (AGI) that would take over humanity. It made me dig deeper into the limits of current Artificial Intelligence (AI) approaches and the natural next step for me was to look into neuroscience. My first book on the subject, "*On Intelligence*" by Jeff Hawkins, profoundly piqued my curiosity and I rapidly became addicted to neuroscience books, reading dozens of them in the last few years. More recently, the very inspirational book "*The Brain from Inside Out*" by György Buzsaki changed my perspective on how to understand the brain and fueled new ideas on my side.

My fascination grew to the level that I decided to take a sabbatical leave dedicated to neuroscience. I intensified my readings of scientific papers, went to seminars, exchanged with people in the field, and finally felt the need to make some order in my notes. Wearing my consultant hat, I chose to digest this complex knowledge by making visual and synthetic slides. It is a good way to identify, collect, adapt and assemble the scattered existing pieces of the giant brain puzzle. Information exposed here is certainly not new, but I hope that this presentation format, in sharp contrast with classic scientific literature, can serve as a useful and more accessible document by the neuroscience & AI community.

Researchers are vigorously looking for new algorithms beyond *deep learning* to model real intelligence - still an elusive concept without widely adopted universal definition. I see two possible roads: the very hard one and the hard one. The very hard road is like finding blindly one of the exits of an unknown giant multi-dimensional maze. Virtually all directions will lead to a dead-end without knowing it until the long-lasting complete exploration of each segment. This is the current fundamental and abstract approach taken, by trying to integrate symbol-manipulation and causality principles in artificial neural networks. The hard road is the biological one. Evolution has done an incredible job to come up with living examples of intelligent agents. We can use insights from our brain to rapidly eliminate all biologically-incompatible segments in the maze. That's why I am convinced that understanding the brain is the quickest road towards Machine Intelligence, even if it may take hundreds of years.

I learnt a lot during this fantastic and intense journey. I now have more questions than I started with, but those are more informed, and they better motivate me to continue my own investigations.

Matthieu Thiboust

Acknowledgements

Writing this document turned out to be more difficult than initially thought. Putting aside the time investment it required, its content and its format were challenging, and I was lucky to found encouragement and help from the online neuroscience & AI communities.

Because the neuroscience content I addressed in this document is outside my initial area of expertise, I had first to read and digest a broad knowledge scattered among many books, scientific papers, blog posts, video lectures, podcasts and forums. Some highly debated topics were hard to sum up in a way to popularize complex neuroscience ideas to a classic AI practitioner, my primary target audience. I hope I have fulfilled this task without making misleading approximations.

The innovative format I chose also proved to be demanding in term of global coherence (at the level of pages and chapters), synthesis (one topic per page), illustrations (half self-made, half reuse with author’s or publisher’s permission), page layout (no editor/publisher) and language (English is not my native language as you probably guessed).

Even if I do not necessarily agree with all their ideas, I would like to thank the people who inspired me. I listed their names at the beginning of each chapter. I discovered the work of some of them a few years ago, while others popped out recently during the writing of this document. In alphabetical order, those inspiring people are *Subutai Ahmad, Lisa Barrett Feldman, Yoshua Bengio, Mark Browne, György Buzsáki, William Calvin, François Chollet, Paul Cisek, Antonio Damasio, Stanislas Dehaene, John E. Dowling, Karl Friston, Sten Grillner, Demis Hassabis, Jeff Hawkins, Geoffrey Hinton, David Hubel, Ole Jensen, Christof Koch, Matthew E. Larkum, Yann LeCun, Joseph Ledoux, Gary Marcus, Kevin Mitchell, Edvard Moser, May-Britt Moser, Vernon Mountcastle, Randall O’Reilly, Carlos E. Perez, Giovanni Pezzulo, Luis Puelles, Santiago Ramon y Cajal, Blake Richards, Gordon M. Shepherd, Murray Sherman, David Silver, Richard S. Sutton and Torsten Wiesel.*

I would like to address a special thanks to Mark Browne and Casey Walsh-Warder from the HTM community. Mark took the time to review the first draft versions of this document. In addition to his thoughtful comments and numerous reading advice, he regularly challenged the global organization and the content of this document. So did Casey with his in-depth review before publication.

Another thanking goes to Numenta for openly sharing their research ideas and animating an active community interested in neuroscience & AI on its forum (special thanks to its community manager Matt Taylor). Many of my ideas were firstly developed on this forum thanks to active forum members. I also feel grateful to my different proofreaders (with a special mention for Falco).

Lastly, I am extremely grateful to my fiancee, my family and my close friends who deeply supported my project.

References

References of materials that inspired me

Matthieu Thiboust

References & Recommended materials

Introduction

Artificial Intelligence needs a new momentum. Why not look at the brain?

- *Marcus, Gary, et Ernest Davis. Rebooting AI: building artificial intelligence we can trust. First edition, Pantheon Books, 2019.*

- *Chollet, François. « On the Measure of Intelligence ». arXiv:1911.01547 [cs], novembre 2019. arXiv.org, http://arxiv.org/abs/1911.01547.*
- *Heaven, Douglas. « Why Deep-Learning AIs Are so Easy to Fool ». Nature, vol. 574, no 7777, octobre 2019, p. 163-66. DOI.org (Crossref), doi:10.1038/d41586-019-03013-5.*
- *Marcus, Gary. « Deep Learning: A Critical Appraisal ». arXiv:1801.00631 [cs, stat], 1, janvier 2018. arXiv.org, http://arxiv.org/abs/1801.00631.*

- *Ganguli, Surya. The intertwined quest for understanding biological intelligence and creating artificial intelligence. Stanford HAI. https://neuroscience.stanford.edu/news/intertwined-quest-understanding-biological-intelligence-and-creating-artificial-intelligence*
- *Lomonaco, Vincenzo. Towards Neuroscience-Grounded Artificial Intelligence. https://towardsdatascience.com/towards-neuroscience-grounded-artificial-intelligence-9d592ace4314*
- *Thompson, Clive. How to Teach Artificial Intelligence Some Common Sense. Wired. https://www.wired.com/story/how-to-teach-artificial-intelligence-common-sense/*

References & Recommended materials

Brains and cognitive abilities

The primary function of a brain is not to think but to efficiently control complex behavior
This control is supported by abilities that were progressively acquired and refined through evolution

- *Damasio, Antonio R. The Strange Order of Things: Life, Feeling, and the Making of Cultures. 2019.*
- *LeDoux, Joseph E., et Caio Sorrentino. The deep history of ourselves: the four-billion-year story of how we got conscious brains. Viking, 2019.*
- *Mitchell, Kevin J. Innate: how the wiring of our brains shapes who we are. Princeton University Press, 2018.*

- *Cisek, Paul. « Resynthesizing Behavior through Phylogenetic Refinement ». Attention, Perception, & Psychophysics, vol. 81, no 7, octobre 2019, p. 2265-87. DOI.org (Crossref), doi:10.3758/s13414-019-01760-1.*
- *Grillner, Sten, et Abdeljabbar El Manira. « Current Principles of Motor Control, with Special Reference to Vertebrate Locomotion ». Physiological Reviews, vol. 100, no 1, janvier 2020, p. 271-320. DOI.org (Crossref), doi:10.1152/physrev.00015.2019.*
- *Puelles, Luis. « Thoughts on the Development, Structure and Evolution of the Mammalian and Avian Telencephalic Pallium ». Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, édité par T. Schilling et S. Wilson, vol. 356, no 1414, octobre 2001, p. 1583-98. DOI.org (Crossref), doi:10.1098/rstb.2001.0973.*

Brains and cognitive abilities

Biological intelligence gradually emerged with active perception and cognition

- *Buzsáki, György. The Brain from Inside Out. 1re éd., Oxford University Press, 2019. DOI.org (Crossref), doi:10.1093/oso/9780190905385.001.0001.*
- *Dehaene, Stanislas. Le code de la conscience. O. Jacob, 2014 (in French)*

- *Gershman, Samuel J. « What does the free energy principle tell us about the brain? » arXiv:1901.07945 [q-bio], octobre 2019. arXiv.org, http://arxiv.org/abs/1901.07945.*
- *Pezzulo, Giovanni, et Paul Cisek. « Navigating the Affordance Landscape: Feedback Control as a Process Model of Behavior and Cognition ». Trends in Cognitive Sciences, vol. 20, no 6, juin 2016, p. 414-24. DOI.org (Crossref), doi:10.1016/j.tics.2016.03.013.*
- *Pezzulo, Giovanni, et Cristiano Castelfranchi. « The Symbol Detachment Problem ». Cognitive Processing, vol. 8, no 2, mai 2007, p. 115-31. DOI.org (Crossref), doi:10.1007/s10339-007-0164-0.*

- *Perez, Carlos E. Bridging the Semantic Gap via Symbol Detachment. https://medium.com/intuitionmachine/solving-the-semantic-gap-via-the-symbol-detachment-problem-5293354d316f*

- *Friston, Karl. Free Energy Principle. Serious Science. https://www.youtube.com/watch?v=NIu_dJGyIQI*

References & Recommended materials

Brain general machinery

Neurons are sophisticated elementary components of the neural "hardware"
Neuron plasticity allows to retain memories of previous neural activity
Interconnected brain structures group neurons into organized network architectures

- *Dowling, John E. Understanding the brain: from cells to behavior to cognition. W.W. Norton & Company, 2018.*

- *Feldman, Daniel E. « The Spike-Timing Dependence of Plasticity ». Neuron, vol. 75, no 4, août 2012, p. 556-71. DOI.org (Crossref), doi:10.1016/j.neuron.2012.08.001.*
- *Irimia, Andrei, et al. « Circular Representation of Human Cortical Networks for Subject and Population-Level Connectomic Visualization ». NeuroImage, vol. 60, no 2, avril 2012, p. 1340-51. DOI.org (Crossref), doi:10.1016/j.neuroimage.2012.01.107*
- *Richards, Blake A., et Timothy P. Lillicrap. « Dendritic Solutions to the Credit Assignment Problem ». Current Opinion in Neurobiology, vol. 54, février 2019, p. 28-36. DOI.org (Crossref), doi:10.1016/j.conb.2018.08.003.*
- *Stuart, Greg J., et Nelson Spruston. « Dendritic Integration: 60 Years of Progress ». Nature Neuroscience, vol. 18, no 12, décembre 2015, p. 1713-21. www.nature.com, doi:10.1038/nn.4157.*

- *Browne, Mark. Grids into Maps. Numenta forum. https://discourse.numenta.org/t/grids-into-maps*
- *Humphries, Mark. Your Cortex Contains 17 Billion Computers. https://medium.com/the-spike/your-cortex-contains-17-billion-computers-9034e42d34f2*
- *Perez, Carlos E. Surprise! Neurons are Now More Complex than We Thought!!. https://medium.com/intuitionmachine/neurons-are-more-complex-than-what-we-have-imagined-b3dd00a1dcd3*
- *Terwilliger, Jack. Biological Neural Networks, Spiking Neurons. Interactive simulations. http://jackterwilliger.com/biological-neural-networks-part-i-spiking-neurons/*

Brain general machinery

Brain activity continuously loops across those structures through parallel pathways

- *Murray, Elisabeth A., et al. The Evolution of Memory Systems: Ancestors, Anatomy, and Adaptations. Oxford University Press, 2016 (Chapters 1&2)*

- *Ahissar, Ehud, et Eldad Assa. « Perception as a closed-loop convergence process ». eLife, édité par David Kleinfeld, vol. 5, mai 2016, p. e12830. eLife, doi:10.7554/eLife.12830.*
- *Buzsáki, György. « Time, Space and Memory ». Nature, vol. 497, no 7451, mai 2013, p. 568-69. www.nature.com, doi:10.1038/497568a.*
- *Cox, Julia, et Ilana B. Witten. « Striatal Circuits for Reward Learning and Decision-Making ». Nature Reviews Neuroscience, vol. 20, no 8, août 2019, p. 482-94. DOI.org (Crossref), doi:10.1038/s41583-019-0189-2.*

References & Recommended materials

Focus on the neocortex

The neocortex is divided into hundreds of functionally specialized but anatomically similar cortical areas

- *Molnár, Z. « Cortical Columns ». Neural Circuit Development and Function in the Brain, Elsevier, 2013, p. 109-29. DOI.org (Crossref), doi:10.1016/B978-0-12-397267-5.00137-0. (book chapter)*

- *Harris, Julie A., et al. « Hierarchical Organization of Cortical and Thalamic Connectivity ». Nature, vol. 575, no 7781, novembre 2019, p. 195-202. DOI.org (Crossref), doi:10.1038/s41586-019-1716-z.*
- *Kast, Ryan J., et Pat Levitt. « Precision in the Development of Neocortical Architecture: From Progenitors to Cortical Networks ». Progress in Neurobiology, vol. 175, avril 2019, p. 77-95. DOI.org (Crossref), doi:10.1016/j.pneurobio.2019.01.003.*
- *Mountcastle, V. « The columnar organization of the neocortex ». Brain, vol. 120, no 4, avril 1997, p. 701-22. DOI.org (Crossref), doi:10.1093/brain/120.4.701.*
- *Puelles, Luis, et al. « Concentric Ring Topology of Mammalian Cortical Sectors and Relevance for Patterning Studies ». Journal of Comparative Neurology, vol. 527, no 10, juillet 2019, p. 1731-52. DOI.org (Crossref), doi:10.1002/cne.24650.*

- *Browne, Mark. Basic system level diagram of the brain. Numenta forum. https://discourse.numenta.org/t/basic-system-level-diagram-of-the-brain*

- *Taylor, Matt. HTM School: Thousand Brains Theory & Hierarchy (Episode 16). Numenta. https://www.youtube.com/watch?v=mP7neeymcUY*

Focus on the neocortex

Cortical areas receive and send information in a laminar-specific way

- *Baker, Arielle, et al. « Specialized Subpopulations of Deep-Layer Pyramidal Neurons in the Neocortex: Bridging Cellular Properties to Functional Consequences ». The Journal of Neuroscience, vol. 38, no 24, juin 2018, p. 5441-55. DOI.org (Crossref), doi:10.1523/JNEUROSCI.0150-18.2018.*
- *Clascá, Francisco, et al. « Unveiling the Diversity of Thalamocortical Neuron Subtypes: Thalamocortical Neuron Diversity ». European Journal of Neuroscience, vol. 35, no 10, mai 2012, p. 1524-32. DOI.org (Crossref), doi:10.1111/j.1460-9568.2012.08033.x.*
- *Guillery, R. W., et S. Murray Sherman. « Branched Thalamic Afferents: What Are the Messages That They Relay to the Cortex? » Brain Research Reviews, vol. 66, no 1-2, janvier 2011, p. 205-19. DOI.org (Crossref), doi:10.1016/j.brainresrev.2010.08.001.*
- *Halassa, Michael M., et S. Murray Sherman. « Thalamocortical Circuit Motifs: A General Framework ». Neuron, vol. 103, no 5, septembre 2019, p. 762-70. www.cell.com, doi:10.1016/j.neuron.2019.06.005.*
- *Harris, Kenneth D., et Gordon M. G. Shepherd. « The Neocortical Circuit: Themes and Variations ». Nature Neuroscience, vol. 18, no 2, février 2015, p. 170-81. DOI.org (Crossref), doi:10.1038/nn.3917.*
- *Tanaka, Y. R., et al. « Local Connections of Excitatory Neurons to Corticothalamic Neurons in the Rat Barrel Cortex ». Journal of Neuroscience, vol. 31, no 50, décembre 2011, p. 18223-36. DOI.org (Crossref), doi:10.1523/JNEUROSCI.3139-11.2011.*
- *Usrey, W. Martin, et S. Murray Sherman. « Corticofugal Circuits: Communication Lines from the Cortex to the Rest of the Brain ». Journal of Comparative Neurology, vol. 527, no 3, février 2019, p. 640-50. DOI.org (Crossref), doi:10.1002/cne.24423.*
- *Young, Hedi, et al. Laminar-Specific Cortico-Cortical Loops in Mouse Visual Cortex. preprint, Neuroscience, 19 septembre 2019. DOI.org (Crossref), doi:10.1101/773085.*

- *Sherman, Murray. Thalamocortical System I. Recorded lecture. https://www.youtube.com/watch?v=aB2M1gg_1sU*
- *Sherman, Murray. Thalamocortical System II. Recorded lecture. https://www.youtube.com/watch?v=KBILhSTpzFI*

References & Recommended materials

Focus on the neocortex

A majority of long-distance projecting pyramidal neurons cohabits with a minority of local inhibitory cells

- *Kubota, Yoshiyuki, et al. « The Diversity of Cortical Inhibitory Synapses ». Frontiers in Neural Circuits, vol. 10, avril 2016. DOI.org (Crossref), doi:10.3389/fncir.2016.00027.*
- *Mohan, Hemanth, et al. « Dendritic and Axonal Architecture of Individual Pyramidal Neurons across Layers of Adult Human Neocortex ». Cerebral Cortex, vol. 25, no 12, décembre 2015, p. 4839-53. DOI.org (Crossref), doi:10.1093/cercor/bhv188.*
- *Tremblay, Robin, et al. « GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits ». Neuron, vol. 91, no 2, juillet 2016, p. 260-92. DOI.org (Crossref), doi:10.1016/j.neuron.2016.06.033.*

Focus on the neocortex

Functional neocortical circuits rely on laminar-specific lateral and radial interactions

- *Calvin, William H. The cerebral code: thinking a thought in the mosaics of the mind. MIT Press, 1996.*

- *Bosking, William H., et al. « Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex ». The Journal of Neuroscience, vol. 17, no 6, mars 1997, p. 2112-27. DOI.org (Crossref), doi:10.1523/JNEUROSCI.17-06-02112.1997.*
- *Martin, Kevan A. C., et al. « A Biological Blueprint for the Axons of Superficial Layer Pyramidal Cells in Cat Primary Visual Cortex ». Brain Structure and Function, vol. 222, no 8, novembre 2017, p. 3407-30. DOI.org (Crossref), doi:10.1007/s00429-017-1410-6.*
- *Narayanan, Rajeevan T., et al. « Beyond Columnar Organization: Cell Type- and Target Layer-Specific Principles of Horizontal Axon Projection Patterns in Rat Vibrissal Cortex ». Cerebral Cortex, vol. 25, no 11, novembre 2015, p. 4450-68. DOI.org (Crossref), doi:10.1093/cercor/bhv053.*
- *Rockland, Kathleen S. « About connections ». Frontiers in Neuroanatomy, vol. 9, mai 2015. DOI.org (Crossref), doi:10.3389/fnana.2015.00061.*
- *Shepherd, Gordon M., et Timothy B. Rowe. « Neocortical Lamination: Insights from Neuron Types and Evolutionary Precursors ». Frontiers in Neuroanatomy, vol. 11, novembre 2017, p. 100. DOI.org (Crossref), doi:10.3389/fnana.2017.00100.*

- *Browne, Mark. HTM Mini-Columns into Hexagonal Grids!. Numenta forum. https://discourse.numenta.org/t/htm-mini-columns-into-hexagonal-grids*

Focus on the neocortex

Sensory stimuli, motor actions and spatial navigation offer a window into the cortical code

- *Gu, Yi, et al. « A Map-like Micro-Organization of Grid Cells in the Medial Entorhinal Cortex ». Cell, vol. 175, no 3, octobre 2018, p. 736-750.e30. DOI.org (Crossref), doi:10.1016/j.cell.2018.08.066.*
- *Kerdels, Jochen, et Gabriele Peters. « A Survey of Entorhinal Grid Cell Properties ». arXiv:1810.07429 [q-bio], octobre 2018. arXiv.org, http://arxiv.org/abs/1810.07429.*

- *Moser, May-Britt. Nobel laureate lecture: Grid cells and cortical maps for space. https://www.youtube.com/watch?v=BEScyWMvSKk*
- *O'Keefe, John. Place Cells in the Hippocampus, Past and Present. https://www.youtube.com/watch?v=5IX5QAfqS2M*
- *Taylor, Matt. HTM School: Grid Cells (Episode 14). Numenta. https://www.youtube.com/watch?v=mP7neeymcUY*

Focus on the neocortex

The dynamics of cortical activity can only be analyzed in relation to brain oscillations

- *Buzsáki, György, et Edvard I. Moser. « Memory, Navigation and Theta Rhythm in the Hippocampal-Entorhinal System ». Nature Neuroscience, vol. 16, no 2, février 2013, p. 130-38. www.nature.com, doi:10.1038/nn.3304.*
- *Buzsáki, György, et Xiao-Jing Wang. « Mechanisms of Gamma Oscillations ». Annual Review of Neuroscience, vol. 35, no 1, juin 2012, p. 203-25. annualreviews.org (Atypon), doi:10.1146/annurev-neuro-062111-150444.*
- *Buzsáki, György, et David Tingley. « Space and Time: The Hippocampus as a Sequence Generator ». Trends in Cognitive Sciences, vol. 22, no 10, octobre 2018, p. 853-69. DOI.org (Crossref), doi:10.1016/j.tics.2018.07.006.*
- *Jensen, Ole, et Ali Mazaheri. « Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition ». Frontiers in Human Neuroscience, vol. 4, 2010. DOI.org (Crossref), doi:10.3389/fnhum.2010.00186.*
- *Klinzing, Jens G., et al. « Mechanisms of Systems Memory Consolidation during Sleep ». Nature Neuroscience, vol. 22, no 10, octobre 2019, p. 1598-610. DOI.org (Crossref), doi:10.1038/s41593-019-0467-3.*
- *Lisman, John E., et Ole Jensen. « The Theta-Gamma Neural Code ». Neuron, vol. 77, no 6, mars 2013, p. 1002-16. DOI.org (Crossref), doi:10.1016/j.neuron.2013.03.007.*
- *Spaak, Eelke, et al. « Layer-Specific Entrainment of Gamma-Band Neural Activity by the Alpha Rhythm in Monkey Visual Cortex ». Current Biology, vol. 22, no 24, décembre 2012, p. 2313-18. DOI.org (Crossref), doi:10.1016/j.cub.2012.10.020.*
- *Staudigl, Tobias, et al. « Saccades Are Phase-Locked to Alpha Oscillations in the Occipital and Medial Temporal Lobe during Successful Memory Encoding ». PLOS Biology, édité par Frank Tong, vol. 15, no 12, décembre 2017, p. e2003404. DOI.org (Crossref), doi:10.1371/journal.pbio.2003404.*

References & Recommended materials

Back to code

General: Interviews of key AI leaders & thinkers

- *Bengio, Yoshua. Deep Learning. Artificial Intelligence (AI) Podcast by Lex Fridman. https://www.youtube.com/watch?v=azOmzumh0vQ*
- *Chollet, François. Keras, Deep Learning, and the Progress of AI. Artificial Intelligence (AI) Podcast by Lex Fridman. https://www.youtube.com/watch?v=Bo8MY4JpiXE*
- *Hawkins, Jeff. Thousand Brains Theory of Intelligence. Artificial Intelligence (AI) Podcast by Lex Fridman. https://www.youtube.com/watch?v=-EVqrDlAqYo*
- *Kahneman, Daniel. Thinking Fast and Slow, Deep Learning, and AI. Artificial Intelligence (AI) Podcast by Lex Fridman. https://www.youtube.com/watch?v=UwwBG-MbniY*
- *LeCun, Yann. Deep Learning, Convolutional Neural Networks, and Self-Supervised Learning. Artificial Intelligence (AI) Podcast by Lex Fridman. https://www.youtube.com/watch?v=SGSOCuByo24*
- *Pearl, Judea. Causal Reasoning, Counterfactuals, Bayesian Networks, and the Path to AGI. Artificial Intelligence (AI) Podcast by Lex Fridman. https://www.youtube.com/watch?v=pEBI0vF45ic*

- *Botvinick, Matt. Neuroscience and AI at DeepMind. Brain-inspired podcast (episode 21). https://braininspired.co/podcast/21/*
- *George, Dileep. Vicarious Robot AI. Brain-inspired podcast (episode 13). https://braininspired.co/podcast/13/*
- *Hadsell, Raia. Robotics and Deep RL. Brain-inspired podcast (episode 45). https://braininspired.co/podcast/45/*
- *Richards, Blake. Deep Learning in the Brain. Brain-inspired podcast (episode 9). https://braininspired.co/podcast/9/*

Back to code

Next-level artificial neural networks model more realistic neurons, architectures and learning rules

- *Ahmad, Subutai, et Jeff Hawkins. « How do neurons operate on sparse distributed representations? A mathematical theory of sparsity, neurons and active dendrites ». arXiv:1601.00720 [cs, q-bio], mai 2016. arXiv.org, http://arxiv.org/abs/1601.00720.*
- *Ahmad, Subutai, et Luiz Scheinkman. How Can We Be So Dense? The Benefits of Using Highly Sparse Representations. mars 2019. arxiv.org, https://arxiv.org/abs/1903.11257v2.*
- *Bengio, Yoshua, et al. Towards Biologically Plausible Deep Learning. février 2015. arxiv.org, https://arxiv.org/abs/1502.04156v3.*
- *Hawkins, Jeff, et al. « A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex ». Frontiers in Neural Circuits, vol. 12, janvier 2019, p. 121. DOI.org (Crossref), doi:10.3389/fncir.2018.00121.*
- *Lotter, William, et al. « Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning ». arXiv:1605.08104 [cs, q-bio], février 2017. arXiv.org, http://arxiv.org/abs/1605.08104.*
- *Marblestone, Adam H., et al. « Toward an Integration of Deep Learning and Neuroscience ». Frontiers in Computational Neuroscience, vol. 10, septembre 2016. DOI.org (Crossref), doi:10.3389/fncom.2016.00094.*
- *O'Reilly, Randall C., et al. « Deep Predictive Learning: A Comprehensive Model of Three Visual Streams ». arXiv:1709.04654 [q-bio], septembre 2017. arXiv.org, http://arxiv.org/abs/1709.04654.*
- *Richards, Blake A., et al. « A Deep Learning Framework for Neuroscience ». Nature Neuroscience, vol. 22, no 11, novembre 2019, p. 1761-70. DOI.org (Crossref), doi:10.1038/s41593-019-0520-2.*
- *Whittington, James C. R., et Rafal Bogacz. « Theories of Error Back-Propagation in the Brain ». Trends in Cognitive Sciences, vol. 23, no 3, mars 2019, p. 235-50. DOI.org (Crossref), doi:10.1016/j.tics.2018.12.005.*

- *Lomonaco, Vincenzo. A Machine Learning Guide to HTM (Hierarchical Temporal Memory). Numenta blog. https://numenta.com/blog/2019/10/24/machine-learning-guide-to-htm*
- *Perez, Carlos E. The Emergence of Inside Out Architectures in Deep Learning. https://medium.com/intuitionmachine/controlled-hallucinations-in-deep-learning-architecture-fd617150d677*
- *Perez, Carlos E. How Human and Deep Learning Perception are Very Different. https://medium.com/intuitionmachine/our-minds-see-and-hear-only-what-we-imagine-dc303056171*

- *Hawkins, Jeff, et Ahmad, Subutai. The Thousand Brains Theory. Talk at Microsoft Research. https://numenta.com/resources/videos/thousand-brains-theory-of-intelligence-microsoft/*

References & Recommended materials

Back to code

The transition from artificial networks to artificial agents is a necessary step towards machine intelligence

- *Pearl, Judea, et Dana Mackenzie. The Book of Why: The New Science of Cause and Effect. 2018.*
- *Sutton, Richard S., et Andrew G. Barto. Reinforcement learning: an introduction. MIT Press, 1998.*

- *Miconi, Thomas, et al. « Backpropamine: training self-modifying neural networks with differentiable neuromodulated plasticity ». arXiv:2002.10585 [cs], février 2020. arXiv.org, http://arxiv.org/abs/2002.10585.*
- *Pearson, Martin J., et al. « Biomimetic Vibrissal Sensing for Robots ». Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 366, no 1581, novembre 2011, p. 3085-96. DOI.org (Crossref), doi:10.1098/rstb.2011.0164.*

- *Juliani, Arthur. The present in terms of the future: Successor representations in Reinforcement learning. https://medium.com/@awjuliani/the-present-in-terms-of-the-future-successor-representations-in-reinforcement-learning-316b78c5fa3*

- *Silver, David. Reinforcement Learning course. https://www.youtube.com/watch?v=2pWv7GOvuf0*
- *Sutton, Richard S. TD Learning. https://www.youtube.com/watch?v=LyCpuLikLyQ*

- *Summerfield, Christopher. How to build a brain from scratch. Department of Experimental Psychology, University of Oxford, UK. https://humaninformationprocessing.files.wordpress.com/2020/01/how-to-build-a-brain-from-scratch_all_lectures.pdf*

References & Recommended materials

Back to code

The potential emergence of machine intelligence already raises existential questions

- *Barrett, Lisa Feldman. How emotions are made: the secret life of the brain. Houghton Mifflin Harcourt, 2017.*
- *Dehaene, Stanislas, et al. La plus belle histoire de l'intelligence: des origines aux neurones artificiels : vers une nouvelle étape de l'évolution. 2018. (in French)*

- *Dehaene, Stanislas, et al. « What Is Consciousness, and Could Machines Have It? » Science (New York, N.Y.), vol. 358, no 6362, octobre 2017, p. 486-92. escholarship.org, doi:10.1126/science.aan8871.*
- *Heaven, Douglas. « Why Faces Don't Always Tell the Truth about Feelings ». Nature, vol. 578, no 7796, février 2020, p. 502-04. DOI.org (Crossref), doi:10.1038/d41586-020-00507-5.*
- *Man, Kingson, et Antonio Damasio. « Homeostasis and Soft Robotics in the Design of Feeling Machines ». Nature Machine Intelligence, vol. 1, no 10, octobre 2019, p. 446-52. DOI.org (Crossref), doi:10.1038/s42256-019-0103-7.*

- *Ball, Philip. Neuroscience Readies for a Showdown Over Consciousness Ideas. Quanta Magazine. 2019. https://www.quantamagazine.org/neuroscience-readies-for-a-showdown-over-consciousness-ideas-20190306/*
- *Chollet, François. The implausibility of intelligence explosion. 2017. https://medium.com/@francois.chollet/the-impossibility-of-intelligence-explosion-5be4a9eda6ec*
- *Cobb, Matthew. Why your brain is not a computer. The Guardian. https://www.theguardian.com/science/2020/feb/27/why-your-brain-is-not-a-computer-neuroscience-neural-networks-consciousness*
- *Koch, Christof. Proust among the Machines. Scientific American. Dec 2019. https://christofkoch.files.wordpress.com/2019/12/proust-among-the-machines-19.pdf*

Legal

License of my work & Illustration credits

Matthieu Thiboust

License

In order to give back to the community and to make my work easier to share, I decided to license my work under the CC BY-NC 4.0 Creative Commons license, a more permissive license than the common copyright notice “All rights reserved”.

In short, feel free to reuse and modify my work for a non-commercial use, as long as you give credit to my work. As a bonus, I would appreciate if you keep me informed of such use.

This license is applicable to all the content of this document (text & illustrations) except to the illustrations which are not entirely mine as mentioned in the “illustration credits” section.

Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)

This license lets you to use, remix, tweak, and build upon my work non-commercially, and although your new works must also acknowledge me and be non-commercial, you don’t have to license your derivative works on the same terms.

Illustration credits (1/5)

My own illustration are in green. If there is a mention of “by reusing” or “adapted”, you may need an additional permission from these authors to reuse the material.

Introduction

- P1: Book cover, By the author (by reusing illustration from Ramon y Cajal (1899))
- P3: Table of content, human and digital brains, unknow authors
- P4: Chapter cover, Brainbow Hippocampus, Greg Dunn Design (with permission)
- P5: Adversarial examples of stop signs, Eykholt (2017) , Robust Physical-World Attacks on Deep Learning Models
- P5: Adversarial examples of the sloth, Deep Mind (2019), Identifying and eliminating bugs in learned predictive models
- P5: Adversarial examples of abstract patterns, Nguyen (2014), Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images
- P5: Adversarial examples of GPT-2, By the author
- P7: Illustrations with robots, By the author (by reusing public domain images)
- P8: Bio-inspired AI vs biological intelligence, By the author (by reusing public domain images)
- P9: AI & neuroscience research fields, By the author

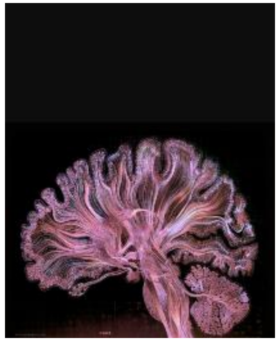
Brains and cognitive abilities

- P10: Chapter cover, Midas and the Bandsaw, Greg Dunn Design (with permission)
- P11: Nervous system, Adapted from Wikipedia
- P12: Brain structures along the phylogenetic tree, Adapted from Cisek (2019) and wikipedia, Resynthesizing behavior through phylogenetic refinement (with permission)
- P13: Action / Perception / Cognition, By the author
- P15: Brain evolution, Cisek (2019), Resynthesizing behavior through phylogenetic refinement (with permission)
- P16: Ancestral vertebrate & mammalian brain, Cisek (2019) (with permission)
- P17: Brain development, Mitchell (2018), Innate: How the Wiring of Our Brains Shapes Who We Are (with permission)
- P18: Brain development, By the author (by reusing illustrations from Mitchell 2018 and Kold 2009), Innate: How the Wiring of Our Brains Shapes Who We Are (with permission)
- P20: Perception in brain, By the author (by reusing public domain images)
- P20: Optical illusion, Adapted from Wikipedia
- P20: Suppression of footstep noise, By the author (by reusing public domain images)
- P21: Eye saccades, By the author (by reusing wikipedia images)
- P22: Grounding perception, By the author (by reusing public domain images)
- P23: Perception to cognition, By the author (by reusing public domain images)
- P24: Perception & cognition mechanisms, By the author

Illustration credits (2/5)

My own illustration are in green. If there is a mention of “by reusing” or “adapted”, you may need an additional permission from these authors to reuse the material.

Brain general machinery



- P25: Chapter cover, Self Reflected, Greg Dunn Design (with permission)
- P26: Neuron diversity, Braini (2016), adapted from Ramon y Cajal, Biophysical approach of neuronal shapes
- P27: Neural communication, By the author (by reusing image from Dowling, 2018), Understanding the Brain (with permission)
- P28: Neurotransmitter & neuromodulator, Adapted from Hirase (2014), Volume transmission signalling via astrocytes (with permission)
- P29: Dentritic computations, Adapted from Stuart & Spruston (2015), Dendritic integration: 60 years of progress (with permission)
- P30: Voltage spatial map, By the author
- P30: AP voltage time profile, Wikipedia
- P31: Rate coding, Adapted from Dowling (2018), Understanding the Brain (with permission)
- P31: Firing patterns, Izhikevich (2004) (with permission)
- P34: Synapse, Kessels (2015)
- P35: STDP, Feldman (2012) (with permission)
- P36: Non synaptic plasticity, By the author (Adapted from Dowling (2018), Understanding the Brain) (with permission)
- P37: Credit assignment, Richards and Lillicrap (2019), Dendritic solutions to the credit assignment problem (with permission)
- P39: Brain structures, The human protein atlas (2019) (with permission)
- P40: Drawings of brain structures, Ramon y Cajal (1900)
- P41: Brain connectome, Courtesy of the Laboratory of Neuro Imaging and Martinos Center for Biomedical Imaging, Consortium of the Human Connectome Project – www.humanconnectomeproject.org (with permission)
- P43: Neural loops, By the author (adapted from Buzsáki (2013), Time, space and memory) (with permission)
- P44: General sensortimotor diagram, By the author
- P44: Sensorimotor loops of the mouse vibrissal system, Ahissar (2016), Perception as a closed-loop convergence process (with permission)
- P45: Simplified view of parallel cortical basal ganglia thalamo cortical loops, By the author
- P45: List of basal ganglia loops, Adapted from Graham, Marray, Wise (2017), The Evolution of Memory Systems

Illustration credits (3/5)

My own illustration are in green. If there is a mention of “by reusing” or “adapted”, you may need an additional permission from these authors to reuse the material.

Focus on the neocortex

- P46: Chapter cover, Gold Cortex, Greg Dunn Design (with permission)
- P47: Pictures of animal brains, Courtesy of Kinser (2000) (with permission)
- P47: Cortical lobes, By the author
- P47: Unfolded macaque cortex, NeupsyKey, adapted from Van Essen (2001) (with permission)
- P48: Cortical sheet, Jolygon, Shutterstock
- P48: Macrocolumn, Unknow author
- P48: Minicolumn, Moutcastles (1997), The columnar organization of the neocortex.
- P48: Inside-out migration, Hippenmeyer (2014), Molecular Pathways Controlling the Sequential Steps of Cortical Projection Neuron Migration (with permission)
- P49: Layered composition of cortex, By the author (adapted from Ranson (1959) and Moutcastle (1997)
- P50: Cortical organization in other animals, Dugas-Ford (2012), Cell-type homologies and the origins of the neocortex (with permission)
- P50: Cytoarchitecture different in cortex, von Economo (1929), The Cytoarchitectonics of the Human Cerebral Cortex
- P50: Mapping cytoarchitecture on cortical regions, Fukutomi (2018), Neurite imaging reveals microstructural variations in human cerebral cortical gray matter (with permission)
- P51: Processing flow at a laminar level, By the author
- P51: Processing flow at cortex level, Adapted from Beul (2015) with added arrows, Towards a “canonical” agranular cortical microcircuit (with permission)
- P52: Perception in brain, By the author (by reusing public domain images)
- P52: Connectome of visual areas, Rees (2002), Neural correlates of consciousness in humans (with permission)
- P53: System level diagram of cortex, Browne (2019), Numenta forum (with permission)
- P55: Canonical cortical microcircuit, By the author
- P56: Corticocortical interactions, By the author
- P57: Subcortical inputs, By the author
- P58: Subcortical outputs, By the author
- P59: Thalamocortical interactions, By the author
- P61: Pyramidal neurons and inhibitory interneurons, Kubota (2016), The Diversity of Cortical Inhibitory Synapses (with permission)
- P61: L4 reconstruction, Motta (2019), Dense connectomic reconstruction in layer 4 of the somatosensory cortex (with permission)
- P62: Dendrites of pyramidal neurons, By the author (from an unknown pyramidal neuron image)
- P62: Diversity of pyramidal neurons, Ledergerber and Larkum (2010), Properties of Layer 6 Pyramidal Neuron Apical Dendrites (with permission)

Illustration credits (4/5)

My own illustration are in green. If there is a mention of “by reusing” or “adapted”, you may need an additional permission from these authors to reuse the material.

Focus on the neocortex

- P63: Distant axonal arbors, MouseLight Janelia (neuron AA0949)
- P63: Local Axonal arbors: Axons and Brain Architecture, 1st Edition, ISBN: 978-0-12-801393-9, adapted from Narayanan et al. (2015) (By permission of Oxford University Press)
- P64: Diverse morphologies, Sultan et al (2017), Generation of diverse cortical inhibitory interneurons
- P64: Diverse postsynaptic targets, Favuzzi et al (2019), Distinct molecular programs regulate synapse specificity in cortical inhibitory circuits (with permission)
- P64: Diverse firing patterns, Sultan et al (2017), Generation of diverse cortical inhibitory interneurons
- P66: Radial and lateral cortical interactions at different levels, By the author (from the minicolumn illustration of Mountcastles)
- P67: Radial interactions schema, By the author (from the minicolumn illustration of Mountcastles)
- P67: Radial interactions illustration, Adapted from Tanaka et al (2011), Local Connections of Excitatory Neurons to Corticothalamic Neurons in the Rat Barrel Cortex
- P68: Lateral and conical connections schema, By the author (from the minicolumn illustration of Mountcastles)
- P68: Lateral and conical connections illustration, Adapted from Jiang et al (2015), Principles of connectivity among morphologically defined cell types in adult neocortex (with permission)
- P69: Lateral recurrent connection schema, By the author (from the minicolumn illustration of Mountcastles)
- P69: Lateral recurrent connection illustration, Adapted from Martin et al (2017) , A biological blueprint for the axons of superficial layer pyramidal cells in cat primary visual cortex (with permission)
- P69: Orientation columns, Bosking et al (1997), Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex
- P72: Experimental setup, Purves (2001) (with permission)
- P72: Recording of activity, Orientation columns, Bosking et al (1997), Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex
- P72: Map of V1 orientation columns, Afgoustidis (2015), Monochromaticity of Orientation Maps in V1 Implies Minimum Variance for Hypercolumn Size (with permission)
- P73: Recording of head direction cells, Adapted from Page, 2017 (with permission) and Sharp 2001
- P74: Place field, Moser (2008)
- P74: Schematic illustration, Bottom part by the author, top part by Wagatsuma (2007)
- P75: Grid cell fields, Moser (2008)
- P75: Schematic scaling, orientation and phase, By the author
- P75: Grid cell fields at different scales, Moser (2008)
- P77: Brain waves, By the author (by reusing public domain images)
- P78: Layer specific cortical activity in A1, Sakata and Harris (2010), Laminar Structure of Spontaneous and Sensory-Evoked Population Activity in Auditory Cortex (with permission)
- P78: Layer specific cortical activity in V1, Spaak et al (2012), Layer-Specific Entrainment of Gamma-Band Neural Activity by the Alpha Rhythm in Monkey Visual Cortex (with permission)
- P79: Phase precession of place cells, By the author
- P80: Brain waves during wakefulness and sleep, By the author

Illustration credits (5/5)

My own illustration are in green. If there is a mention of “by reusing” or “adapted”, you may need an additional permission from these authors to reuse the material.

Back to code

- P81: Chapter cover, Cortical Circuitboard, Greg Dunn Design & Brian Edwards (with permission)
- P82: HTM neuron, Hawkins et al (2016), Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex (with permission)
- P82: Spiking neuron, Unknow author
- P83: Various network architectures, By the author
- P84: Various illustration, Ahmad et al (2019), How Can We Be So Dense? The Benefits of Using Highly Sparse Representations (with permission)
- P85: Hierarchy and network architecture, Adapted from Lin, 2017 (dense connections and ZigZag Net added), Feature Pyramid Networks for Object Detection (with permission)
- P86: STDP, Li, Miao et al, Scientific Reports (2014), Activity-Dependent Synaptic Plasticity of a Chalcogenide Electronic Synapse for Neuromorphic Systems (with permission)
- P86: Competitive learning, McClelland (2015), Feature Discovery by Competitive Learning
- P87: Contrasive hebbian learning, Detorakis (2018), Contrastive Hebbian Learning with Random Feedback Weights (with permission)
- P87: Target propagation, Bengio (2014), How Auto-Encoders Could Provide Credit Assignment in Deep Networks via Target Propagation (with permission)
- P87: Feedback alignment, Nøkland (2016), Direct Feedback Alignment Provides Learning in Deep Neural Networks (with permission)
- P89: Schematic illustration of RL, By the author
- P90: Temporal difference, By the author
- P90: Successor representation, Stachenfeld (2017), The hippocampus as a predictive map (with permission)
- P91: PredNet architecture, Lotter et al (2016) (with permission)
- P92: Starcraft video game, Blizzard entertainment
- P92: Shrewbot, Prescott (2011), Biomimetic vibrissal sensing for robots (with permission)
- P92: Robot arm, Princeton website
- P93: Curriculum training example, Juliani (2017), Introducing ML-Agents Toolkit v0.2: Curriculum Learning, new environments, and more (with permission)
- P93: Learning graphs, By the author
- P94: Example of network architecture optimization via evolutionary algorithms , Adapted from Wistuba (2018)
- P96: SinNNaker, Stromatias et al (2015), Scalable Energy-Efficient, Low-Latency Implementations of Trained Spiking Deep Belief Networks on SpiNNaker (with permission)
- P96: TrueNorth DARPA, Wikipedia (public domain)
- P98: HLMI predictions, Grace et al (2018), When Will AI Exceed Human Performance? Evidence from AI Experts (with permission)
- P99: Intelligence takeoff, adapted from Bostrom 2014
- P100: Virtual world example, Gao et al (2019), VRKitchen: an Interactive 3D Virtual Environment for Task-oriented Learning (with permission)
- P101: Stereotyped facial expressions, Pass, 1821, after Charles Le Brun
- P102: Consciousness theories, Adapted from Reading-Ikkanda, Quanta Magazine, 2019 , Neuroscience Readies for a Showdown Over Consciousness Ideas

Glossary

Making your way through the neuroscience & AI jargon

Matthieu Thiboust

Glossary (1/10)

An **Action Potential (AP)** is a propagating spike along its axon from the axon initial segment (near the soma) towards axon terminals. An Action Potential is a spike, but not every spike is an Action Potential.

Adaptative Learning is a paradigm in which the model adjusts its parameter in order to find a compromise between two goals: adapting to new tasks and enforcing stability to preserve knowledge from previous tasks.

Affects are basic biological signals that help living organisms to maintain their body budget in their quest for survival and reproduction.

The **Agranular Cortex** refers to the part of the cerebral cortex that does not contain a granular layer (for instance, the motor cortex).

The **Amygdala** is a brain structure often associated with quick reactive behaviors in response to potential threats.

An **Apical Dendrite** is an ascending dendritic branch extending towards the pial surface, contrary to common dendritic arbors around the soma (basal dendrites). They constitute a key characteristic of pyramidal neurons.

An **Apical Tuft** is the extremity of some Apical Dendrites that terminates in a tuft morphology.

Artificial General Intelligence (AGI): See HLMI.

Artificial Intelligence (AI): See Machine Intelligence.

Artificial Neural Networks are connectionist AI methods that attempt to mimic biological neural networks even if they remain far from their biological counterparts. They are organized in layers of artificial neurons.

An **Axon** (or Nerve Fiber) is the long and slender neuron part that conducts Action Potentials to the next connected neurons. It can be as long as one meter. The axon spreads from the Axon Initial Segment (near the soma) to many axon terminals at the other end.

Axon Collaterals are forks from a neuron main axon. They transmit the same neuronal signal to other brain structures.

Axon Initial Segment: See Axon.

Axonal terminals: See Axon.

Backpropagating Action Potential (bAP) are propagating spikes of dendrites travelling in the reverse direction: from the soma to dendrite terminals. Interactions between dendritic spikes and bAP are believed to be involved in synapse learning mechanisms.

The **Backpropagation Algorithm** is a widely used gradient-based algorithm in training feedforward ANN in supervised learning. It is not biologically-inspired and the neuroscience-grounded AI community is looking for alternatives.

Glossary (2/10)

Basal ganglia are a collection of nuclei in the brain often associated with motor control, motor learning and decision making.

Brain Waves: See Neural Oscillations.

Brodmann Areas are subparts of the Cerebral Cortex. There are 52 Brodmann areas per hemisphere in the human brain.

The **Cerebellum** is the brain structure containing most neurons (around 80% of neurons in human). It is often associated with motor control and motor learning.

The **Cerebral Cortex** is a two-dimensional thin sheet of neural tissue covering the outside of the brain in two hemispheres. All vertebrates possess a cerebral cortex, but its significance greatly increased in mammals, with the expansion of the part that is called neocortex (or isocortex to avoid the misconception of a mammalian innovation).

Cognition is an extension of perception for more abstract constructed mental representations. It adds the ability to form internal representations and use them to guide complex behaviors requiring abilities such as planning, thinking long term, building upon other's knowledge, making rational choices…

Common Sense is a sound judgment in practical matters that depends on a shared background knowledge inside a community.

Competitive Learning is a form of unsupervised learning in which neurons compete for the right to respond to a subset of the input data: if one neuron responds more strongly to a particular input it inhibits the output of the other neurons in the group.

Compositionality is the capacity to understand and produce novels combination from known components.

Connectionist AI regroups a collection of AI methods based on networks of relatively simple elements organized in a typical topology (like Artificial Neural Networks)

A **Connectome** is a macroscopic connectivity map between brain regions in the brain.

Consciousness is still an evasive concept referring to what we experience. It is related to concepts such as awareness, self-awareness, awareness of awareness, feeling, private thought, introspection…

Contrastive Hebbian Learning (CHL) is an alternative to the backpropagation algorithm to solve the credit assignment problem.

Convolutional Neural Network (CNN) are a class of ANN with shared-weights architecture and translation invariance characteristics. They are commonly used in image recognition tasks.

Corollary Discharge: See Efference Copy.

"**Cortex**" refers to the Cerebral Cortex (when the term is used alone). See Cerebral Cortex.

Cortical Areas are subparts of the Cerebral Cortex. There are around 180 cortical areas per hemisphere in the human brain.

Cortical Lobes are subparts of the Cerebral Cortex. There are 4 cortical lobes per hemisphere: frontal, temporal, parietal and occipital.

Cortical Sheet: See Cerebral Cortex.

The **Credit Assignment** is a process that computes the contribution of each neurons to the overall error. It answers to the following question: in a network of neurons, how to know which synapses to strengthen and which synapses to weaken when the outcome turned out to be bad?

Curriculum Learning is a paradign in which the model is trained by gradually more difficult tasks in order to increase the learning performance.

The **Cytoarchitecture** is the cellular composition of a brain tissue that can be observed under the microscope.

Deep Learning is a field of AI focused on the use of Deep Neural Networks.

Deep Neural Networks are Artificial Neural Networks with many layers of neurons (usually dozens of layers).

Dendrites are neuron parts that conduct electrical potentials generated by other neurons. In each neuron, there are different dendritic segments. Each dendritic segment has a dendritic arbor with many dendritic terminals.

Dendritic Segment: See Dendrite.

A **Dendritic Spike** is a propagating spike along some of its dendrites from axon terminals towards the soma. Dendritic spikes increase the probability of AP firing in the axon, but they do not assure it. NMDA spikes are examples of dendritic spikes.

Dendritic Terminals: See Dendrite.

An **Efference Copy** (or Corollary Discharge) is a copy of a motor command signal, going directly from motor to sensory brain areas. It is an essential information in order to predict the next sensory stimuli by taking into account the upcoming self-generated movements, in addition to the flow of sensory inputs.

The expression "**Embodied AI**" has been coined in reference to the expression “embodied cognition” that underlines the strong intertwinement of the mind and the body. In the AI community, embodied AI refers to those intelligent agents that learn from their own perspective with sensorimotor interactions (for example, a visual representation within an environment). The acquired knowledge of these agents is grounded in their artificial embodiment.

Emotions like fear and happiness are mental concepts that also help us to navigate through our social life. They are evolutionary more recent than affects.

Evolutionary Algorithms are a class of computational algorithms that use mechanisms inspired by biological evolution like natural selection, reproduction, mutation and recombination.

Evolutionary Tree: See Phylogenetic Tree.

An **Excitatory PostSynaptic Potential (EPSP)** is a postsynaptic potential that increases the probability of an action potential occurring in a postsynaptic neuron. EPSP are triggered by Excitatory Synapses.

Excitatory Synapse: See Excitatory PostSynaptic Potential (EPSP).

Feedback Alignment (FA) is an alternative to the backpropagation algorithm to solve the credit assignment problem.

Feedback Connections are neuronal projections in the opposite direction of the main processing flow (to previous layers). This vocabulary can be confusing because the same terms are also used to describe connections between areas of different hierarchical level (but main processing flow does not necessarily follow the level of abstraction)

Feedforward Connections are neuronal projections in the direction of the main processing flow (to next layers). This vocabulary can be confusing because the same terms are also used to describe connections between areas of different hierarchical level (but main processing flow does not necessarily follow the level of abstraction)

Fiber: See Axon.

Gated Recurrent Unit (GRU) networks are RNN with feedback connections adding greater memory abilities.

Our **Genome** is a collection of genes that encode developmental rules like a recipe specifying how to make a mature brain from neural stem cells. Those rules are executed in each cell by the sequential expression of specific genes depending on the cell surroundings, thanks to other genes ruling those conditional gene expressions (depending on chemical gradients).

The **Granular Cortex** refers to the part of the cerebral cortex that contains a granular layer (for instance the sensory cortices).

The **Granular Layer** refer to the fourth layer of the cerebral cortex (L4).

Grounding is an active process that attaches a meaning to a stimuli-induced neural activity that becomes a meaningful percept. Grounding is realized via sensorimotor interactions through time.

Glossary (5/10)

The historical **Hebbian Learning** rule postulates that when one neuron drives the activity of another neuron, the connection between these neurons is potentiated (often summarized as “cells that fire together wire together”).

High-Order Areas refer to areas representing abstract concepts compared to Low-Order Areas. For example. sensory areas are lower in the hierarchy than associative and motor areas. This classification can be refined at a finer level: for example, the primary visual area has a lower hierarchy level than the secondary visual area.

The **Hippocampus** is a brain structure often associated with memory consolidation and spatial memory.

High-Level Machine Intelligence (HLMI) is a still-innocent term referring to superhuman artificial abilities of machines. I prefer to use this term instead of Artificial General Intelligence (AGI) which has become a strongly loaded expression.

Homeostasis is a self-regulating process by which biological systems tend to maintain stability. It can be seen as an internal drive for survival of individual living organisms and their species as a whole.

Hypercolumn: See Macrocolumn.

The **Hypothalamus** is a brain structure deeply involved in the regulation of basic vital needs of the body like hunger, temperature, thirst, fatigue, sleep, circadian rhythms.

Infragranular Layers refer to deep layers L5 & L6 of the cerebral cortex, below the granular layer.

An **Inhibitory PostSynaptic Potential (IPSP)** is a postsynaptic potential that decreases the probability of an action potential occurring in a postsynaptic neuron. IPSP are triggered by Inhibitory Synapses.

Inhibitory Synapse: See Inhibitory PostSynaptic Potential (IPSP).

Innate Priors refer to hardcoded materials in our genome (our developmental recipe). We, humans, are not born with a blank state. We possess numerous innate priors – selected by natural selection over millions of generations and likely embedded in our DNA – that speed up our intellectual development during a lifetime. Innate Priors also refer to hardcoded characteristics and constraints in ANN.

Intelligence is still an elusive concept related to advanced abilities. There is no widely adopted universal definition in the scientific community.

The term "**Interneuron**" seems to have different meanings. I use the term "Interneuron" to refer to a neuron that influence activity within a limited, localized brain region (contrary to a projection neuron). Inhibitory neurons in the cerebral cortex are interneurons.

Laminar: See Layer.

Glossary (6/10)

Layers reflect an organizational design. Biological neurons are something grouped in layers (like the neurons in the cerebral cortex have a laminar organization). Artificial neurons are conceptually organized in layers (for instance, a feedforward connection links a neuron from a layer to another neuron in the next layer layer).

Long Short-Term Memory (LSTM) networks are RNN with feedback connections adding greater memory abilities.

Long Term Depression (LTD) produces long-lasting decreases in synaptic efficacy of excitatory synapses using the glutamate neurotransmitter (most excitatory synapses use glutamate).

Long Term Potentiation (LTP) produces long-lasting increases in synaptic efficacy of excitatory synapses using the glutamate neurotransmitter (most excitatory synapses use glutamate).

Low-Order Area: See High-Order Area.

Machine Intelligence is a still-innocent term referring to advanced artificial abilities of machines. I prefer to use this term instead of Artificial Intelligence (AI) which has become a strongly loaded expression.

Macrocolumns (also called Hypercolumn in some cortical areas) are ensembles of minicolumns (around 500 µm of diameter). The existence of this structure is not always clear in the cerebral cortex.

Membrane Potential (also called membrane voltage) is the difference of electric potential between the inner and outer part of a biological cell like a neuron.

Minicolumns are fundamental units that constitutes the cerebral cortex (around 2 mm long and 50 µm of diameter).

A **Myelinated Axon** is an axon covered with specific cells that strongly accelerate the propagation of Action Potentials.

Neocortex: See Cerebral Cortex

Nerve fiber: See Axon

Nerve Tracts (also called Fiber Tracts) are bundles of axons that form massive interconnections between brain areas.

Neural Oscillations (or Brain Waves) are rhythmic patterns of various frequencies that constitute neural activity. At the level of neural ensembles, synchronized activity of large numbers of neurons gives rise to macroscopic oscillations, which can be observed with non invasive methods like electro-encephalography (EEG) or magneto-encephalography (MEG).

Neurogenesis is the process by which new neurons are produced by neural stem cells. There is a significant neuron proliferation during the last months of human embryos before birth.

Glossary (7/10)

A **Neuromodulator** is a molecule that conveys slow and long-lasting point-to-many chemical signals at the synapse level. They induce biochemical changes in the postsynaptic neuron.

The "**Neuron**" term can refer to biological or artificial neurons. A biological neuron is an electrically excitable cell that transmits nerve impulses to other neurons (or muscles & gland cells). An artificial neuron is a simplified model of a biological neuron used in artificial neural networks.

Neuroscience-Grounded AI is an AI approach that attempts to make ANN more biologically realistic. This approach has the human brain as a reliable and invaluable guide to progress incrementally towards Machine Intelligence.

A **Neurotransmitter** is a molecule that conveys fast and ephemeral point-to-point chemical signals at the synapses level

NMDA Spike: See Dendritic Spike.

NREM Sleep is a sleep phases that alternates with REM Sleep phases. NREM sleep stands for non-REM sleep. It groups the other phases of sleep. See REM Sleep.

A **Nucleus** (plural "nuclei") is a structure grouping neurons. Neurons are segregated along a radial organization that is sometimes described as concentric layers.

The **Pallidum** is a part of basal ganglia.

Perception is our sensory experience of the world around us. They are constructed mental representations, not the veridical representations of the objective world. Organisms that perceive are able to associate a valence (goodness scale) to situations in order to select an appropriate behavior and flexibly adapt its execution.

Phase Coding is a neural code encoding information with the precise timing of spikes regarding a time reference based on slower oscillations. Some neurons fire individual action potentials that are precisely timed at a specific phase of neural oscillations in the surrounding cells (a process referred to as phase precession)

Phase Precession is a process by which some neurons fire individual action potentials that are precisely timed at a specific phase of neural oscillations in the surrounding cells.

A **Phylogenetic Tree** (also called Evolutionary Tree) is a branching diagram showing the evolutionary relationships among various biological species.

A **Postsynaptic Neuron** in a neuron that receives the neurotransmitter after it has crossed the synapse and may fire an action potential if the neurotransmitter is strong enough.

A **Presynaptic Neuron** is a neuron that releases the neurotransmitter at the synapse as a result of an action potential entering its axon terminal.

A **Projection Neuron** is a neuron that send its axon to distant brain targets. Projection neurons in the cerebral cortex are pyramidal neurons.

Pyramidal Neurons are excitatory neurons that constitutes around 75% of the neurons of the cerebral cortex. They have a characteristic apical dendritic arbor and project their axon over long distances to cortical and subcortical targets. They form an extensive network mainly among themselves.

Rate Coding is a neural code encoding information with the spike frequency. Typically, the intensity or salience of a feature is represented by the rate of firing.

Recurrent Connections are neuronal projections to areas of the same level (to same layer).

Recurrent Neural Networks (RNN) are a class of ANN where the output from previous step are fed as input to the current step (recurrent connections). They are commonly used in speech recognition and natural language processing (NLP).

Reinforcement Learning (RL) is a paradigm in which software agents learn to take actions in an environment so as to maximize a cumulative reward.

REM Sleep is a sleep phases that alternates with NREM Sleep phases. REM stands for Rapid-Eye-Movement. It is recognizable by rapid movements of the eyes, low muscle tone and a propensity of the sleeper to dream vividly.

Saccades are fast eye motions that direct the fovea which has much better acuity than the rest of the retina (around 5 saccades per second).

The **Self-Organizing-Map (SOM)** algorithm is a competitve learning algorithm that produces a low-dimensional & discretized representation of the input space that is called a map.

Self-supervised Learning is a paradigm in which the model use labels that are naturally part of the input data, instead of separate external labels. Streams of temporal data can be used to “self-train” a model by continuously predicting the future from the past, and then comparing the prediction vs the outcome at the next timestep. Note that it does not require pre-labelled data.

Sensorimotor Interaction is a process by which an agent actively interacts with its environment in order to gain knowledge via self-induced stimuli.

Short Term Plasticity produces short-lasting effects in synaptic strength of synapses.

The **Singularity** is an hypothetical point in time when an intelligent agent slightly surpasses human intelligence, and then recursively designs more intelligent agents in such a way that machines overtake human intelligence by orders of magnitude in a short period of time.

Skip Connections are neuronal projection bypassing neighboring layers. They acts as shortcut across the hierarchy.

The **Soma** refers to the cell body of a neuron cell.

Sparse Distributed Representations (SDR) are data structures enforcing the sparsity of the encoded data. They mimic the sparse activity occurring in the brain.

A **Spike** is a propagating depolarization of neuron membrane potential (=voltage) along its axon or some of its dendrites. Axonal spikes are called Action Potentials.

Spike Timing Dependent Plasticity (STDP) is a synaptic plasticity mechanism that involves both pre and postsynaptic mechanisms. The precise temporal order of activity between the two neurons matters.

Spiking Neural Networks are ANN that closely mimic natural neural networks for neuroscience research purposes. They are computationally-intensive.

The **Striatum** is a part of basal ganglia.

Supervised Learning is a paradigm in which the model is fed with labelled training data so that it can learn the association between input-output pairs. It requires a significant human intervention.

Supragranular Layers refer to upper layers L1 & L2/3 of the cerebral cortex, above the granular layer.

Symbolic AI regroups a collection of AI methods based on abstract logical operations upon symbols explicitly representing human knowledge in a declarative form.

The “**Synapse**" term can refer to biological or artificial synapses. A biological synapse is the junction between two neurons (or between a neuron and a muscle/grand cell). A synapse involve an axon terminal (on the presynaptic neuron) and a dendritic terminal (on the postsynaptic neuron). An artificial synapse represent the connection between two artificial neurons. Each synapse has a synaptic weight.

Synaptic Plasticity is a biological mechanism that induces adaptions in synaptic characteristics by weakening/strengthening synaptic strength and creating/pruning synapses.

Synaptic Pruning is the process by which synapses are eliminated. It mainly occurs between early childhood and the onset of puberty in many mammals.

The **Synaptic Strength** (also called synaptic weight) characterizes the impact level of a synapse on the postsynaptic neuron.

Glossary (10/10)

Synaptic Weight: See Synaptic Strength.

Synaptogenesis is the process by which new synapses are formed between neurons. There is a significant synaptogenesis during early childhood in many mammals, followed by a major synaptic pruning.

Target Propagation (TP) is an alternative to the backpropagation algorithm to solve the credit assignment problem.

Temporal Difference Learning (TD Learning) is “learning a prediction from another, later, learned prediction”. This scalable & online learning algorithm is well adapted to real-life general multi-step prediction problems where the feedback is delayed or even not reached

The **Thalamus** is the gateway to the neocortex. It routes and gates the inputs it receives from nearly all brain structures.

Tuft: See Apical Tuft.

Unsupervised Learning is a paradigm in which the model is only fed with unlabeled data, with no human intervention.

The **White Matter** refers to areas that are mainly made up of axons. There are significant volumes of white matter underlying the cerebral neocortex.

The **Winner-Take-All (WTA)** algorithm is a competitive learning algorithm by which neurons of a layer compete with each other for activation. Only the best players win the right to stay active while the other neurons are shut down.

Versioning

History of modifications

Matthieu Thiboust

Dec 2019	Draft	First complete draft
April 2020	v1	Initial release
June 2020	v1.0.1	Typos & corrections (“place cells” → “grid cells”, p75)
Nov 2020	v1.0.2	Typos & corrections (“LTD” → “LTP”, p34)
Feb 2021	v1.0.3	Typos & corrections (“supragranular” → “granular”, p78)

Insights from the brain:

The road towards Machine Intelligence

Matthieu Thiboust

April 2020

To be continued…

I would be happy to read your comments, answer your questions, correct the errors that you may have spotted, add key missing elements to the document, or just discuss machine intelligence & neuroscience with you.

Do not hesitate to contact me:

✉ *matthieu.thiboust@gmail.com*

@mthiboust

in *Matthieu Thiboust*