Cognition and Neuroscience (Module 2)

Last update: 02 June 2024

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

Contents

1 Object recognition 1

1.1 Pathways 2

1.2 Neuron receptive field 3

1.3 Retina cells 5

1.4 Area V1 cells 7

1.4.1 Simple cells 7

1.4.2 Complex cells 7

1.4.3 End-stopped (hypercomplex) cells 8

1.4.4 Ice cube model 8

1.5 Extrastriate visual areas 8

1.5.1 Area V4 10

1.5.2 Inferior temporal cortex (IT) 10

1.5.3 Local vs distributed coding 12

2 Object recognition emulation through neural networks 18

2.1 Convolutional neural networks 18

2.2 Recurrent neural networks 20

2.2.1 Object recognition 20

2.2.2 Visual pattern completion 22

2.3 Unsupervised neural networks 25

3 Dopamine in reinforcement learning 29

3.1 Decision making 29

3.2 Reinforcement learning 31

3.2.1 RL classes 31

3.3 Dopaminergic system 32

3.4 Reward prediction error (RPE) theory of dopamine 35

3.4.1 Dopamine is not fully model-free 35

3.4.2 Dopamine as generalized prediction error 39

3.4.3 Successor representation 40

3.5 Distributional reinforcement learning 41

4 Deep reinforcement learning 45

4.1 Deep reinforcement learning applications 45

4.1.1 Inefficiency factors 46

4.2 Episodic reinforcement learning 47

4.3 Meta-learning 50

Bibliography 53

1 Object recognition

Vision Process that, from images of the external world, produces a description without irrelevant information (i.e. interference) useful to the viewer.

Vision

This description includes information such as what is in the world and where it is.

> **Remark.** Vision is the most important sense in primates (i.e. in case of conflicts between senses, vision is usually prioritized).

> **Remark.** Vision is also involved in memory and thinking.

> **Remark.** The two main tasks for vision are:
>
> - Object recognition.
> - Guiding movement.
>
> These two functions are mediated by (at least) two pathways that interact with each other.

Vision Bayesian modeling Vision can be modeled using Bayesian theory.

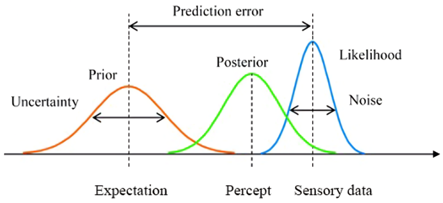
Bayesian modeling

Given an image I and a stimulus S, an ideal observer uses some prior knowledge (expectation) of the stimulus ($\mathcal{P}(S)$) and input sensory data ($\mathcal{P}\left(I|S\right)$) to infer the most probable interpretation of the stimulus in the image:

$$\mathcal{P}(S|I) = \frac{\mathcal{P}(I|S)\mathcal{P}(S)}{\mathcal{P}(I)}$$

> **Remark.** Prior knowledge is learned from experience. It could be related to the shape of the object, the direction of the light or the fact that objects cannot overlap.

> **Remark.** If the image is highly ambiguous, prior knowledge contributes more to disambiguate it.



Feed-forward processing Involves the likelihood $\mathcal{P}\left(I|S\right)$.

Feed-back processing Involves the prior $\mathcal{P}(S)$.

> **Remark.** Perception integrates both feed-forward and feed-back processing.

Vision levels A visual scene is analyzed at three levels:

Vision levels

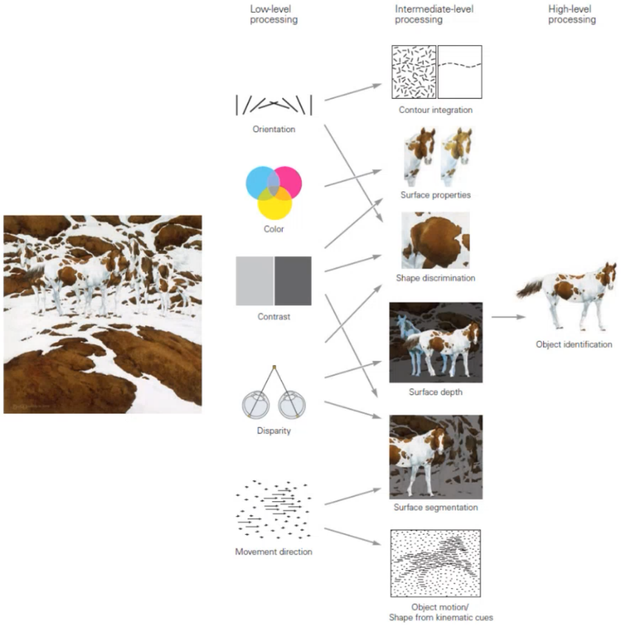
Low level Processes simple visual attributes captured by the retina such as local contrast, orientation, color, depth and motion.

Intermediate level Low-level features are used to parse the visual scene (i.e. local features are integrated into the global image). This level is responsible for identifying boundaries and surfaces belonging to the same object and discriminating between foreground and background objects.

High level Responsible for object recognition.

Once the objects have been recognized, they can be associated with memories of shapes and meaning.

> **Case study** (Agnosia)**.** Patients with agnosia have their last level of vision damaged. They can see (e.g. avoid obstacles) but cannot recognize objects or get easily confused.

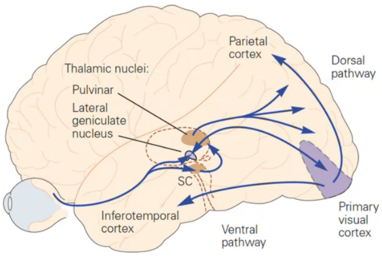


1.1 Pathways

Retino-geniculo-striate pathway Responsible for visual processing. It includes the:

Retino-geniculo-striate pathway

- Retina.
- Lateral geniculate nucleus (LGN) of the thalamus.
- Primary visual cortex (V1) or striate cortex.
- Extrastriate visual areas (i.e. beyond the area V1).



Ventral pathway Responsible for object recognition. It extends from the area V1 to the temporal lobe (feed-forward processing).

Ventral pathway

> **Remark.** This pathway emphasizes color.

> **Remark.** The connection from the frontal lobe encodes prior knowledge (feed-back processing).

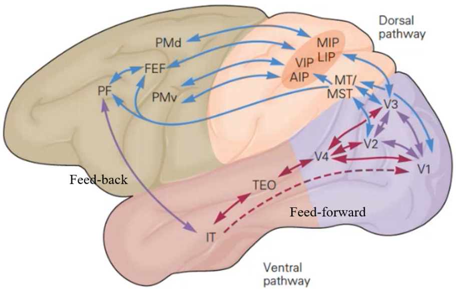
Dorsal pathway Responsible for movement guiding. It connects the V1 area with the parietal lobe and then with the frontal lobe.

Dorsal pathway

> **Remark.** This pathway is colorblind.

> **Remark.** The ventral and dorsal pathways are highly connected and share information.

> **Remark.** All connections in the ventral and dorsal pathways are reciprocal (i.e. bidirectional).



1.2 Neuron receptive field

Single-cell recording Technique to record the firing rate of neurons. A fine-tipped electrode is inserted into the animal's brain to record the action potential of a single neuron.

Single-cell recording

This method is highly invasive but allows to obtain high spatial and temporal readings of the neuron firing rate while distinguishing excitation and inhibition.

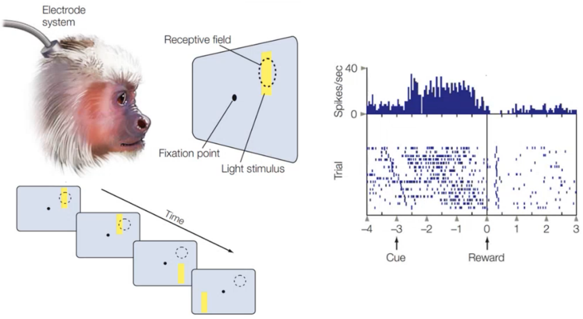
> **Remark.** On a theoretical level, neurons can fire a maximum of 1000 times per second. This may actually happen in exceptional cases.

Receptive field Region of the visual scene at which a particular neuron will respond if a stimulus falls within it.

Receptive field

> **Remark.** The receptive field of a neuron can be described through a Gaussian.

> **Case study.** A monkey is trained to maintain fixation at a point on a screen. Then, stimuli are presented in various positions of the visual field.
> It has been seen that a particular neuron fires vigorously only when a stimulus is presented in a particular area of the screen.
> The response is the strongest at the center of the receptive field and gradually declines when the stimulus moves away from the center.



> **Remark.** Neurons might only react to a specific type of stimuli in the receptive field (e.g. color, direction, . . .).
>
> > **Case study.** It has been seen that a neuron fires only if a stimulus is presented in its receptive field while moving upwards.

Retinotopy Mapping of visual inputs from the retina (i.e. receptive field) to the neurons.

Retinotopy

There is a non-random relationship between the position of the neurons in the visual areas (V1, V2, V4): their receptive fields form a 2D map of the visual field in such a way that neighboring regions in the visual image are represented by adjacent regions of the visual cortical area (i.e. the receptive fields are spatially ordered).

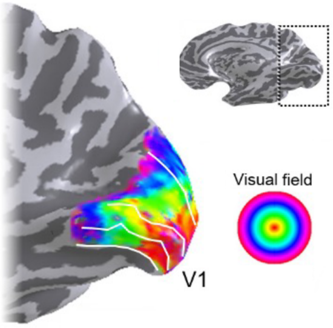


Figure 1.1: Mapping from the visual field to the neurons in the primary visual cortex (V1)

Eccentricity The diameter of the receptive field is proportional to the wideness of the visual angle.

Eccentricity

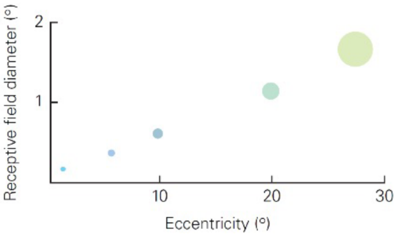


Figure 1.2: Relationship between visual angle and receptive field diameter

Cortical magnification The neurons (retinal ganglion cells, RGCs) responsible for the center of the visual field (fovea) have a visual angle of about 0.1° while the neurons at the visual periphery reach up to 1° of visual angle.

Cortical magnification

Accordingly, more cortical space is dedicated to the central part of the visual field. This densely packed amount of smaller receptive fields allows to obtain the highest spatial resolution at the center of the visual field.

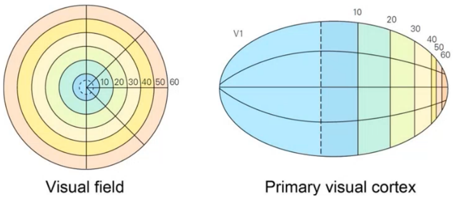


Figure 1.3: Cortical magnification in V1

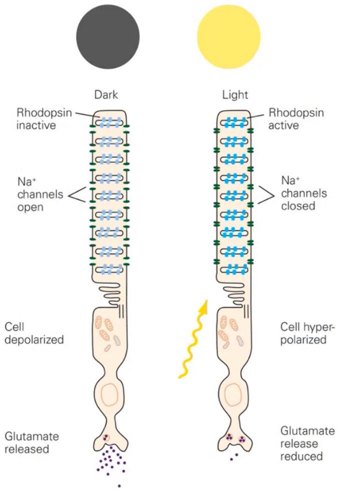
> **Remark.** The brain creates the illusion that the center and the periphery of vision are equal. In reality, the periphery has less resolution and is colorblind.

Hierarchical model of receptive field The processing of some information in the visual image is done through an incremental convergence of information in a hierarchy of receptive fields of neurons. Along the hierarchy the size of the receptive field increases.

1.3 Retina cells

Photoreceptor Specialized neurons that are hyperpolarized in bright regions and depolarized in dark regions.

Photoreceptor



> **Remark.** They are the first layer of neurons in the retina.

Retinal ganglion cell (RGC) Neurons of the visual cortex with a circular receptive field. They are categorized into:

Retinal ganglion cell (RGC)

- **ON-center** RGCs that are activated in response to a bright stimulus in the center of the receptive field.

ON-center cells

- **OFF-center** RGCs that are activated in response to a dark stimulus in the center of the receptive field.

OFF-center cells

Remark. They are the last layer of neurons in the retina, before entering LGN of the thalamus.

The receptive field of RGCs is composed of two concentric areas. The inner one acts according to the type of the cell (ON-center/OFF-center) while the outer circle acts antagonistically to the inner area.

Remark. A uniform stimulus that covers the entire receptive field of an RGC produces a weak or no response.

Remark. RGCs are not responsible for distinguishing orientation or lines.

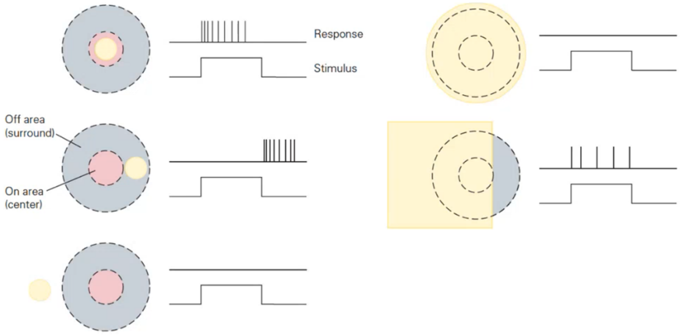
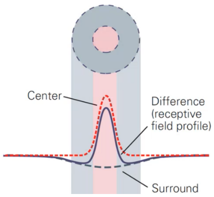


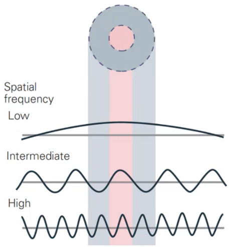
Figure 1.4: Responses of an ON-center RGC

Remark. The response of RGCs can be described by two Gaussians: one is positive with a narrow peak and represents the response at the center while the other is negative with a wide base and covers both the inner and outer circles. Their difference represents the response of the cell (receptive field profile).



Band-pass behavior The visual system of humans has a band-pass behavior: it only responds to a narrow band of intermediate frequencies and is unable to respond to spatial frequencies that are too high or too low (as they get canceled by the two antagonistic circles).

Band-pass behavior

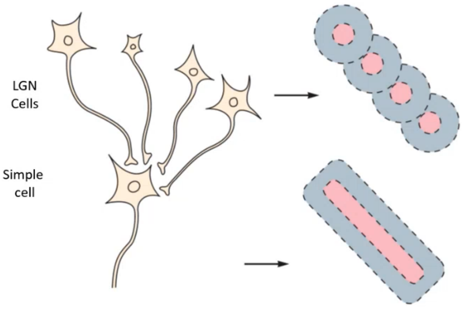


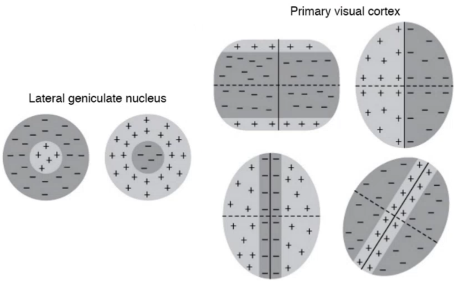
1.4 Area V1 cells

1.4.1 Simple cells

Simple cells

Neurons that respond to a narrow range of orientations and spatial frequencies.
This is the result of the alignment of different circular receptive fields of the LGN cells (which in turn receive their input from RGCs).

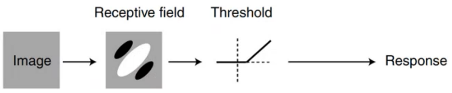




Case study. In monkeys, the neurons of the LGN have non-oriented circular receptive fields. Still, simple cells are able to perceive specific orientations.

Simple cells model The stages of computation in simple cells are:

1. Linear filtering through a weighted sum of the image intensities done by the receptive field (i.e. convolutions).
2. Rectification (i.e. thresholding with non-linearity) to determine if the neuron has to fire.

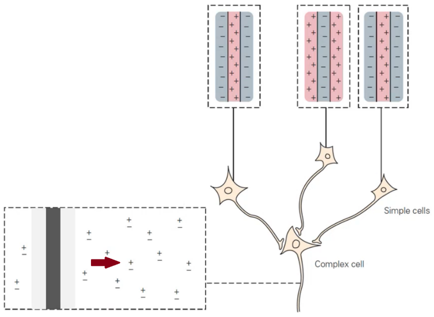


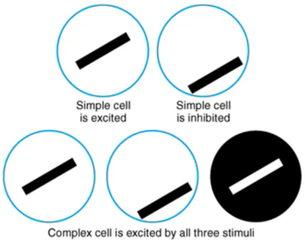
1.4.2 Complex cells

Complex cells

Neurons with a rectangular receptive field larger than simple cells. They respond to linear stimuli with a specific orientation and with a specific movement direction (position invariance).

Remark. At this stage, the position of the stimulus is not relevant anymore as the ON and OFF zones of the previous cells are mixed.

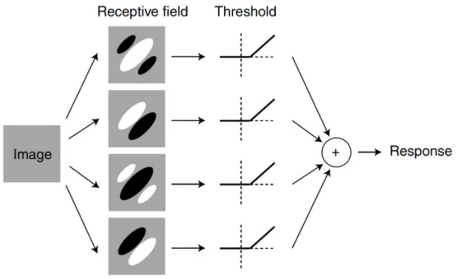




Complex cell model The stages of computation in complex cells are:

1. Linear filtering of multiple receptive fields.

2. Rectification for each receptive field.
3. Summation of the response.



1.4.3 End-stopped (hypercomplex) cells

End-stopped cells

Neurons whose receptive field has an excitatory region surrounded by one or more inhibitory regions all with the same preferred orientation.
This type of cell responds to short segments, long curved lines (as the tail of the curve that ends up in the inhibition region is not the preferred orientation) or to angles.

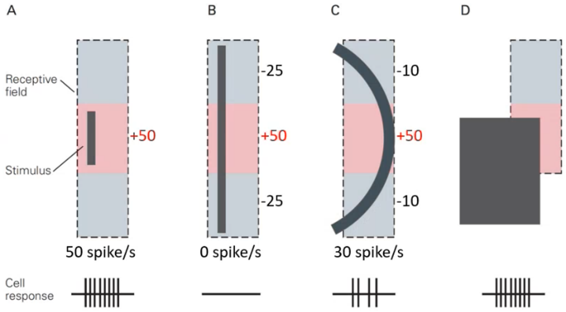
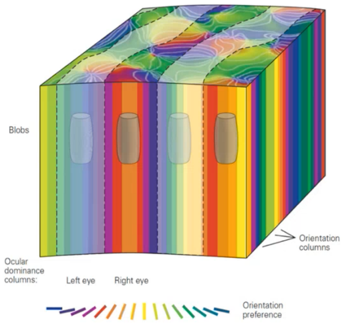


Figure 1.7: End-stopped cell with a vertical preferred orientation

1.4.4 Ice cube model

Each 1 mm of the visual cortex can be modeled through an ice cube module that has all the neurons for decoding all the information (e.g. color, direction, ...) in a specific location of the visual scene (i.e. each cube is a block of filters).



1.5 Extrastriate visual areas

Extrastriate visual areas Areas outside the primary visual cortex (V1). They are responsible for the actual object recognition task.

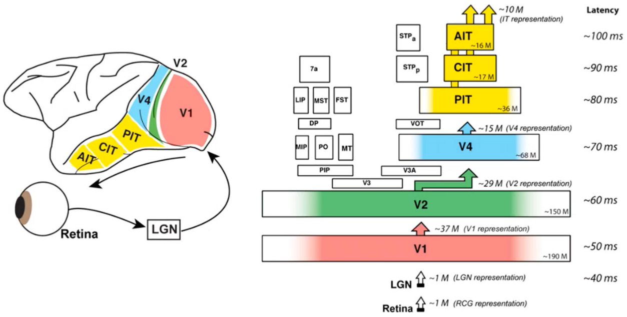


Figure 1.8: Ventral pathway

Visual object Set of visual features (e.g. color, direction, orientation, . . .) perceptually grouped into discrete units.

Visual object

Visual recognition Ability to assign a verbal label to objects in the visual scene.

Identification Recognize the object at its individual level.

Identification

Categorization Recognize the object as part of a more general category.

Categorization

> **Remark.** In humans, categorization is easier than identification. Stimuli originating from distinct objects are usually treated as the same on the basis of past experience.

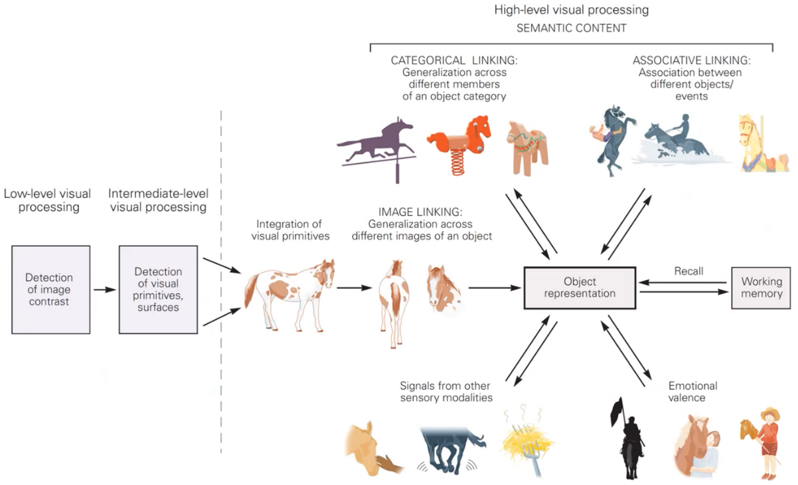


Figure 1.9: Processes that start after an object has been recognized

> **Remark.** For survival reasons, after an object has been recognized, its emotional valence is usually the first thing that is retrieved to determine if the object is dangerous.

Object recognition requires both the following competing properties:

Selectivity Different responses to distinct specific objects.

Selectivity

Consistency Similar responses to transformations (e.g. rotation) of the same object (generalization).

Consistency

Core object recognition Ability to rapidly (< 200 ms) discriminate a given visual object from all the other possible objects.

Core object recognition

Remark. Primates perform this task exceptionally well even if the object is transformed.

Remark. 200 ms is the time required to move the eyes. Experiments on core object recognition don't want candidates to move their eyes. Moreover, it prevents feedback processing from starting.

1.5.1 Area V4

Area V4

Intermediate cortical area responsible for visual object recognition and visual attention. It facilitates figure-ground segmentation of the visual scene enabling both bottom-up and top-down visual processes.

1.5.2 Inferior temporal cortex (IT)

Inferior temporal cortex (IT)

Responsible for object perception and recognition. It is divided into three areas:

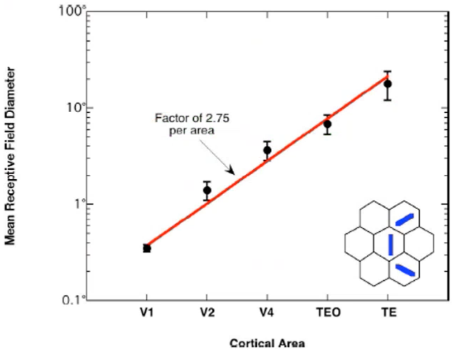
- Anterior IT (AIT).
- Central IT (CIT).
- Posterior IT (PIT).

Remark. It takes approximately 100 ms for the signal to arrive from the retina to the IT.

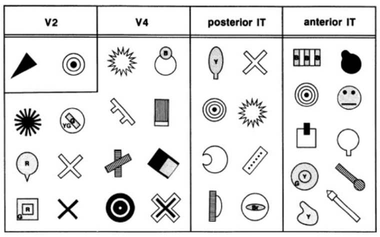
Remark. The number of neurons decreases from the retina to the IT. V1 can be seen as the area that sees everything and decides what to further process.

Remark. Central IT and anterior IT do not show clear retinotopy. Posterior IT shows some sort of pattern.

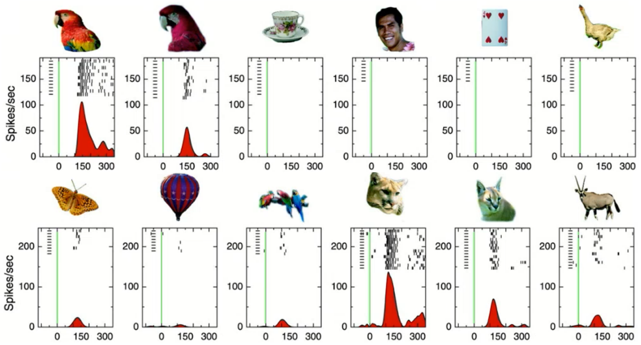
Remark. Receptive field scales by a factor of ~ 3 after passing through each cortical area.



Remark. It is difficult to determine which stimuli trigger the neurons in the IT and what actual stimuli trigger the IT is still unclear.
Generally, neurons in this area respond to complex stimuli, often biologically relevant objects (e.g. faces, hands, animals, ...).

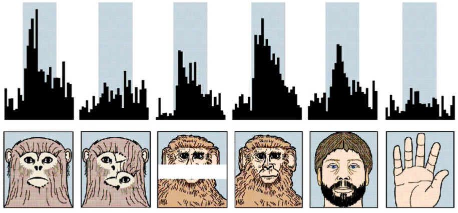


(a) Stimuli that trigger specific neurons of the ventral pathway

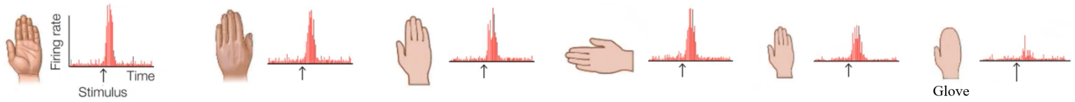


(b) Responses of a specific IT neuron to different stimuli

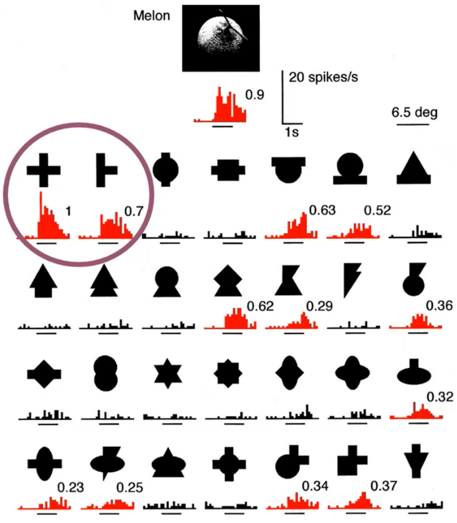
Case study (IT neurons in monkeys)**.** Several researchers observed that a group of IT neurons in monkeys respond selectively to faces. The response is stronger when the full face is visible and gets weaker if it is incomplete or malformed.



It also has been observed that an IT neuron responds to hands presented at various perspectives and orientations. A decrease in response is visible when the hand gets smaller and it is clearly visible when a glove is presented.



Case study (IT neuron response to a melon)**.** An IT neuron responds to a complex image of a melon. However, it has been shown that it also responds to simpler stimuli that represent the visual elements of the melon.

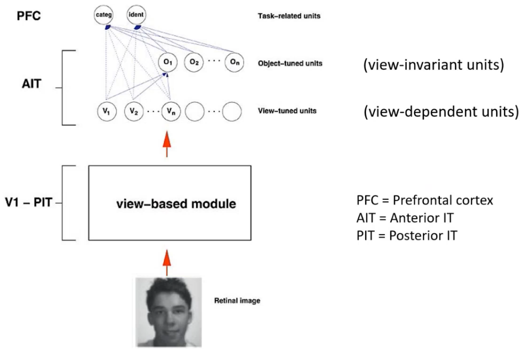


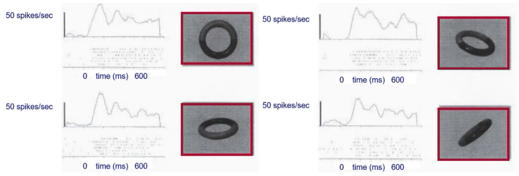
View-dependent unit The majority of IT neurons are view-dependent and respond only to objects at specific points of view.

View-dependent unit

View-invariant unit 10% of IT neurons are view-invariant and respond regardless of the position of the observer.

View-invariant unit



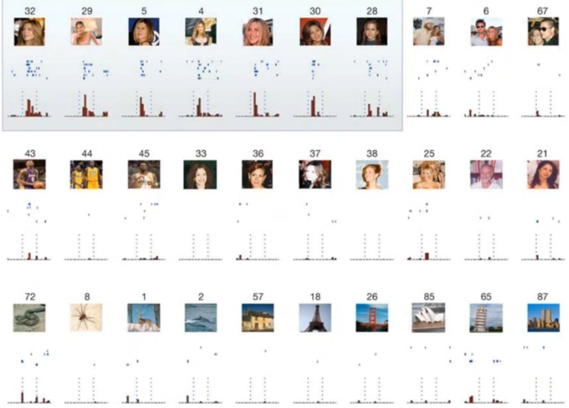


(a) View-invariant unit

Gnostic unit Neuron in the object detection hierarchy that gets activated by complex stimuli (i.e. objects with a meaning).

Gnostic unit

Case study (Jennifer Aniston cell)**.** An IT neuron of a human patient only responded to pictures of Jennifer Aniston or to its written name.



1.5.3 Local vs distributed coding

Local coding hypothesis IT neurons are gnostic units that are activated only when a particular object is recognized.

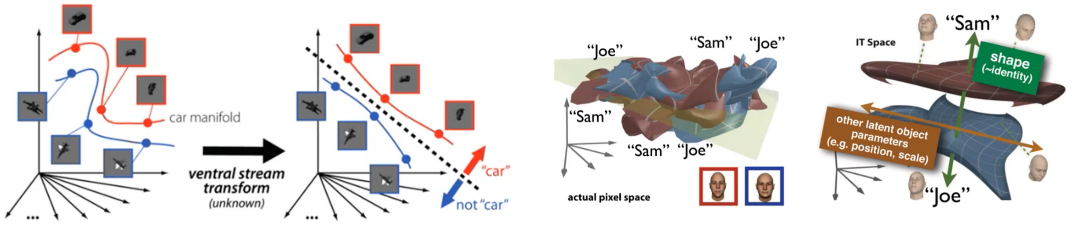
Local coding hypothesis

Distributed coding hypothesis Recognition is due to the activation of multiple IT neurons.

Distributed coding hypothesis

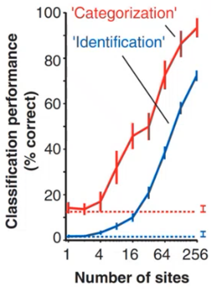
Remark. This is the most plausible hypothesis.

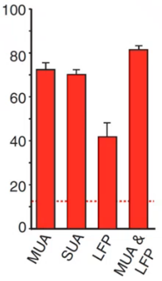
Case study (Neurons in vector space)**.** The response of a population of neurons can be represented in a vector space. It is expected that transformations of the same object produce representations that lie on the same manifold.
In the first stages of vision processing, various manifolds are tangled. Object recognition through the visual cortex aims to untangle the representations of the objects.



Case study (Classifier from monkey neurons [8]). An animal maintains fixation at the center of a screen on which images of different categories are presented very quickly (100 ms + 100 ms pause) at different scales and positions.
The responses of IT neurons are taken with some offset after the stimulus (to give them time to reach the IT) and converted into vector form to train one binary classifier (SVM) for each category (one-vs-all).
Once trained, testing was done on new stimuli. Results show that the performance increases linearly with the logarithm of the number of sites (measured neurons). It can be concluded that:

- Categorization is easier than identification.
- The distributed coding hypothesis is more likely.

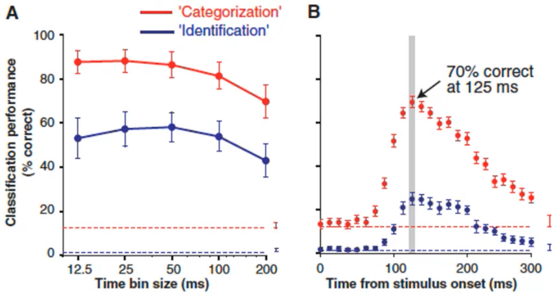




(a) Accuracy on multiple units (MUA), a single unit (SUA) and readings made on the cortex, not inside (LFP)

Time-wise, it has been observed that:

- Performance gets worse if the measurement of the neurons spans for too long (no explanation was given in the original paper, probably noise is added up to the signal for longer measurements).
- The best offset from the stimulus onset at which the measures of the IT neurons should be taken is 125 ms.



It has also been observed that the visual ventral pathway, which is responsible for object recognition, also encodes information on the size of the objects. This is not strictly useful for recognition, but a machine learning algorithm is able to extract this information from the neural readings. This hints at the fact that the ventral pathway also contributes to identifying the location and size of the objects.

Case study (Artificial neural network to predict neuronal activity [27])**.** Different neural networks are independently trained on the task of image recognition. Then, the resulting networks are compared to the neuronal activity of the brain.
The network should have the following properties:

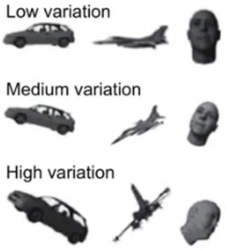
- Provide information useful to support behavioral tasks (i.e. act as the IT neurons).
- Layers of the network should have a corresponding area on the ventral pathway (mappable).
- It should be able to predict the activation of single and groups of biological neurons (neurally predictive).

Dataset A set of images is divided into two sets:

Train set To train the neural networks.

Test set To collect neuronal data and evaluate the neural networks.

Images have different levels of difficulty (low to high variation) and are presented on random backgrounds.

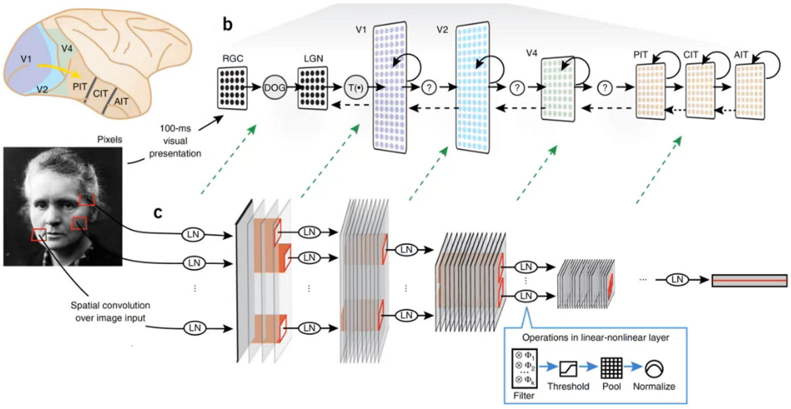


Neuronal data Neuronal data are collected from the area V4 and IT in two macaque monkeys. They are tasked to maintain fixation at the center of a screen on which images are presented for 100 ms followed by a 100 ms blank screen.

For each stimulus, the firing rate is obtained as the average of the number of spikes in the interval 70 ms - 170 ms after the stimulus onset.

Neural network training Hierarchical convolutional neural networks (HCNN) are used for the experiments. They are composed of linear-nonlinear layers that do the following:

1. Filtering through linear operations of the input stimulus (i.e. convolutions).
2. Activation through a rectified linear threshold or sigmoid.
3. Mean or maximum pooling as nonlinear aggregation operation.
4. Divisive normalization to output a standard range.


V1
V4
V2
PIT
CIT
AIT
b
RGC
DOG
LGN
T(•)
V1
V2
?
V4
PIT
CIT
AIT
Pixels
100-ms visual presentation
c
LN
Spatial convolution over image input
Operations in linear-nonlinear layer
Filter
Threshold
Pool
Normalize

The HCNNs have a depth of 3 or fewer layers and are trained independently from the neuronal measurements. For evaluation, models are divided into groups following three different criteria:

- Random sampling.
- Selection of models with the highest performance on the high-variation images.
- Selection of models with the highest IT neural predictivity.

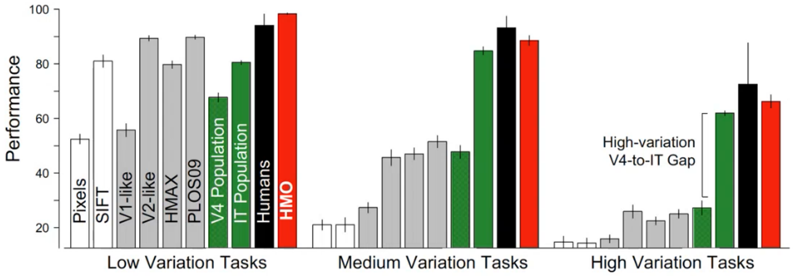
Resulting HCNNs are also used to create a new high-performance architecture through hierarchical modular optimization (HMO) by selecting the best-performing modules from the trained networks (as each layer is modular).

Evaluation method Evaluation is done using the following approach:

- Object recognition performances are assessed using SVM classifiers:
  - For neural networks, the output features of a stimulus are obtained from the activations at the top layers.
  - For neuronal readings, the output features of a stimulus are obtained by converting the firing rates into vector form.
- To measure the ability of a neural network to predict the activity of a neuron, a partial least squares regression model is used to find a combination of weights at the top layers of the network that best fits, using as metric the coefficient of determination (R^2), the activity of the neuron on a random subset of test images.
- An ideal observer is used as a baseline. It has all the categorical information to make correct predictions but it does not use a layered approach.

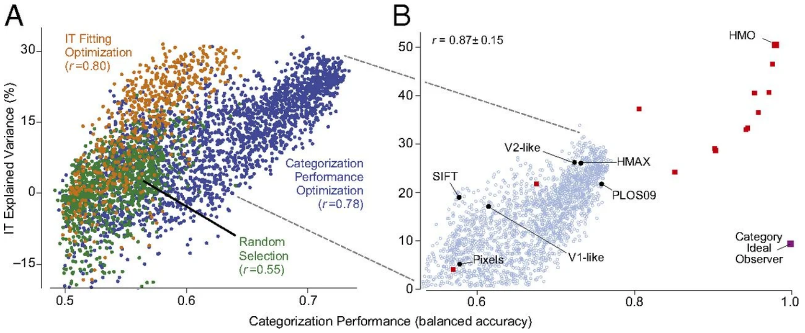
Results It has been observed that:

- The HMO model has human-like performances.

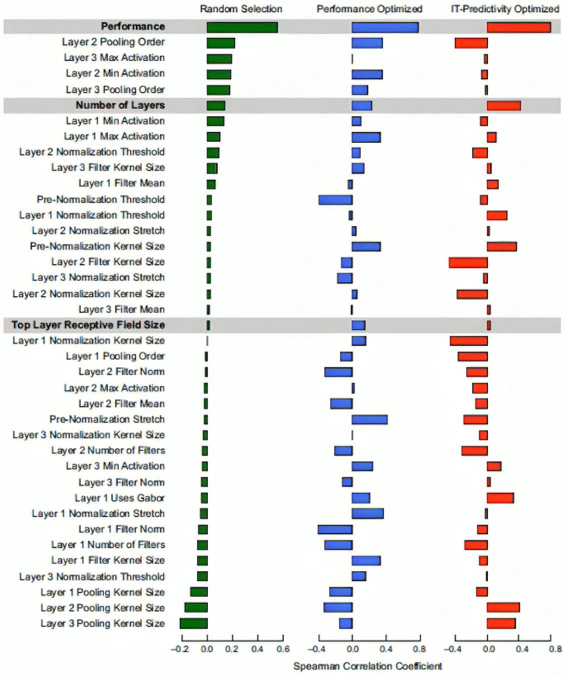


- The higher the categorization accuracy, the better the model can explain the IT.

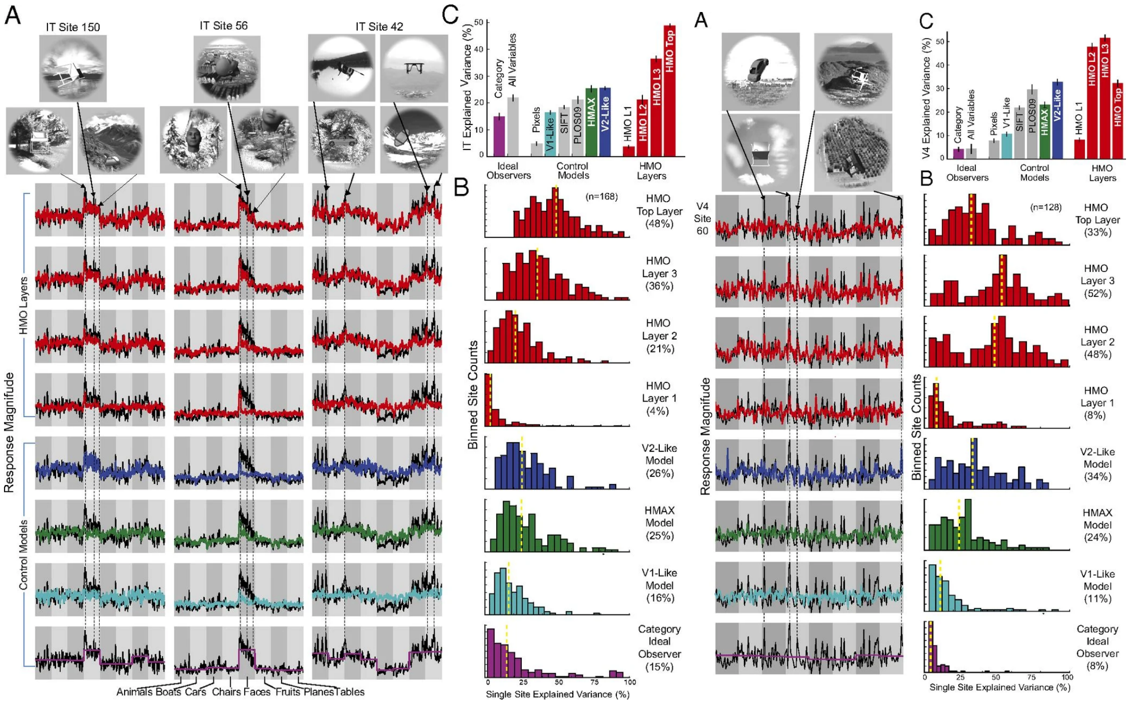
  Moreover, forcefully fitting a network to predict IT as the main task predicts the neuronal activity worse than using a model with high categorization accuracy.



- None of the parameters of the neural networks can independently predict the IT better than performance (i.e. the network as a whole).



- Higher levels of the HMO model yield good prediction capabilities of IT and V4 neurons. More specifically:
  - The fourth (last) layer of the HMO model predicts well the IT.
  - The third layer of the HMO model predicts well the area V4.



(a) Results on IT

(b) Results on V4

Figure 1.14: (A) Actual neuronal activity (black) and predicted activity (colored).
(B) R^2 value over the population of single IT neurons.
(C) Median R^2 over the population of IT neurons.

2 Object recognition emulation through neural networks

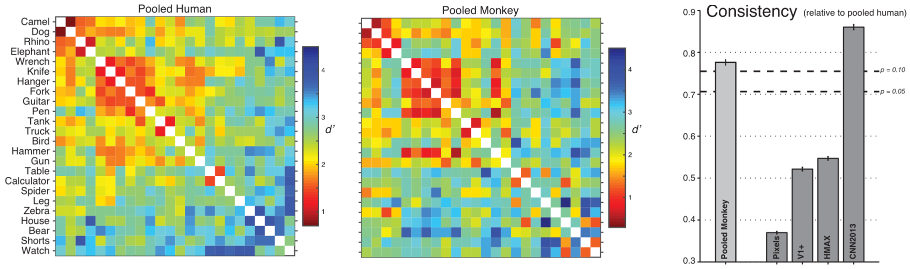
2.1 Convolutional neural networks

Convolutional neural networks

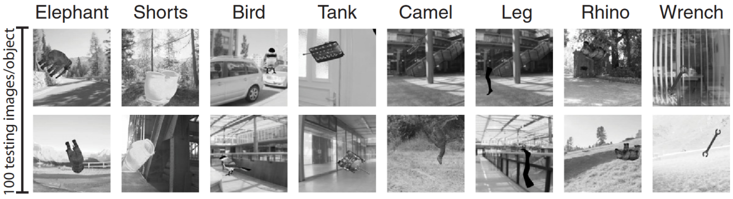
Deep convolutional neural networks (DCNNs) show an internal feature representation similar to the representation of the ventral pathway (primate ventral visual stream). Moreover, object confusion in DCNNs is similar to the behavioral patterns in primates.
However, on a higher resolution level (i.e. not object but image level), the performance of DCNNs diverges drastically from human behavior.

Remark. Studies using HCNN have also been presented in the previous chapter.

Case study (Humans and monkeys object confusion [15])**.** It has been seen that monkeys show a confusion pattern correlated to that of humans on the task of object recognition. Convolutional neural networks also show this correlation while low-level visual representations (V1 or pixels, a baseline computed from the pixels of the image) correlate poorly.



Case study (Primates and DCNNs object recognition divergence [16])**.** Humans, monkeys and DCNNs are trained for the task of object recognition.
To enforce an invariance recognition behavior, each image has an object with a random transformation (position, rotation, size) and has a random natural background.



- For humans, a trial starts with fixation. Then, an image is displayed for 100 ms followed by a binary choice. The human has to make its choice in 1000 ms.
- For monkeys, a trial starts with fixation. Then, an image is displayed for 100 ms followed by a binary choice. The monkey has up to 1500 ms to freely view the response images and has to maintain fixation on its choice for 700 ms.
- DCNNs are trained as usual.

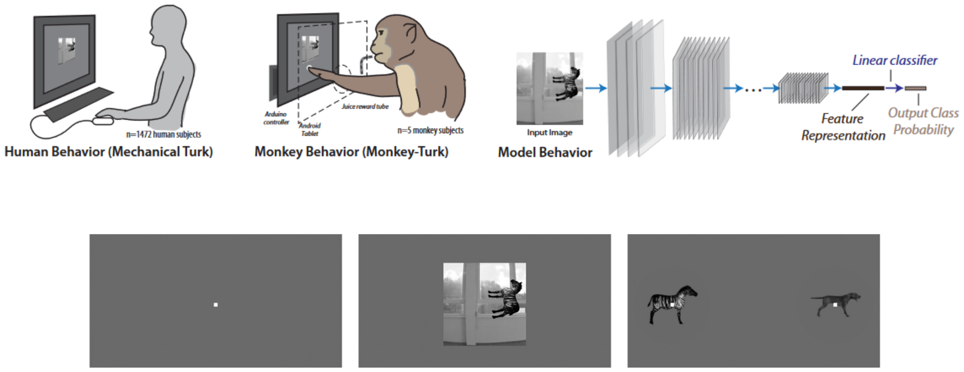


Figure 2.1: Steps of a trial

Performance is measured using behavioral metrics. Results show that:

Object-level Object-level measurements are obtained as an average across all images of that object.

Recognition confusion of primates and DCNNs are mostly correlated.

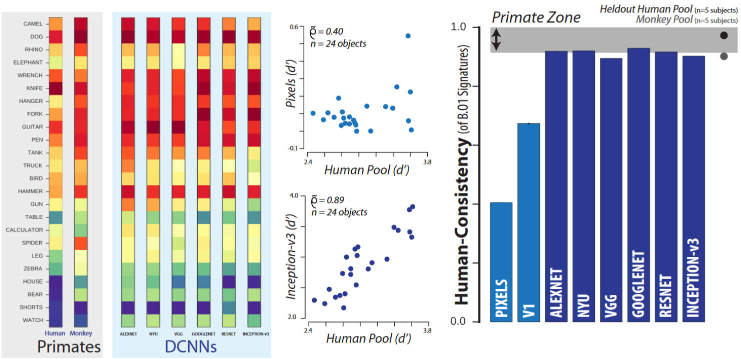
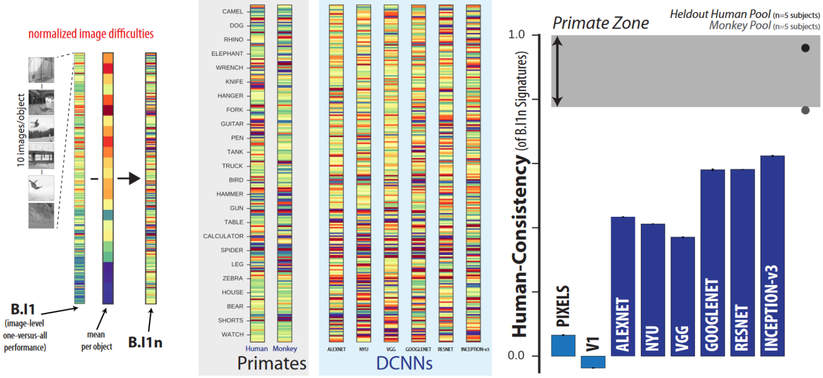


Figure 2.2: Object-level results. In the first part, warmer colors indicate a better classification.

Image-level Image-level measurements are obtained by normalizing the raw classification results.

All DCNNs fail to replicate the behavioral signatures of primates. This hints at the fact that the architecture and/or the training process is limiting the capability of the models.



2.2 Recurrent neural networks

2.2.1 Object recognition

Recurrent neural networks

The short duration for which candidates of the previous experiments were exposed to an image suggests that recurrent computation is not relevant for core object recognition. However, the following points are in contrast with this hypothesis:

- DCNNs fail to predict primate behavior in many cases.
- Specific image instances (e.g. blurred, cluttered, occluded) are easy for primates but difficult for DCNNs.

This hints at the fact that recurrent computation might be involved, maybe at later stages of the recognition process.

Case study (Primates recognition reaction time [9])**.**

Recognition training and evaluation Humans, macaques and DCNNs are trained for the task of object recognition on images with two levels of difficulty:

Control images Easier to recognize.

Challenge images Harder to recognize.

Results show that primates outperform DCNNs on challenge images.

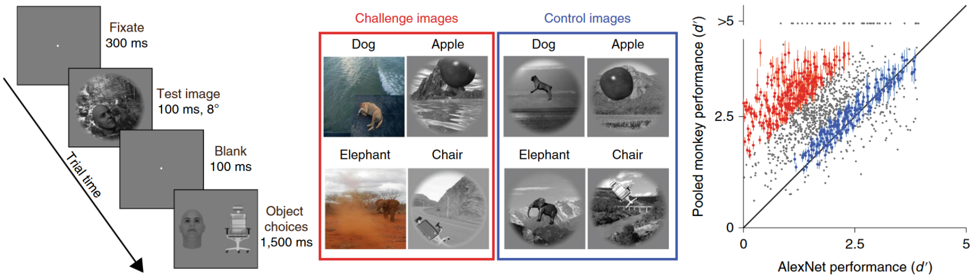
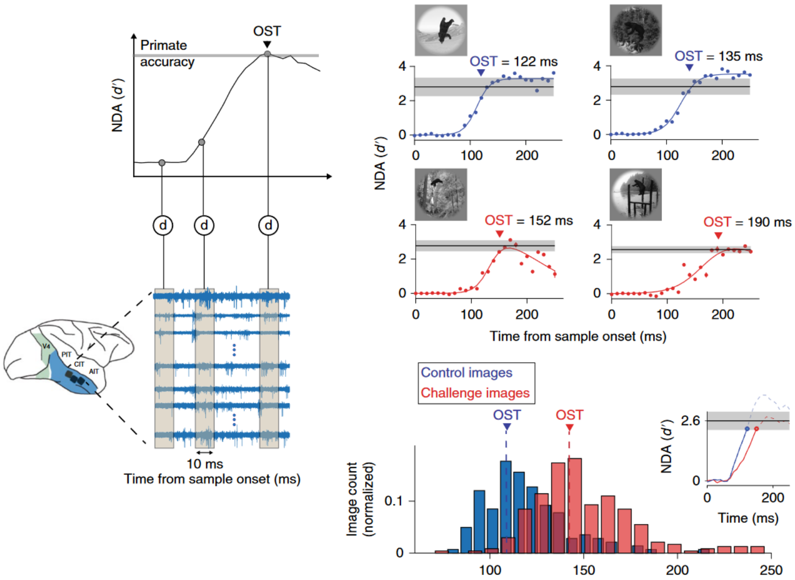


Figure 2.3: Trial steps, example images and behavioral comparison between monkeys and DCNNs. Red and blue points in the graph are challenge and control images, respectively.

Reaction time It also has been observed that the reaction time of both humans and monkeys for challenge images is significantly higher than the reaction for control images (ΔRT $=$ 11.9 ms for monkeys and ΔRT $=$ 25 ms for humans).

To determine the time at which the identity of an object is formed in the IT cortex, the neural activity is measured every 10 ms after the stimulus onset and a linear classifier (decoder) is trained to determine the **neural decode accuracy (NDA)** (i.e. the best accuracy that the classifier can achieve with the information in that time slice). We refer with **object solution time (OST)** the time at which the NDA reached the primate accuracy (i.e. high enough).

It has been observed that challenge images have a slightly higher OST ($\sim$ 30 ms) whether the animal was actively performing the task or passively viewing the image.

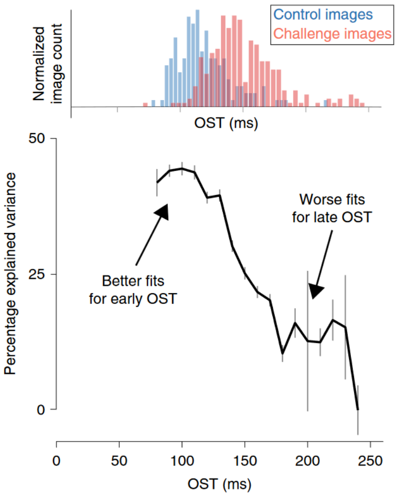


DCNN IT prediction The IT neuronal response for a subset of challenge and control images has been measured across 10 ms bins to obtain two sets R^{train} and R^{test} (50/50).

During training, the activation F^{train} of a layer of the DCNN is used to predict R^{train} through partial least square regression (i.e. a linear combination of F^{train}).

During testing, the activation of the same layer of the DCNN is transformed using the found parameters and compared to R^{test}.

Results show a higher predictivity for early responses (which are mainly feedforward) and a significant drop over time. The drop coincides with the OST of challenge images, hinting at the fact that later phases of the IT might involve recurrence.



CORnet IT prediction The previous experiment has also been done using deeper CNNs that showed better predictivity. This can be explained by the fact that deeper networks simulate the unrolling of a recurrent network and are therefore an approximation of them.

Deeper networks are also able to solve some of the challenge images but those that remained unsolved are those with the longest OSTs among the challenge images.

CORnet, a four-layer recurrent neural network, has also been experimented. Results show that the first layers of CORnet are good predictors of the early IT phases while the last layers are good at predicting the late phases of IT.

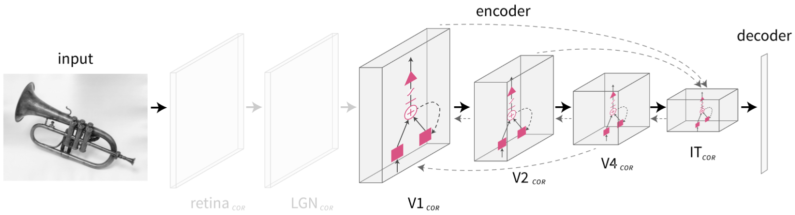
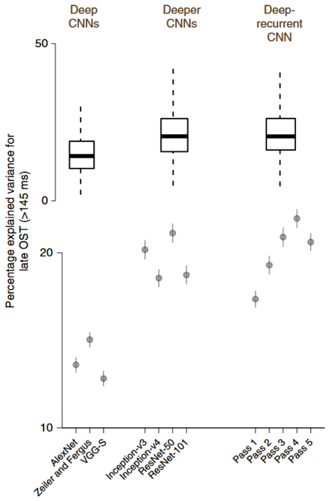
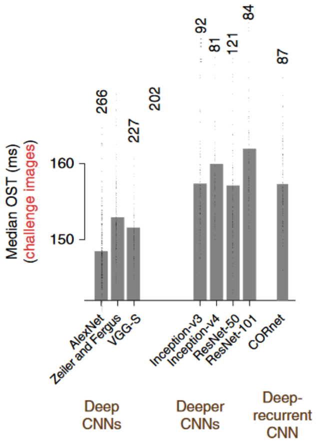


Figure 2.4: Architecture of CORnet



(a) Predictivity for deep, deeper and recurrent CNNs



(b) Number (on top) and median OST (bars) of the unsolved images for each model

> **Remark.** Recurrence can be seen as additional non-linear transformations in addition to those of the feed-forward phase.

2.2.2 Visual pattern completion

Pattern completion Ability to recognize poorly visible or occluded objects.

Pattern completion

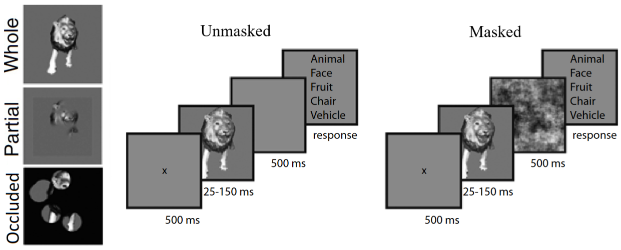
> **Remark.** The visual system is able to infer an object even if only 10-20% of it is visible.
> It is hypothesized that recurrent computation is involved.

Case study (Human and RNN pattern completion [22]).

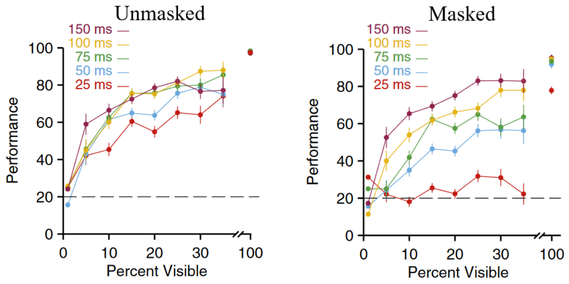
Trial structure Whole and partial images are presented to humans through two types of trials:

Unmasked After fixation, an image is displayed for a short time followed by a blank screen. Then, a response is required from the candidate.

Backward masking After fixation, an image is displayed for a short time followed by another image. Then, a response is required from the candidate. The second image aims to interrupt the processing of the first one (i.e. interrupt recurrent processing).



Human results Results show that subjects are able to robustly recognize whole and partial objects in the unmasked case. In the masked case, performances are instead worse.



Moreover, measurements show that the neural response to partially visible objects is delayed compared to whole images, hinting at the fact that additional computation is needed.

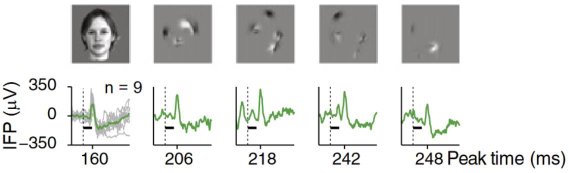
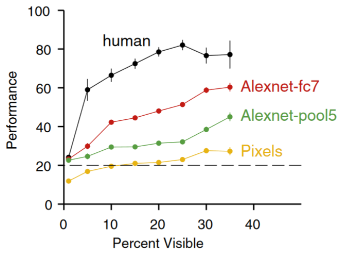
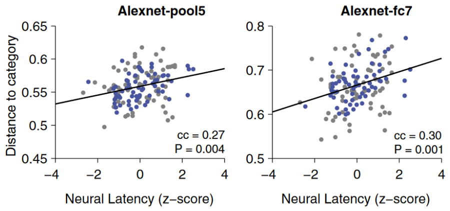


Figure 2.6: Activity (IFP) of a neuron that responds to faces

CNN results Feed-forward CNNs have also been trained on the task of object recognition.

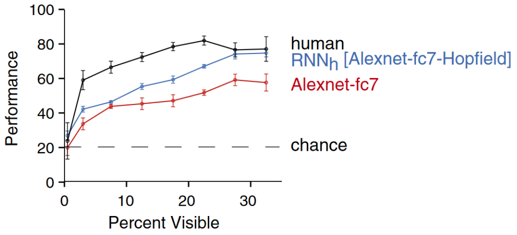
- Performances are comparable to humans for whole images but decline for partial images.
- There is a slight correlation between the latency of humans' neural response and the distance of the internal representation in the CNNs of each partial object to its whole image.



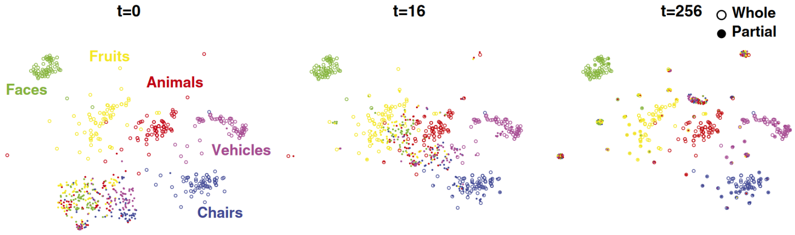


(a) Representation and latency correlation. The color of the dots depends on the electrode that measured the latency.

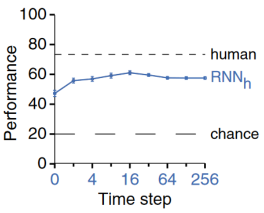
RNN results Recurrent neural networks have also been tested by using existing CNNs enhanced through attractor networks[a] (Hopfield network, RNNh). Results show that RNNh has higher performance in pattern completion.

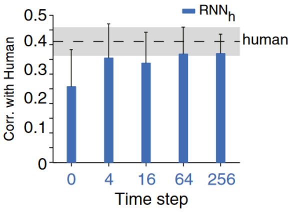


Moreover, by plotting the temporal evolution of the internal representation of partial objects, it can be seen that, at the beginning, partial images are more similar among themselves than their corresponding attractor point, but, over time, their representation approaches the correct cluster.

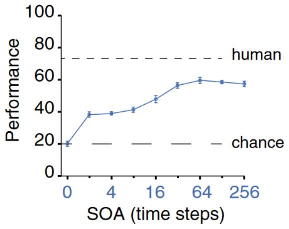


Time-wise, RNNh performance and correlation with humans increase over the time steps and saturates at around 10-20 steps. This is consistent with the physiological delays of the human ventral visual stream.





By backward masking the input of the RNNh (i.e. present the image for a few time steps and then change it), performance drops from 58 ± 2% to 37 ± 2%.



[a] Recurrent network with multiple attractor points each representing a whole image. By processing the same partial image for multiple time steps, its representation should converge to an attractor point.

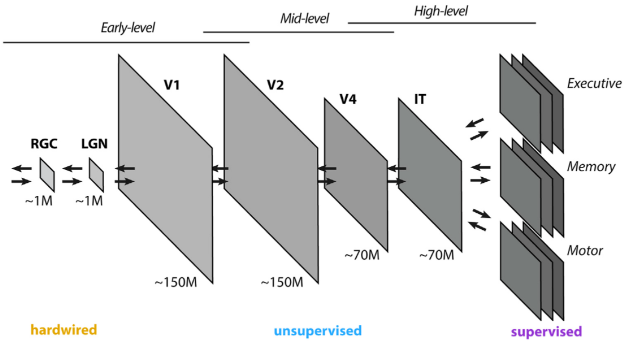
2.3 Unsupervised neural networks

Unsupervised neural networks

Most of the models to simulate the visual cortex are trained on supervised datasets of millions of images. Such supervision is not able to explain how primates learn to recognize objects as processing a huge amount of category labels during development is highly improbable. Possible hypotheses are:

- Humans might rely on different inductive biases for a more efficient learning.
- Humans might augment their initial dataset by combining known instances.

Unsupervised learning might explain what happens in between the representations at low-level visual areas (i.e. the retina), which are mostly hardcoded from evolution, and the representations learned at higher levels.



Case study (Unsupervised embedding [28]). Different unsupervised embedding methods are used to create a representation for a dataset of images that are then assessed on various tasks.

Contrastive embedding Unsupervised embedding method that uses a DCNN (which simulates low-level visual areas) to create the representation of an image in a low dimensional space and then optimize it by pushing each embedding closer to its close neighbors and far from its background neighbors.

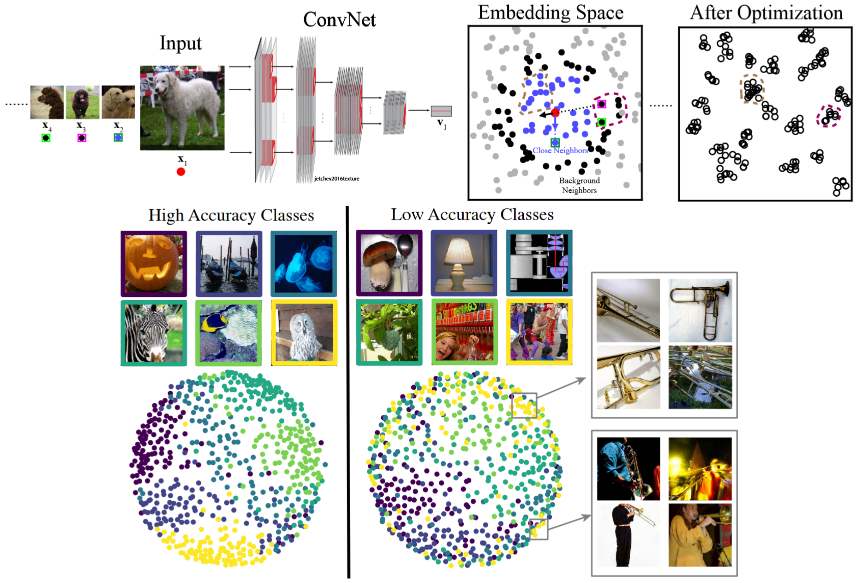


Figure 2.8: Workflow and visualization of the local aggregation algorithm

Results on object recognition tasks To solve the tasks, unsupervised embeddings are used in conjunction with a linear classifier. A supervised DCNN is also used as a baseline.

Results show that:

- Among all the unsupervised methods, contrastive embeddings have the best performances.
- Unsupervised methods equaled or outperformed the DCNN on tasks such as object position and size estimation.
- The DCNN outperforms unsupervised models on categorization tasks.

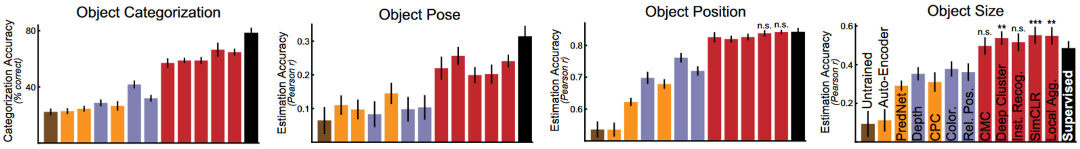


Figure 2.9: Evaluation accuracy of an untrained model (brown), predictive encoding methods (orange), self-supervised methods (blue), contrastive embeddings (red) and a supervised DCNN (black).

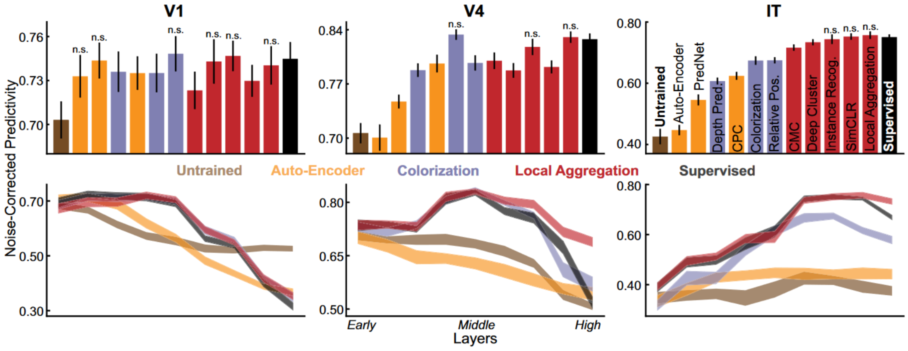
Results on neural data Techniques to map the responses of an artificial network to real neural responses have been used to evaluate unsupervised methods.

Results show that:

Area V1 None of the unsupervised methods are statistically better than the DCNN.

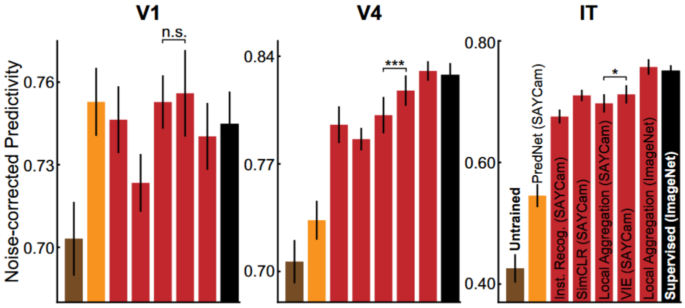
Area V4 A subset of methods equaled the DCNN.

Area IT Only contrastive embeddings equaled the DCNN.



Results on video data As training on single distinct images (ImageNet) is significantly different from real biological data streams, a dataset containing videos (SAYCam) has been experimented with. A contrastive embedding, the VIE algorithm, has been employed to predict neural activity.

Results show that embeddings learned from videos are comparable to those learned from only images.



Semi-supervised learning Semi-supervised embedding aims to find a representation using a small subset of labeled data points and a large amount of unlabeled data.

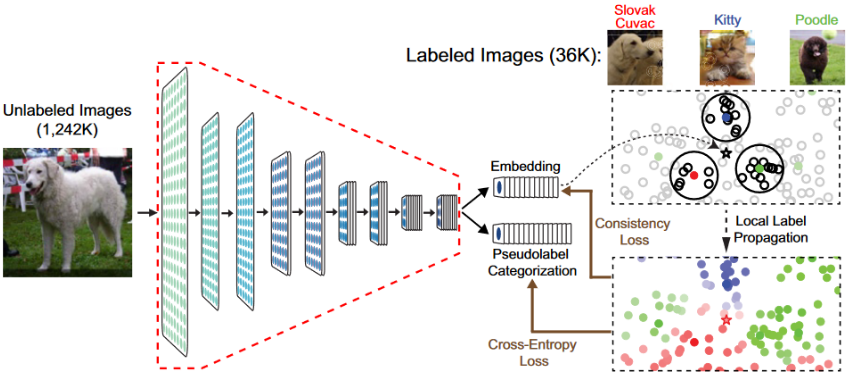
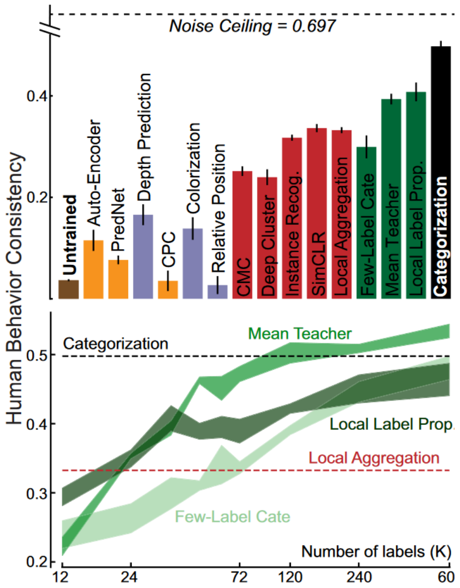


Figure 2.10: Workflow of the local label propagation algorithm

Results show that semi-supervised embeddings with only a 3% of supervision are substantially more consistent than purely unsupervised methods. Although, the gap between them and the DCNN still remains.

Nevertheless, a significant gap is also present between the results of all the models and the noise ceiling of the data, indicating that there still are inconsistencies between artificial networks and the human visual system.



Noise Ceiling = 0.697
Human Behavior Consistency
0.4
0.2
Untrained
Auto-Encoder
PredNet
Depth Prediction
CPC
Colorization
Relative Position
CMC
Deep Cluster
Instance Recog.
SimCLR
Local Aggregation
Few-Label Cate
Mean Teacher
Local Label Prop.
Categorization
Mean Teacher
Categorization
0.5
0.4
0.3
0.2
Local Label Prop.
Local Aggregation
Few-Label Cate
Number of labels (K)
12
24
72
120
240
60

3 Dopamine in reinforcement learning

3.1 Decision making

Decision-making Voluntary process that leads to the selection of an action based on sensory information.

Decision-making

Decisions are inherently non-deterministic as:

- Agents make inconsistent choices.
- Agents make choices unaware of the full consequences (uncertainty).
- Internal and external signals are noisy.

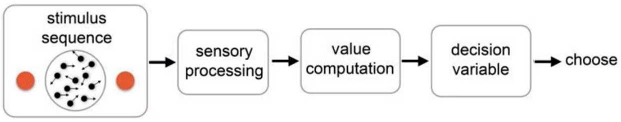
> **Remark.** From an evolutionary point of view, stochasticity in decisions increases the chances of survival.

> **Remark.** Decision-making studied within the neuroscience field is actually studying cognition. It involves studying the neural processes underlying a variety of mental functions.

Perceptual decision-making An agent selects between action A and B based on weak or noisy external signals (e.g. "do you see A or B?").

Perceptual decision-making

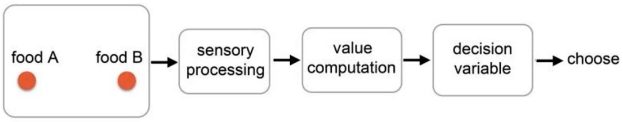
In this case, uncertainty comes from the external stimulus.



Value-based decision-making An agent selects between action A and B based on its subjective preferences (e.g. "do you prefer A or B?").

Value-based decision-making

In this case, uncertainty comes from the value associated with the action.



Decision-making processes Decision-making involves the following processes:

- **Representation** States, actions, and internal and external factors are identified.

Representation

- **Valuation** A value is assigned to the possible alternatives.

Valuation

- **Choice** Values are compared and a proper action is selected.

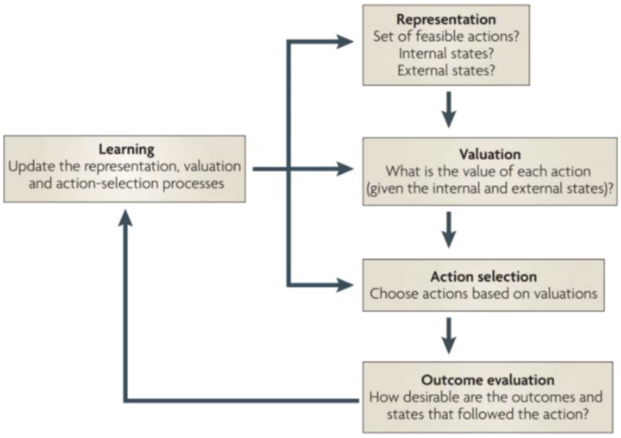
Choice

- **Outcome evaluation** After performing the action, the desirability of the outcome is measured (reward prediction error).

Outcome evaluation

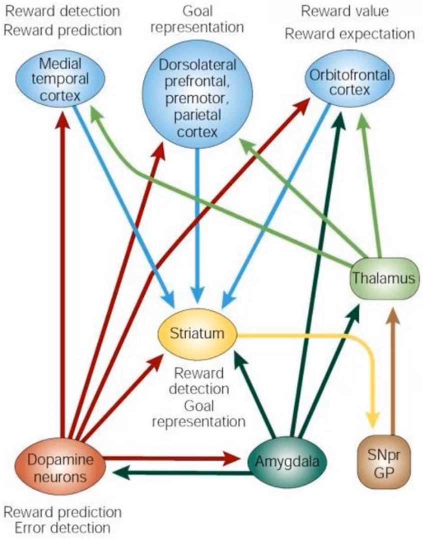
- **Learning** Feedback signals are used to update the processes and improve the quality of future decisions.

Learning


Representation
Set of feasible actions?
Internal states?
External states?
Learning
Update the representation, valuation and action-selection processes
Valuation
What is the value of each action (given the internal and external states)?
Action selection
Choose actions based on valuations
Outcome evaluation
How desirable are the outcomes and states that followed the action?

Valuation circuitry Involves neurons sensitive to reward value. They are spread in the brain, both in the cortical and subcortical regions.

Valuation circuitry



Decision-making theories

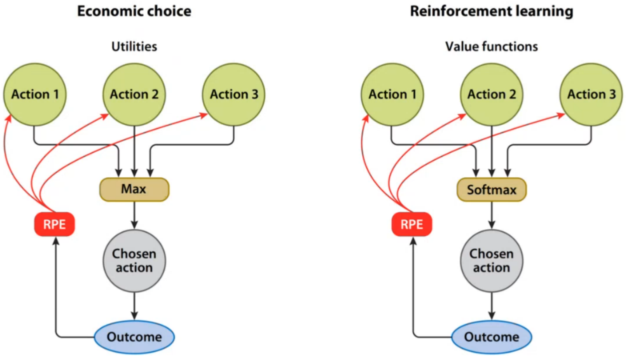
Decision-making theories

Economic learning Decision-making involving the selection of an action with the maximum utility.

Economic learning

Reinforcement learning Decision-making involving the probabilistic selection of an action.

Reinforcement learning



3.2 Reinforcement learning

Reinforcement learning (RL)

Reinforcement learning (RL) Learn a mapping between states and actions aiming to maximize the expected cumulative future reward.

Markov conditional independence At any time step, all future states and rewards only depend on the current state and action.

Bellman equation

Bellman equation Given an action a_t performed in the state s_t following a policy π, the expected future reward is given by the following equation:

$$Q_\pi(s_t, a_t) = r_t + \gamma \sum_{s_{t+1}} \mathcal{P}\left(s_{t+1} | s_t, a_t\right) Q_\pi(s_{t+1}, \pi(s_{t+1}))$$

where γ is a discount factor.

3.2.1 RL classes

Model-based

Model-based Aims to learn the right-hand side of the Bellman equation. This requires knowing the state transition distribution $\mathcal{P}$ which is costly.

Model-free

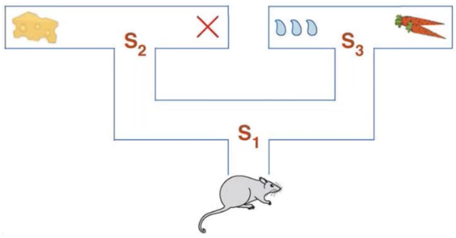
Model-free Aims to directly learn the left-hand side of the Bellman equation by estimating Q_π from experience. Agents use states, actions and rewards they experienced by averaging them to update a table of long-run reward predictions that approximate the right-hand side of the Bellman equation.

Temporal difference learning

Temporal difference learning The reward prediction error at time t is obtained by comparing the expected reward at time t and at the next time step $t + 1$:

$$\delta_t = r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$$

Example (Rat in maze)**.** A rat has to navigate a maze with two crossroads and two different outcomes.

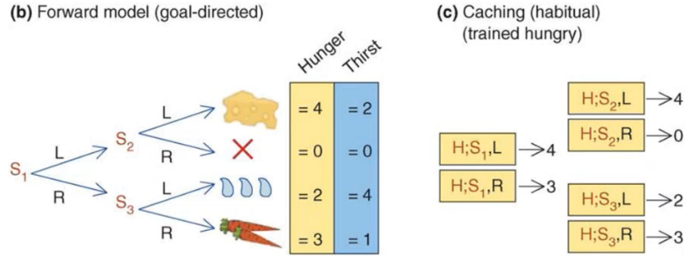


Two strategies can be developed:

Model-based By learning the model of the environment, the rat can decide its path by using a search tree. The path can be changed depending on its motivational state (e.g. hungry or thirsty) showing a goal-directed behavior.

Model-free The value of each state-action pair is stored and action selection consists of choosing the highest cached value at the current state. Values do not consider the identity of the outcome and are therefore decoupled from the motivational state of the animal.

Nevertheless, if the motivational state is stored as part of the environmental state, the animal would be able to account for it.



3.3 Dopaminergic system

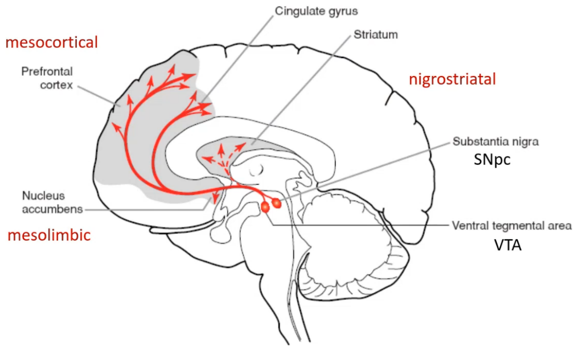
There is strong evidence showing that the dopaminergic system is highly involved in reinforcement learning for predicting both natural rewards and addictive drugs.

Dopamine pathways

Dopamine pathways Dopamine projections include:

Nigrostriatal system Mostly associated with motor functions (action policy).

Meso-cortico-limbic system Mostly associated with motivation (value function).



Actor/critic architecture

Actor/critic architecture Model with two components:

Critic Takes as input a state and is responsible for learning and storing state values. It also receives the reward from the environment and computes, through a temporal difference module, the prediction error δ_t that is used to update its own state values and train the actor.

Actor Takes as input a state and maps it to an action policy $\pi(a, s)$ that is used to determine the action to perform.

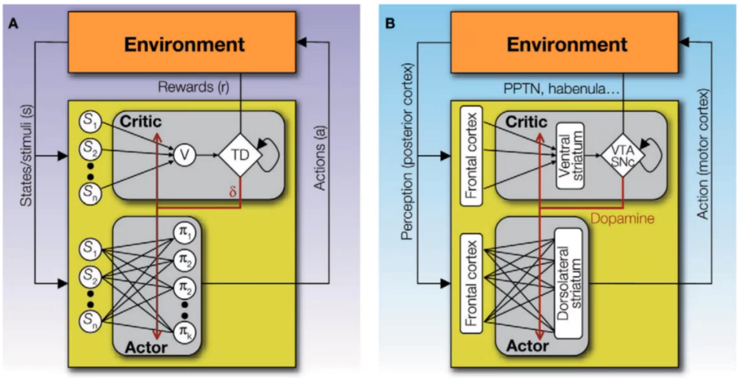


Figure 3.1: Actor/critic architecture (A) and a possible mapping of the architecture onto neural substrates (B)

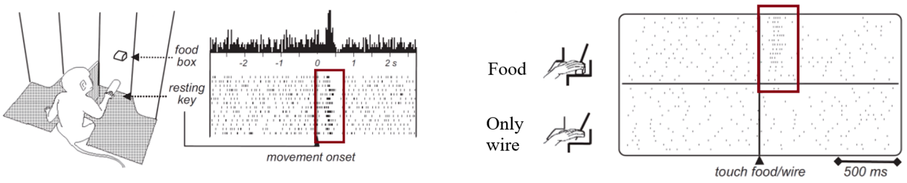
Dopamine properties

Phasic response Depending on the stimulus, dopamine neurons can show excitatory or inhibitory responses. This can be interpreted as a reward prediction error.

Dopamine phasic response

> **Remark.** About 75% of dopamine neurons are activated when there is a rewarding stimulus and about 14% of dopamine neurons are activated in response to an aversive stimulus.

> **Case study** (Dopamine as reward prediction error [17])**.** A monkey is required to touch the content of a box for which it does not have vision.
> It has been seen that dopamine neurons respond differently based on the content of the box. This is consistent with the fact that dopamine is used as a prediction error signal.



Bidirectional prediction Dopamine captures both an improvement (positive prediction error) and a worsening (negative prediction error) of the reward.

Dopamine bidirectional prediction

> **Case study** (Dopamine bidirectional prediction error [24])**.** It has been observed that the dopaminergic response differs depending on the amount of reward.

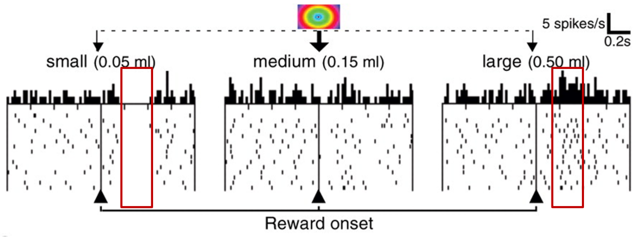
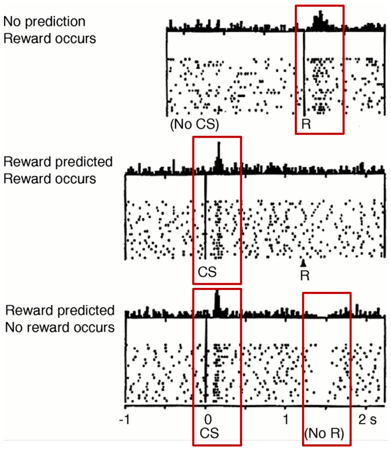


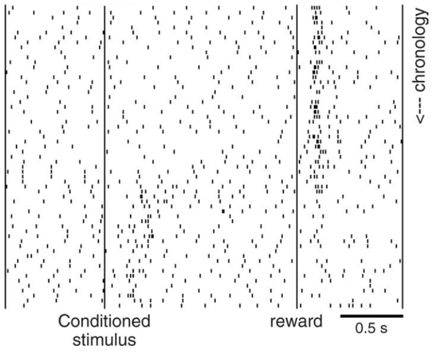
Figure 3.3: Dopamine response of a monkey trained on a medium amount of reward

Transfer Dopaminergic activity shifts from responding to the reward to responding to the conditioned stimulus that predicts it.

Dopamine transfer

> **Case study** (Dopamine transfer [20, 19])**.** It has been seen that the dopaminergic response transfers from the moment of receiving the reward to the stimuli associated with it (CS). This is in line with the temporal difference model.

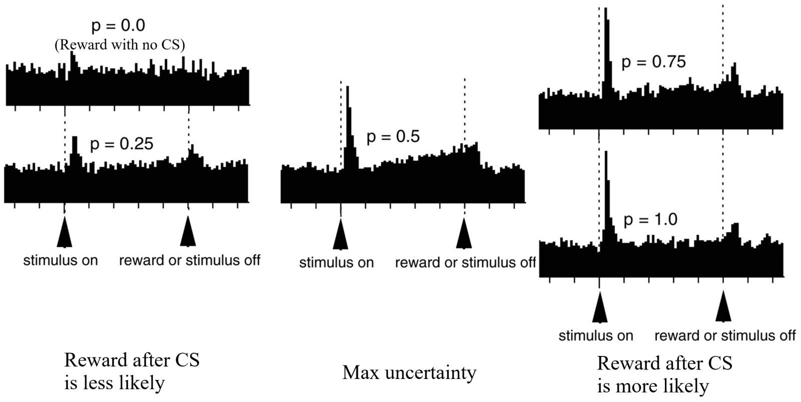




Probability encoding The dopaminergic response varies with the reward probability.

Probability encoding

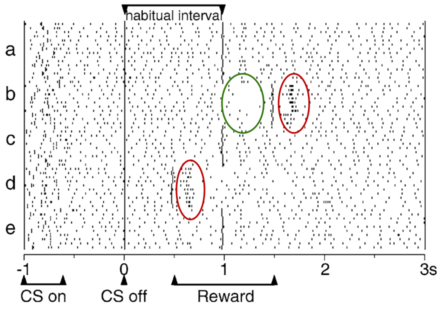
Case study (Dopamine probability encoding [4])**.** It has been shown that dopamine responds differently based on the probability of receiving a reward. For high uncertainty (50% probability of reward), a tonic response that starts from the CS and grows up to the reward time has been observed.



Temporal prediction Apart from encoding the unexpectedness of an event occurring, dopamine also accounts for the time the reward is expected to be delivered, and responds accordingly if the delivery happens earlier or later.

Dopamine temporal prediction

Case study (Dopamine temporal prediction [7])**.** It has been shown that dopamine responds differently based on the time the reward is delivered.
If the delivery happens earlier, the dopaminergic response increases. On the other hand, if the delivery happens later, dopamine neurons first pass a depressed phase.



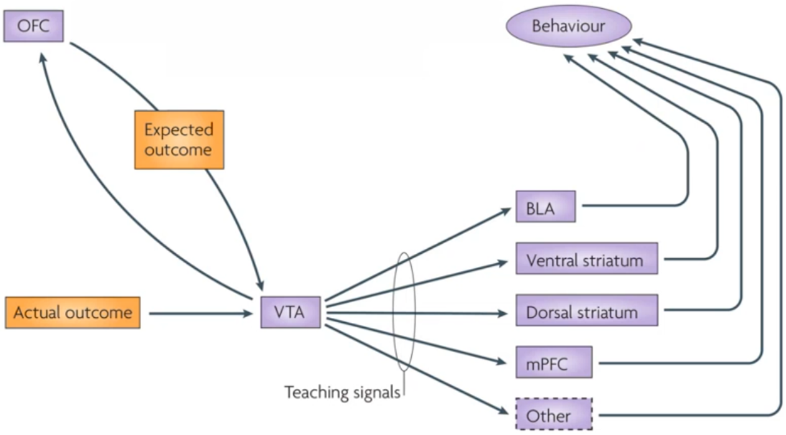
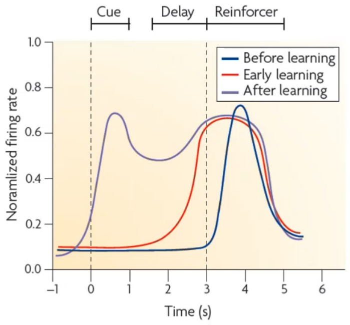
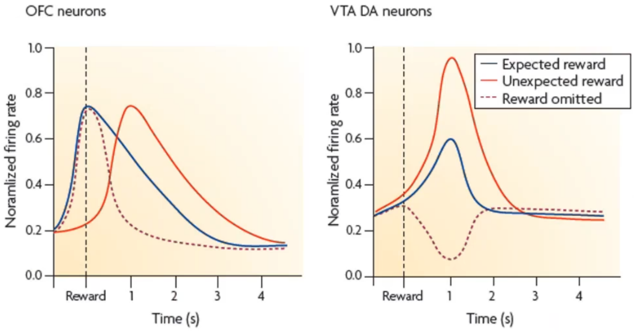


Figure 3.5: Dopamine flow in the dopaminergic system

Remark (PFC neurons)**.** Differently from dopamine neurons, PFC neurons during learning fire in response to the reward and progressively start to fire also at the CS. In addition, the strength of the response does not depend on the expectedness of the reward (i.e. it only acts as a prediction for the reward and not for the error).



(a) PFC response during learning



(b) PFC vs dopamine (DA)

3.4 Reward prediction error (RPE) theory of dopamine

3.4.1 Dopamine is not fully model-free

There is strong evidence that midbrain dopamine is used to report RPE as in model-free RL. RPE theory of dopamine states that:

- Dopamine reflects the value of the observable state, which can be seen as a quantitative summary of future reward.
- State values are directly learned through experience.
- Dopamine only signals surprising events that bring a reward.
- Dopamine does not make inferences on the model of the environment.

However, individuals also learn about the model of the world (e.g. cognitive map) and this knowledge can affect the neuronal prediction error, but, there is evidence that this acquisition involves cortical signals (e.g. PFC) rather than dopamine. Despite that, dopamine still seems to integrate predictive information from models.

Case study (Monkey saccade [13])**.** Monkeys are required to solve a memory-guided saccade task where, after fixation, a light is flashed in one of the four directions indicating the saccade to be made after the fixation point goes off.

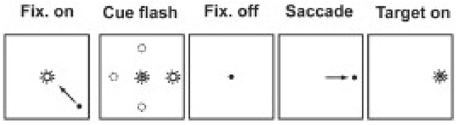


Figure 3.7: Structure of the task

Experiments are done in sub-blocks of four trials, each corresponding to a direction. Moreover, the probability of reward increases with the number of non-rewarded trials (post-reward trial number, `PNR`). If `PNR` $= 1$, the probability of reward is the lowest (0.0625), while if `PNR` $= 7$, the reward probability is 1.0.

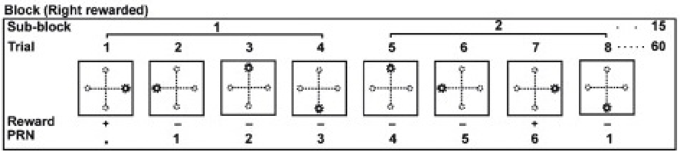
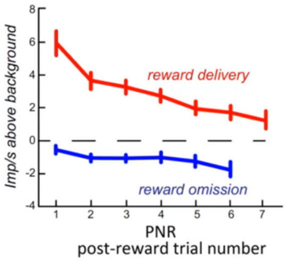


Figure 3.8: Structure of a block

It is expected that the animal's reward prediction increases after each non-rewarded trial. In other words, as the reward is more likely after each non-rewarded trial, positive prediction error should decrease and negative prediction error should be stronger.
Results show that dopamine neurons are less active if the reward is delivered later and more depressed if the reward is omitted after each non-rewarded trial.



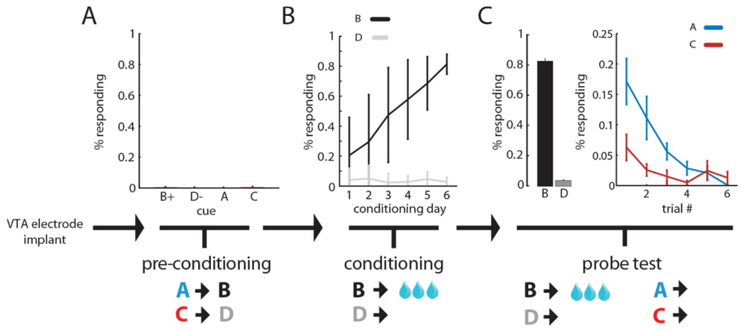
The results are in contrast with an exclusive model-free view of dopamine as, if this were the case, learning would only involve past non-rewarded trials causing positive prediction error to decrease and negative prediction error to be weaker. Therefore, dopamine might process prediction error in both model-free and model-based approaches.

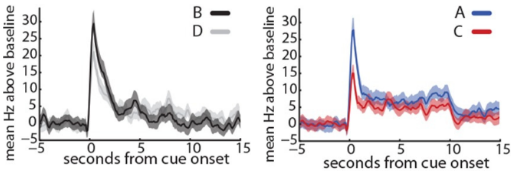
Case study (Dopamine indirect learning [18])**.** Rats are exposed to:

Pre-conditioning Sound stimuli $A \mapsto B$ and $C \mapsto D$ are paired together.

Conditioning The stimulus B is paired with a reward.

Results show that rats respond to both B and A in a correlated manner.





The results show that dopamine might also reflect values learned indirectly. This is in contrast with the temporal difference learning of model-free RL in which only directly experienced states are learned.

Case study (Dopamine RPE reflects inference over hidden states [21])**.** Rats are trained to associate odors with rewards. Two types of tasks are considered:

Task 1 Odors are always associated with a reward. Odor A is delivered with a delay sampled from a Gaussian distribution, odors B and C are deterministic and odor D is for control.

Task 2 As above, but odors are associated with a reward 90% of the time.

The period in which no reward is expected is called ITI, while the period in which the animal expects a reward is the ISI (i.e. after the odor onset).

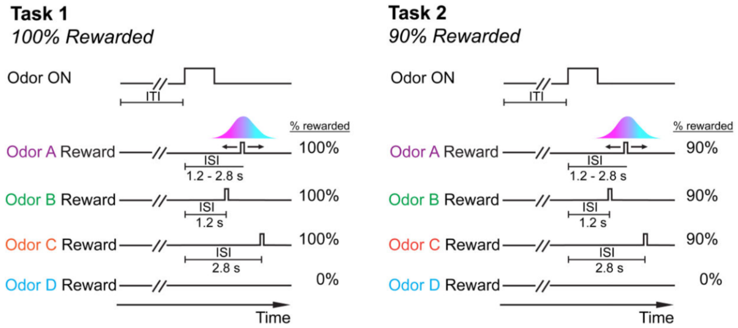


Figure 3.9: Tasks representation

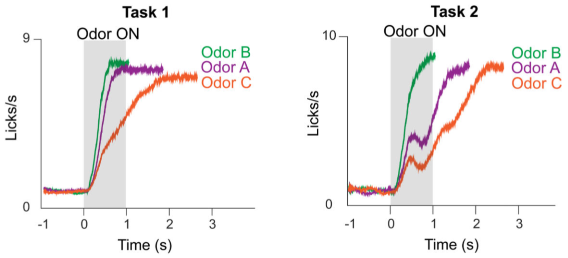
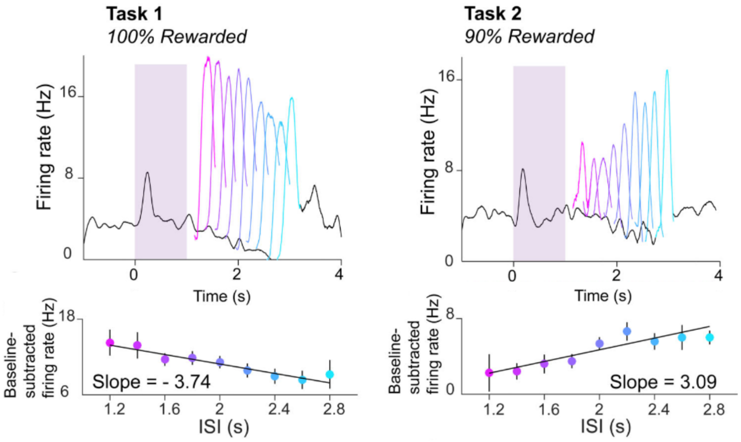


Figure 3.10: Licking behavior in the two tasks. It can be seen that, for each odor, licking depends on the time of arrival of the reward. On task 2, licking is more uncertain.

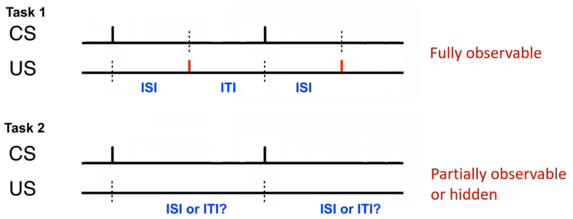
Considering odor A, results show that:

- For task 1, the dopaminergic response gets smaller over time within the ISI
- For task 2, the dopaminergic response grows, hinting at the fact that some sort of inference about the state is being made.



An explanation is that the animal has to solve the problem of determining whether it is in an ITI or ISI period:

- For task 1, the rat can easily determine in which period it is.
- For task 2, as the reward is not always delivered, being in ISI and ITI is not always clear.



The dopaminergic activity for the two tasks can be explained as follows:

- For task 1, after the stimulus, the probability of receiving a reward increases over time. Therefore, the RPE is increasingly suppressed.

- For task 2, as the reward fails to arrive, the belief state progressively shifts towards the ITI state. Therefore, if the reward is delivered later, the RPE is high as it was unexpected.

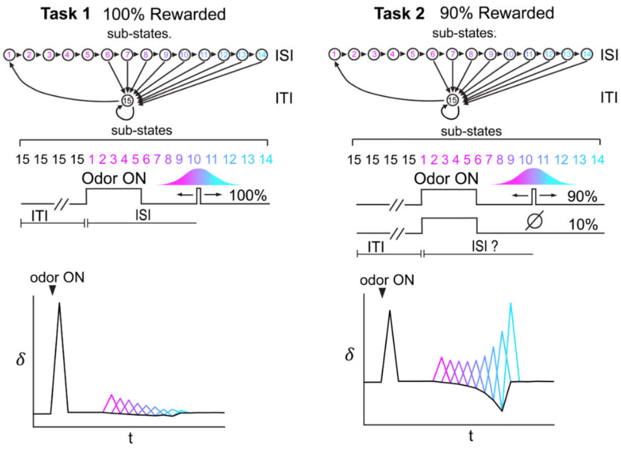


Figure 3.11: Experiment represented as sub-states (top) and RPE to a reward over time (bottom)

These results hint at the fact that:

- The brain is not limited to passively observing the environment but also makes latent inferences.
- The results can be modeled using a temporal difference model that incorporates hidden-state inference.

3.4.2 Dopamine as generalized prediction error

Dopamine might not be limited to only predicting reward error but is also involved in a more general state prediction error.

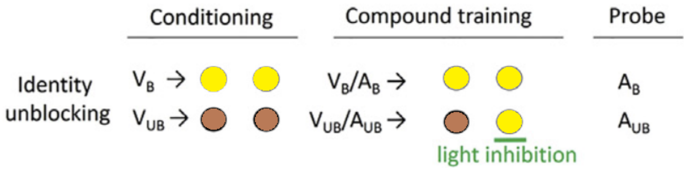
Case study (Dopamine state change prediction [2]). Rats are exposed to the following training steps:

Conditioning The following stimuli are associated with some rewards (it must be ensured that rewards are liked in the same way, i.e. same value but different identity):

- V_B is associated with two units of banana.
- V_{UB} is associated with two units of chocolate.

Compound training New stimuli are paired with the previously learned ones:

- A_B is paired with V_B. Because of blocking, A_B should not be learned as a rewarding stimulus.
- A_{UB} is paired with V_{UB}. The reward is changed to achieve identity unblocking.



It has been shown that the animal learns the new CS A_{UB} and dopamine responds to the change even if only the identity of the reward changed and the value remained the same.

3.4.3 Successor representation

Sensory prediction error (SPE) Generalized prediction error over sensory features that estimates the successor representation of a state.

Sensory prediction error (SPE)

Successor representation (SR) The SR of a state s is a mapping $M(s, \cdot)$ where $M(s, s')$ indicates the expected occupancy of s' by starting from s (i.e. how likely it is to end up in s' from s).

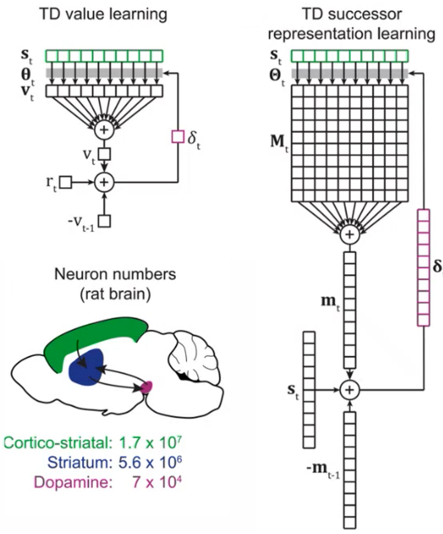
Successor representation (SR)

A SR learner predicts the value of a state by taking into account the reward R and the successor representation M:

$$V(s_t) = R(s_t)M(s_t, s_{t+1}) \qquad \left([12] \text{ says } V(s_t) = \sum_{t+1} R(s_{t+1})M(s_t, s_{t+1}) \right)$$

Remark. SR learning might be a middle ground between model-free and model-based methods. SR computes the future reward by combining the efficiency of model-free approaches and some flexibility from model-based RL (i.e. caching of state transitions). This is suited for tasks where the states are more or less stable but rewards and goals change frequently.

Remark. A criticism against this method is that it is space-wise expensive as it involves a prediction error for states (the matrix M). This space requirement finds a mismatch in the number of neurons available in the brain as dopaminergic neurons, which are responsible for updating the values, might not be enough.



Case study (SR in humans [12])**.** The experiment is divided into the following phases:

Learning phase Candidates are exposed to a sequence of three stimuli, where the third is associated with a reward. After a bunch of trials (with different stimuli), they are asked to indicate which starting stimulus is the one leading to the greater future reward.

Re-learning phase Two types of revaluation are considered:

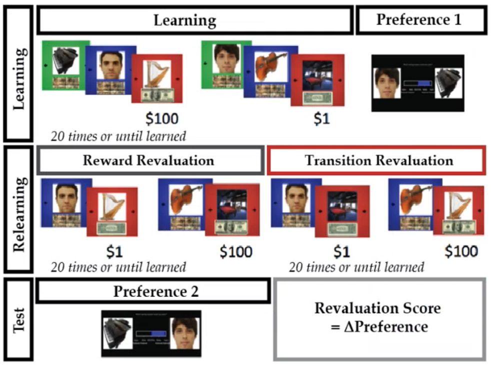
Reward revaluation Final rewards are swapped with the stimuli unchanged (i.e. value change).

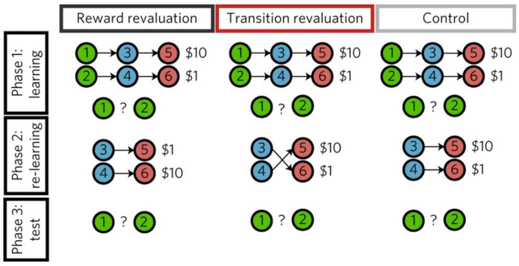
Transition revaluation Third stimuli are swapped with the rewards unchanged (i.e. state change).

Candidates are again exposed to the sequence of stimuli starting from the middle one (i.e. the first stimulus is dropped).

Expected results are:

- Model-free approaches should fail on both changes.
- Model-based approaches should succeed in both changes.
- SR-based approaches should succeed in the reward change but not in the transition change.

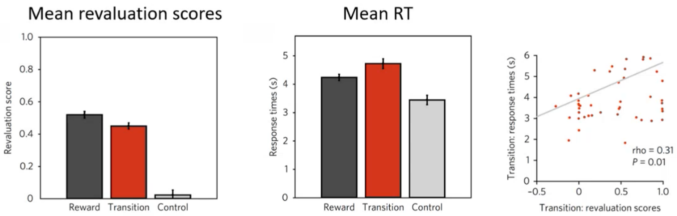




Results show that:

- Revaluation scores (i.e. change in preference) on reward revaluation are slightly better than transition revaluation.
- Reaction time for reward revaluation is faster (cached rewards can be easily updated) but slower for the transition revaluation (cannot rely on cached states as they require more time to be updated).

This suggests that humans might use successor representation learning with some form of model-based approach. Because of the differences in score and reaction time, learning cannot be fully model-based.



3.5 Distributional reinforcement learning

Distributional reinforcement learning RL methods that aim to learn the full distribution of the expected reward instead of the mean expected reward.

Distributional reinforcement learning

Remark. Certain deep RL algorithms improve with distributional RL.

Remark. In traditional temporal-difference learning, predictors learn similar values. In distributional temporal-difference learning, there are optimistic and pessimistic predictors with different scaling that expect larger and smaller future rewards, respectively.

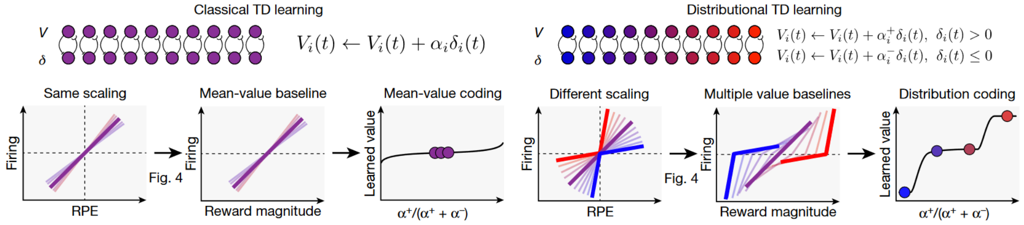


Figure 3.13: Traditional RL (left) and distributional RL (right). In distributional RL, red nodes are optimistic and blue nodes are pessimistic.

Reversal point r_0 is the reversal point of a dopaminergic neuron if:

Reversal point

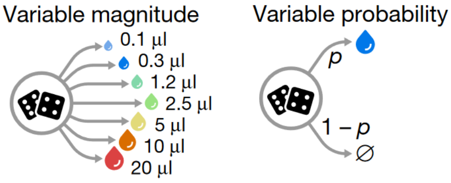
- A reward $r < r_0$ expresses a negative error.
- A reward $r > r_0$ expresses a positive error.

Remark. In traditional temporal-difference learning, the reversal point of individual neurons should be approximately identical.

Case study (Distributional RL in dopamine response [3]). Single dopaminergic neurons are recorded in rats. Rats are trained on two different tasks:

Variable-magnitude A random amount of reward is given to the rat. The reward is anticipated by an odor stimulus in half of the trials.

Variable-probability Three odor stimuli are each associated with a probability of reward (90%, 50%, 10%). A control odor is associated with no reward.



Results on the variable-magnitude task show that:

- Neurons in simulated classical RL carry approximately the same RPE signal for each magnitude and have similar reversal points (~ 0).
- Neurons in simulated distributional RL have different reversal points and there is more variety in responses (e.g. RPEs of optimistic neurons are positive only for large magnitudes and vice versa for pessimistic neurons).
- Measured neural data are more similar to the simulated distributional RL data.

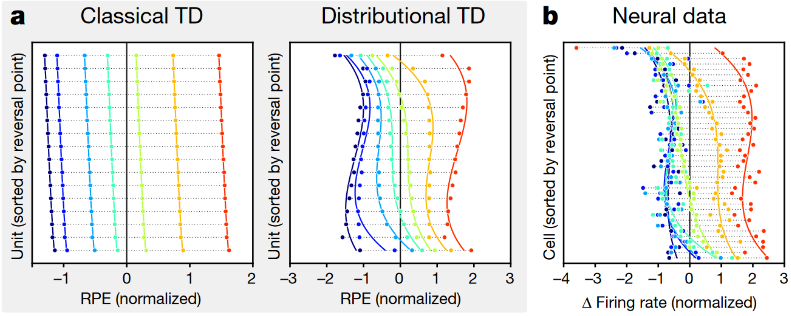


Figure 3.14: Simulated (a) and measured (b) neurons. Points at the same y-axis represent the same neuron and are sorted by reversal point. The color of the dots represents the magnitude of the reward.

Results on the variable-probability task show that:

- Neurons in simulated classical RL do not show differences when comparing the stimulus with 50% reward against the 10% and 90% responses.
- Neurons in simulated distributional RL vary a lot when responding to the 50% reward probability stimulus.
- Measured neural data are more similar to the simulated distributional RL data.

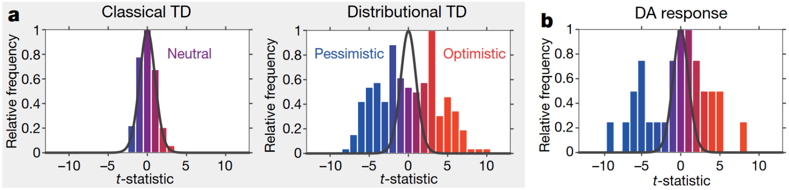


Figure 3.15: Simulated (a) and measured (b) neurons. T-statistics comparing each cell's response to the stimulus associated with the 50% reward against the mean stimulus response across cells.

Responses of dopamine neurons show that some cells are in fact more optimistic and some more pessimistic depending on how they respond to the 50% stimulus.

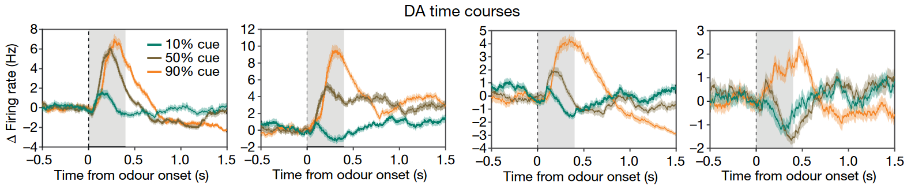
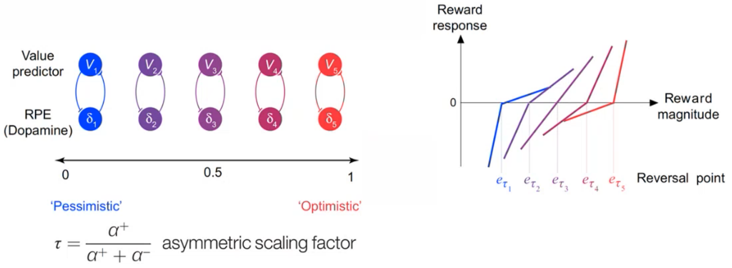
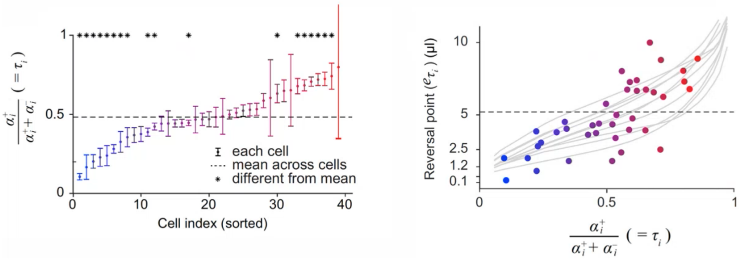


Figure 3.16: Activities of four dopaminergic neurons

An explanation for this behavior of having different reversal points is that the weights for positive (α^+) and negative (α^-) RPEs are different, or, more specifically, the asymmetric scaling factor $\frac{\alpha^+}{\alpha^+ + \alpha^-}$ is different. This creates a disequilibrium that can be rebalanced by changing the reversal points of the neurons.

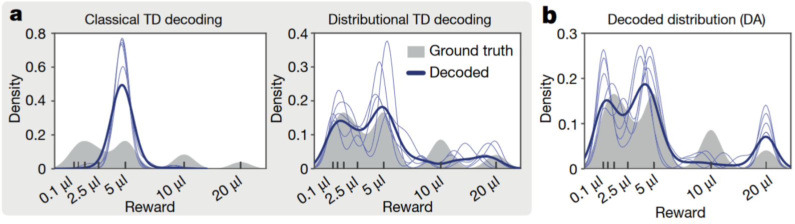


Indeed, measurements show that the reversal point and the asymmetric scaling factor are correlated indicating the need to shift the reversal point to reach equilibrium.



By decoding reward distributions from neural responses, it can be seen that:

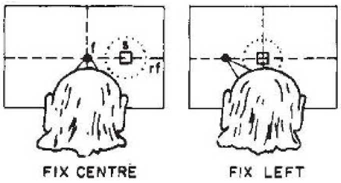
- Classical RL is not able to predict the correct distribution.
- Distributional RL and neuronal data are able to approximate the reward distribution.



4 Deep reinforcement learning

Deep learning has a large impact on neuroscience. For instance:

Case study (Neural network to represent parietal neurons [29])**.** A monkey is tasked to maintain fixation at a point with a stimulus visible within its receptive field. The fixation point and the stimulus are moved together so that, for the retina, the stimulus is at the same retinal coordinates.



In the parietal cortex, the firing of neurons depends on:

- The position of the stimulus in the receptive field.
- The position of the eyes in the orbit.

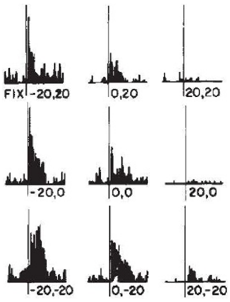


Figure 4.1: Readings of a parietal neuron at different fixation points

Using the parietal readings, the authors were able to create a network to predict the spatial position of objects.

Remark. This is one of the first successful applications of neural networks in neuroscience.

4.1 Deep reinforcement learning applications

Reinforcement learning Given a state and a set of stimuli, find the best policy to maximize future reward.

Reinforcement learning

Remark. Differently from supervised learning, in reinforcement learning:

- The agent does not know the correctness of its actions.
- The agent has to learn the best policy in the distribution of policies. In contrast, in supervised learning the model needs to learn to generalize different

distributions (e.g. different images of the same class).

Remark. In early RL work, a tabular state representation was adopted. This approach was unable to generalize as it was not possible to distinguish similar states. A workaround is to use function approximations to encode states into a feature space.

Deep reinforcement learning Use a neural network to learn the representation of the environment and the policy solving an RL problem.

Deep reinforcement learning

Case study (TD-Gammon [23])**.** TD-Gammon is a neural network trained using a form of temporal difference learning to play the game of backgammon. Due to instability, the results were not satisfying.

Case study (Atari deep Q-learning [11, 10])**.** It has been shown that deep Q-learning networks (DQN) can successfully learn to play classic Atari games at human-level performance. However, the learning method is significantly different from human's:

- Professional gamers are able to reach good scores in approximately 2 hours.
- DQN requires 200 million frames ($\sim$ 942 hours) with experience replay where each frame is replayed approximately eight times during learning.

Comparison with different variants of DQN shows that agents require more time to reach human-like performance but are also able to outperform them given enough time.

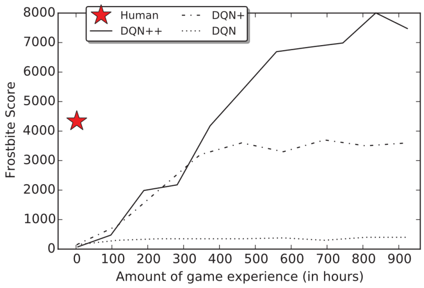


Figure 4.2: Comparison on the Frostbite game

Remark. Speculations on possible ways to improve artificial agents are:

- Provide causal models of the world to support explanation and understanding (at this stage, agents are similar to humans with neural lesions that make them unable to understand causal relationships).
- Ground learning on intuitive theories of physics and psychology.
- Learn how to learn.

4.1.1 Inefficiency factors

There are at least two factors that cause the sample inefficiency of deep reinforcement learning networks.

Slow parameter adjustments Parameters of a neural network are updated by small steps as larger steps might cause the problem of ”catastrophic interference”.

Slow parameter adjustments

Bias-variance tradeoff

Bias-variance tradeoff Neural network have a weak inductive bias:

- A learning procedure with a weak inductive bias (and large variance) is able to learn a wide range of patterns but is generally less sample-efficient.
- A learning procedure with a strong inductive bias can use its prior hypotheses to rapidly solve the task, provided that the hypotheses are suited for that specific task.

4.2 Episodic reinforcement learning

Episodic RL

Episodic RL Non-parametric (examplar-based) RL approach that learns from past experience.

The agent explicitly records past events that are used for future decision-making.

> **Remark.** This can be seen as a third action selection method together with model-based and model-free approaches.

Long-term memory

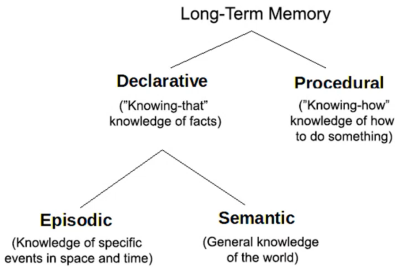
Long-term memory Inactive past memories that have to be reactivated in order to use them. In humans, it comes in different forms:

Declarative/Explicit Memory that can be expressed in a propositional form (i.e. by words).

Episodic Memory related to a specific experience in space and time.

Semantic Memory related to the shared aspects of many episodic memories (can be seen as an average).

Procedural/Implicit Memory involving practical procedures (knowing-how). It can be related to model-free approaches.


Long-Term Memory
Declarative
("Knowing-that" knowledge of facts)
Procedural
("Knowing-how" knowledge of how to do something)
Episodic
(Knowledge of specific events in space and time)
Semantic
(General knowledge of the world)

> **Remark.** On the other spectrum, there is short-term and working memory. The former, once formed, is sort of read-only (e.g. repeat a sequence of numbers). The latter allows the manipulation of its content (e.g. repeat a sequence of numbers in reverse order).

Complementary learning systems (CLS) theory

Complementary learning systems (CLS) theory Theory stating that intelligent agents possess two learning systems involving two parts of the brain:

Neocortex Slowly acquires structured knowledge of the environment that is shaped based on the structure of past experience.

Hippocampus Allows to rapidly learn spatial and non-spatial features of a particular experience (episodic memory). It forms the initial memory of an episode and interacts with the cortex by replaying it multiple times to eventually form a lasting memory (reinstatement).

Remark. Patients without the hippocampus cannot remember events that happened close in time.

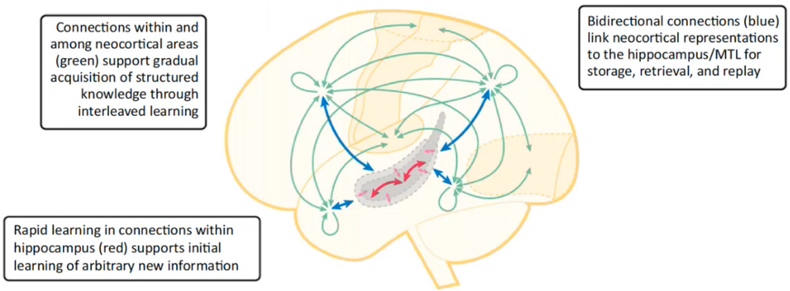
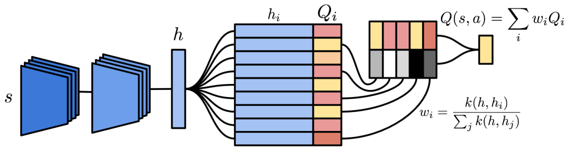

Connections within and among neocortical areas (green) support gradual acquisition of structured knowledge through interleaved learning
Bidirectional connections (blue) link neocortical representations to the hippocampus/MTL for storage, retrieval, and replay
Rapid learning in connections within hippocampus (red) supports initial learning of arbitrary new information

Figure 4.3: Pathways of the CLS theory

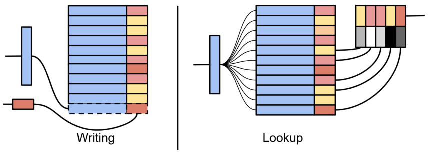
Case study (Neural episodic control [14])**.** In neural episodic control (NEC), an agent stores each encountered state with the discounted sum of rewards of n future steps to form an episodic memory associated with the state and the received reward. NEC consists of three components:

- A convolutional network to process the image at state s and convert it into a key to query the memory,
- A set of memory modules (differentiable neural dictionary),
- A final network that converts action memories into $Q(s, a)$ values.

To estimate the value of a state, the agent computes the sum of the stored discounted rewards weighted depending on the similarity between the stored states and the new state.


s
h
h_i
Q_i
$Q(s,a) = \sum_i w_i Q_i$
$w_i = \frac{k(h,h_i)}{\sum_j k(h,h_j)}$

(a) NEC structure



(b) Possible memory operations

This allows to make the system faster and more efficient.

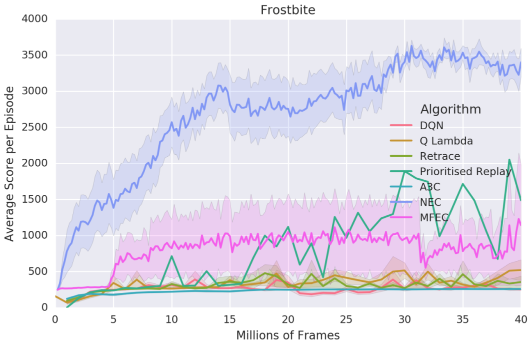
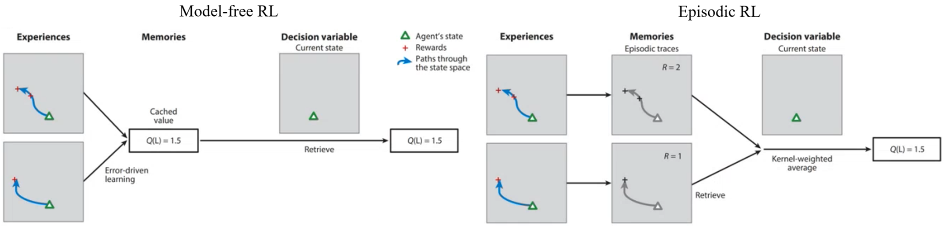


Figure 4.5: Results on Frostbite

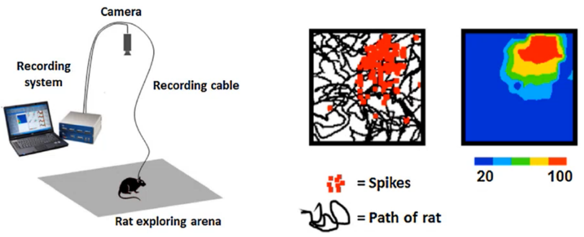
Remark. The difference between model-free and episodic RL is how they store experiences:

- In model-free RL, each individual experience is integrated into a cached value and stored in memory. The cached value is used to make decisions.
- In episodic RL, each individual experience and its reward is stored in memory. Each episodic trace is weighted depending on the similarity to the current state to make a decision.

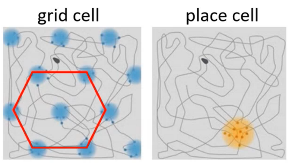


Remark. Episodic control and replay are not the same thing.

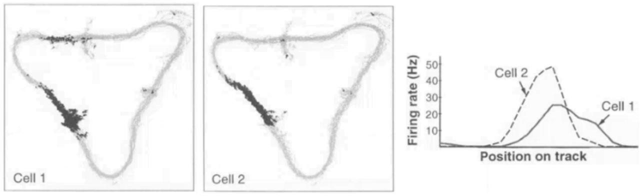
Case study (Replay in rats)**.** The movements and the activity of a neuron (place cell) in the hippocampus of a rat are recorded. It has been observed that the place cell fires only at a specific spatial area (which changes if the rat is moved to another environment).



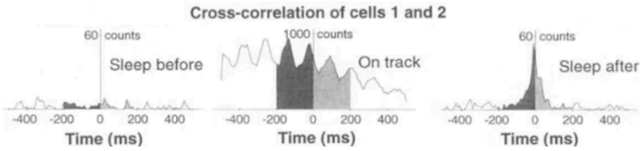
Remark.
Grid cells are another type of cells in the medial entorhinal cortex that fire based on spatial features. Differently from place cells, this type of cell is active in multiple zones arranged in a regular manner (i.e. triangular or hexagonal units).
In other words, place cells are based on landmarks and grid cells are based on self-motion.



By placing the rat on a triangular track where some reward is available, a correlation between two place cells have been recorded.

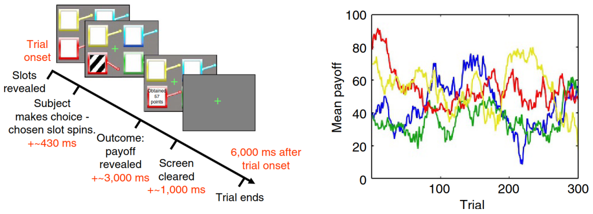


It has been observed that the rat replays the episode as the activation of the two cells becomes correlated also when asleep.



Case study (Memory in human decisions [1]).

Experiment 1 In each trial, candidates are asked to choose a slot machine to spin (bandit problem). Each slot has a different reward in points that changes at each trial.



Two models are used to attempt to fit the behavior of humans:

1. A temporal difference learning approach with a running average estimate of action value.
2. A model that samples from previous experiences to estimate action value. More recent experiences are more likely to be sampled.

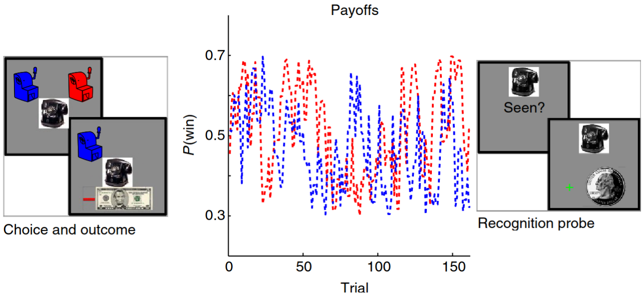
It has been observed that each subject is more similar to a sampling model.

Experiment 2 In the previous experiment, it was not possible to determine which individual trial a participant sampled to make a decision.

The experiment is therefore simplified with two slots with two possible outcomes (±5 dollars) and in some of the trials (32) the choice is also paired with a probe image whose purpose is to bring to mind that specific trial.

After the first round of trials, candidates are asked to perform a second round. Now, before making a choice, one of the probe images is shown to the candidate.

Results show that, by showing the image, decisions are more biased favoring the winning slot and avoiding the losing slot.



4.3 Meta-learning

Meta-learning Use past experience to speed up new learning.

Meta-learning

Remark. Animals use meta-learning as learning does not happen from scratch.

Case study (Meta-learning in monkeys [5])**.** Monkeys are presented with two unfamiliar objects and are asked to grab one of them. Under an object lays either food or nothing. The same procedure is repeated for six trials and the position of the two objects can be swapped. After the first round of trials, new rounds are repeated with new objects.
It has been observed that after some rounds, monkeys are able to learn the task in one shot.

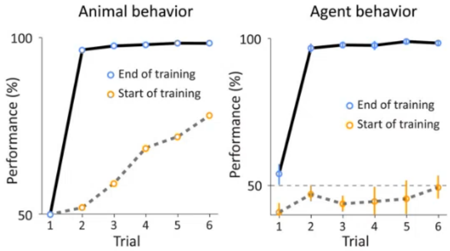
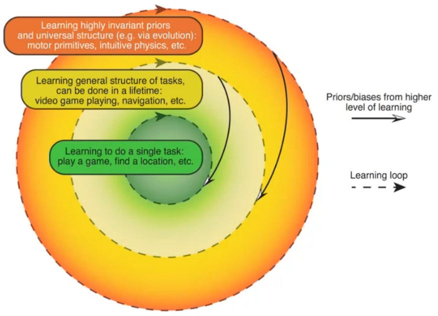


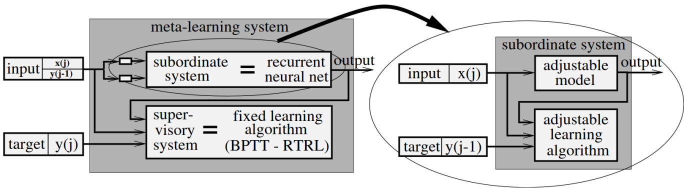
Figure 4.6: Monkeys (left) and machine simulated (right) results

Remark. Learning is nested at multiple scales:

- At the highest level, learning involves evolution and aims to learn highly invariant universal structures (e.g. intuitive physics).
- At the middle level, learning involves the general structure of different tasks (e.g. how to play video games).
- At the innermost level, learning involves a fast adaptation to specific tasks (e.g. how to play a new video game).


Learning highly invariant priors and universal structure (e.g. via evolution): motor primitives, intuitive physics, etc.
Learning general structure of tasks, can be done in a lifetime: video game playing, navigation, etc.
Learning to do a single task: play a game, find a location, etc.
Priors/biases from higher level of learning
Learning loop

Case study (Meta-learning RNN model [6])**.** Meta-learning can be modeled using an RNN that is used to solve a task. At each step, the network takes, apart from the input, the correct output of the previous step that can be seen as a form of supervision.



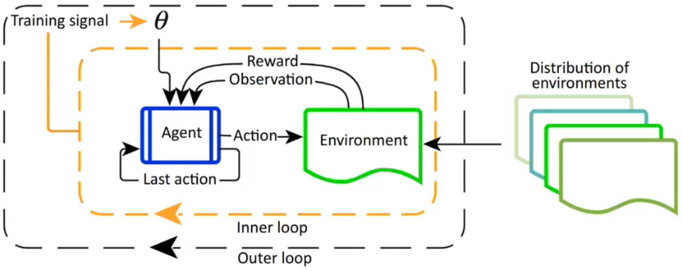
Meta-reinforcement learning Learning system that adjusts another learning system. It can be described with two loops:

Meta-reinforcement learning

Outer-loop System that uses its experience over many task contexts to adjust the parameters of the inner-loop. It is a slow system as it learns multiple tasks.

Inner-loop System that is tuned by the outer-loop. It is a fast system that needs to adapt to a specific task.

Example. An RNN that tasks as input the last action, reward and state.

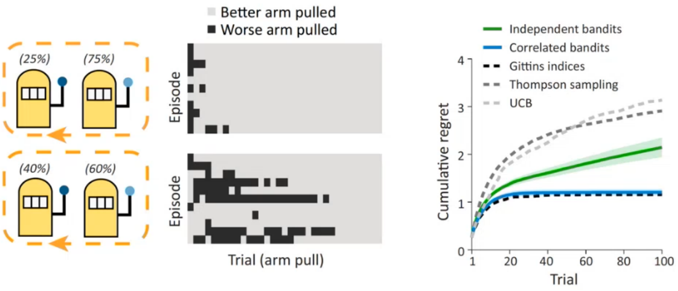


Remark. Meta-learning emerges spontaneously from two basic conditions:

- A learning system with some form of short-term memory.
- A training environment that exposes the learning system to interrelated tasks.

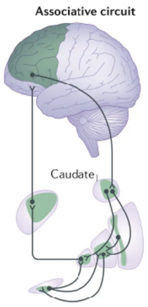
If the two conditions are satisfied, the system slowly learns to use the short-term system as a basis for fast learning.

Case study (RNN meta-learning [25])**.** A recurrent neural network is trained on a series of interrelated RL tasks (choose between two slot machines, i.e. bandit problem) varying only on their parametrization. The network interacts with one bandit problem (consisting of two slots) at a time for a fixed number of steps before moving to another one.
Results show that the agent learns to balance exploration and exploitation: for harder instances (i.e. similar probability of winning) it explores more while easier instances are exploited more.
Moreover, after some training, the network with its weights fixed is able to explore new bandit problems.



Case study (Meta-learning neuronal mapping [26])**.**

Researchers hypothesized that meta-learning involves the prefrontal cortex where the inner-loop resides. On the other hand, dopamine acts as the outer-loop and is responsible for tuning the inner-loop.
Machine simulations are in favor of this hypothesis.



`<end of course>`

Bibliography

[1] Aaron M. Bornstein et al. “Reminders of past choices bias decisions for reward in humans”. In: *Nature Communications* 8.1 (2017), p. 15958. ISSN: 2041-1723. DOI: 10.1038/ncomms15958.

[2] Chun Yun Chang et al. “Optogenetic blockade of dopamine transients prevents learning induced by changes in reward features”. In: *Curr. Biol.* 27.22 (2017), 3480–3486.e3. DOI: 10.1016/j.cub.2017.09.049.

[3] Will Dabney et al. “A distributional code for value in dopamine-based reinforcement learning”. In: *Nature* 577.7792 (2020), pp. 671–675. DOI: 10.1038/s41586-019-1924-6.

[4] Christopher D. Fiorillo, Philippe N. Tobler, and Wolfram Schultz. “Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons”. In: *Science* 299.5614 (2003), pp. 1898–1902. DOI: 10.1126/science.1077349.

[5] H F Harlow. “The formation of learning sets”. In: *Psychol Rev* 56.1 (1949), pp. 51–65. DOI: 10.1037/h0062474.

[6] Sepp Hochreiter, A. Steven Younger, and Peter R. Conwell. “Learning to Learn Using Gradient Descent”. In: *Artificial Neural Networks — ICANN 2001*. Springer Berlin Heidelberg, 2001, pp. 87–94. DOI: 10.1007/3-540-44668-0_13.

[7] Jeffrey R. Hollerman and Wolfram Schultz. “Dopamine neurons report an error in the temporal prediction of reward during learning”. In: *Nature Neuroscience* 1.4 (1998), pp. 304–309. DOI: 10.1038/1124.

[8] Chou P. Hung et al. “Fast Readout of Object Identity from Macaque Inferior Temporal Cortex”. In: *Science* 310.5749 (2005), pp. 863–866. DOI: 10.1126/science.1117593.

[9] Kohitij Kar et al. “Evidence that recurrent circuits are critical to the ventral stream’s execution of core object recognition behavior”. In: *Nature Neuroscience* 22.6 (2019), pp. 974–983. DOI: 10.1038/s41593-019-0392-5.

[10] Brenden M. Lake et al. “Building machines that learn and think like people”. In: *Behavioral and Brain Sciences* 40 (2017), e253. DOI: 10.1017/S0140525X16001837.

[11] Volodymyr Mnih et al. “Human-level control through deep reinforcement learning”. In: *Nature* 518.7540 (2015), pp. 529–533. ISSN: 1476-4687. DOI: 10.1038/nature14236.

[12] I. Momennejad et al. “The successor representation in human reinforcement learning”. In: *Nature Human Behaviour* 1.9 (2017), pp. 680–692. ISSN: 2397-3374. DOI: 10.1038/s41562-017-0180-8.

[13] Hiroyuki Nakahara et al. “Dopamine neurons can represent context-dependent prediction error”. In: *Neuron* 41 (2004), pp. 269–280. DOI: 10.1016/S0896-6273(03)00869-9.

[14] Alexander Pritzel et al. *Neural Episodic Control*. 2017. arXiv: 1703.01988 [cs.LG].

[15] Rishi Rajalingham, Kailyn Schmidt, and James J. DiCarlo. "Comparison of Object Recognition Behavior in Human and Monkey". In: *Journal of Neuroscience* 35.35 (2015), pp. 12127–12136. DOI: 10.1523/JNEUROSCI.0573-15.2015.

[16] Rishi Rajalingham et al. "Large-Scale, High-Resolution Comparison of the Core Visual Object Recognition Behavior of Humans, Monkeys, and State-of-the-Art Deep Artificial Neural Networks". In: *Journal of Neuroscience* 38.33 (2018), pp. 7255–7269. DOI: 10.1523/JNEUROSCI.0388-18.2018.

[17] R. Romo and W. Schultz. "Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements". In: *Journal of Neurophysiology* 63.3 (1990), pp. 592–606. DOI: 10.1152/jn.1990.63.3.592.

[18] Brian F Sadacca, Joshua L Jones, and Geoffrey Schoenbaum. "Midbrain dopamine neurons compute inferred and cached value prediction errors in a common framework". In: *eLife* 5 (2016), e13665. ISSN: 2050-084X. DOI: 10.7554/eLife.13665.

[19] Wolfram Schultz. "Behavioral Theories and the Neurophysiology of Reward". In: *Annual Review of Psychology* 57.Volume 57, 2006 (2006), pp. 87–115. DOI: 10.1146/annurev.psych.56.091103.070229.

[20] Wolfram Schultz, Peter Dayan, and P. Read Montague. "A Neural Substrate of Prediction and Reward". In: *Science* 275.5306 (1997), pp. 1593–1599. DOI: 10.1126/science.275.5306.1593.

[21] Clara Kwon Starkweather et al. "Dopamine reward prediction errors reflect hidden-state inference across time". In: *Nature Neuroscience* 20.4 (2017), pp. 581–589. ISSN: 1546-1726. DOI: 10.1038/nn.4520.

[22] Hanlin Tang et al. "Recurrent computations for visual pattern completion". In: *Proceedings of the National Academy of Sciences* 115.35 (2018), pp. 8835–8840. DOI: 10.1073/pnas.1719397115.

[23] Gerald Tesauro. "TD-Gammon, a Self-Teaching Backgammon Program, Achieves Master-Level Play". In: *Neural Computation* 6.2 (1994), pp. 215–219. ISSN: 0899-7667. DOI: 10.1162/neco.1994.6.2.215.

[24] Philippe N. Tobler, Christopher D. Fiorillo, and Wolfram Schultz. "Adaptive Coding of Reward Value by Dopamine Neurons". In: *Science* 307.5715 (2005), pp. 1642–1645. DOI: 10.1126/science.1105370.

[25] Jane X Wang et al. *Learning to reinforcement learn.* 2017. arXiv: 1611.05763 [cs.LG].

[26] Jane X. Wang et al. "Prefrontal cortex as a meta-reinforcement learning system". In: *Nature Neuroscience* 21.6 (2018), pp. 860–868. ISSN: 1546-1726. DOI: 10.1038/s41593-018-0147-8.

[27] Daniel L. K. Yamins et al. "Performance-optimized hierarchical models predict neural responses in higher visual cortex". In: *Proceedings of the National Academy of Sciences* 111.23 (2014), pp. 8619–8624. DOI: 10.1073/pnas.1403112111.

[28] Chengxu Zhuang et al. "Unsupervised neural network models of the ventral visual stream". In: *Proceedings of the National Academy of Sciences* 118.3 (2021), e2014196118. DOI: 10.1073/pnas.2014196118.

[29] David Zipser and Richard A. Andersen. "A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons". In: *Nature* 331.6158 (1988), pp. 679–684. ISSN: 1476-4687. DOI: 10.1038/331679a0.