

Centre for Cognitive Science

Bratislava



Computational cognitive neuroscience:

7. Motor Control and Reinforcement Learning

Lubica Beňušková
Centre for Cognitive Science, FMFI
Comenius University in Bratislava

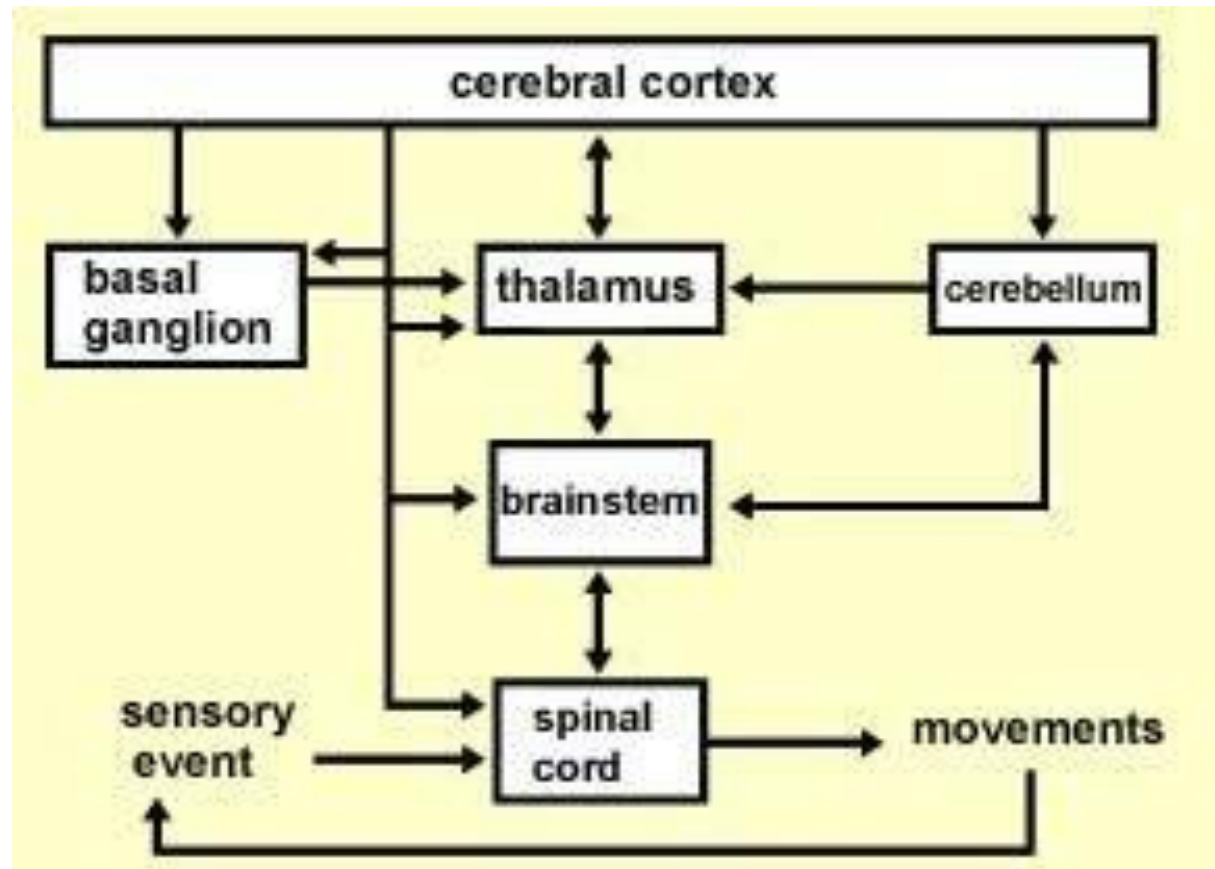
Sensory-motor loop

- The essence of behaviour is built upon the so-called **sensory-motor loop** – i.e. processing sensory inputs to determine which motor action to perform next.
- This is the most basic function of any nervous system – from worms to humans, from sensation to action.



- The human brain has a number of such loops, from the most primitive reflexes in the peripheral nervous system, up to the most abstract plans, such as the decision to apply to and attend University...

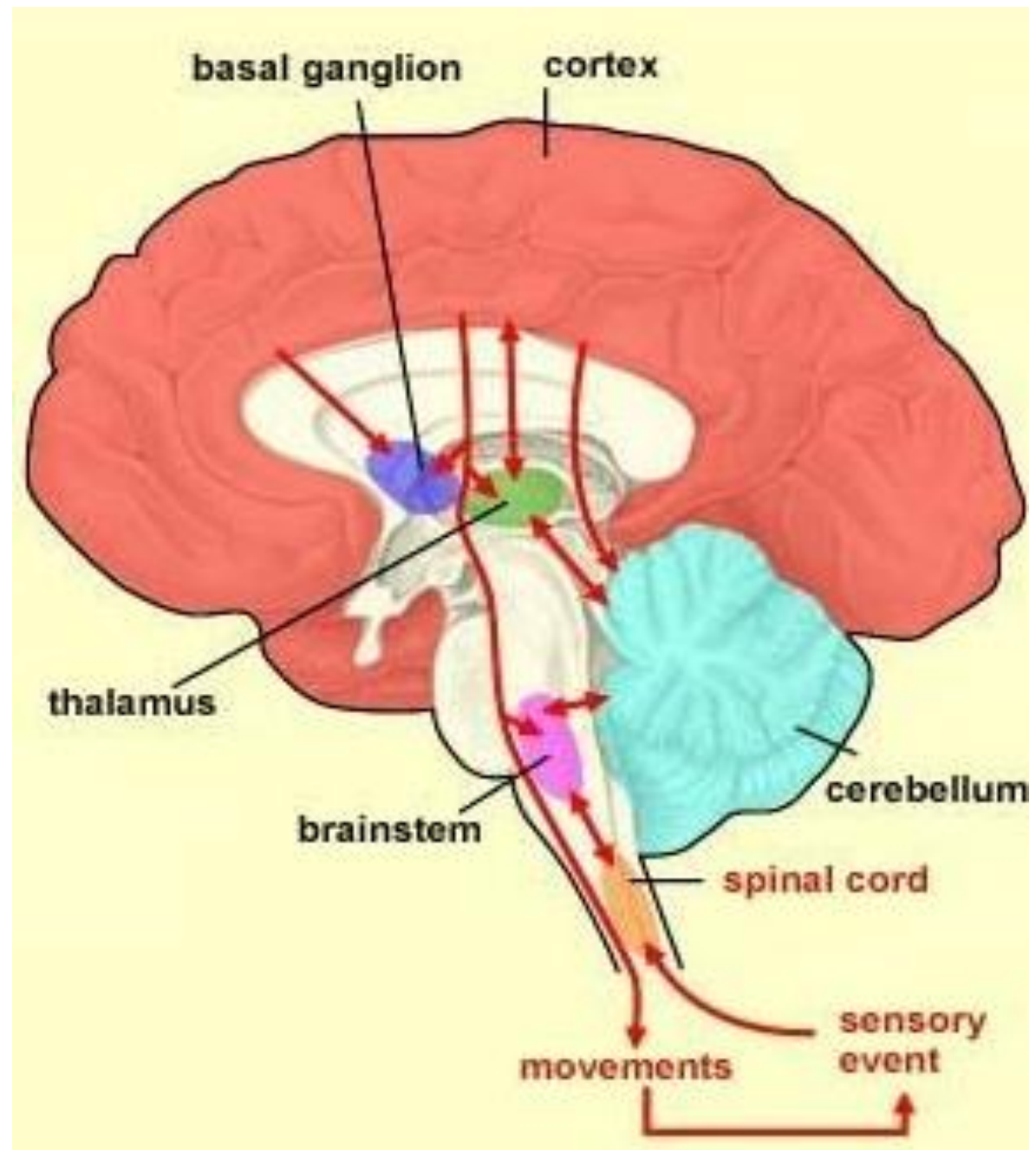
Sensory-motor loop: the block scheme



- Somatosensory events are relayed through the spinal cord then brainstem and through the thalamus to the cerebral cortex, where they are processed.
- Commands for action are then relayed from the cortex to the basal ganglia and cerebellum, from there back to the brainstem and the spinal cord from where they result in concrete muscle actions (movements).

Sensory-motor loop: the brain scheme

- Somatosensory events are relayed through the spinal cord then brainstem and through the thalamus to the cerebral cortex, where they are processed.
- Commands for action are then relayed from the cortex to the basal ganglia and cerebellum, from there back to the brainstem and the spinal cord from where they result in concrete muscle actions (movements).



Basal Ganglia and the Cerebellum

- The most important motor output and control systems at the subcortical level are the *cerebellum* and *basal ganglia*, each of which has specially adapted learning mechanisms.
- **Basal ganglia** are specialized for learning from reward/punishment signals, in comparison to expectations for reward/punishment, respectively. This learning then guides the **action selection**, selecting the most rewarding actions and avoiding punishing ones. This form of learning is called the **reinforcement learning**.
- **The cerebellum** is specialized for **learning from errors**, specifically errors between the sensory outcomes associated with motor actions, relative to expectations for these sensory outcomes associated with those motor actions. Thus, the cerebellum can refine the **implementation** of a given motor plan, to make it more accurate, efficient, and well-coordinated.

Basal ganglia, cerebellum and the cortex

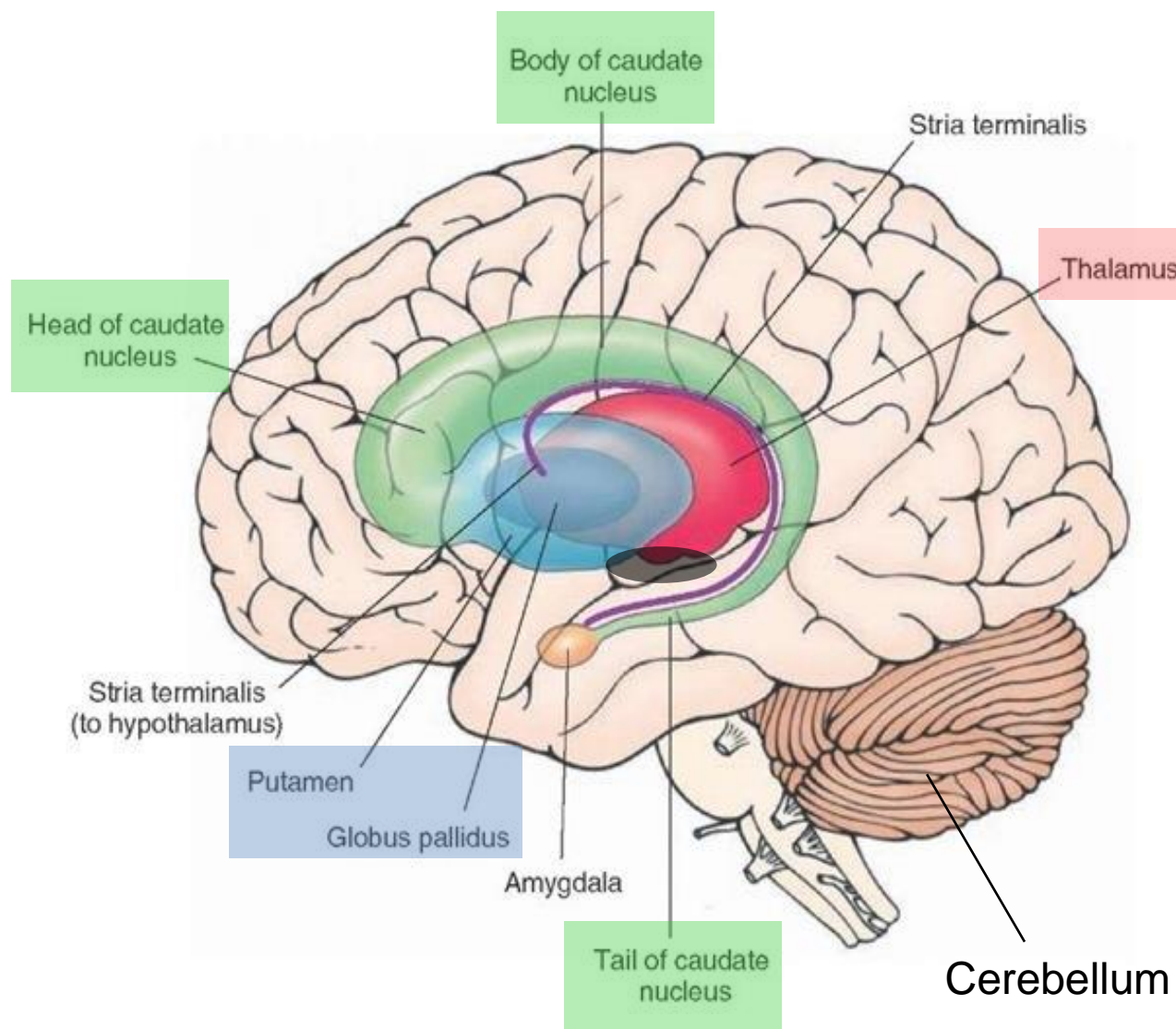
- In short, the basal ganglia select one out of many possible actions to perform, and the cerebellum then makes sure the selected action is performed well.
- In particular, parietal representations (i.e., the **“where” pathway**), drive motor action execution as coordinated by the **cerebellum**, and cerebellum is also densely interconnected with parietal cortex.
- In contrast, the **basal ganglia** are driven to a much greater extent by the ventral **“what” pathway**, which indicates the kinds of rewarding objects that might be present in the environment.
- Interestingly, there are no direct connections between the basal ganglia and cerebellum -- instead, each operates in interaction with various areas in the cortex, where the action plans are formulated and coordinated.

Basal ganglia, cerebellum and the cortex

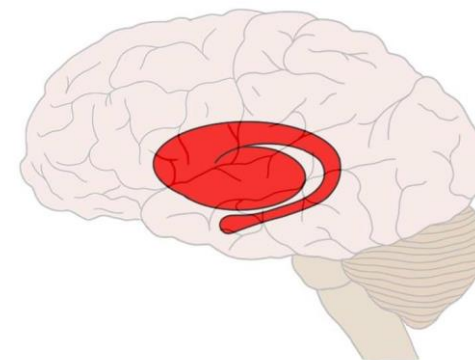
- Both the cerebellum and basal ganglia have a complex dis-inhibitory output dynamics, which produces a gating-like effect on the brain areas they control.
- For example, the basal ganglia can **disinhibit** neurons in specific nuclei of the thalamus, which have bidirectional excitatory circuits through frontal and prefrontal cortical areas. The net effect of this disinhibition is to **enable an action to proceed**, without needing to specify any of the details for how to perform that action. This is what is meant by a **gate** -- something that broadly modulates the flow of other forms of activation.
- The **cerebellum similarly disinhibits parietal and frontal neurons** to affect its form of precise control over the shape of motor actions. It also projects directly to motor outputs in the brain stem, something that is not true of most basal ganglia areas.

Parts of Basal ganglia (BG)

- **Striatum**, which is comprised of
 - **caudate nucleus**
 - and **putamen**
- **Globus pallidus**
 - internal segment (GPi)
 - external segment (GPe)
- **Thalamus and subthalamic nucleus**
- **Substantia nigra pars compacta (SNc)**

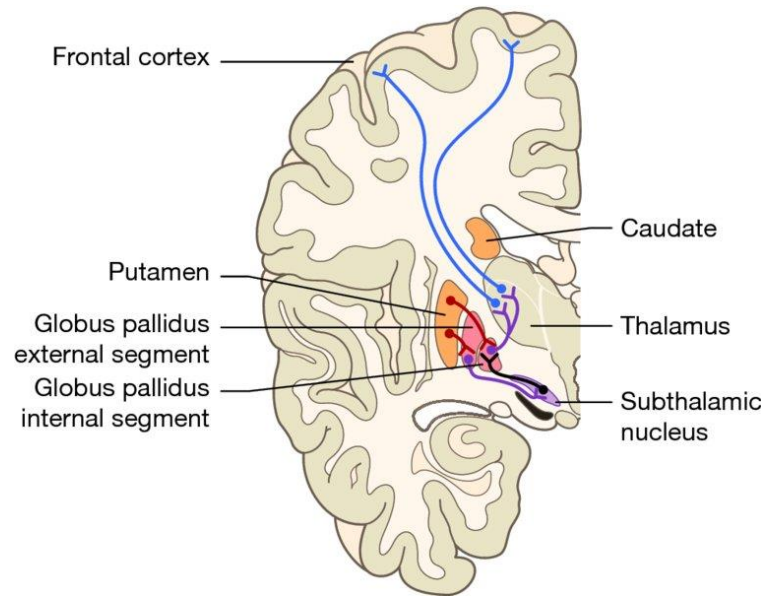


Striatum



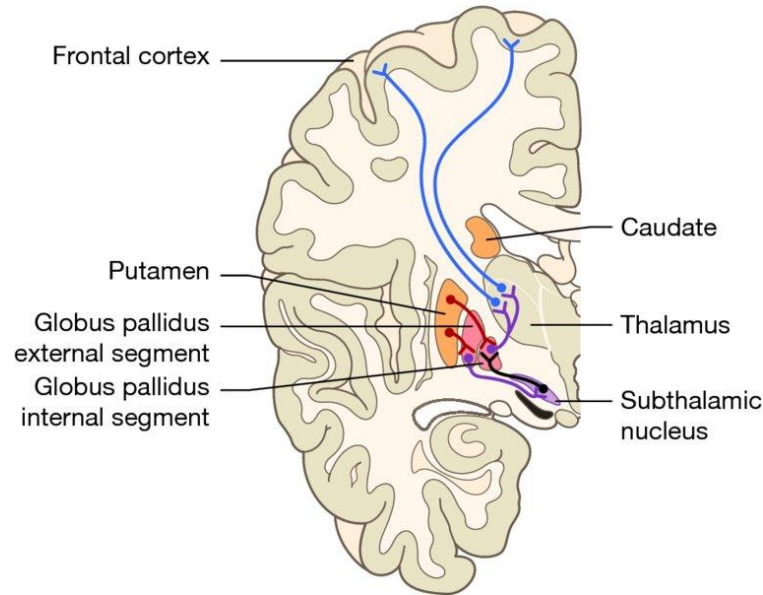
- The **striatum** is the major **input region** of BG.
- Consists of the **caudate and putamen**.
- The striatum is anatomically subdivided into many small clusters of neurons, with two major types of clusters: **patch/striosomes** and **matrix/matrisomes**.
- The **matrix** clusters contain **direct (Go) and indirect (NoGo)** pathway medium spiny neurons, which together make up 95% of striatal cells, both of which receive **excitatory inputs** from all over the cortex **but are inhibitory** on their downstream targets in the globus pallidus as described next.
- The **patch cells** project to the **dopaminergic system**, and thus appear to play a more indirect role in modulating learning signals.
- There are also a relatively few widely spaced tonically active neurons (TAN's), which release acetylcholine as a neurotransmitter and appear to play a modulatory role, and inhibitory interneurons, which likely perform the same kind of dynamic gain control that they play in the cortex.

Globus pallidus: internal segment (GPi)



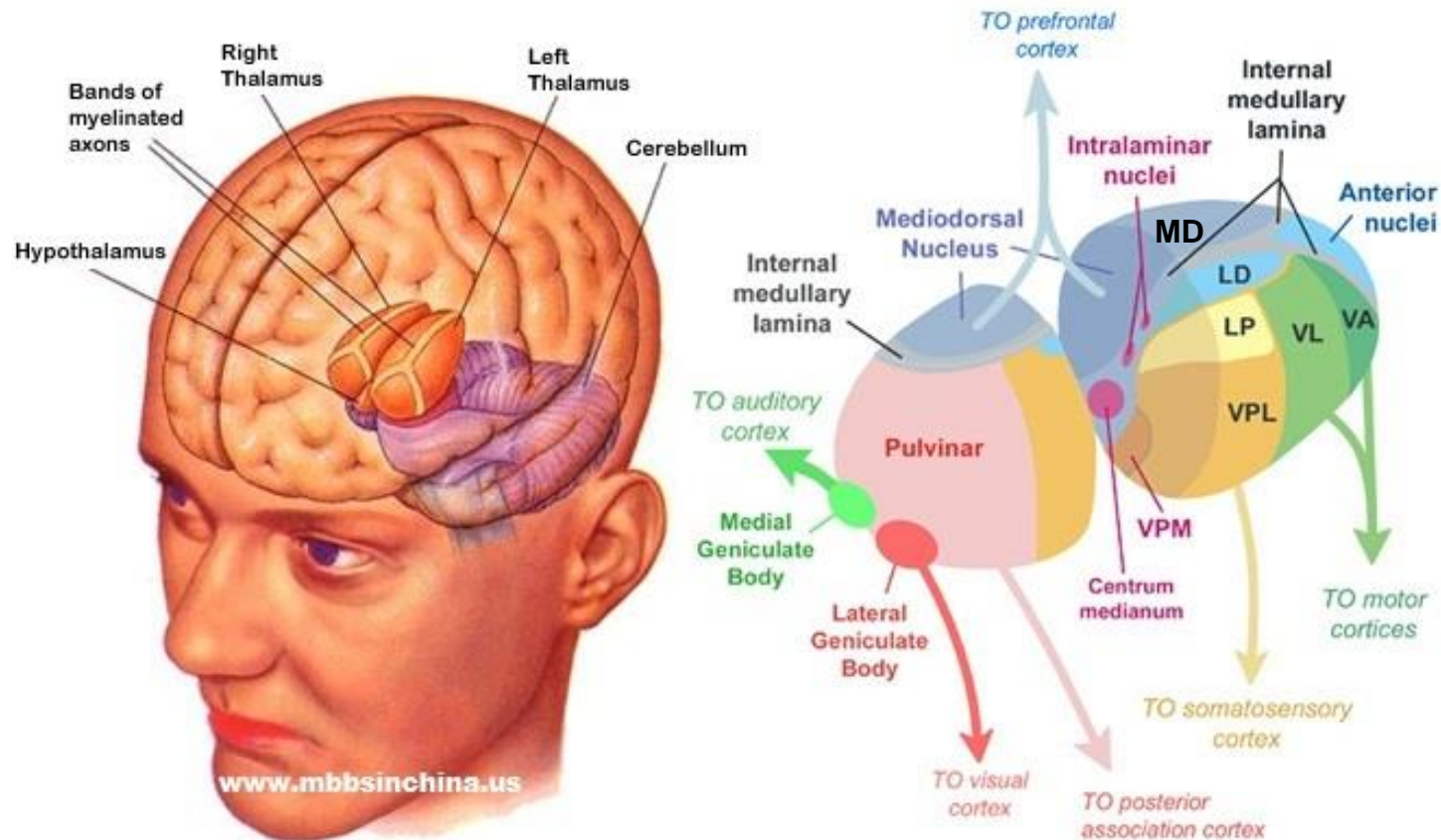
- Contains neurons that are constantly i.e. **tonically active** even with no input. These neurons send **inhibition** to specific nuclei in the thalamus.
- When the direct/**Go** pathway striatum neurons fire, they inhibit these GPi neurons, and thus **disinhibit the thalamus**, resulting ultimately in the **initiation** of a specific motor (or cognitive) action (such as a thought).
- In another fronto-basal ganglia circuits, the role of the GPi is taken up by the substantia nigra pars reticulata (SNr), which receives input from other areas of striatum and projects to outputs regulating other actions (e.g., eye movements in the superior colliculus).

Globus pallidus: external segment (GPe)



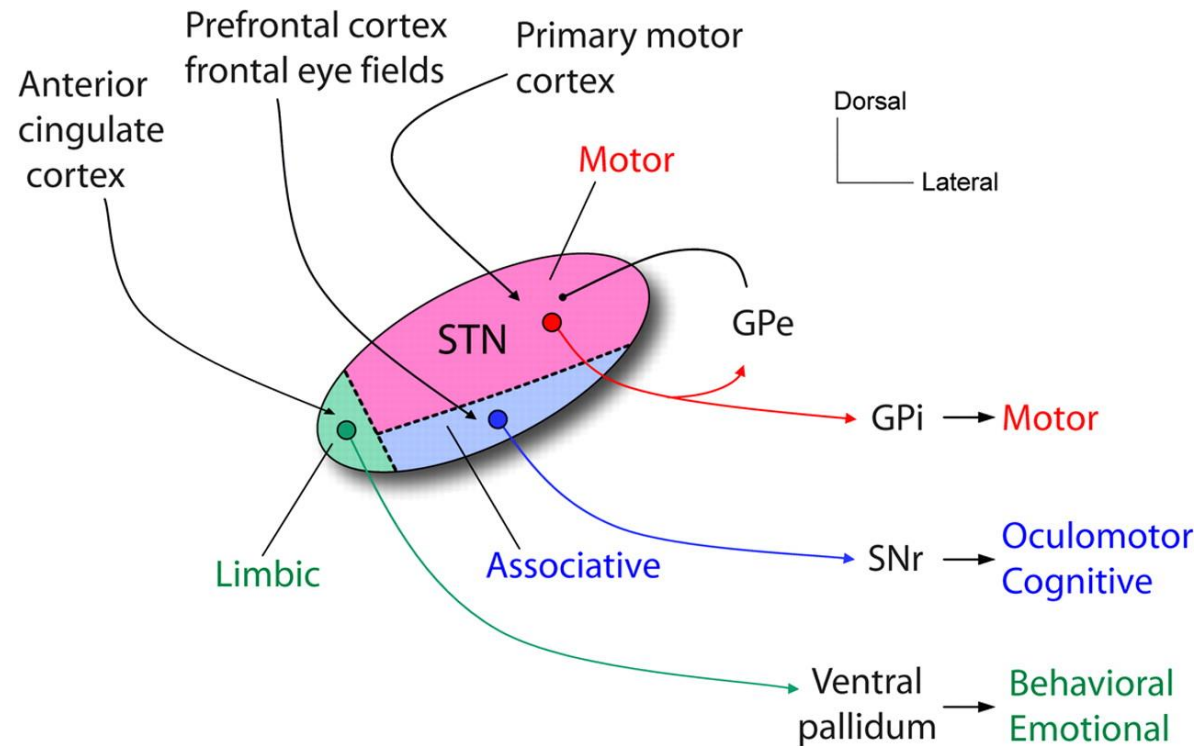
- Contains **tonically active** neurons that send focused **inhibitory projections** to corresponding GPi neurons.
- When the indirect/**NoGo** pathway neurons in the striatum fire, they inhibit the GPe neurons, and thus disinhibit the GPi neurons, causing them to provide even greater inhibition onto the thalamus.
- This **blocks the initiation** of specific actions coded by the population of active **NoGo** neurons.

Thalamus



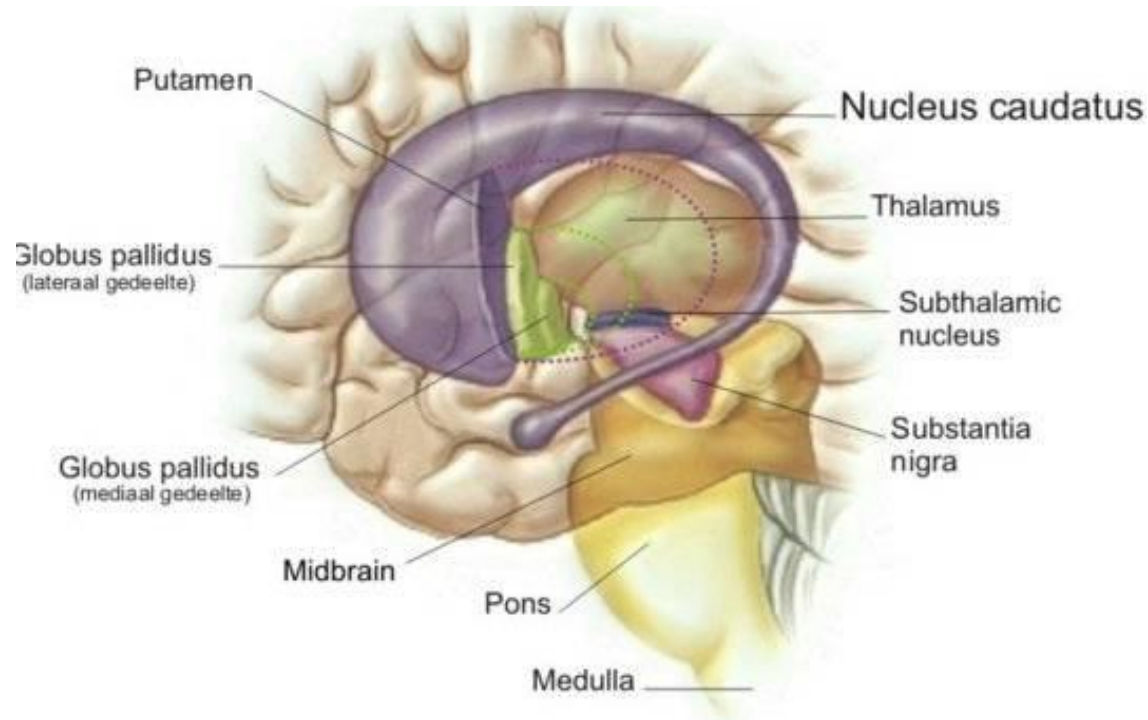
- The thalamus, specifically the medial dorsal (MD), ventral anterior (VA), and ventrolateral (VL) nuclei. When the thalamic neurons get disinhibited by the **Go** pathway firing, they can fire, but only when driven by top-down excitatory input from the frontal cortex. In this way, the basal ganglia serve as a gate on the thalamocortical circuit – **Go** firing opens the gate, while **NoGo** firing closes it.

The sub-thalamic nucleus (STN)



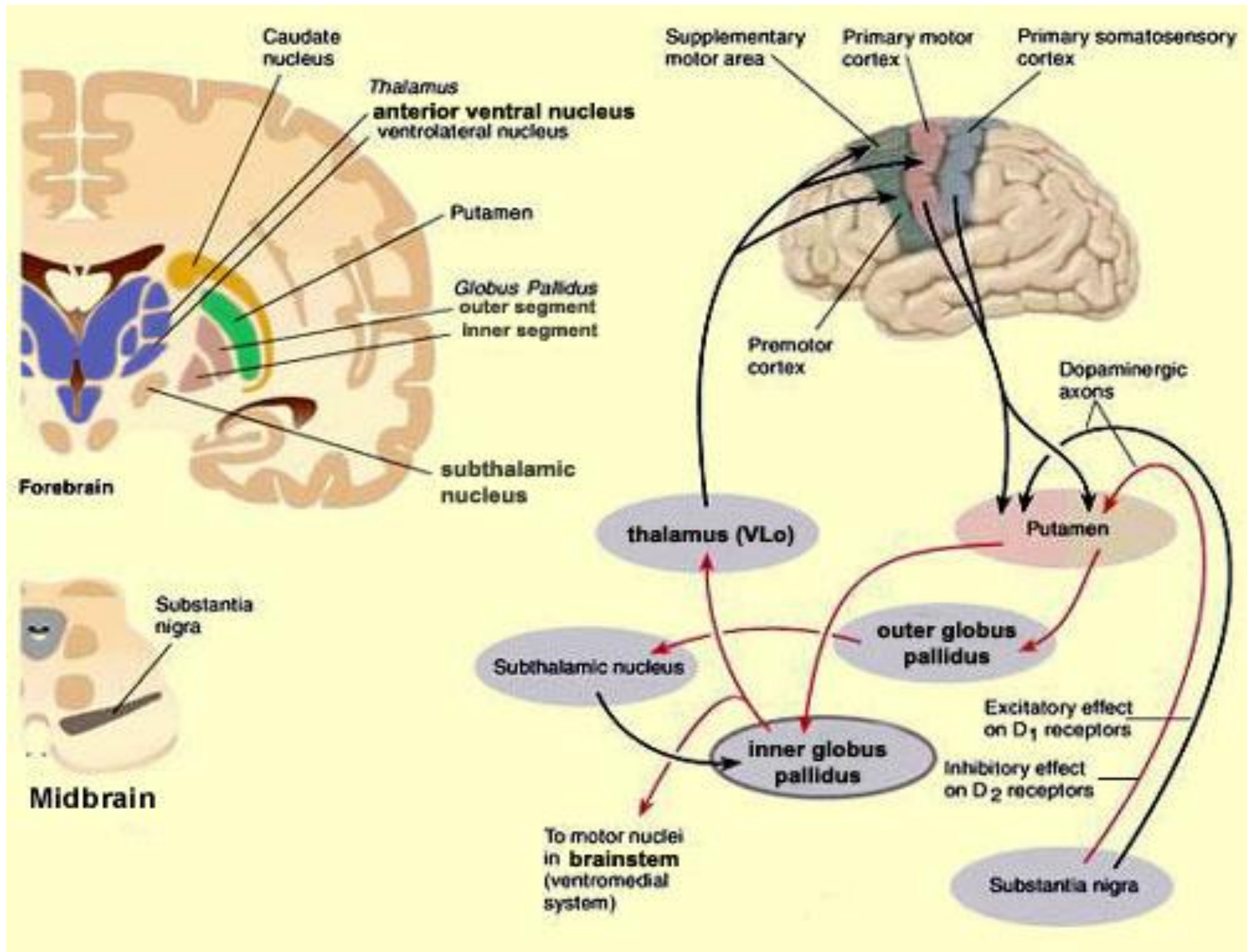
- Acts as the so-called hyperdirect pathway, because it receives input directly from frontal cortex and sends excitatory projections directly to BG output (GPI), bypassing the striatum altogether.
- A single STN neuron projects broadly to many GPI neurons, and as such the STN is thought to provide a global **NoGo** function that prevents gating of any motor or cognitive action (technically, it raises the threshold for gating).

Substantia nigra pars compacta (SNc)

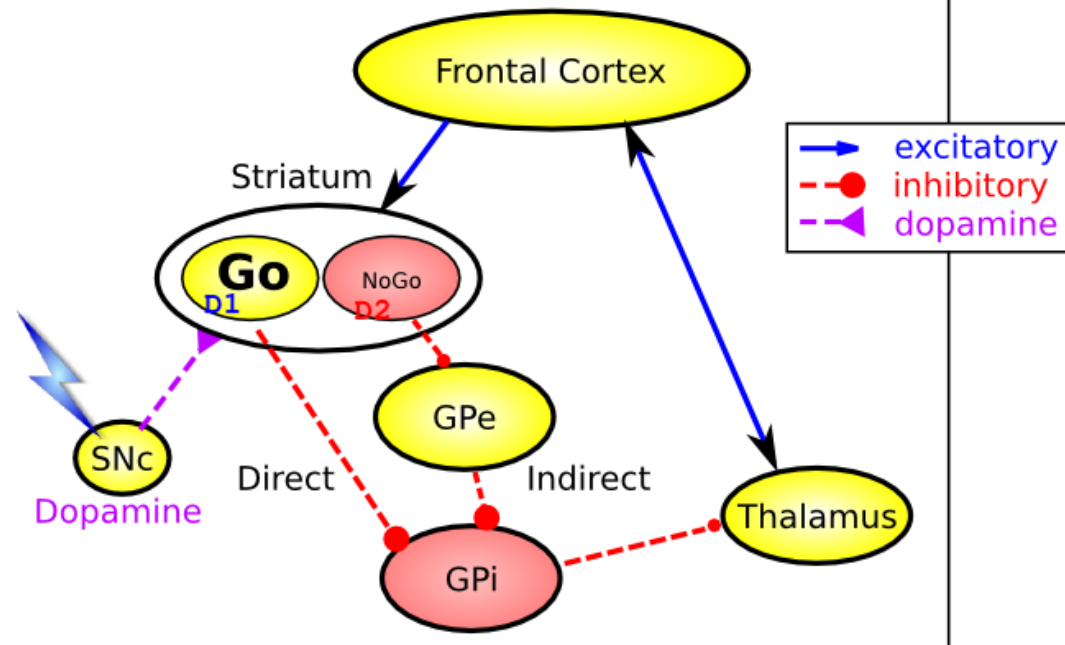


- SNc has neurons release the neuromodulator dopamine and innervate the striatum.
- There are two different kinds of dopamine receptors in the striatum.
 - **D1** receptors are prevalent in **Go** pathway neurons, and dopamine has an excitatory effect on neurons with D1 receptors.
 - In contrast, **D2** receptors are prevalent in **NoGo** pathway neurons, and dopamine has an inhibitory effect via the D2 receptors.

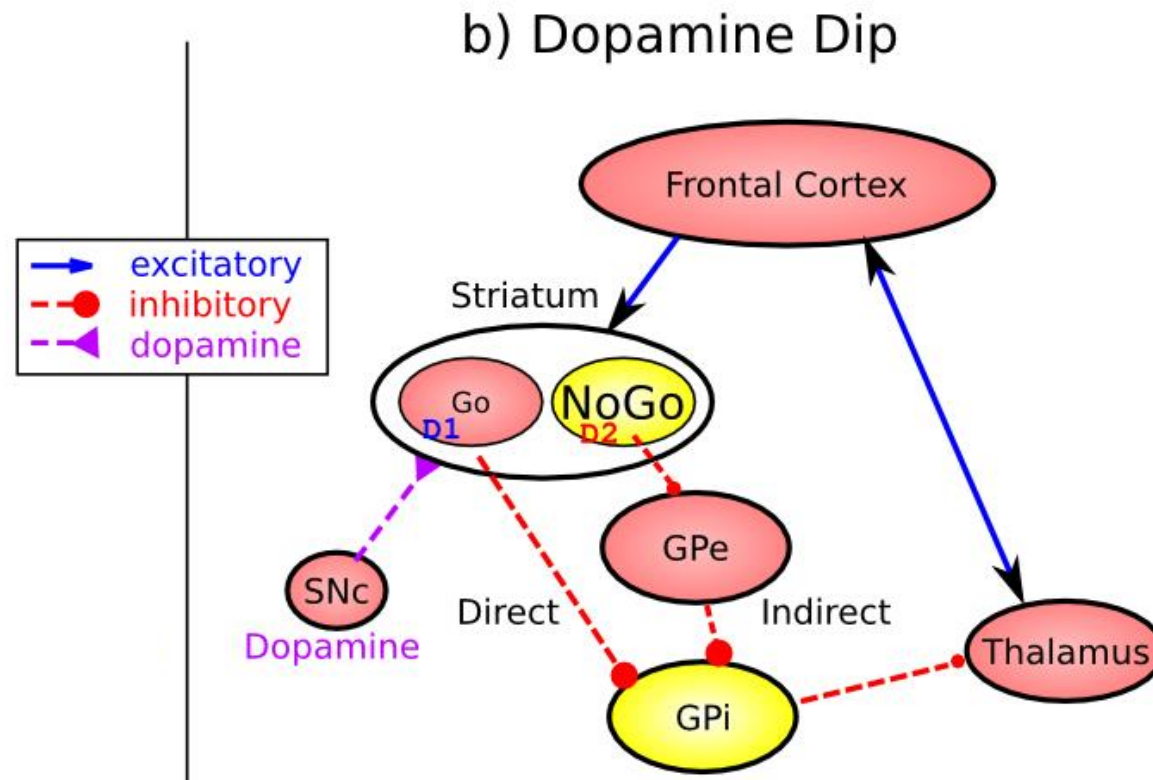
The basal ganglia – cortex loop



a) Dopamine Burst



- DA burst activity drives the direct "**Go**" pathway neurons in the striatum, which then inhibit the tonic activation in the globus pallidus internal segment (GPI), which releases specific nuclei in the thalamus from inhibition, allowing them to complete a bidirectional excitatory circuit with the frontal cortex, resulting in the initiation of a motor action.
- The increased **Go** activity during DA bursts results in potentiation of cortico-striatal synapses, and hence learning to select actions that tend to result in positive outcomes.



- DA dip (decrease in tonic DA neuron firing in SNc) leads to preferential activity of indirect "**NoGo**" pathway neurons in the striatum, which inhibit the external segment globus pallidus neurons (GPe), that inhibit the GPi.
- Increased **NoGo** activity thus results in disinhibition of GPi, making it more active and thus inhibiting the thalamus, preventing initiation of the corresponding motor action.
- The DA dip results in potentiation of corticostriatal **NoGo** synapses, and hence learning to avoid selection of actions that tend to result in negative outcomes.

Which inputs influence SNc

- How the dopamine neurons in the SNc come to exhibit their reward prediction error firing.
- **Lateral hypothalamus (LHA)** provides a primary reward signal for basic rewards like food, water, etc.
- Patch-like neurons in the **ventral striatum (VS-patch)** have direct inhibitory connections onto the dopamine neurons in the SNc, and likely play the role of cancelling out the influence of primary reward signals when these rewards have successfully been predicted.
- Central nucleus of the **amygdala (CNA)** is important for driving dopamine firing to the onset of conditioned stimuli. It receives broadly from the cortex, and projects directly and indirectly to the SNc. Neurons in the CNA exhibit CS-related firing.

Dopamine (DA) and Reinforcement Learning

- The DA neurons in SNc encode **the difference** between reward received versus an expectation of reward.
- Prior to conditioning, when a reward is delivered, the dopamine neurons fire a burst of activity.
- After the animal has learned to associate a conditioned stimulus (CS) (e.g., a tone) with the reward, the dopamine neurons now fire to the onset of the CS, and *not to the reward itself*.
- If a reward is withheld after the CS, there is a dip or pause in DA firing, indicating that there was a prediction of the reward, and when it failed to arrive, there was a negative prediction error.
- This pattern of firing is consistent with reinforcement learning applied to cortico-striatal synapses based on reward prediction error.

Reinforcement Learning

- Computationally, the simplest model of reward prediction error is the Rescorla-Wagner conditioning model, which is mathematically identical to the delta rule for perceptron learning, and is simply the difference between the actual reward r and the expected reward R :

$$\Delta w = \delta x$$

$$\delta = r - R$$

$$\delta = r - \sum xw$$

- Where Δw is the change of the weight, x is the input, w is the weight.
- Thus, δ (“delta”) is the prediction error and $R = \sum xw$ is the amount of expected reward, which is computed as a weighted sum over stimuli x over weights w .

Reinforcement Learning

- The weights adapt to try to accurately predict the actual reward values, and in fact this delta value specifies the direction in which the weights should change:

$$\Delta w = \delta x$$

- When the reward prediction is correct, i.e. $r = R$, then the actual reward value is cancelled out by the prediction, i.e. $\delta = 0$ and then $\Delta w = 0$.
- If the actual reward **r is smaller than R** , then $\delta < 0$ and the weights from cortex to striatum decrease in strength.
- If the actual reward **r is bigger than R** , then $\delta > 0$ and the weights from cortex to striatum increase in strength.

Temporal Difference (TD) Reinforcement Learning

- As the reward may occur later in time, relative to Rescorla-Wagner, TD just adds one additional term to the delta equation, representing the future reward values that might come later in time:

$$\delta = (r + f) - R$$

- where **f represents the future rewards**, and now the reward expectation R has to try to anticipate both the current reward r and this future reward f .
- In a simple conditioning task, where the CS reliably predicts a subsequent reward, the onset of the CS results in an increase in this f value, because once the CS arrives, there is a high probability of reward in the near future.
- Furthermore, this f itself is not predictable, because the onset of the CS is not predicted by any earlier cue (and if it was, then that earlier cue would be the real CS, and drive the dopamine burst). Therefore, the R expectation cannot cancel out the f value, and a dopamine burst lasts.

Temporal Difference (TD) Reinforcement Learning

- Although the f value explains CS-onset dopamine firing, it raises the question of how can the system know what kind of rewards are coming in the future?
- Like anything having to do with the future, you fundamentally just **have to use the past** as your guide to guess the future.
- TD does this by specifying something known as a value function, $v(t)$ that is a sum of all present and future rewards, with the future rewards discounted by a γ "gamma" factor, which captures the intuitive notion that rewards further in the future are worth less than those that will occur sooner.
- Thus if $0 < \gamma < 1$, then

$$V(t) = r(t) + \gamma^1 r(t + 1) + \gamma^2 r(t + 2) + \gamma^3 r(t + 3) + \dots$$

TD Reinforcement Learning

- Computationally, if the actual reward r and the expected reward R , then

$$\Delta w = \delta x$$

$$\delta = (r + f) - R$$

$$\delta = (r(t) + \gamma V(t+1)) - V(t)$$

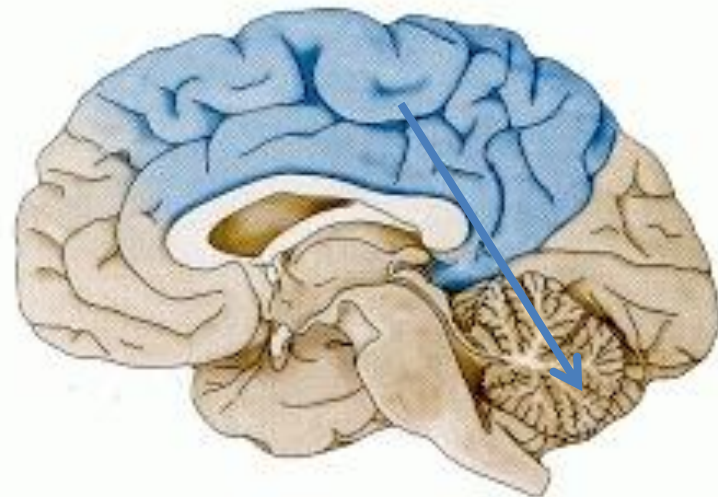
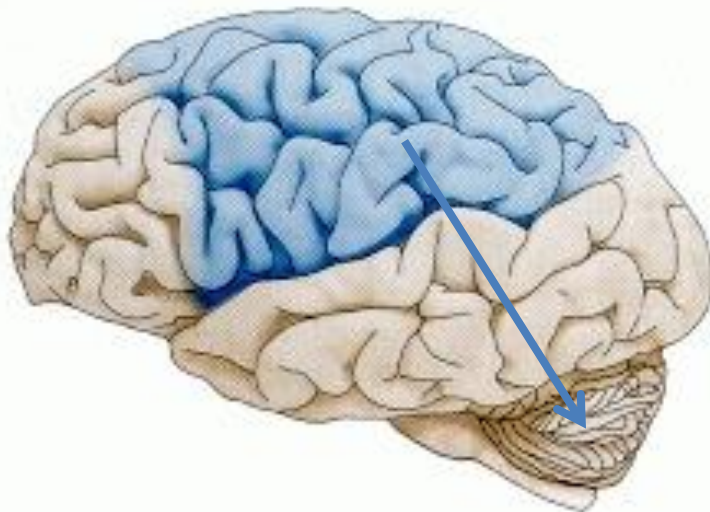
- Thus

$$f = \gamma V(t+1)$$

- Thus, our reward expectation is now a "value expectation" instead (replacing the R with V).
- As with Rescorla-Wagner ruler, the TD delta value here drives learning of the value expectations of the cortico-striatal synapses. ***This rule has been very successful in modelling.***

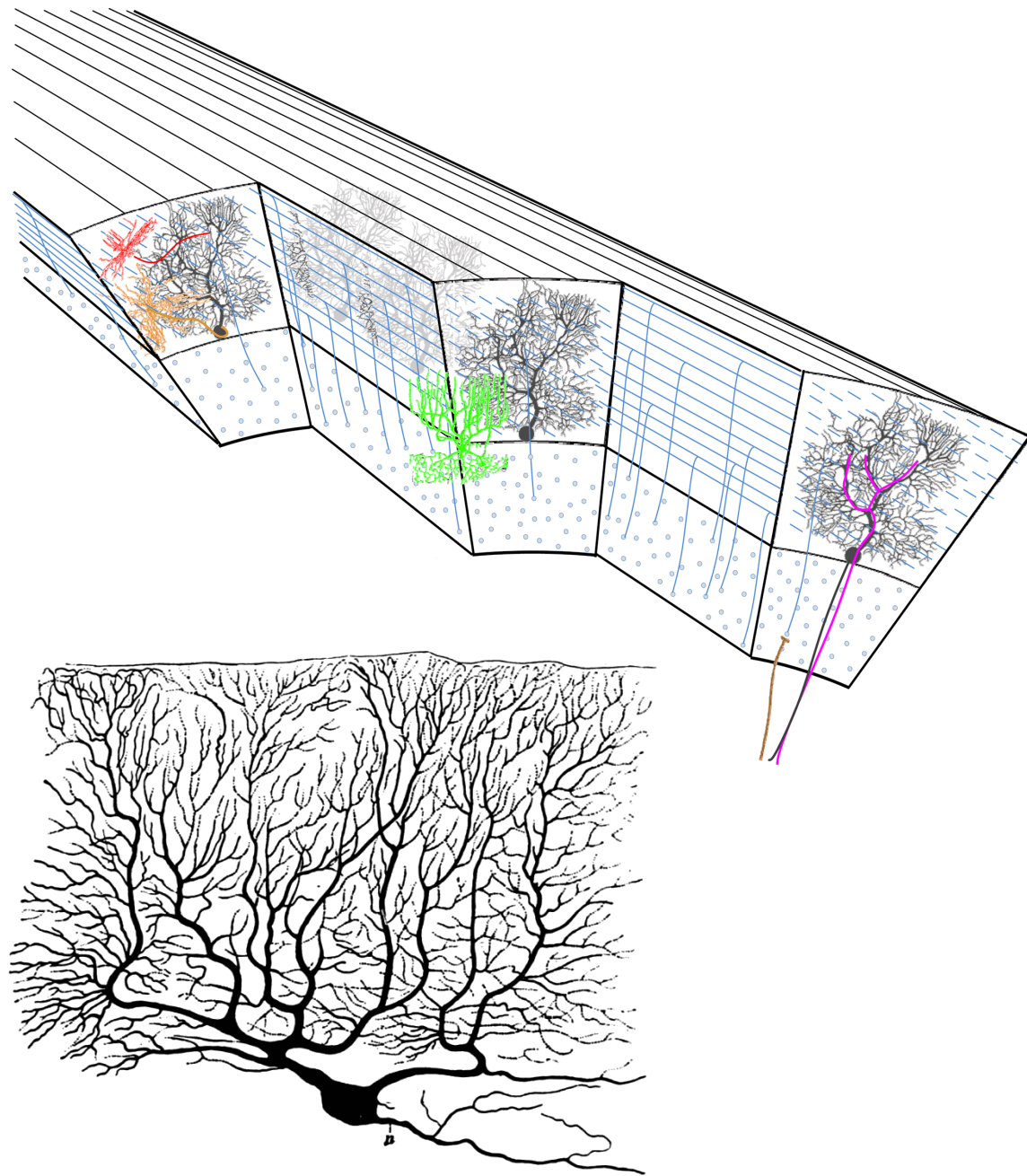
The Cerebellum: inputs

- Once the basal ganglia elect an action to perform based on the reinforcement learning, the cerebellum takes over once the action has been initiated and uses an error-driven learning to shape the performance of the action so that it is accurate and well-coordinated.
- The cerebellum only receives inputs from cortical areas directly involved in the motor production, including the parietal cortex and the motor areas of frontal cortex.
- Unlike the basal ganglia, it does not receive inputs from prefrontal cortex or temporal cortex



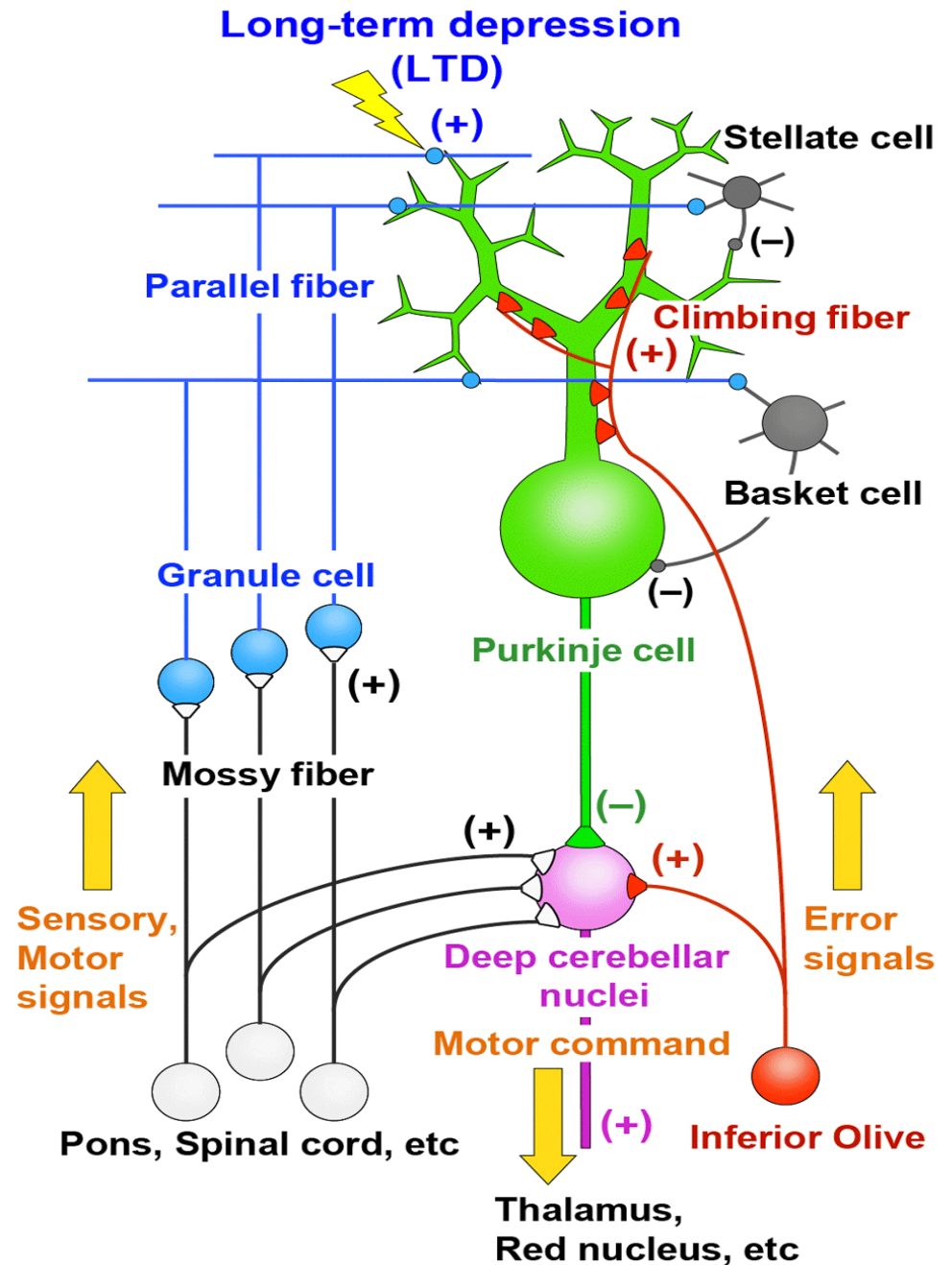
Cerebellum: anatomy

- The cerebellum has a very characteristic anatomy, with the same basic circuit replicated million times over and over.
- There are several neuron types among which the **Purkinje cells** with extremely large dendritic trees are the most dominant.
- These cells were discovered by the Czech scientist Jan Evangelista Purkyně in the 19th century.



Cerebellum: elementary circuitry

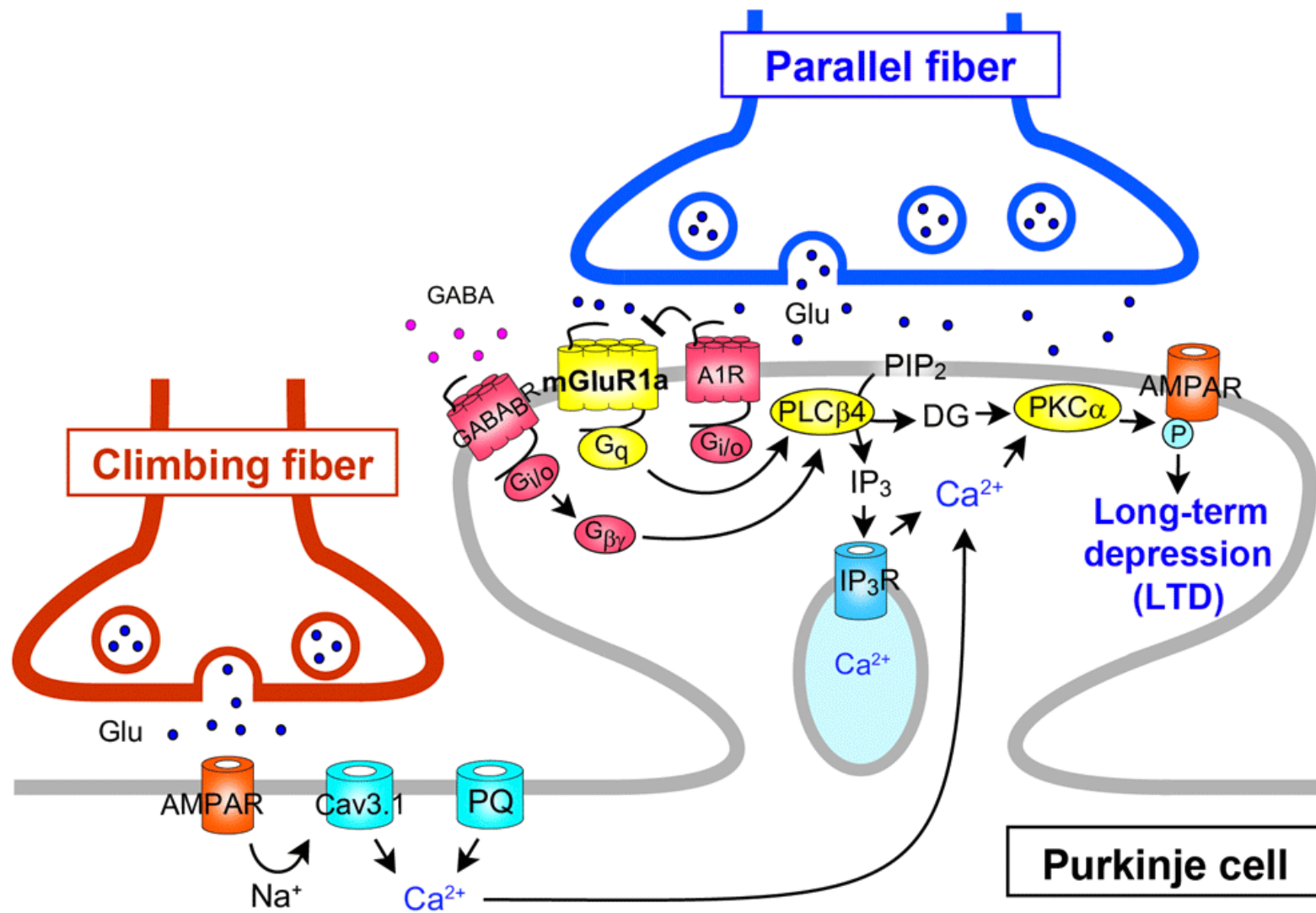
- There are cca 15 million Purkinje cells (PC) and these cells produce output from cerebellum.
- There are cca 40 billion granule cells (GC).
- Each PC receives as many as 200,000 inputs from GCs via their axons called parallel fibers, but only one input climbing fiber.



Cerebellum: learning

- It is thought that climbing fiber inputs relay a training or error signal to the Purkinje cells, which then drives synaptic plasticity in its associated granule cell inputs.
- One prominent idea is that this synaptic plasticity tends to produce LTD (weight decrease) for synaptic inputs where the granule cells are active, which then makes these neurons less likely to fire the Purkinje cell in the future.
- This would make sense given that the Purkinje cells are inhibitory on the deep cerebellar nuclei neurons, so to produce an output from them, the Purkinje cell needs to be turned off.
- David Marr and James Albus have become famous for developing this theory of cerebellum.

mGluR-mediated long-term depression (LTD)



Cerebellum: look-up table

- The goal of this machinery is to associate stimulus inputs with motor output commands, under the command of the climbing fiber inputs.
- One important principle of cerebellar function is the projection of inputs into a very high-dimensional space over the granule cells – computationally this achieves the separation form of learning, where each combination of inputs activates a unique pattern of granule cell neurons.
- This unique pattern can then be associated with a different output signal from the cerebellum, producing something approximating a lookup table of input/output values (for each input pattern x there is only one output $f(x)$).
- A lookup table provides a very robust solution to learning very complex, arbitrary functions – it will always be able to encode any kind of function. The drawback is that it does not generalize to novel input patterns very well.