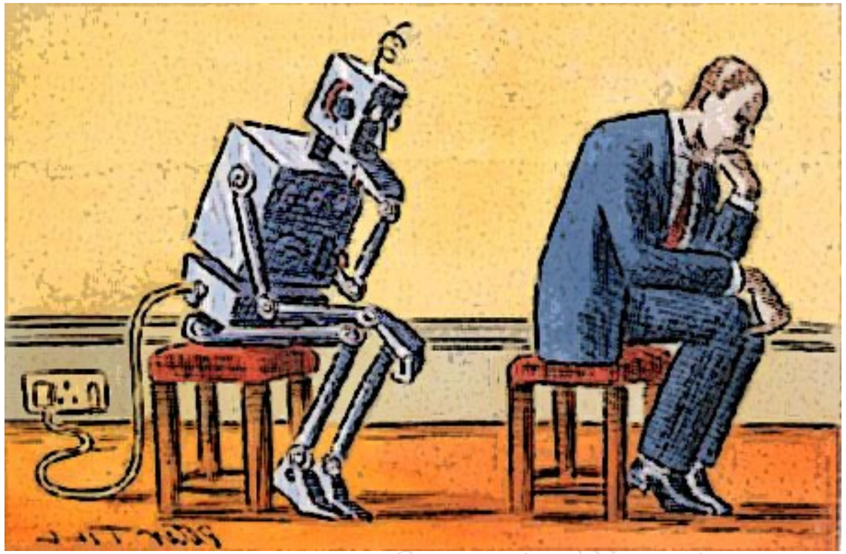
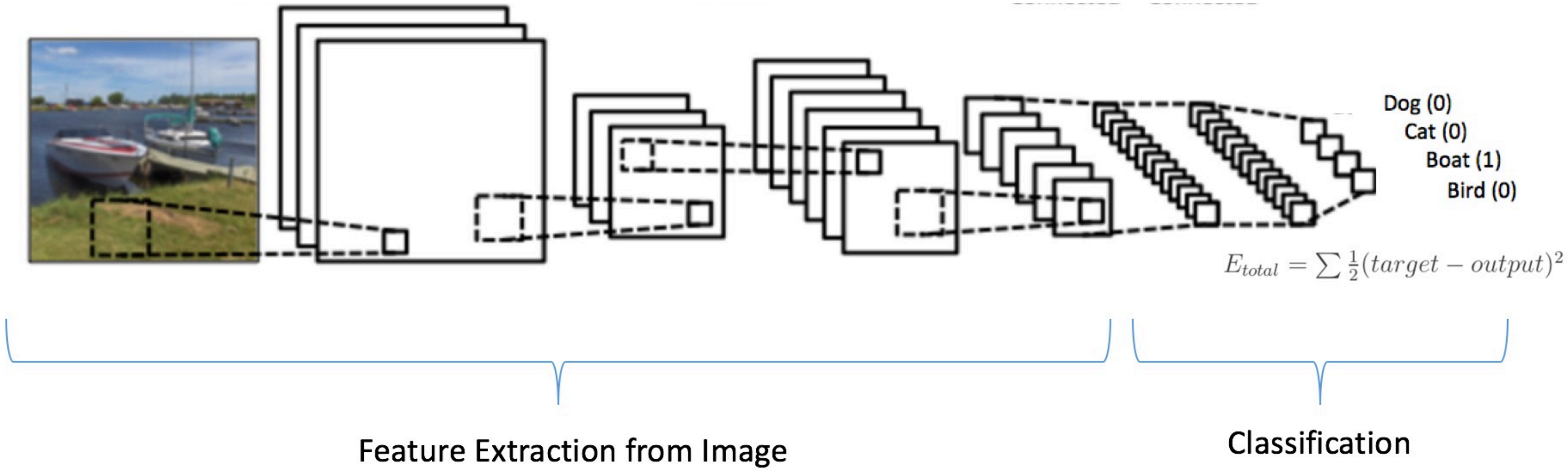
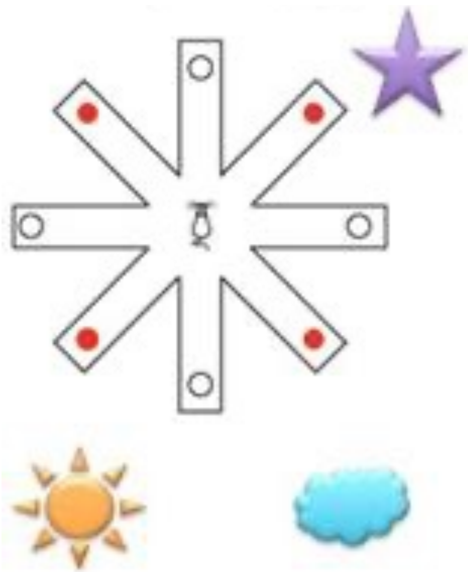
What Cognitive Neuroscience tells Artificial Intelligence

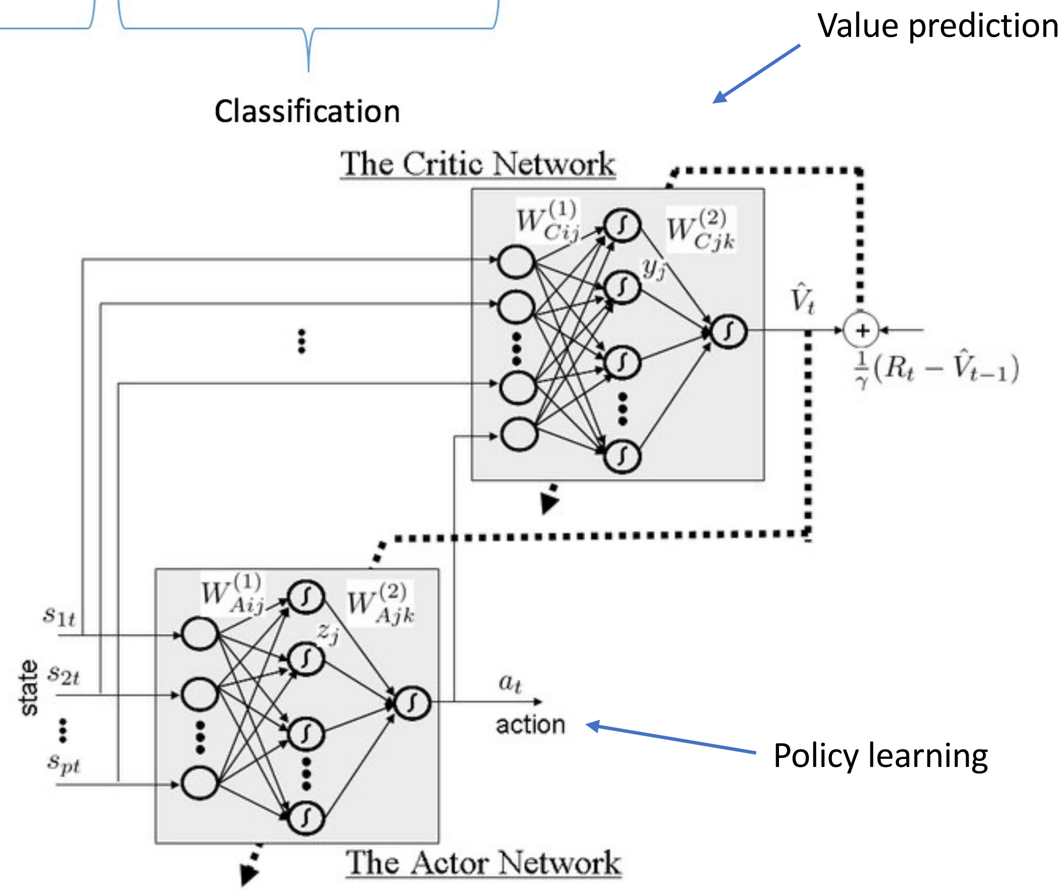


Frederic.Alexandre@inria.fr

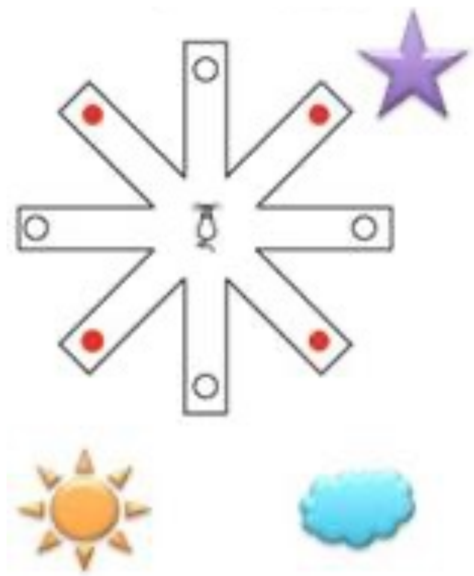
Mnemosyne

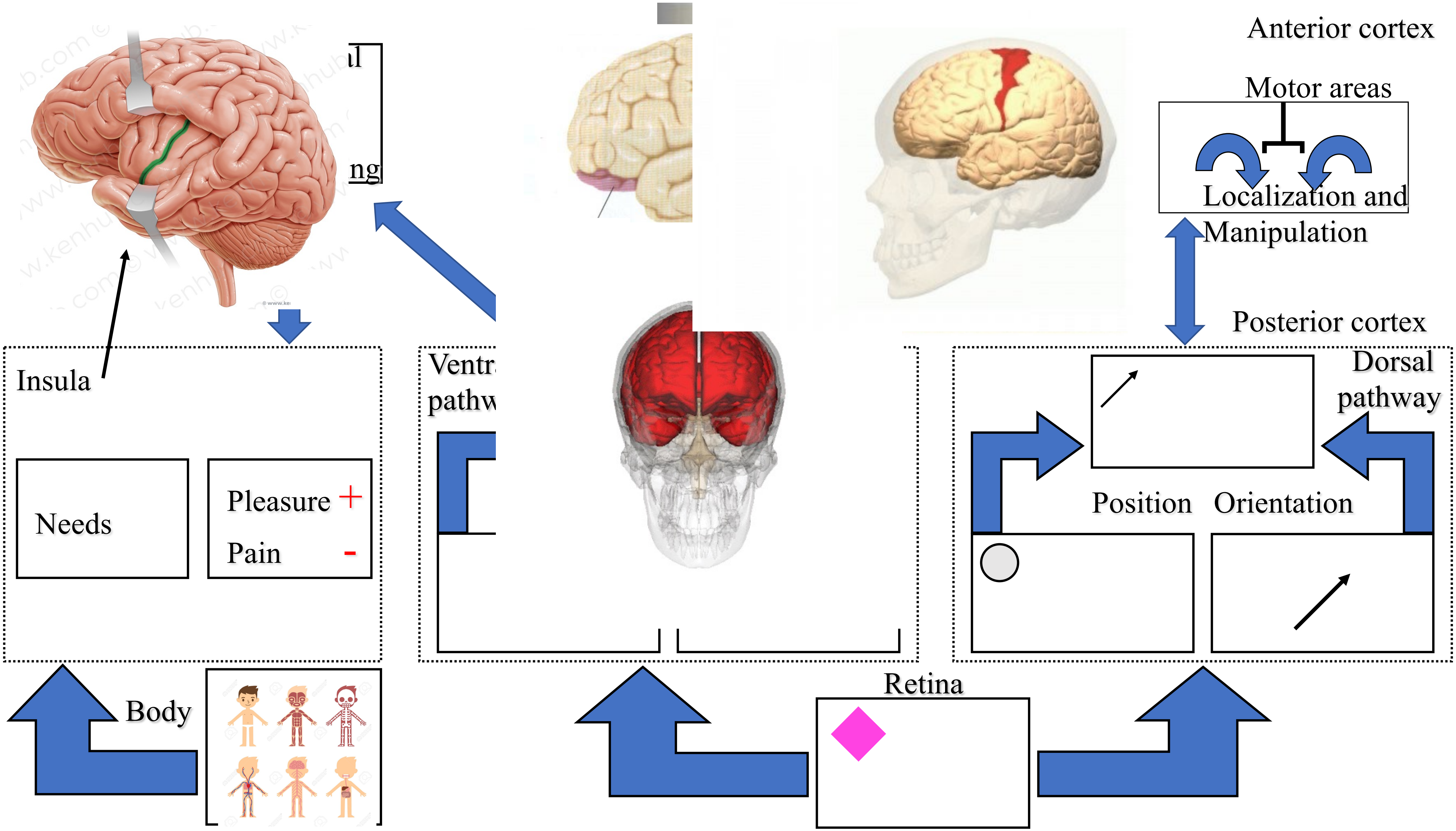

Dog (0)
Cat (0)
Boat (1)
Bird (0)
$E_{total} = \sum \frac{1}{2}(target - output)^2$
Feature Extraction from Image
Classification

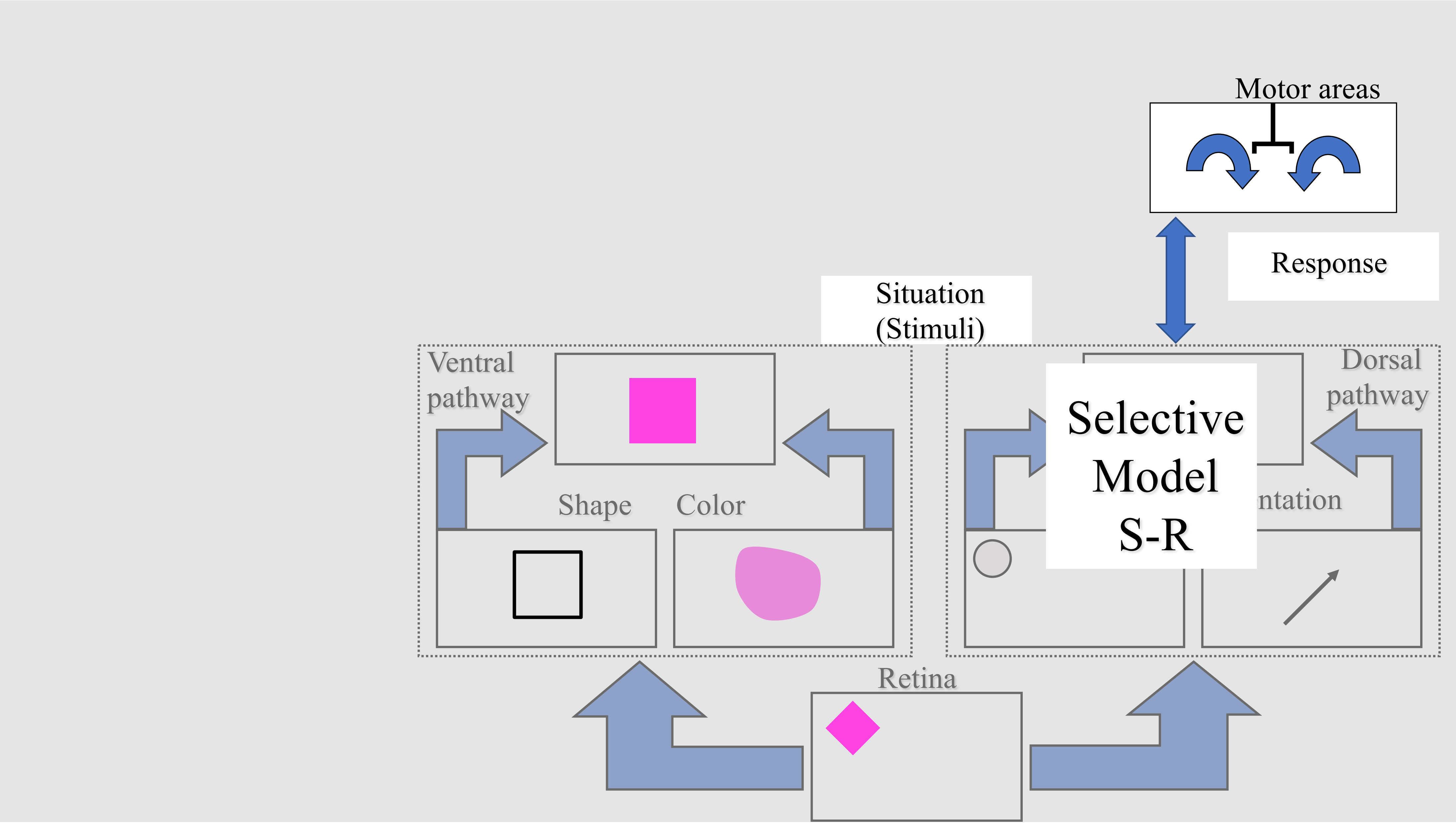


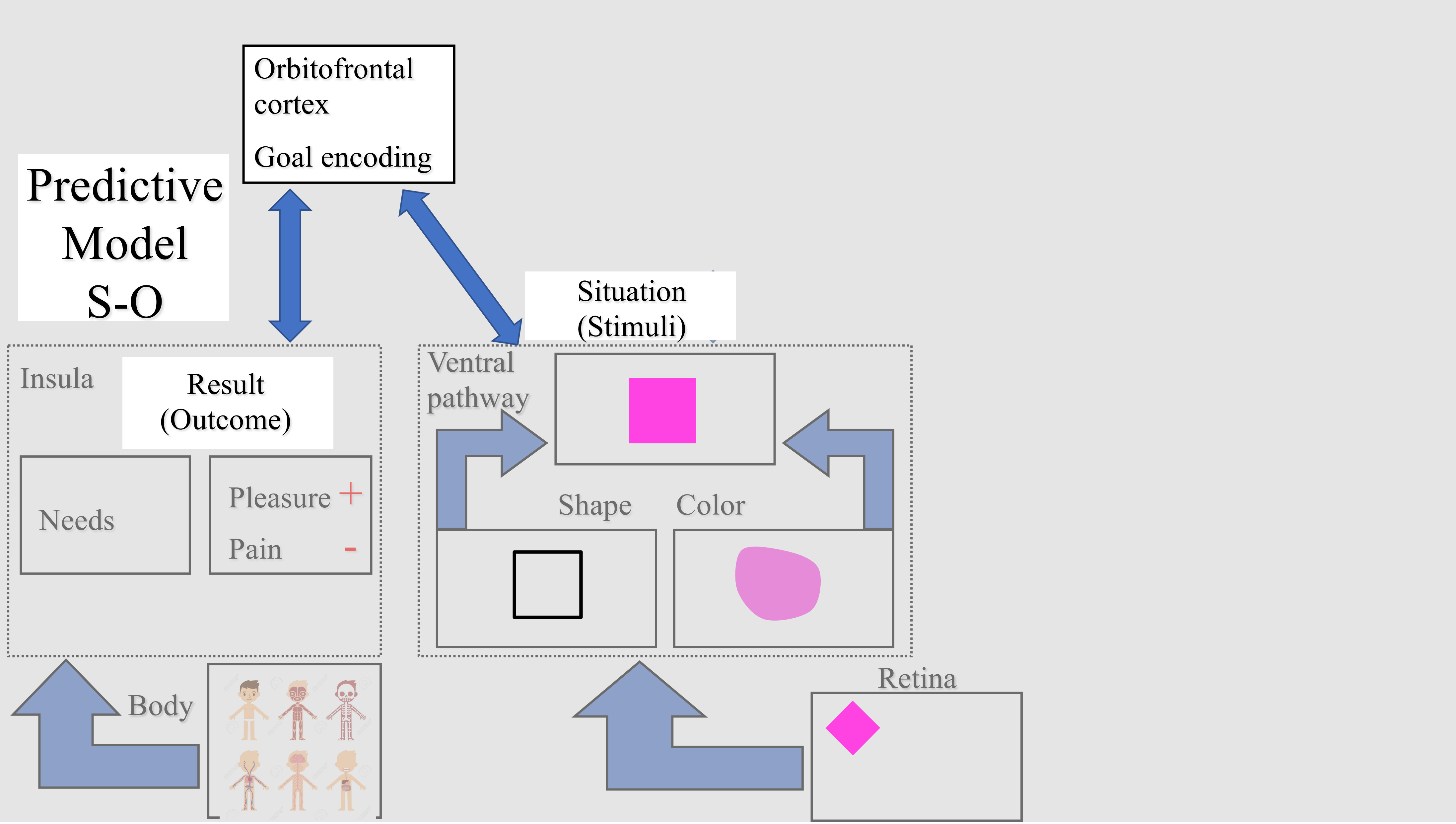

Value prediction
The Critic Network
$W^{(1)}_{Cij}$
$W^{(2)}_{Cjk}$
y_j
$\hat{V}_t$
$\frac{1}{\gamma}(R_t - \hat{V}_{t-1})$
$W^{(1)}_{Aij}$
$W^{(2)}_{Ajk}$
z_j
s_{1t}
s_{2t}
s_{pt}
state
a_t
action
Policy learning
The Actor Network

How does the brain solve this kind of task ?

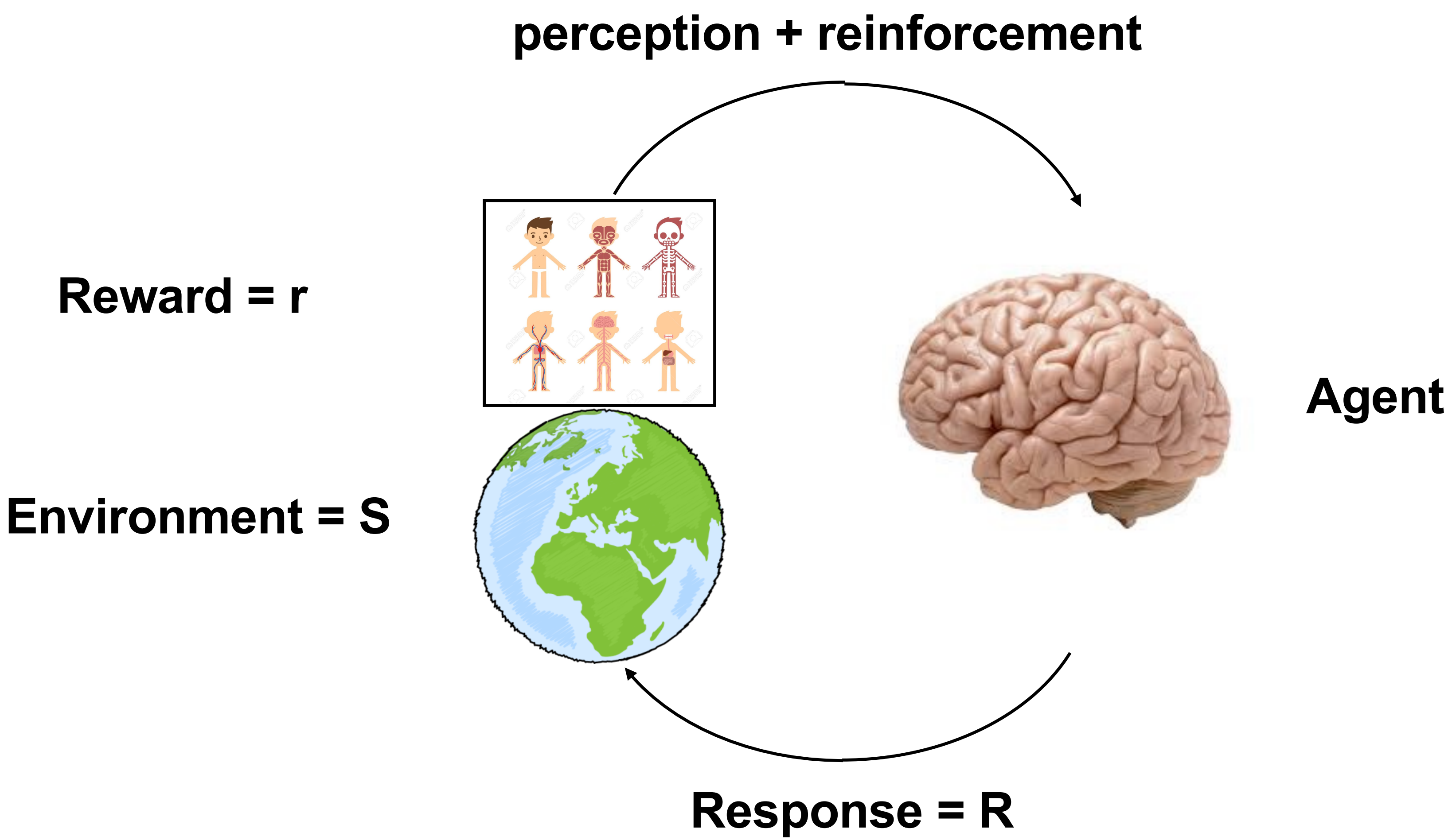



Anterior cortex
Motor areas
Localization and
Manipulation
Posterior cortex
Dorsal
pathway
Position
Orientation
Ventr
pathw
Insula
Needs
Pleasure +
Pain -
Body
Retina
ng

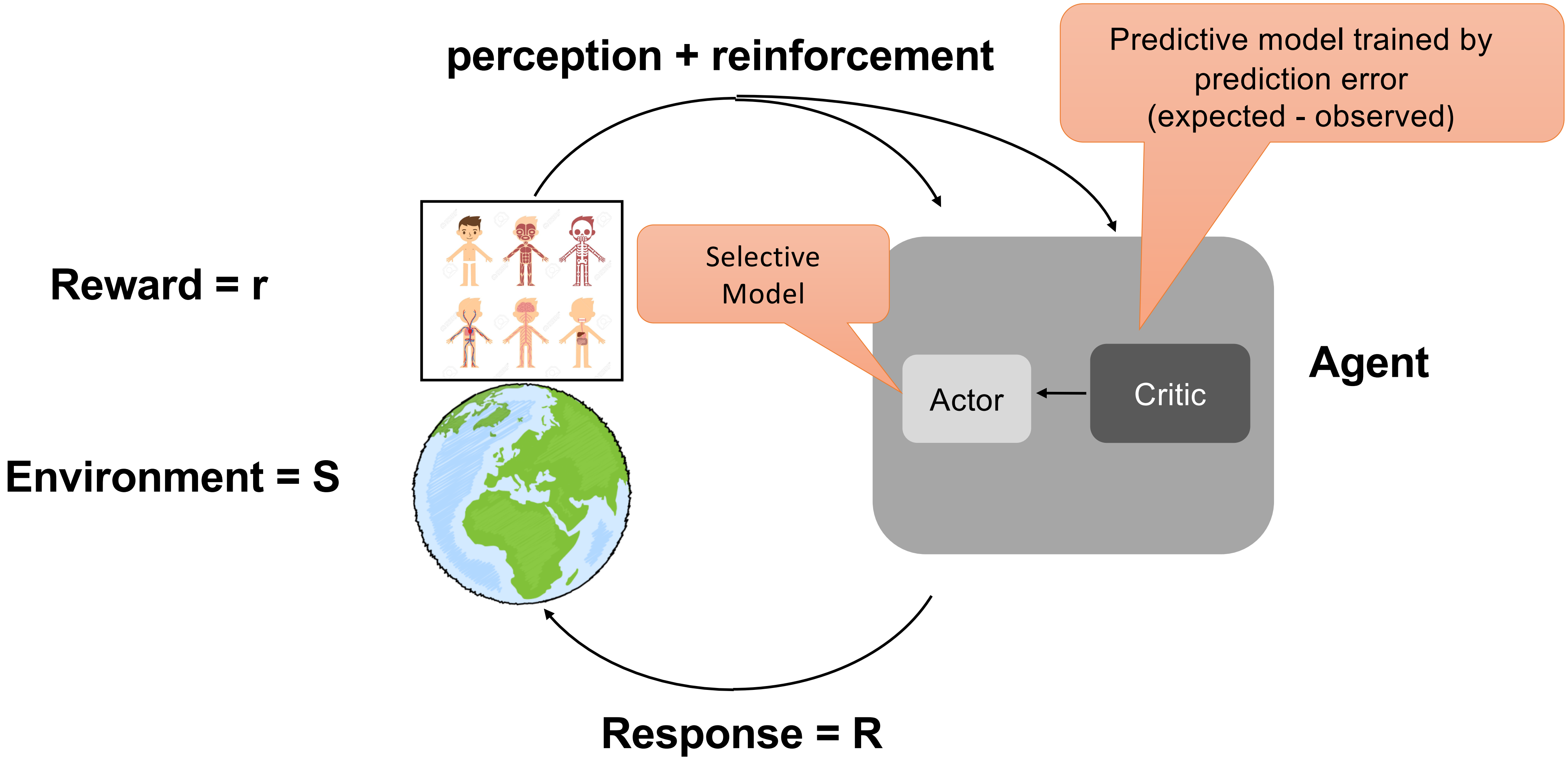

Motor areas
Response
Situation
(Stimuli)
Ventral
pathway
Shape
Color
Dorsal
pathway
ntation
Selective
Model
S-R
Retina


Orbitofrontal cortex
Goal encoding
Predictive Model S-O
Situation (Stimuli)
Insula
Result (Outcome)
Needs
Pleasure +
Pain -
Ventral pathway
Shape
Color
Body
Retina

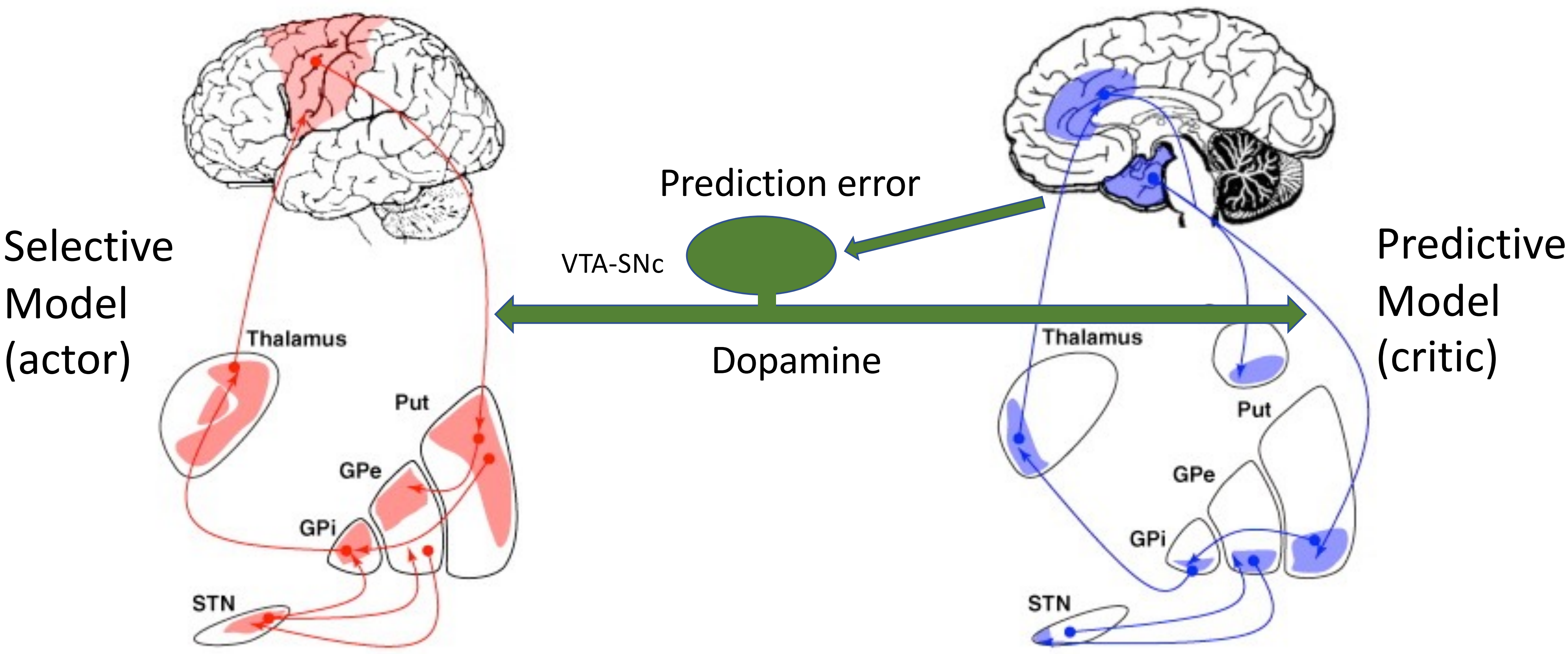
Slow <u>implicit</u> learning similar to Reinforcement Learning (RL) *(Sutton and Barto, 1998)*



Slow **implicit** learning similar to Reinforcement Learning (RL) *(Sutton and Barto, 1998)*



Role of the loops between **Basal Ganglia** and **Anterior Cortex** to implement Predictive et Selective Models, trained by an error signal carried by **Dopamine**



The three curses of Artificial Intelligence

- **Symbol grounding :**
how to access the meaning of the task

- **Combinatorics**:
in high dimensional spaces, learning is very slow

- **Uncertainties**:
need of flexibility in the changing world

What Artificial Intelligence proposes:

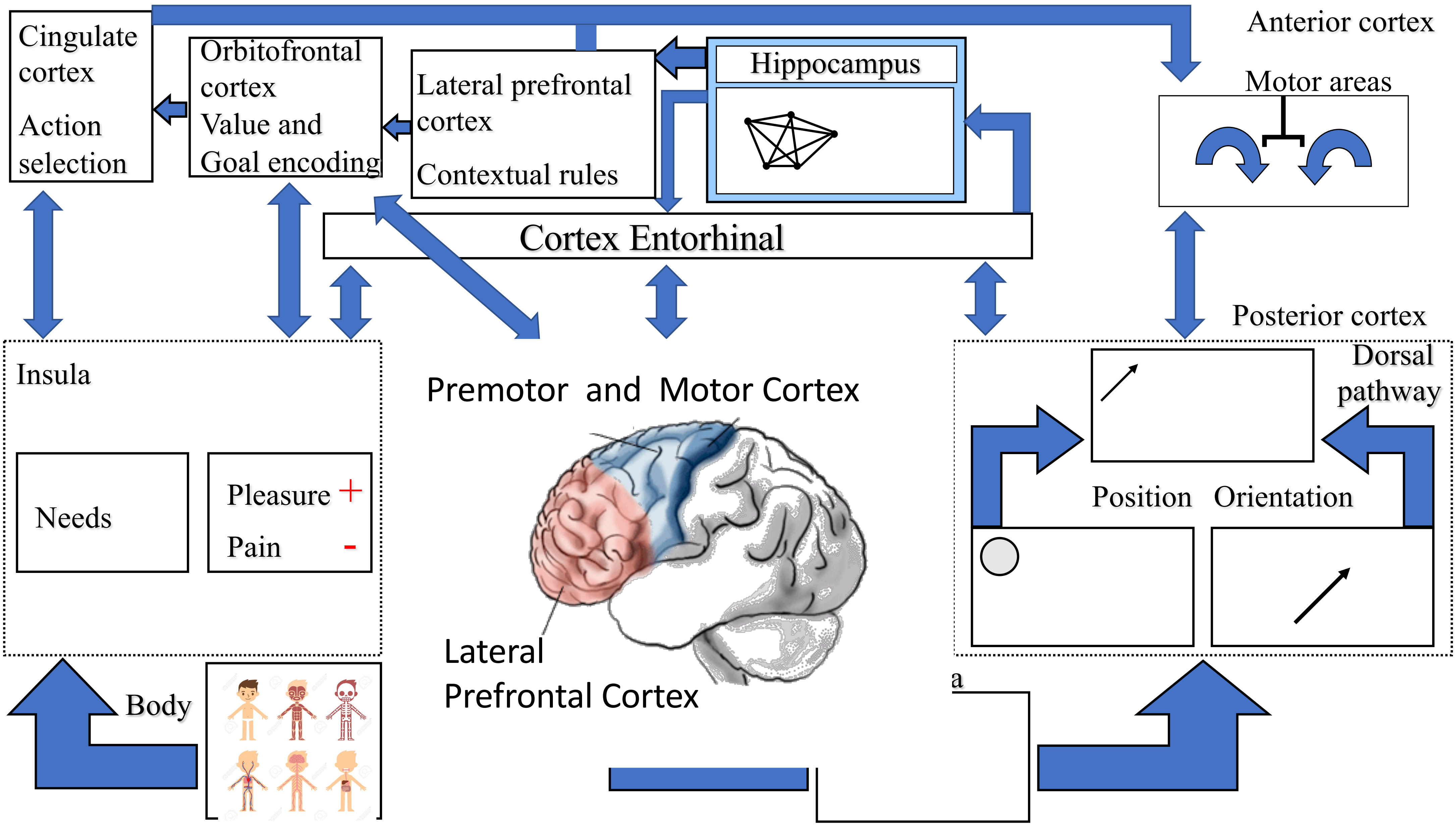
- **Symbol grounding** : develop ontologies
- **Combinatorics** : use computing power
- **Uncertainties** : use probabilistic models

What the brain proposes:

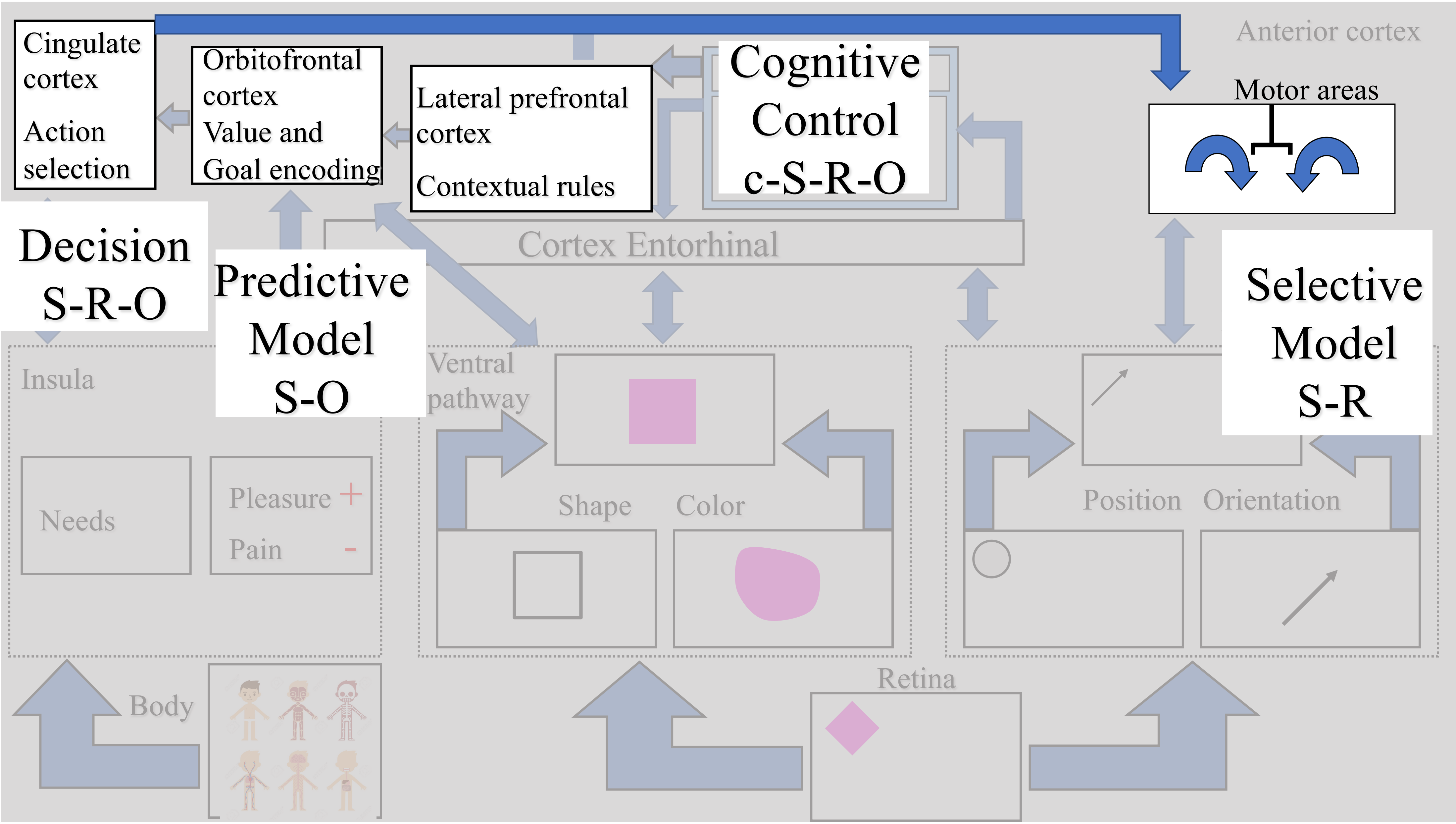
- **Symbol grounding** : feel with our body (role of the insula)

- **Combinatorics** : better use our explicit memory (role of the hippocampus)

- **Uncertainties** : make hypotheses (role of the prefrontal cortex)



Cingulate cortex
Action selection
Orbitofrontal cortex
Value and Goal encoding
Lateral prefrontal cortex
Contextual rules
Hippocampus
Anterior cortex
Motor areas
Cortex Entorhinal
Insula
Needs
Pleasure +
Pain -
Premotor and Motor Cortex
Lateral Prefrontal Cortex
Posterior cortex
Dorsal pathway
Position
Orientation
Body
a



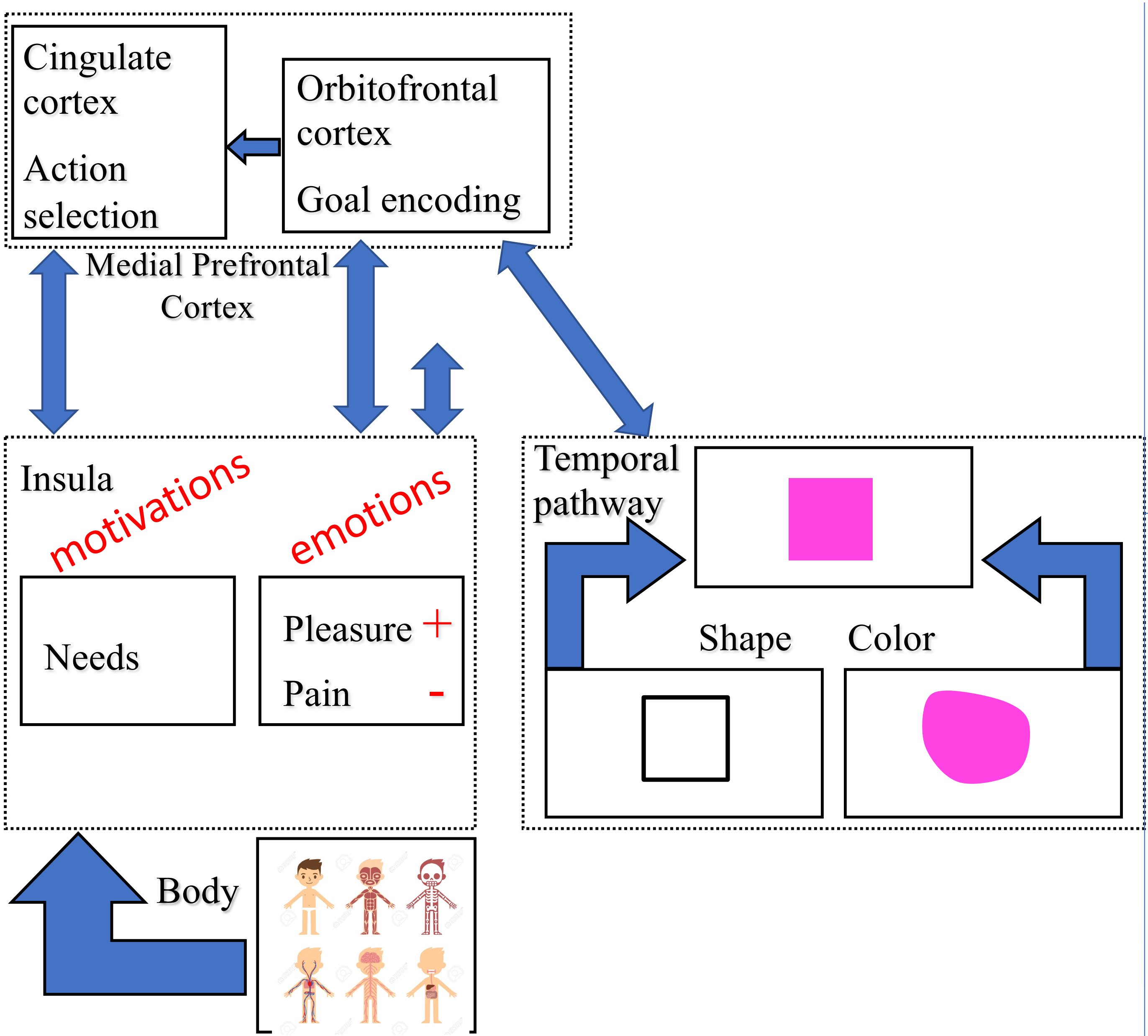
Anterior cortex
Cingulate cortex
Action selection
Orbitofrontal cortex
Value and Goal encoding
Lateral prefrontal cortex
Contextual rules
Cognitive Control
c-S-R-O
Motor areas
Decision
S-R-O
Predictive Model
S-O
Cortex Entorhinal
Selective Model
S-R
Insula
Needs
Pleasure +
Pain -
Ventral pathway
Shape
Color
Position
Orientation
Body
Retina

What the brain proposes:

- Symbol grounding : feel with our body (role of the insula)

- Combinatorics : better use our explicit memory (role of the hippocampus)

- Uncertainties : make hypotheses (role of the prefrontal cortex)



Role of the body:

- Propose richer reinforcement signals
- Anticipate pain and pleasure: emotions
- Compare uncomparable aspects: somatic markers

- Generate one's own external goals: extrinsic motivations
- Intrinsic motivations : novelty, curiosity, game
- Do not only obey environment and have behaviors directed by your goals

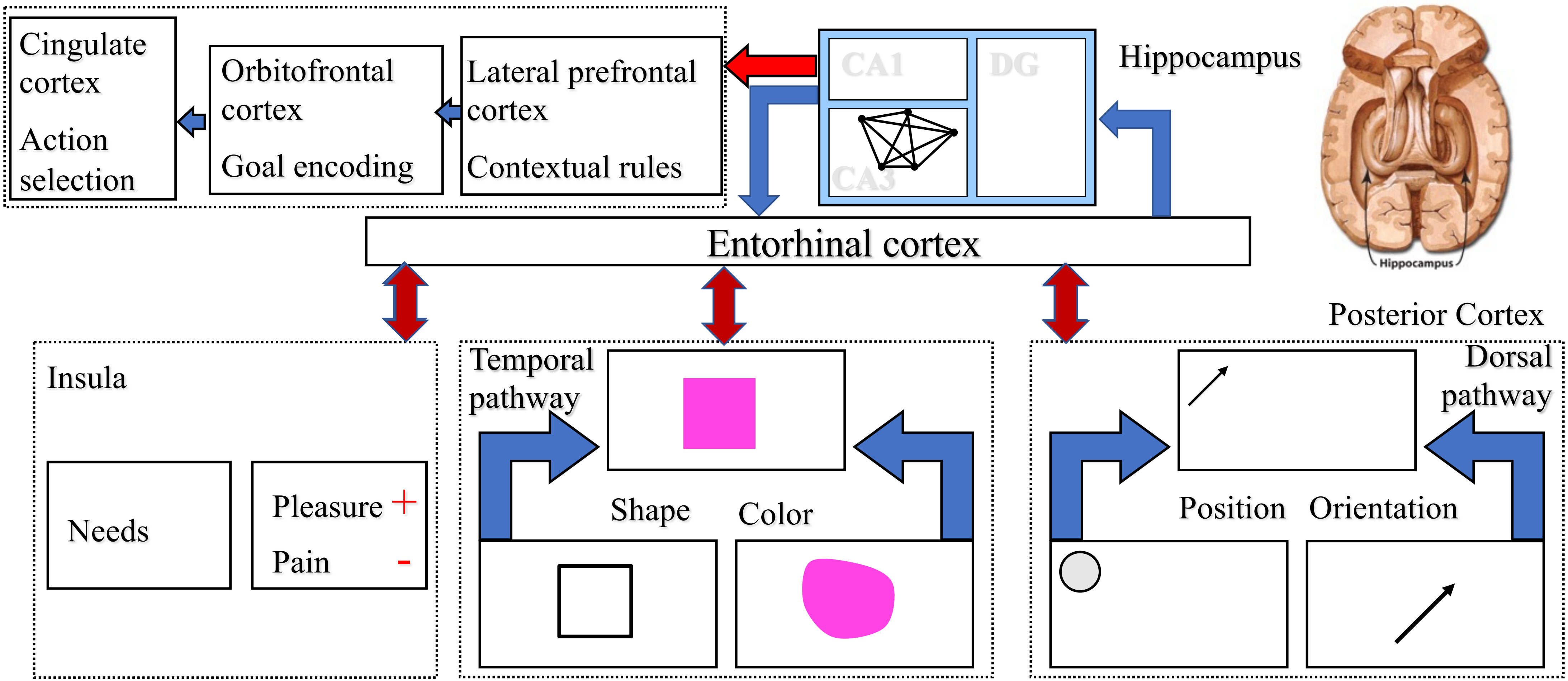
What the brain proposes:

- Symbol grounding : feel with our body (role of the insula)

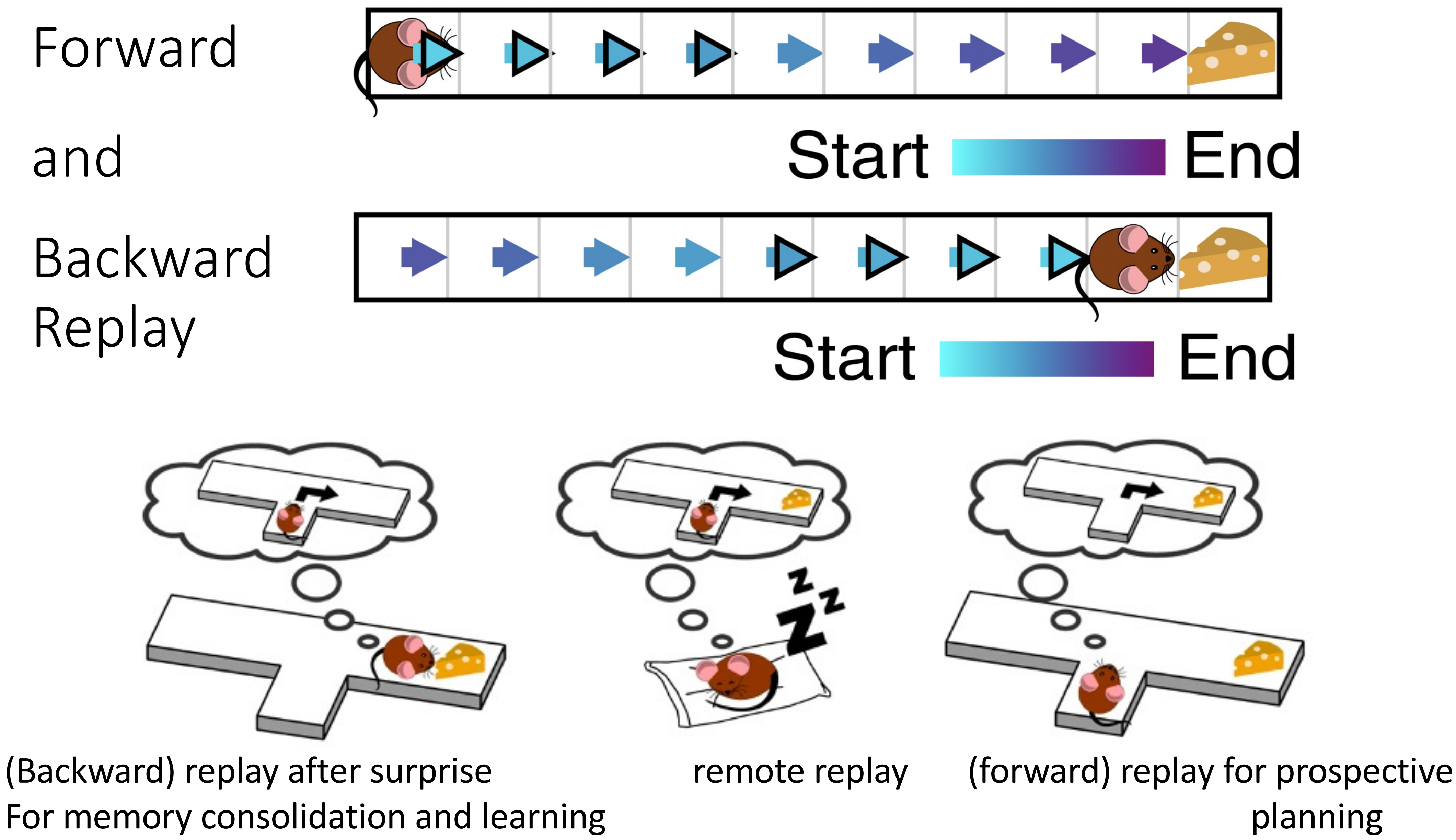
- Combinatorics : better use our explicit memory (role of the hippocampus)

- Uncertainties : make hypotheses (role of the prefrontal cortex)

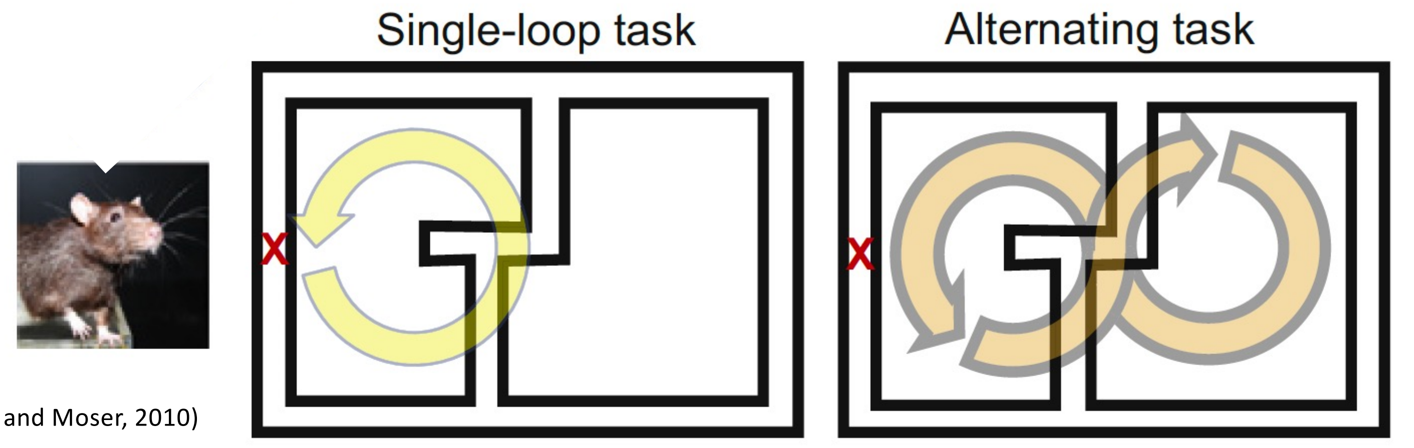
Combinatorics: Hippocampus-Cortex: Complementary Learning System



The Hippocampus can learn quickly and store episodes and, by replay, accelerate learning of concepts and planning in the Cortex

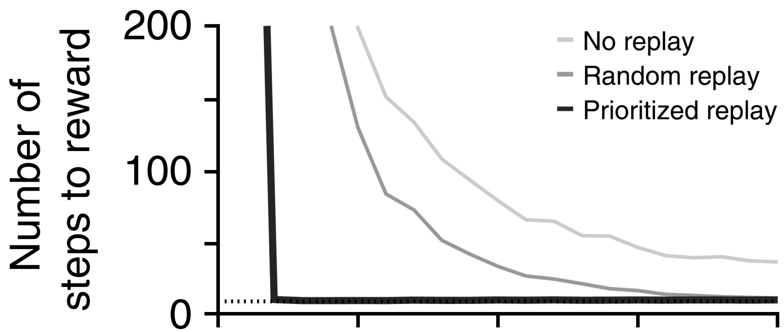

Forward
and
Backward
Replay
Start
End
Start
End
(Backward) replay after surprise
For memory consolidation and learning
remote replay
(forward) replay for prospective
planning

Trajectories Replay by the hippocampus at rest

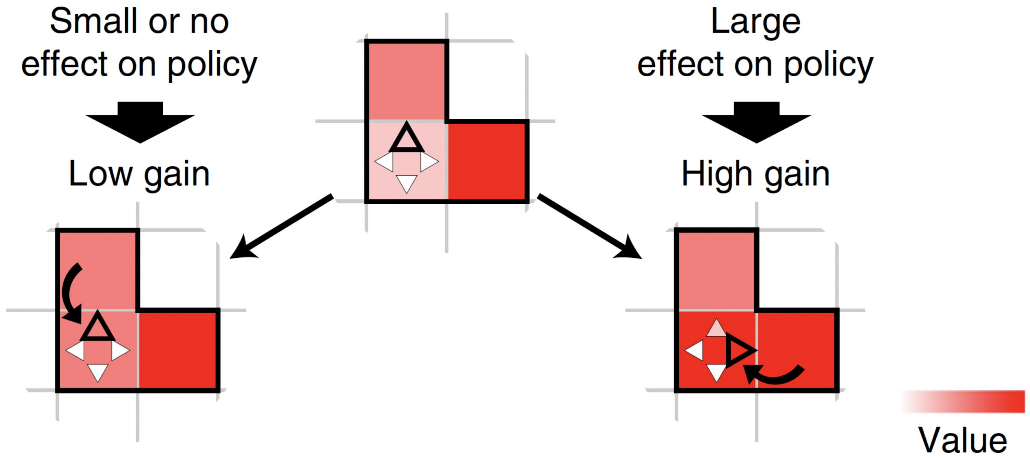


(from Derdikman and Moser, 2010)

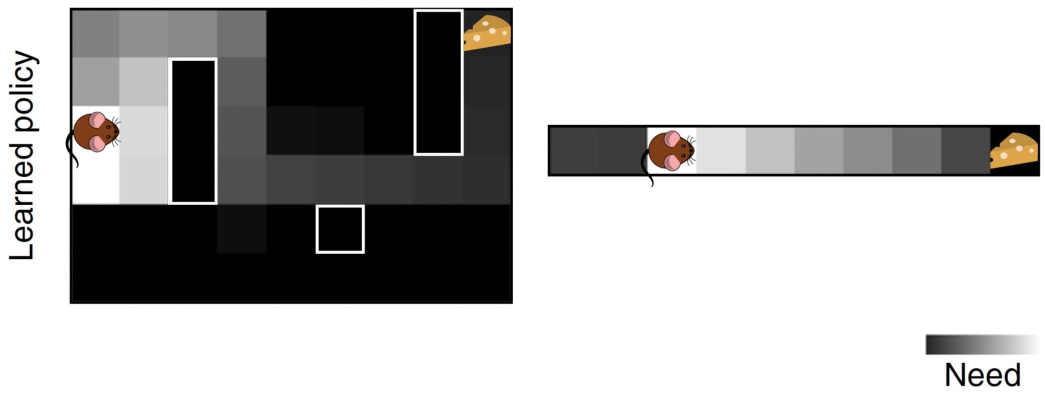
- What is the best strategy of replay ?
  Which episodes to replay?



- The most informative (surprising) episodes (backward replay for memory consolidation, more frequent in goal locations or during sleep)



- The most adapted to the situation (forward replay for prospective planning, More frequent at initial positions)



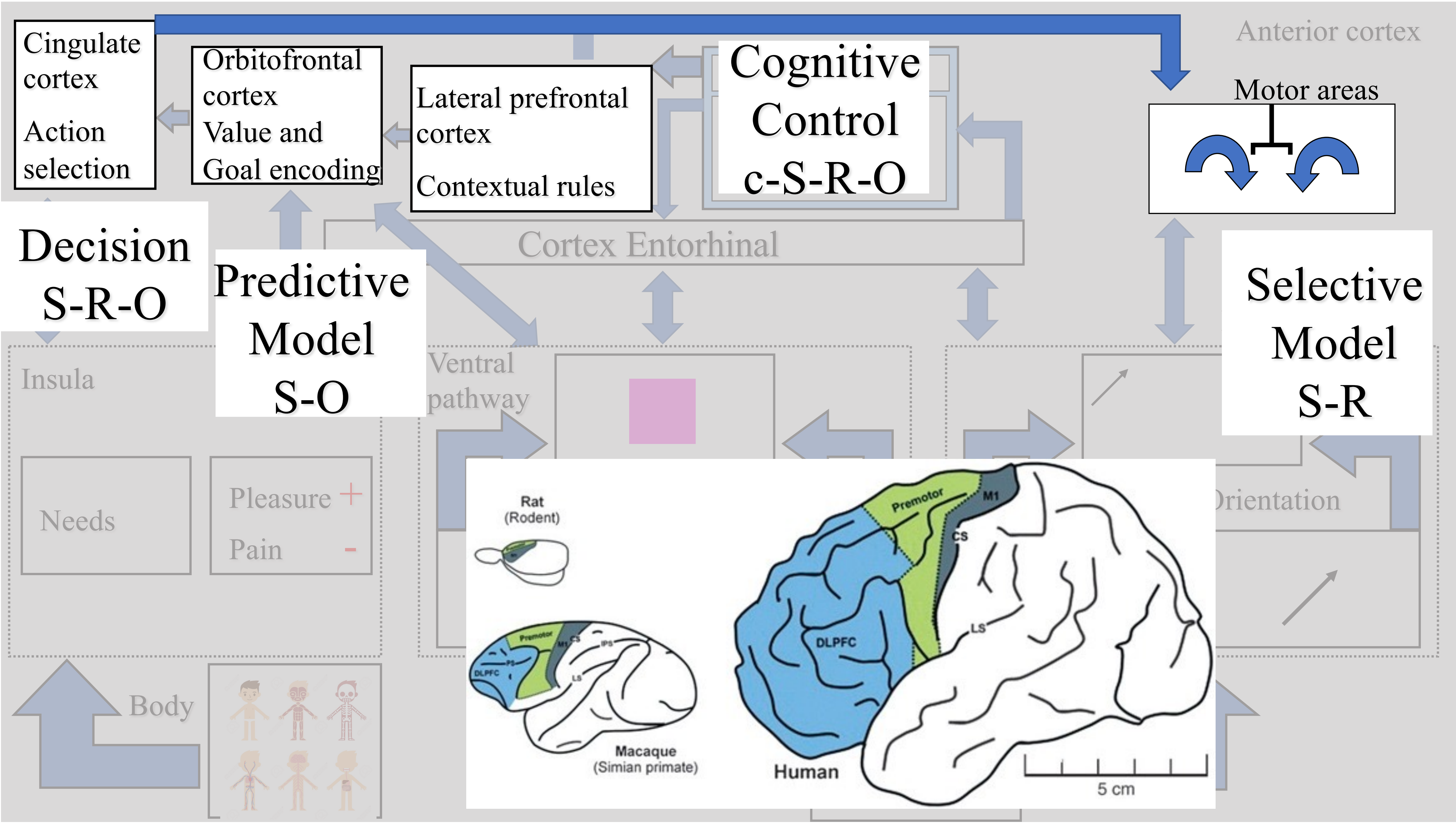
Episodic Reinforcement Learning
Curriculum Learning

What the brain proposes:

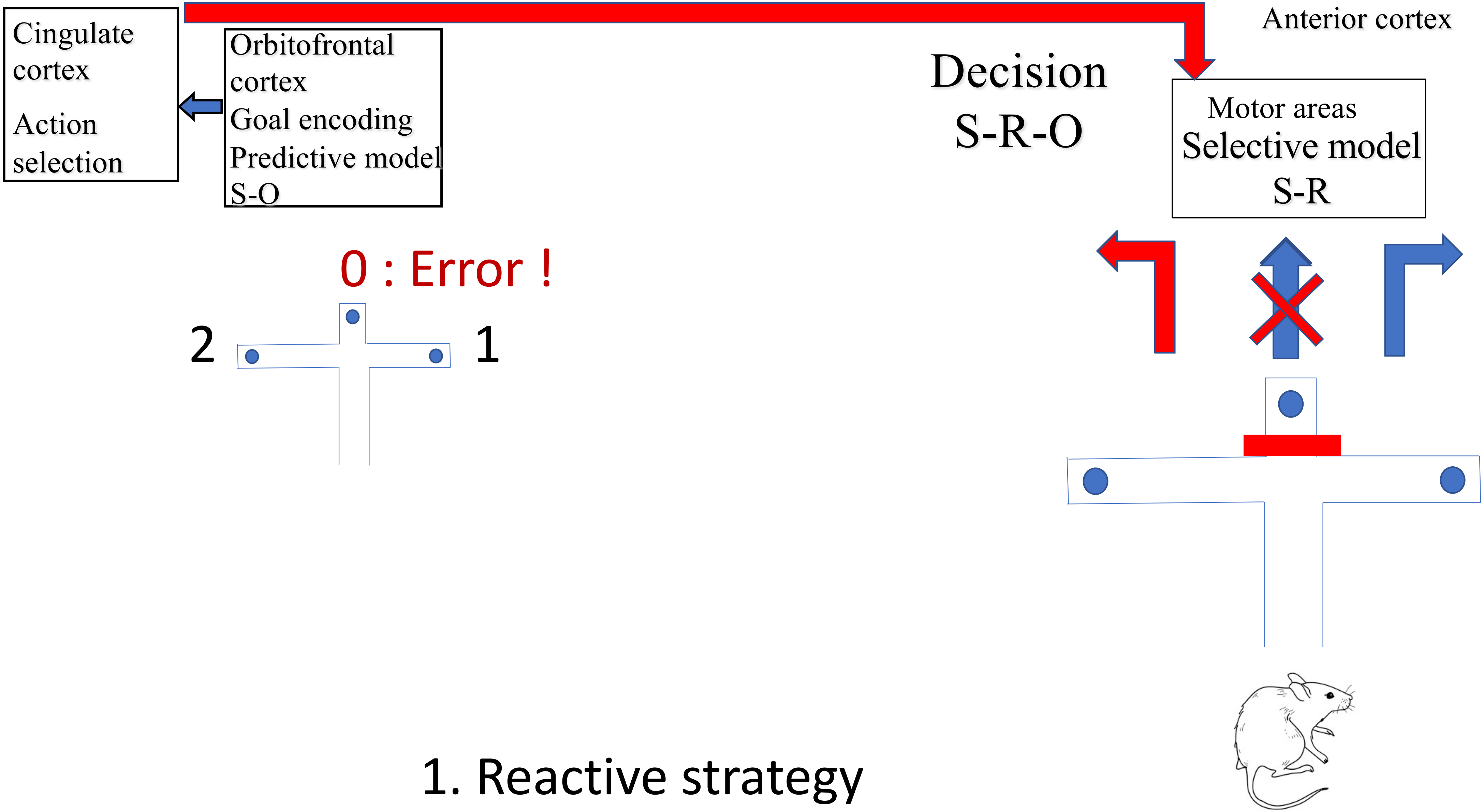
- Symbol grounding : feel with our body (role of the insula)

- Combinatorics : better use our explicit memory (role of the hippocampus)

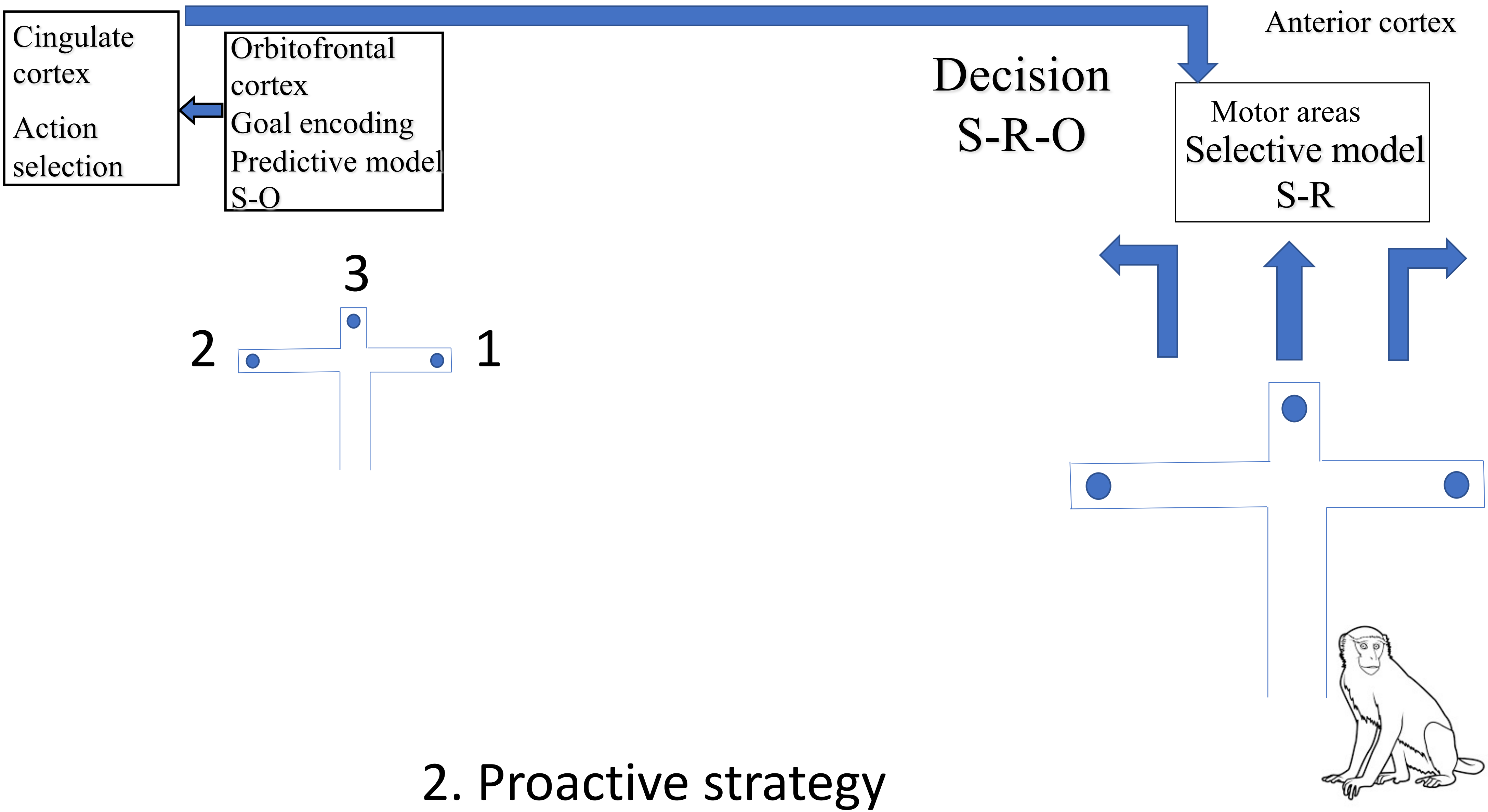
- Uncertainties : make hypotheses (role of the prefrontal cortex)



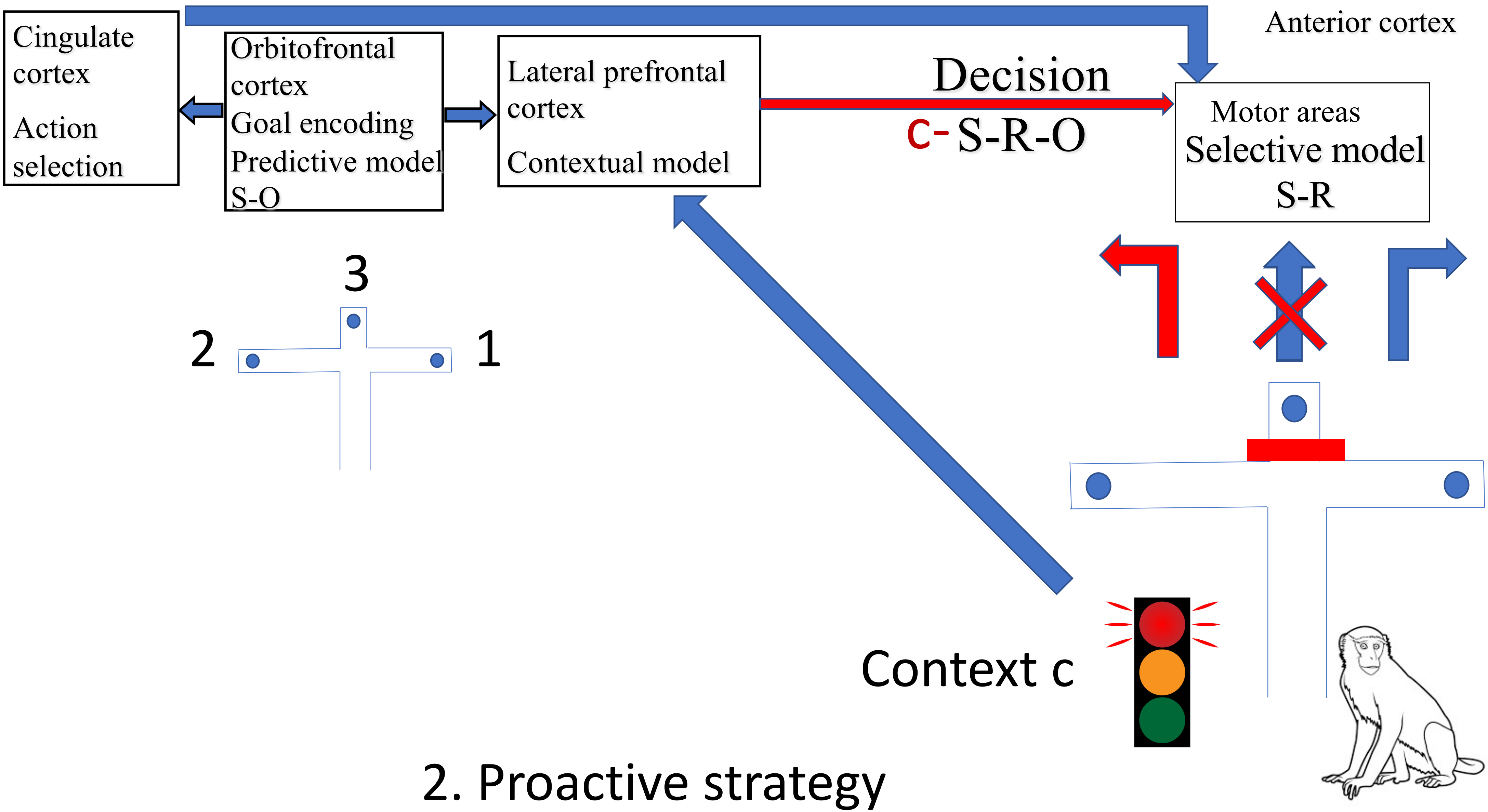
Cingulate cortex
Action selection
Orbitofrontal cortex
Value and Goal encoding
Lateral prefrontal cortex
Contextual rules
Cognitive Control
c-S-R-O
Anterior cortex
Motor areas
Decision
S-R-O
Predictive Model
S-O
Cortex Entorhinal
Selective Model
S-R
Insula
Ventral pathway
Needs
Pleasure +
Pain -
Body
Orientation
Rat
(Rodent)
Macaque
(Simian primate)
Human
Premotor
M1
CS
LS
IPS
PS
DLPFC
5 cm



1. Reactive strategy



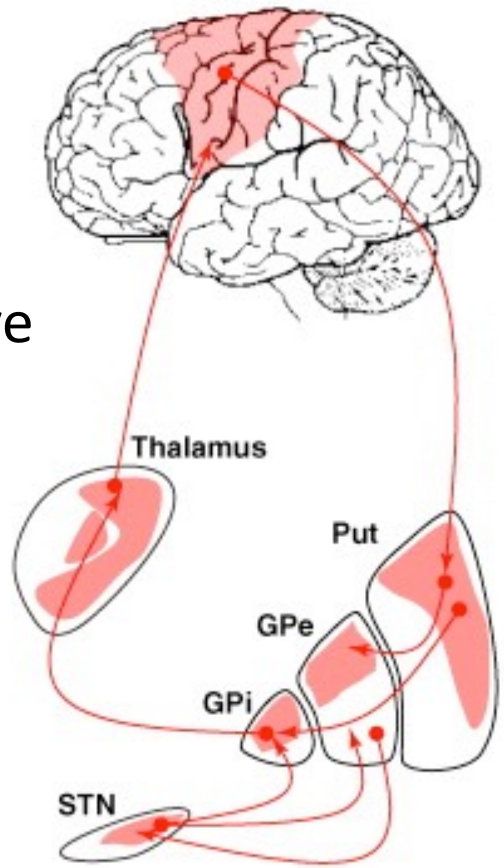
2. Proactive strategy

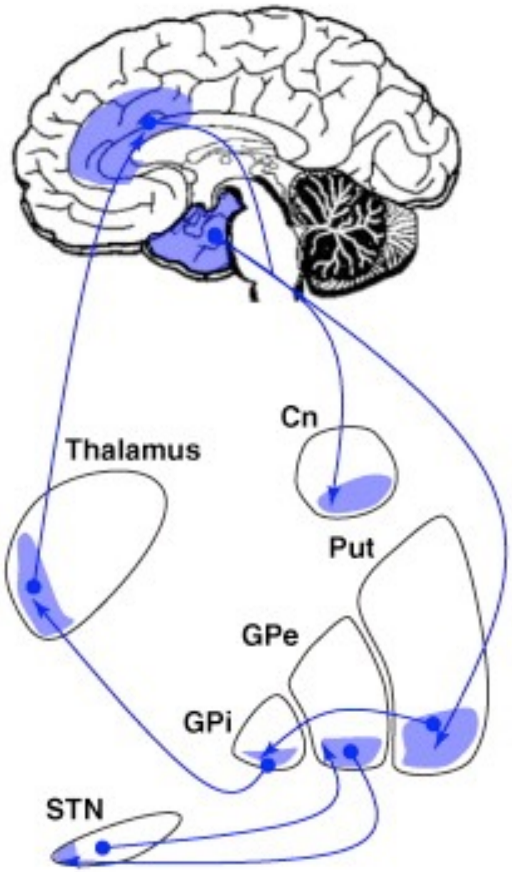


2. Proactive strategy

Double role of DA for learning and Task maintenance/switching

Selective model (actor)





Predictive model (critic)

Contextual model (rules)

Uncertainties: Role of the prefrontal cortex

Based on contextual information and history of performance (confidence level, costs):

- The Cingulate cortex can inhibit the usual behavior in the motor cortex
- The Lateral PFC can <u>explicitly</u> trigger a new contextual behavior by selective attention

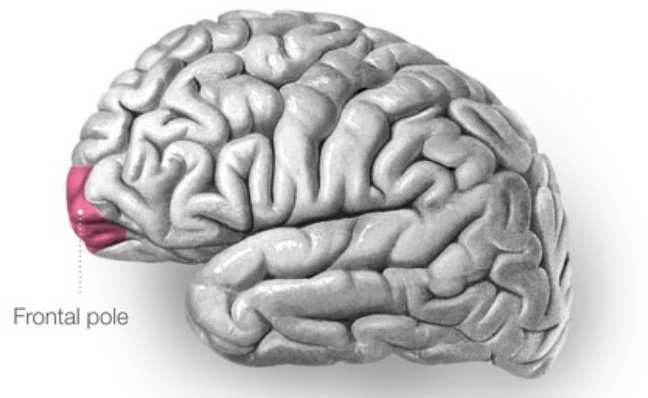
➡ Meta Reinforcement Learning

Based on information replayed by the hippocampus:

- The same system can <u>explicitly</u> imagine a prospective scenario

Using their frontopolar cortex:

- Humans can compare several scenarios



Discussion

Important characteristics:

- Embodied cognition
- Major role of episodic memory and replay for consolidation
- Association of implicit and explicit memories (slow and fast thinking)

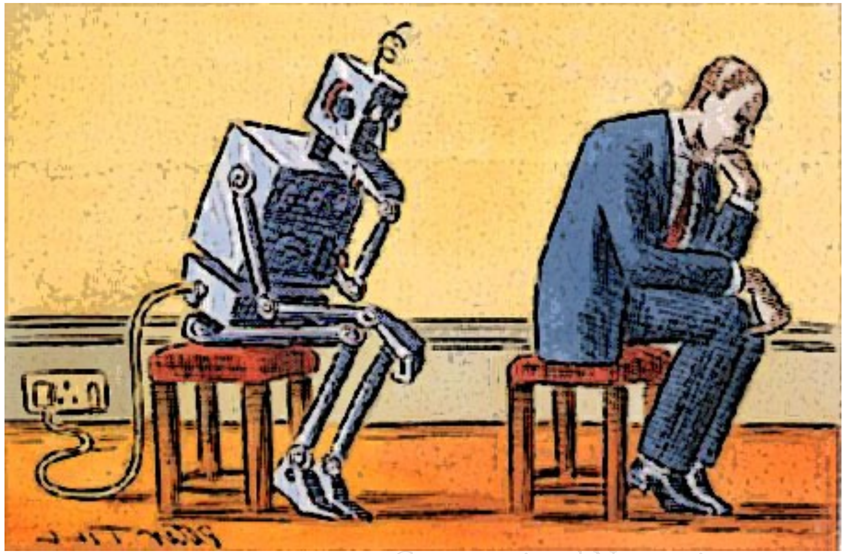
Link to social sciences and humanities:

- Can explain cognitive biases (e.g., risk aversion, bias of reference, bias of confirmation): we solve another problem

Ongoing work and new tasks to consider:

- Problem solving, deliberation, creativity, judgement, human rationality
- Collective decision and negotiation

Is the brain a good model for artificial intelligence?



Frederic.Alexandre@inria.fr

Mnemosyne