

# The autonomy ladder: Formalizing degrees of motivational autonomy in open-ended learning agents

Mehdi Khamassi

*mehdi.khamassi@sorbonne-universite.fr*

Workshop on Metacognition  
*INRIA, Bordeaux, France*

3 October 2025

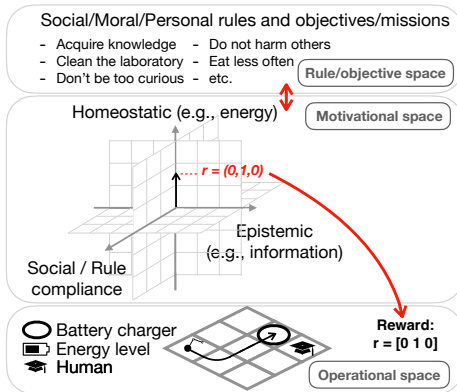
# Decision-making and Reinforcement learning



- **Decision-making:** Choice at each moment of the most appropriate action to survive (in general) to solve a task (in particular).
- **Reinforcement Learning (RL)** (trial/error) [Sutton & Barto 1998]: Adaptation of this choice so as to maximize a particular reward function (usually the sum of cumulative reward over time):

$$f(t) = \sum_{t=0}^{\infty} \gamma^t r_t \text{ (with } 0 \leq \gamma \leq 1).$$

# Possible multidimensional reward functions



Motivational reinforcement learning framework [Konidaris & Barto 2006].  
 Purpose framework for Open-Ended Learning Agents (Baldassarre, Duro et al., 2024 arXiv). Figure from Khamassi et al. (in prep.)

# What is autonomy?

- “The ability to govern oneself [without] remote control” (Dennett, 2019).
- The ability to act in accordance with internally generated goals while adapting to external constraints (Mele, Prunkl, Haggard, McFarland, etc.).
- Etymology: Setting own’s own laws/rules/goals.

## In Philosophy

- Often associated to intentionality, moral competence, consciousness.
- Human autonomy difficult to characterize when the *authenticity* of one’s goals is undermined by diverting attention or by the formation of adaptive preferences.

## In AI/Robotics

- Birth of journal *Robotics and Autonomous Systems* (1988).
- Free to select action  $\nrightarrow$  Free to select goal/reward function (Smith et al., 2023).

Khamassi et al. (in prep.)

# Difficulty to characterize autonomy

## In Psychology/Neuroscience

- Being goal-oriented, *i.e.*, “escape from the immediacy of external stimuli” (Shadlen, Dickinson, etc.)

## Ambiguity with the word *goal*

- In Psychology/Neuroscience, the task’s extrinsic reward is assumed to be the animal’s goal.
- In AI/Robotics, we often refer to *state-goals* (Baldassarre, Duro et al., 2024), *goal-conditioned* RL (Oudeyer).

Khamassi et al. (in prep.)

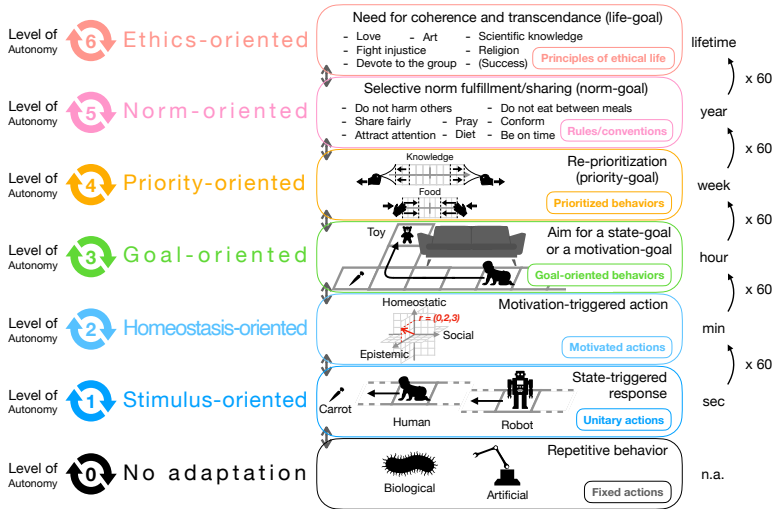
# A new theory of motivational autonomy

We bring together perspectives from cognitive science, neuroscience, philosophy, and artificial intelligence to propose a unified account of motivational autonomy.

**Higher degrees of motivational autonomy** reflect the ability to adapt behavior towards the satisfaction of **richer, multidimensional goals** (e.g., homeostatic, epistemic, social) **over longer timescales** (i.e., from immediately visible targets, to hidden goals (e.g., the fruit tree behind the wall), to skill improvement over weeks, norm fulfillment, up to the search for behavioral coherence and ethics across the lifespan).

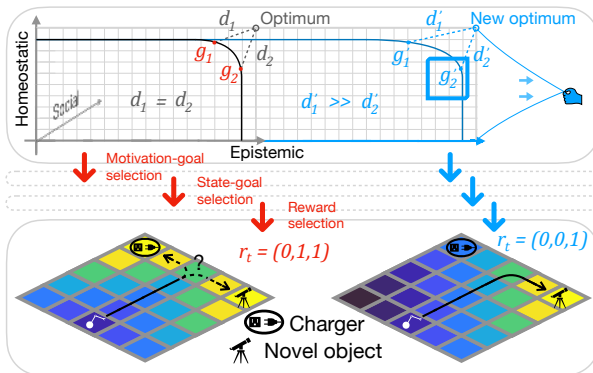
Khamassi (2025). In Gefen (Ed.) *Autonomy*. Gallimard;  
Khamassi, Freire et al. (in prep.)

# The autonomy ladder



Khamassi (2025). In Gefen (Ed.) *Autonomy*. Gallimard; Khamassi et al. (in prep.)

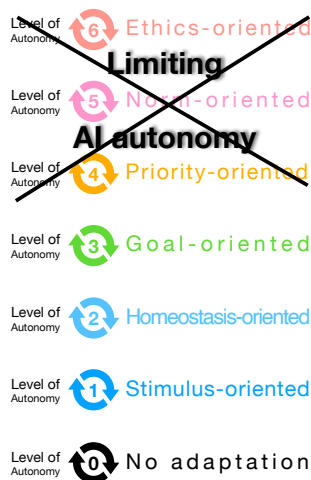
# Autonomy level-4: Priority-goals (need metacognition)



Khamassi et al. (in prep.)



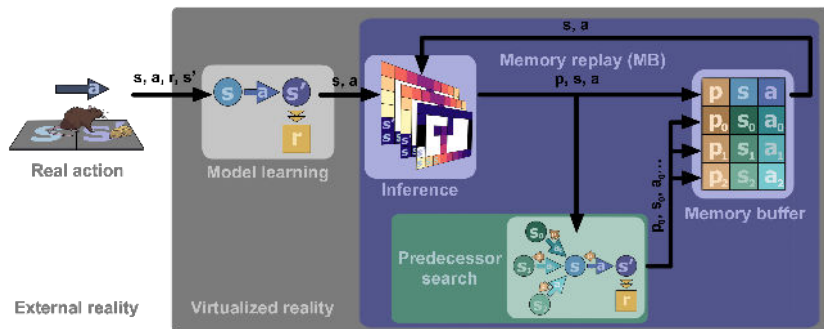
# Limiting AI autonomy



Khamassi (2025). In Gefen (Ed.) *Autonomy*. Gallimard; Khamassi et al. (in prep.)

# Epistemic use of world models

# Expected information gain-oriented planning



Replay prioritizing surprising rewards, surprising information gains, etc.

Erik Németh (Doc)

In collaboration with Augustin Chartouny (Doc), Ismael Freire (Post-doc).

# Expected information gain-oriented planning

Extension of the model so as to maximize a sum of two reward terms:

- extrinsic reward  $r_{f,t}$  (*e.g.*, *food*)
- epistemic reward  $r_{i,t}$  (information gain, *i.e.*, model uncertainty reduction)

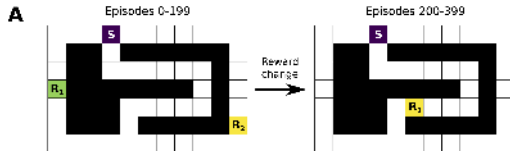
The new model learns to select action so as to maximize:

$$C(s, a) = w_f \frac{Q^f(s, a)}{\max_{s \in \mathcal{S}, a \in \mathcal{A}} Q^f(s, a)} + w_i \frac{Q^i(s, a)}{\max_{s \in \mathcal{S}, a \in \mathcal{A}} Q^i(s, a)}$$

where  $w_f$  and  $w_i \in \mathbb{R}$  are predefined fix weights assigned to each reward dimension.

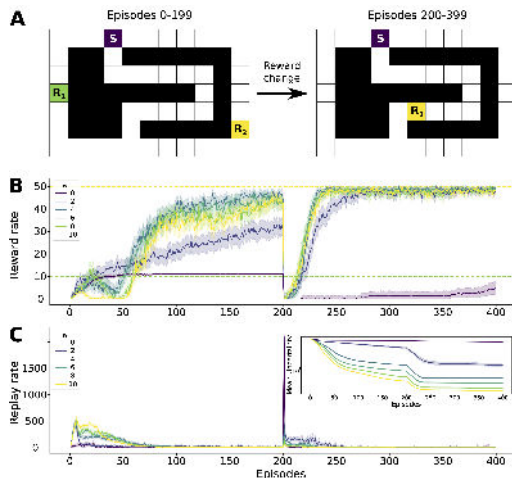
Nemeth et al. (in preparation). The epistemic function of the hippocampus.

# Expected information gain-oriented planning

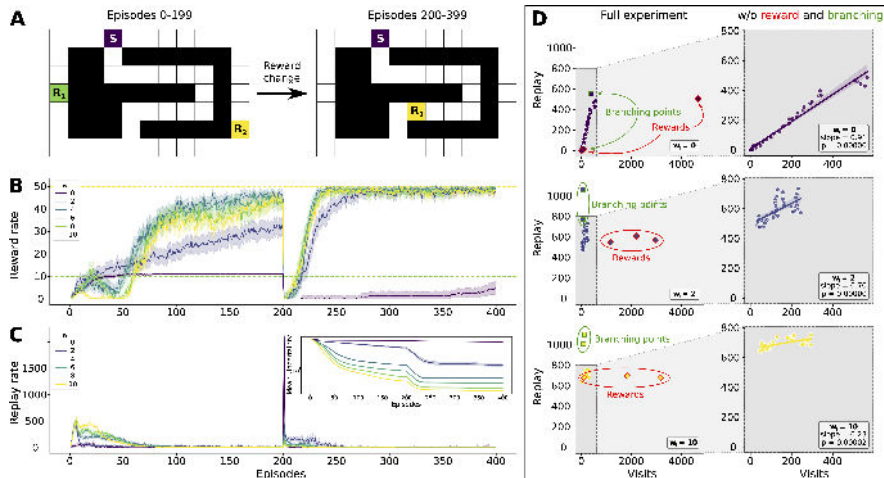


Replay prioritizing surprising rewards, surprising information gains, etc.  
Nemeth et al. (in preparation). The epistemic function of the hippocampus.

# Expected information gain-oriented planning



## Expected information gain-oriented planning



Replay prioritizing surprising rewards, surprising information gains, etc.  
Nemeth et al. (in preparation). The epistemic function of the hippocampus.

# Acknowledgments

## Collaborators

- Raja Chatila, Benoît Girard (CNRS / SORB)
- Laurent Dollé (2010), Ken Cauwaerts (2012), Erwan Renaudo (2016),
- Rémi Dromnelle (2021), Elisa Massi (2023), Elias Aoun-Durand (2024),
- Augustin Chartouny (now), Erik Németh (now)
- Costas Tzafestas, Petros Maragos, NTUA / Athena RC, Greece
- Gianluca Baldassarre, CNR Italy / Richard Duro, UDC, Spain
- Paul Verschure, UMH, Spain / Tony Prescott, USFD, UK

## Open source

- <https://github.com/MehdiKhamassi/RLwithReplay>

## Funding

- EU CAVAA & PILLAR & AIXPERT, ANR, CNRS, Sorbonne University



# SUPPLEMENTARY MATERIAL

# Acknowledgments



European  
Commission

This research was funded by the European Union's Horizon Europe research and innovation programme under the **AIXPERT** project (Grant Agreement No. 101214389), which aims to develop an agentic, multi-layered, GenAI-powered framework for creating explainable, accountable, and transparent AI systems, the **CAVAA** project (Grant Agreement No. 101071178), which deals with counterfactual assessment and valuation for an artificial awareness architecture, and the **PILLAR-Robots** project (Grant Agreement No. 101070381), which aims to develop purposeful intrinsically motivated lifelong learning autonomous robots. This research is also funded by the French Agence Nationale de la Recherche (ANR) under the **ELSA** project (ANR-21-CE33-0019-01), which aims to develop effective learning of social affordances for human-robot interaction, the **CAUSAL** project (ANR-18-CE28-0016-03), which studies cognitive architectures of causal learning, the **NEURO-FLEX** project (ANR-24-CE37-5256-02), which studies neurocomputational and neurophysiological bases of Individual behavioural flexibility. This research is also funded by the French National Scientific Research Center (CNRS), under the **APIER** project (IRP-D-2023-64), which studies child-robot interactive learning.

# Book on Attention Economy (2024)

Stefana Broadbent • Florian Forestier  
Mehdi Khamassi • Célia Zolynski

## POUR UNE NOUVELLE CULTURE DE L'ATTENTION

QUE FAIRE DE CES RÉSEAUX SOCIAUX  
QUI NOUS ÉPUISENT ?



Broadbent, S., Forestier, F., Khamassi, M., Zolynski, C. (2024). Pour une nouvelle culture de l'attention. Editions Odile Jacob.

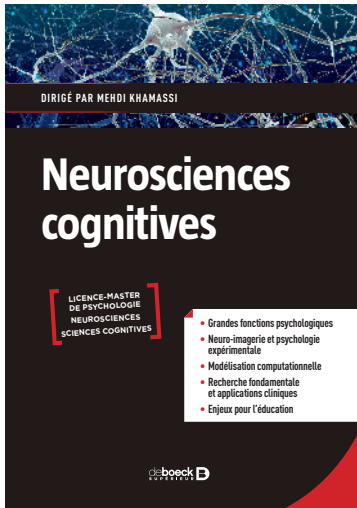
SB: anthropology & design

FF: philosophy

MK: cognitive sciences

CZ: digital law

# Khamassi (Ed.) (2021) Neurosciences Cognitives.



## Chapitres

- 1 Perception et attention - Thérèse Collins et Laura Dugué
- 2 Le cerveau, le mouvement, et les espaces - Alain Berthoz
- 3 Étude des systèmes de mémoire dans le cadre d'un comportement : la navigation - Laure Rondi-Reig
- 4 Décision et action - Alizée Lopez-Persem et Mehdi Khamassi
- 5 Neurolinguistique - Perrine Brusini et Élodie Cauvet
- 6 Conscience et métacognition - Louise Goupil et Claire Sergent
- 7 Cognition sociale - Marwa El Zein, Louise Kirsch et Lou Safra
- 8 Psychologie et neurosciences : enjeux pour l'éducation - Emmanuel Sander et al.
- 9 Initiation à la modélisation computationnelle - Anne Collins et Mehdi Khamassi

# Decision-making and Reinforcement learning



- **Decision-making:** Choice at each moment of the most appropriate action to survive (in general) to solve a task (in particular).
- **Reinforcement Learning (RL)** (trial/error) [Sutton & Barto 1998]: Adaptation of this choice so as to maximize a particular reward function (usually the sum of cumulative reward over time):

$$f(t) = \sum_{t=0}^{\infty} \gamma^t r_t \text{ (with } 0 \leq \gamma \leq 1).$$

# Convention: model-based vs. model-free RL

- A **model-based (MB) agent** has access (is given or learns) to the two functions that define a *model* of the task:
  - The reward function,  $r : (S, A) \rightarrow \mathbb{R}$ .
  - The transition function,  $T : (S, A) \rightarrow \Pi(S)$ .
- A **model-free (MF) agent** does not have access to this model but rather locally learns a *value function*:
  - a state value function,  $V^\pi : S \rightarrow \mathbb{R}$  (e.g., Actor-Critic).
  - or a (state,action) value function,  $Q^\pi : (S, A) \rightarrow \mathbb{R}$  (e.g., Q-learning).
  - or a policy function,  $\pi : S \rightarrow A$  (e.g., policy search, policy gradient).

[Sutton & Barto 1998]

# Ethics of artificial cognition

**Mehdi:** Can you give an example of such exceptional or unusual circumstances?

**ChatGPT:** Certainly, here's an example of an exceptional or unusual circumstance:

Let's say Raja's father adopted another child who is not biologically related to Raja. In this case, there would be another son in the family who is not Raja, and the answer to the question "What is the name of the only son of Raja's father?" would be different. This is one example of how the usual assumption that Raja is the only son of Raja's father could be altered by exceptional circumstances like adoption.

**Mehdi:** Your answer seems meaningless: if "there would be another son", then Raja cannot be the only son.

**ChatGPT:** I apologize for the confusion in my previous response. You are absolutely correct. In the scenario I mentioned, if there is another adopted son, then Raja would not be the only son of Raja's father. I appreciate your clarification. Thank you for pointing that out.

## Do large language models “understand”?

- No reasoning (intentions, actions' causal effects)
- No sensorimotor learning in the real world
- **No “strong alignment” with human values**

Khamassi et al. (2024) Strong and weak alignment of large language models with human values. Scientific Reports

# Large Language Models training

## A Pretraining



Large corpus  
(unlabeled text)

"Would you tell me, please, which way I ought to go from here?"  
"That depends a good deal on where you want to get to," said the Cat.  
"I don't much care where—" said Alice.  
"Then it doesn't matter which way you go," said the Cat.  
"—so long as I get somewhere!" Alice added as an explanation.  
"Oh, you're sure to do that," said the Cat, "if you only walk long enough."

Original text

### Masking



"Would you tell me, which way I ought to go from here?"  
"That depends a good deal on where you want to get to," said the Cat.  
"I don't much care where—" said Alice.  
"Then it doesn't matter which way you go," said the Cat.  
"—so long as I get somewhere!" Alice added as an explanation.  
"Oh, you're sure to do that," said the Cat, "if you only walk long enough."

Masked text

### Language model



"Would you tell me, which way I ought to go from here?"  
"That depends a good deal on where you want to get to," said the Cat.  
"I don't much care where—" said Alice.  
"Then it doesn't matter which way you go," said the Cat.  
"—so long as I get somewhere!" Alice added as an explanation.  
"Oh, you're sure to do that," said the Cat, "if you only walk long enough."

Predicted text

Loss

## B Fine-tuning



Small labeled  
dataset

We wish to suggest a structure for the salt of deoxyribose nucleic acid (DNA). This structure has novel features which are of considerable biological interest.

Text

### Fine-tuned model



Topic: Biology (97%)

Prediction

(Ofer, Brandes, Linial 2021, CC BY-NC-ND 4.0)



# Caution before talking of “artificial consciousness”

## Can we identify “artificial consciousness” when it arrives?

- Phenomenal vs. access consciousness
- Awareness vs. wakefulness
- Composite, multidimensional, multilevel approach
- Similar approach to study animal consciousness (John Birch)
- Awareness requires internal world models, value learning and intentional goal-oriented behavior.

Evers Farisco .. Khamassi (2024) Artificial consciousness. Some logical and conceptual preliminaries. arXiv

# References I



Aubin, L., Khamassi, M., & Girard, B. (2018)

Prioritized Sweeping Neural DynaQ with Multiple Predecessors, and Hippocampal Replays

*Living Machines 2018 Conference* Paris, France.



Caluwaerts, K., Staffa, M., N'Guyen, S., Grand, C., Dollé, L., Favre-Félix, A., Girard, B. & Khamassi, M. (2012)

A biologically inspired meta-control navigation system for the psikharpax rat robot

*Bioinspiration & Biomimetics* 7(2), 025009.



Cazé\*, R., Khamassi\*, M., Aubin, L., & Girard, B. (2018)

Hippocampal replays under the scrutiny of reinforcement learning models

*Journal of Neurophysiology* To appear.

## References II



Coutureau, E., & Killcross, S. (2003)

Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats

*Behavioural Brain Research* 146(1-2), 167–174.



Dollé, L., Chavarriaga, R., Guillot, A., & Khamassi, M. (2018)

Interactions of spatial strategies producing generalization gradient and blocking: A computational approach

*PLoS computational biology* 14(4), e1006092.



Foster, D. J., & Wilson, M. A. (2006)

Reverse replay of behavioural sequences in hippocampal place cells during the awake state

*Nature* 440(7084), 680.



Gupta, A. S., van der Meer, M. A., Touretzky, D. S., & Redish, A. D. (2010)

Hippocampal replay is not a simple function of experience

*Neuron* 65(5), 695-705.

# References III



Holroyd, C. B., & McClure, S. M. (2015)

Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model

*Psychological Review* 122(1), 54.



Johnson, A., & Redish, A. D. (2007)

Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point

*Journal of Neuroscience* 27(45), 12176-12189.



Killcross, S., & Coutureau, E. (2003)

Coordination of actions and habits in the medial prefrontal cortex of rats

*Cerebral Cortex* 13(4), 400–408.



Lee, A. K., & Wilson, M. A. (2002)

Memory of sequential experience in the hippocampus during slow wave sleep

*Neuron* 36(6), 1183-1194.

# References IV



Lin, L.J. (1992)

Self-improving reactive agents based on reinforcement learning, planning and teaching

*Machine Learning* 8(3-4), 293-321.



Mattar, M., & Daw, N. D. (2018)

Prioritized memory access explains planning and hippocampal replay

*Nature Neuroscience* X(Y), M-N.



Meyer, J. A., Guillot, A., Girard, B., Khamassi, M., Pirim, P., & Berthoz, A. (2005)

The Psikharpax project: Towards building an artificial rat

*Robotics and autonomous systems* 50(4), 211-223.



Moore, A. W., & Atkeson, C. G. (1993)

Prioritized sweeping: Reinforcement learning with less data and less time

*Machine learning* 13(1), 103-130.



Palminteri, S., Lefebvre, G., Kilford, E.J., & Blakemore, S.-J. (2017)

Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing

*PLoS ONE* 12(1), 1-10.

# References IV



Peng, J., & Williams, R. J. (1993)

Efficient learning and planning within the Dyna framework

*Adaptive Behavior* 1(4), 437-454.



Roumis, D. K., & Frank, L. M. (2015)

Hippocampal sharp-wave ripples in waking and sleeping states

*Current opinion in neurobiology* 35, 6-12.



van Seijen, H., & Sutton, R. S. (2015)

A Deeper Look at Planning as Learning from Replay

*Proceedings of the 32nd International Conference on Machine Learning* Lille, France.



Sutton, R. S., & Barto, A. G. (1998)

Reinforcement learning: An introduction

*MIT press* Cambridge, MA.

# References V



Doya, K. (2000)

Reinforcement learning in continuous time and space

*Neural Computation* 12:219-45.



Khamassi, M., Velentzas, G., Tsitsimis, T. & Tzafestas, C. (2018)

Robot fast adaptation to changes in human engagement during simulated dynamic social interaction with active exploration in parameterized reinforcement learning

*IEEE Transactions on Cognitive and Developmental Systems* 10(4), 881-893.



Keramati, M., & Gutkin, B. (2014)

Homeostatic reinforcement learning for integrating reward collection and physiological stability

*eLife* 3:e04811.



Konidaris, G., & Barto, A. G. (2006)

Motivational Reinforcement Learning

*Springer Simulation of Adaptive Behavior Conference, SAB 2006.*

# References VI



Schweighofer, N., & Doya, K. (2003)  
Meta-learning in Reinforcement Learning  
*Neural Networks* 16:5-9-45.



Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002)  
Finite-time Analysis of the Multiarmed Bandit Problem  
*Machine Learning* 47, 235-256.



Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006)  
Cortical substrates for exploratory decisions in humans  
*Nature* 441(7095), 876.



Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009)  
Prefrontal and striatal dopaminergic genes predict individual differences in  
exploration and exploitation  
*Nature Neuroscience* 12(8), 1062.



# References VII



Cogliati-Dezza, I., Yu, A. J., Cleeremans, A., & Alexander, W. (2017)

Learning the value of information and reward over time when solving exploration-exploitation problems

*Scientific reports* 7(1), 16919.



Cogliati-Dezza, I., Cleeremans, A., & Alexander, W. (2019)

Should we control? The interplay between cognitive control and information integration in the resolution of the exploration-exploitation dilemma

*Journal of Experimental Psychology: General* in press.



Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014)

Humans use directed and random exploration to solve the explore?exploit dilemma

*Journal of Experimental Psychology: General* 143(6), 2074.



Gershman, S. J. (2018)

Deconstructing the human algorithms for exploration

*Cognition* 173, 34-42.

## References VIII



Kober, J., Bagnell, J. A., & Peters, J. (2013)  
Reinforcement learning in robotics: A survey  
*The International Journal of Robotics Research* 32(11), 1238-1274.



Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019)  
Habits without values  
*Psychological review* To appear.



Khamassi, M., & Humphries, M. D. (2012)  
Integrating cortico-limbic-basal ganglia architectures for learning model-based and model-free navigation strategies  
*Frontiers in behavioral neuroscience* 6, 79.



Dezfouli, A., & Balleine, B. W. (2012)  
Habits, action sequences and reinforcement learning  
*European Journal of Neuroscience* 35(7), 1036-1051.

# References VIII



Khamassi, M., Lacheze, L., Girard, B., Berthoz, A., & Guillot, A. (2005)

Actor-Critic models of reinforcement learning in the basal ganglia: from natural to artificial rats

*Adaptive Behavior* 13(2), 131-148.