



# Reinforcement Learning

Alessandro LAZARIC (*Facebook AI Research / on leave Inria Lille*)

*ENS Cachan - Master 2 MVA*

FAIR / Inria

# From Psychology to Machine Learning

# Decision-making Under Uncertainty

# Decision-making Under Uncertainty



# Decision-making Under Uncertainty



# Decision-making Under Uncertainty



# Decision-making Under Uncertainty



Google   [Advanced Search](#) [Feedback](#)

Web: [Show options...](#) Results 1 - 19 of about 22,100,000 for **bird houses** (0.15)

**Bird Houses** [Sponsored Links](#)  
[www.Scotts.com](#) Scotts attracts colorful birds to your backyard!

**Specialty Bird Houses**  
[www.MidwestBack.com](#) Roosting Boxes, Purple Martins, Bat Chalets.

**Bird Houses at BestPrice**  
[www.BestPrice.com](#) Over 225 different houses in stock. Free shipping!

**Learn More About Bird Houses** [1/10](#)  
Bird House - JustBirdHouses.net learn more about blue bird houses and birdhouses, along with our blue bird house and purple martin bird houses.  
[www.justbirdhouses.net/](#) - [Cached](#) - [Similar](#) - [1/10](#)

**Bird Houses** [1/10](#)  
Bird Houses and Bird Feeders for north american bird species.  
[www.birdhouses101.com/](#) - [Cached](#) - [Similar](#) - [1/10](#)

**Bird Feeders, Bird Houses - The Backyard Bird Company**  
Bird Feeders - The Backyard Bird Company has a variety of bird feeders will attract your landscape and attract wildlife.  
[Bird Houses - BirdFeeders - Decorative Bird Houses](#)  
[www.backyardbird.com/](#) - [Cached](#) - [Similar](#) - [1/10](#)

**Bird Houses** [Sponsored Links](#)  
Find All Types Of Bird Feeders And Houses At Lowest Price Lowest Price.  
[www.Lowes.com](#)

**Bird Houses Sale**  
Avalanche Cedar - New Designs, Low Price Guarantee, Free Shipping.  
[www.CedarBirdHouses.com](#) [Google](#)

**High Quality Bird Houses**  
Roosting boxes & decorative houses. S-Size Service. Free Shipping \$75+.  
[www.birdysitebird.com](#) [Google](#)

**Decorative Bird Houses**  
Beauty Your Garden With Our Wooden Bird Houses at a Discount.  
[BirdHouseStation.com](#) [Google](#)

# Decision-making Under Uncertainty



Google search results for "bird houses".

Search results for "bird houses" (Results 1 - 19 of about 22,100,000 for bird houses, \$0.15)

**Bird Houses**  
www.Scotts.com: Scotts attracts colorful birds to your backyard!

**Specialty Bird Houses**  
www.MidwestBack.com: Roosting Boxes, Purple Martins, Bat Chalets.

**Bird Houses at BestPrice**  
www.BestPrice.com: Over 225 different houses in stock. Free shipping!

**Learn More About Bird Houses** [Link]  
Bird House - JustBirdHouses.net learn more about blue bird houses and birdhouses, along with our blue bird house and purple martin bird houses.  
www.justbirdhouses.net/ - Cached - Similar - >

**Bird Houses** [Link]  
Bird Houses and Bird Feeders for north american bird species.  
www.birdhouses101.com/ - Cached - Similar - >

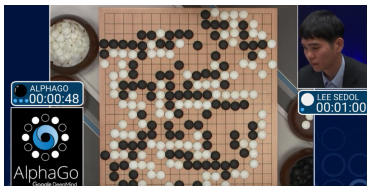
**Bird Feeders, Bird Houses, The Backyard Bird Company**  
Bird Feeders - The Backyard Bird Company has a variety of bird feeders will attract your landscapes and attract wildlife.  
Bird Houses - Bird Feeders - Decorative Bird Houses  
www.backyardbird.com/ - Cached - Similar - >

**Bird Houses**  
Find All Types Of Bird Feeders And Houses At Lowest Price Lowest Price.  
www.Lowes.com

**Bird Houses Sale**  
Authorized Dealer - New Designs, Low Price Guarantee, Free Shipping.  
www.CouponsandDeals.com/ - Google

**High Quality Bird Houses**  
Roosting boxes & decorative houses.  
5-Star Service. Free Shipping \$75+.  
www.backyardbird.com

**Decorative Bird Houses**  
Beautiful Your Garden With Our Wonderful Bird Houses at a Discount.  
BirdHouseStation.com  
Google





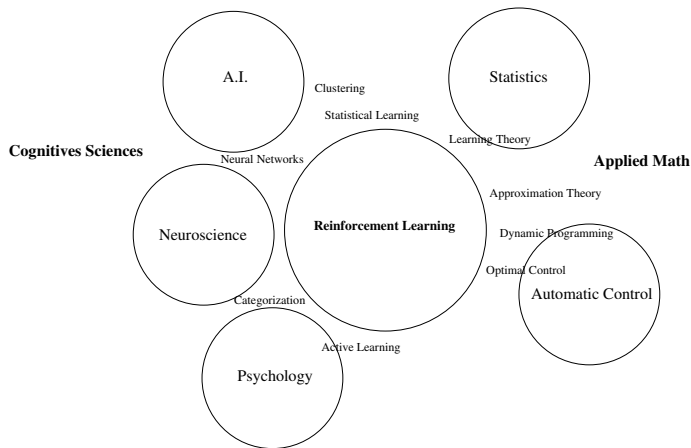
*Invia*

# Reinforcement learning

**Reinforcement learning** is learning what to do – how to map situations to actions – so as to **maximize** a numerical **reward** signal in an **unknown uncertain** environment. The learner is not told which actions to take, as in most forms of machine learning, but she must discover which actions yield the most reward by **trying them** (**trial-and-error**). In the most interesting and challenging cases, actions may affect not only the immediate reward but also the **next situation** and, through that, all **subsequent rewards** (**delayed reward**).

*“An introduction to reinforcement learning”,  
Sutton and Barto (1998).*

# A Multi-disciplinary Field



See the biannual conference “*Reinforcement learning and decision-making (RLDM)*”

# Behavioral (human and animal) Psychology

## The law of effect [Thorndike, 1911]

*“Of several **responses** made to the same situation, those which are accompanied or closely followed by **satisfaction** to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by **discomfort** to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur.*

*The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond.”*

# Inspirations at the Basis of RL

## *Psychology*

- ▶ *Classical (human and) animal conditioning*: “the magnitude and timing of the conditioned response changes as a result of the contingency between the conditioned stimulus and the unconditioned stimulus” [Pavlov, 1927].
- ▶ *Operant conditioning (or instrumental conditioning)*: process by which humans and animals *learn* to behave in such a way as to obtain *rewards* and avoid *punishments* [Skinner, 1938].

# Inspirations at the Basis of RL

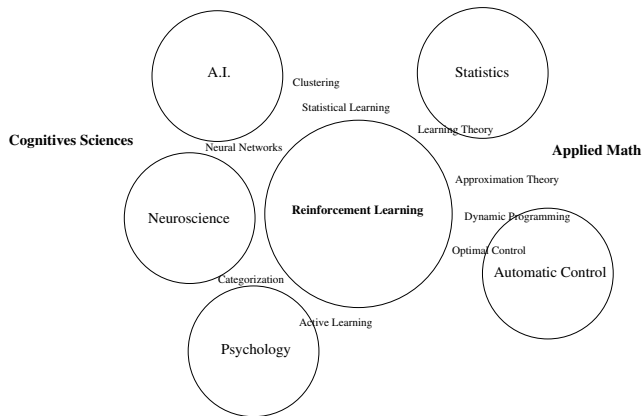
## *Psychology*

- ▶ *Classical (human and) animal conditioning*: “the magnitude and timing of the conditioned response changes as a result of the contingency between the conditioned stimulus and the unconditioned stimulus” [Pavlov, 1927].
- ▶ *Operant conditioning (or instrumental conditioning)*: process by which humans and animals *learn* to behave in such a way as to obtain *rewards* and avoid *punishments* [Skinner, 1938].

## *Neuroscience*

- ▶ *Hebbian learning*: development of formal models of how the synaptic weights between neurons are reinforced by simultaneous activation. “*Cells that fire together, wire together.*” [Hebb, 1961].
- ▶ *Dopamine and basal ganglia model*: direct link with motor control and decision-making (e.g., [Doya, 1999]).

# A Multi-disciplinary Field



*In this course:* RL is a **machine learning** paradigm

# A Machine Learning Paradigm

- ▶ *Supervised learning*: an expert (*supervisor*) provides examples of the right strategy (e.g., classification of clinical images).  
*Supervision is expensive.*



# A Machine Learning Paradigm

- ▶ *Supervised learning*: an expert (*supervisor*) provides examples of the right strategy (e.g., classification of clinical images).  
*Supervision is expensive.*
- ▶ *Unsupervised learning*: different objects are clustered together by similarity (e.g., clustering of images on the basis of their content). *No actual performance is optimized.*

# A Machine Learning Paradigm

- ▶ *Supervised learning*: an expert (*supervisor*) provides examples of the right strategy (e.g., classification of clinical images).  
*Supervision is expensive.*
- ▶ *Unsupervised learning*: different objects are clustered together by similarity (e.g., clustering of images on the basis of their content). *No actual performance is optimized.*
- ▶ *Reinforcement learning*: learning by direct interaction (e.g., autonomous robotics). *Minimum level of supervision (reward) and maximization of long term performance.*

# What is RL Capable of?

**Videos!**

# What Will You Be Capable of After this Course?

# What Will You Be Capable of After this Course?

**None of the Above!**

# What Will You Be Capable of After this Course?

**None of the Above!**

*But hopefully master everything that is at the basis of them!*

# Main Topics Covered in the Course

How to *model* an RL problem

- ▶ *What:* Markov decision process
- ▶ *Tools:* probability, processes, Markov chain

# Main Topics Covered in the Course

How to *model* an RL problem

How to solve *exactly* an RL problem

- ▶ *What*: Dynamic programming
- ▶ *Tools*: fixed point, operators



# Main Topics Covered in the Course

How to *model* an RL problem

How to solve *exactly* an RL problem

How to solve *incrementally* an RL problem

- ▶ *What*: temporal difference, Q-learning
- ▶ *Tools*: stochastic approximation

# Main Topics Covered in the Course

How to *model* an RL problem

How to solve *exactly* an RL problem

How to solve *incrementally* an RL problem

How to *efficiently* explore in an RL problem

- ▶ *What:* multi-armed bandit problem
- ▶ *Tools:* concentration inequalities

# Main Topics Covered in the Course

How to *model* an RL problem

How to solve *exactly* an RL problem

How to solve *incrementally* an RL problem

How to *efficiently* explore in an RL problem

How to solve *approximately* an RL problem

- ▶ *What*: approximate RL
- ▶ *Tools*: supervised learning, optimization

# Main Topics Covered in the Course

How to *model* an RL problem

How to solve *exactly* an RL problem

How to solve *incrementally* an RL problem

How to *efficiently* explore in an RL problem

How to solve *approximately* an RL problem

With (simple!) examples from *resource optimization*, *trade execution*,  
*computer games*, *recommendation systems*.

# Bibliography I



Bellman, R. (2003).

*Dynamic Programming.*

Dover Books on Computer Science Series. Dover Publications, Incorporated.



Damasio, A. R. (1994).

*Descartes' Error: Emotion, Reason and the Human Brain.*

Grosset/Putnam.



Doya, K. (1999).

What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex.

*Neural Networks*, 12:961–974.



Hebb, D. O. (1961).

Distinctive features of learning in the higher animal.

In Delafresnaye, J. F., editor, *Brain Mechanisms and Learning*. Oxford University Press.



Pavlov, I. (1927).

*Conditioned reflexes.*

Oxford University Press.

# Bibliography II



Pontryagin, L. and Neustadt, L. (1962).

*The Mathematical Theory of Optimal Processes.*

Number v. 4 in Classics of Soviet Mathematics. Gordon and Breach Science Publishers.



Skinner, B. F. (1938).

*The behavior of organisms.*

Appleton-Century-Crofts.



Thorndike, E. (1911).

*Animal Intelligence: Experimental Studies.*

The animal behaviour series. Macmillan.

# Reinforcement Learning



*Alessandro Lazaric*

[alessandro.lazaric@inria.fr](mailto:alessandro.lazaric@inria.fr)

[sequel.lille.inria.fr](http://sequel.lille.inria.fr)