Last Module Recap (“Course Overview”)

- Applications of Reinforcement Learning
- Overview of Modules and Labs
- Lab Software Installation

Microsoft

Introduction To Reinforcement Learning

ROLAND FERNANDEZ

Researcher, MSR AI
Instructor, AI School

How do infants learn?

- Tasks: turn over, sit up, crawl, stand
- Learning:
  - Trial-and-error
  - Practice
  - Goal-oriented
  - Interacting with the environment

https://www.youtube.com/watch?v=6jv_Icfa4g4

Outline

- What is Reinforcement Learning?
- Comparisons to other learning types
- Exploration / Exploitation Dilemma
- Elements of the RL problem
- Examples of RL problems
- Approaches to RL
- Fundamental Challenges of RL

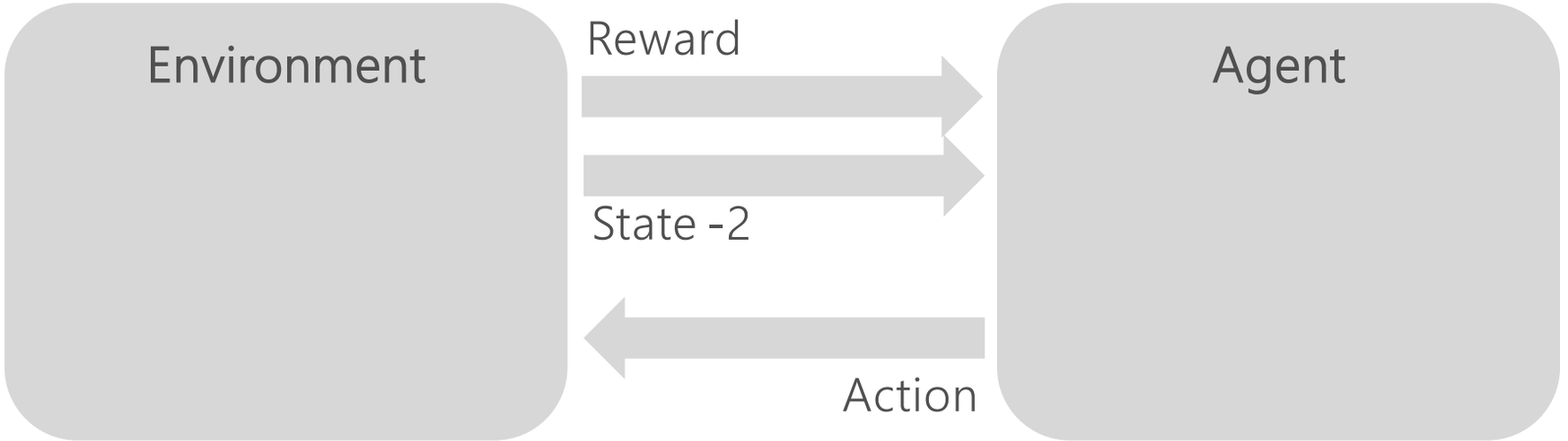
Outline

- What is Reinforcement Learning?
- Comparisons to other learning types
- Exploration / Exploitation Dilemma
- Elements of the RL problem
- Examples of RL problems
- Approaches to RL
- Fundamental Challenges of RL

What is Reinforcement Learning?

- Computational approach to goal-directed learning from interaction with the environment
- It uses idealized learning situations and explores the effectiveness of various learning algorithms

The General RL Problem



Historical Context

- Psychology of Animal Learning
- Puzzle Box
  - Video: https://www.youtube.com/watch?v=fanm--WyQJo

Thorndike Puzzle Box

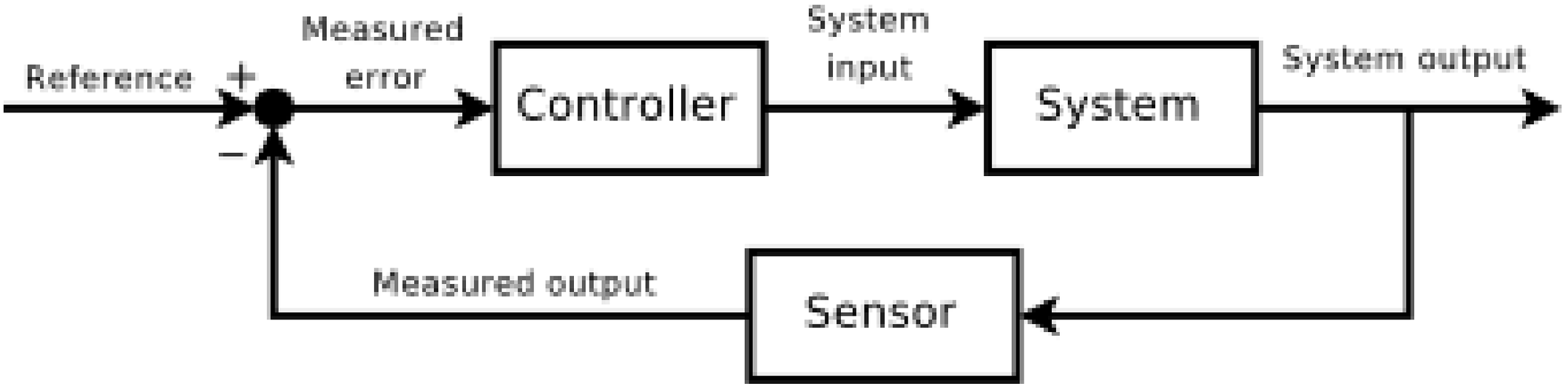
https://terriermandotcom.blogspot.com/2012/05/thorndikes-cat-box.html

Historical Context

- Law Of Effect:
  - Increase behaviors followed by pleasant consequences
  - Decrease behaviors followed by unpleasant consequences
- Trial-and-Error Learning:
  - Selectional
  - Associative

Historical Context

- Optimal Control Theory



https://en.wikipedia.org/wiki/Control_theory

Outline

- What is Reinforcement Learning?
- Comparisons to other learning types
- Exploration / Exploitation Dilemma
- Elements of the RL problem
- Examples of RL problems
- Approaches to RL
- Fundamental Challenges of RL

Supervised Learning

Supervised Learning Data

Example #	X	Y
0	X1	Y1
1	X2	Y2
2	X3	Y3
3	X4	Y4
4	X5	Y5
5	X6	Y6
6	X7	Y7

Unsupervised Learning

Unsupervised Learning

X
X1
X2
X3
X4
X5
X6
X7

- Clustering
- Anomaly Detection
- Low Dim Representation
- Sparse Representation
- Independent Representation

Reinforcement Learning

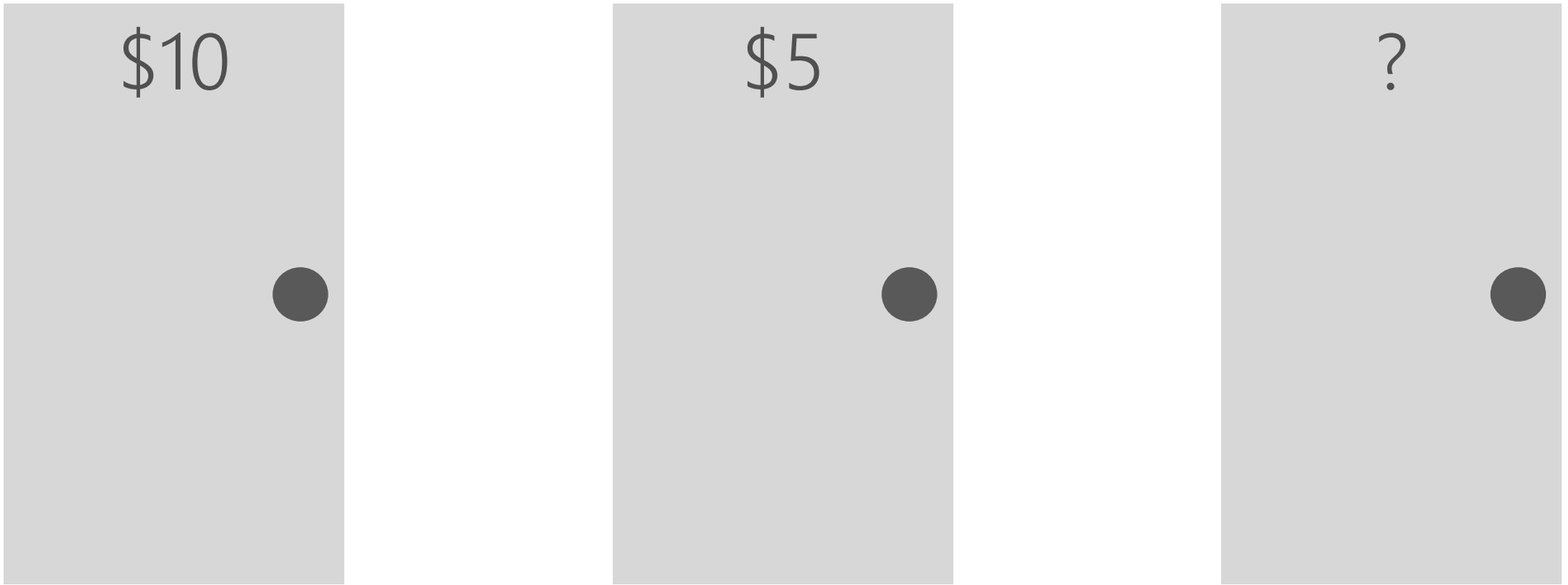
Reinforcement Learning

Timestep #	State	Action	Reward
0	S1	A1	-1
1	S2	A2	0
2	S3	A3	0
3	S4	A4	0
4	S5	A5	0
5	S6	A6	0
6	S7	A7	1

Outline

- What is Reinforcement Learning?
- Comparisons to other learning types
- Exploration / Exploitation Dilemma
- Elements of the RL problem
- Examples of RL problems
- Approaches to RL
- Fundamental Challenges of RL

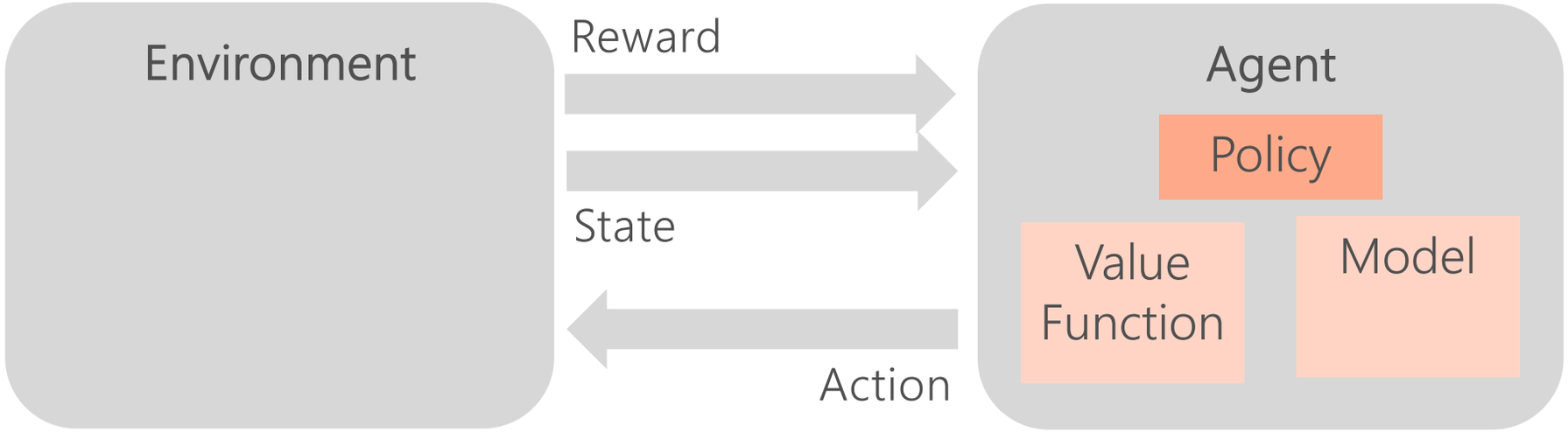
Exploration / Exploitation Dilemma



Outline

- What is Reinforcement Learning?
- Comparisons to other learning types
- Exploration / Exploitation Dilemma
- Elements of the RL problem
- Examples of RL problems
- Approaches to RL
- Fundamental Challenges of RL

Elements of Reinforcement Learning



Time step

- Sets the timing for the basic action selection and state/reward feedback cycle
- Usual notation: t

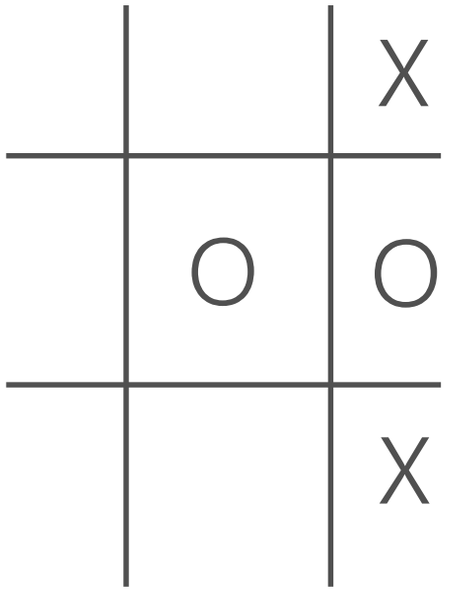
Environment

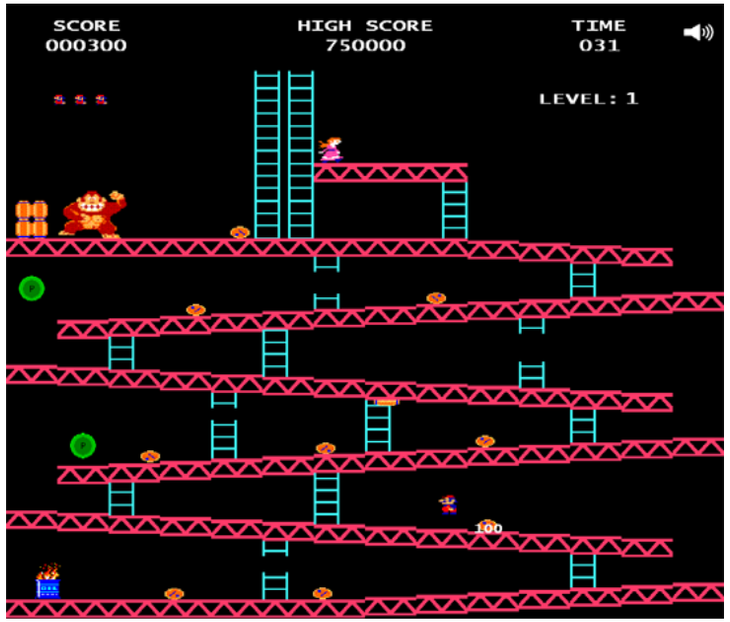
- Defines the world that the agent interacts with
- Basic loop:
  - Produces:
    - States, Rewards for the agent to sense and process
  - Accepts:
    - Actions from the agent

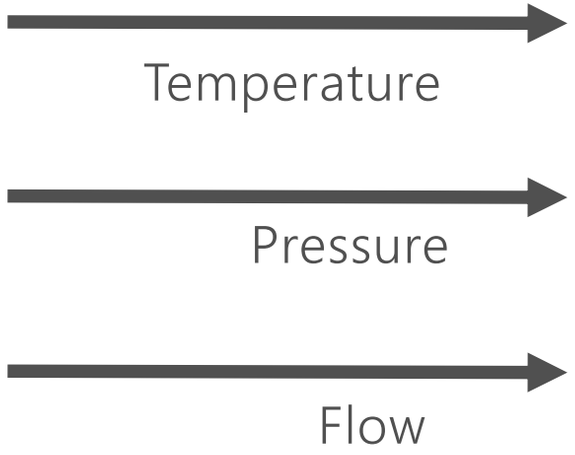
Agent

- Learns to achieve goals by interacting with environment
- Basic loop:
  - Senses: State, Reward
  - Effect: selects an Action

State







https://www.noupe.com/inspiration/showcases/the-evolution-of-mario.html

Reward

- The scalar value (floating point number) returned by the environment when the agent selects an action
- Represents the goal, or goals
- Determined by the RL problem designer
- Usual notation r_t, for reward at time t

Action

- Discrete (1 of N actions)
- Continuous (action as scalar/vector of real values)

Value Function

- Value Function $V^{\pi}(s)$ – a mapping from state s to the long-term accumulation of reward starting from s, following policy π
- Value-State Function $Q^{\pi}(s, a)$ – a mapping from each state/action pair to the expected long-term accumulation of reward starting from state s and taking action a, and thereafter following policy π

Policy

- A mapping from each state to an action to take when in that state
  - Deterministic
  - Stochastic

Models of the Environment

State	Action	Next State	Reward	Probability
S1	A1	S2	-1	.3
S1	A1	S3	0	.7
S1	A2	S4	0	.1
S1	A2	S5	0	.1
S1	A2	S6	0	.8
S2	A1	S7	0	.5
S2	A1	END	1	.5

- Model-Free Methods
- Model-Based Methods:
  - Given model up-front
  - Learn model while exploring

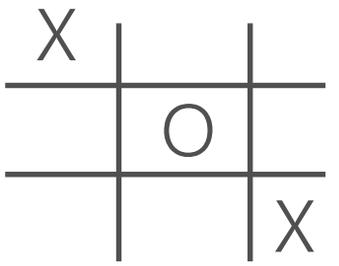
Episodic and Continuing Tasks

- Episodic Tasks
- Continuing Tasks

Outline

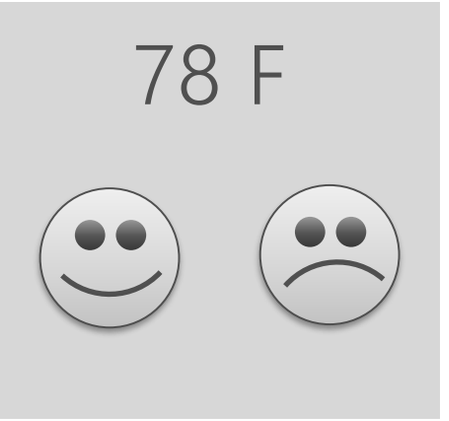
- What is Reinforcement Learning?
- Comparisons to other learning types
- Exploration / Exploitation Dilemma
- Elements of the RL problem
- Examples of RL problems
- Approaches to RL
- Fundamental Challenges of RL

Tic-Tac-Toe (Noughts and Crosses)



- **Agent:** "X" player
- **Environment:** "O" player, general rules
- **State:** 9 x Square occupant: X, O, blank
- **Actions:** 9 x Place "X" in square
- **Rewards:** At end of game: 1=win, 0=tie, -1=loss
- **Task Type**: Episodic

Smart Thermostat



- **Agent:** AC / Heater control
- **Environment:** House and its occupants
- **State:** Current Temperature, Day, Time
- **Actions:** Heat, Cool, Off
- **Rewards:** Smile: +1, Frown: -1, None: 0
- **Task Type**: Continuing

Simulated Robot in Obstacle Course

- **Agent:** Robot brain
- **Environment:** Robot body, Obstacle Course, Physics Simulator
- **State:** Angle/position robot joints, next obstacle distance/type/dims
- **Actions:** Controls for robot muscles/joints
- **Rewards:** X-coordinate of robot
- **Task Type**: Episodic

https://www.youtube.com/watch?v=hx_bgoTF7bs

Outline

- What is Reinforcement Learning?
- Comparisons to other learning types
- Exploration / Exploitation Dilemma
- Elements of the RL problem
- Examples of RL problems
- Approaches to RL
- Fundamental Challenges of RL

Approaches to RL

- **Value function Methods**
- **Direct Policy Search**
- Evolutionary Methods
- Others…

Outline

- What is Reinforcement Learning?
- Comparisons to other learning types
- Exploration / Exploitation Dilemma
- Elements of the RL problem
- Examples of RL problems
- Approaches to RL
- Fundamental Challenges of RL

Fundamental Challenges

- Representation
- Generalization
- Exploration
- Temporal Credit Assignment

Introduction to RL- Summary

- What is Reinforcement Learning?
- Comparisons to other learning types
- Exploration / Exploitation Dilemma
- Elements of the RL problem
- Examples of RL problems
- Approaches to RL
- Fundamental Challenges

Further Reading

- **Thorndike's puzzle boxes and the origins of the experimental analysis of behavior** (Chance, 1999): https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1284753/pdf/jeabehav007200300433.pdf
- **Animal intelligence : an experimental study of the associative processes in animals** (Thorndike, 1898): https://archive.org/details/animalintelligen00thoruoft