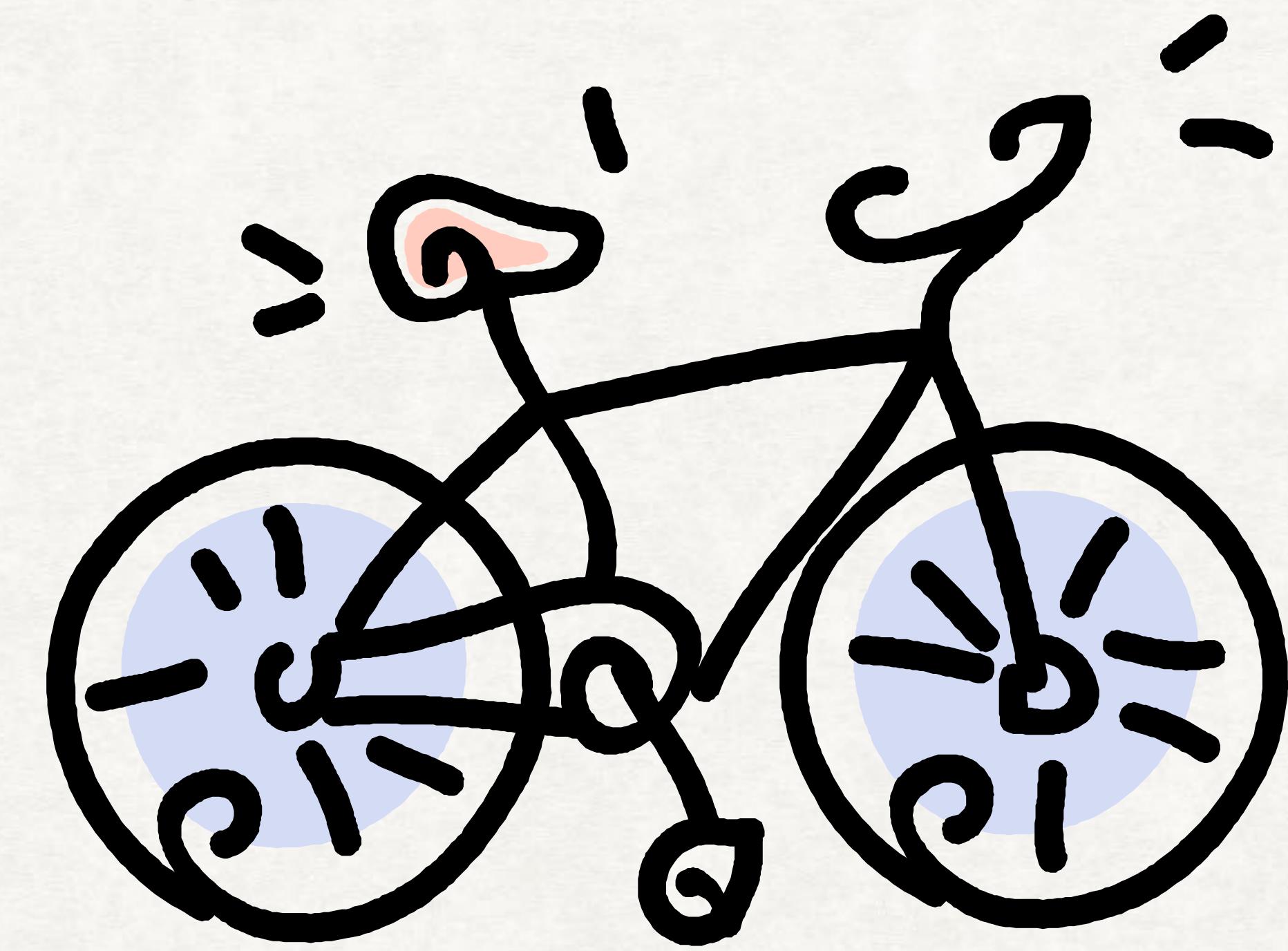
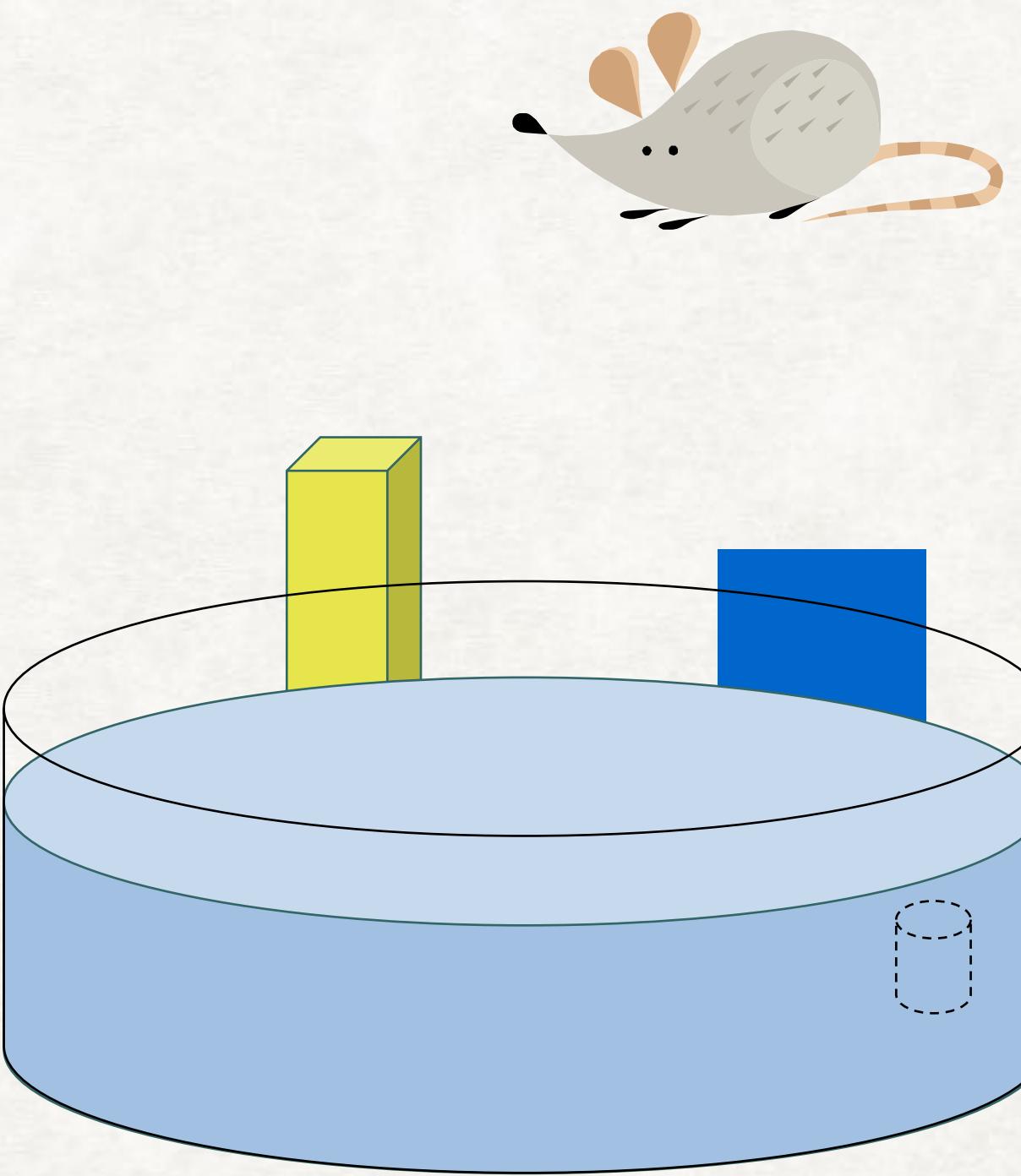


COM3240

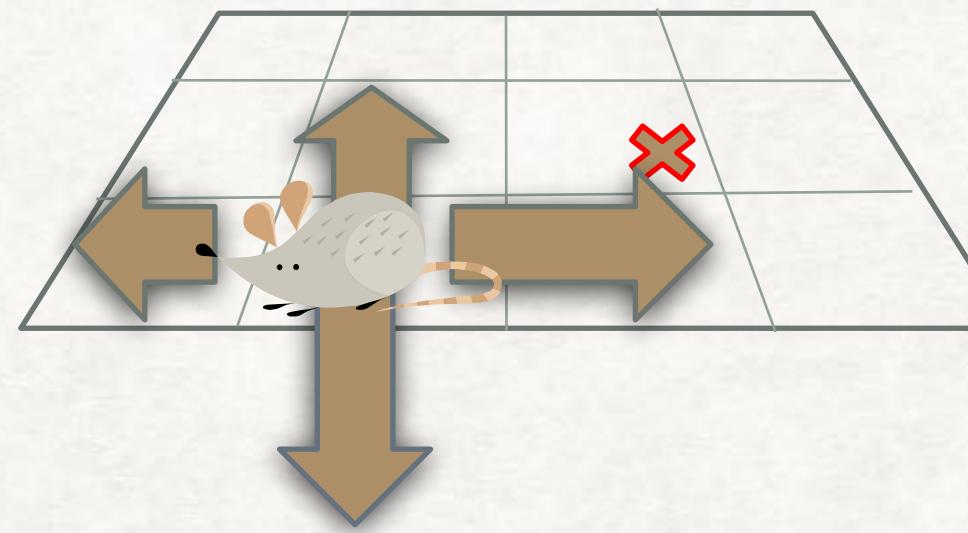
REINFORCEMENT LEARNING

REINFORCEMENT LEARNING

EXAMPLES IN LIFE AND THE LAB



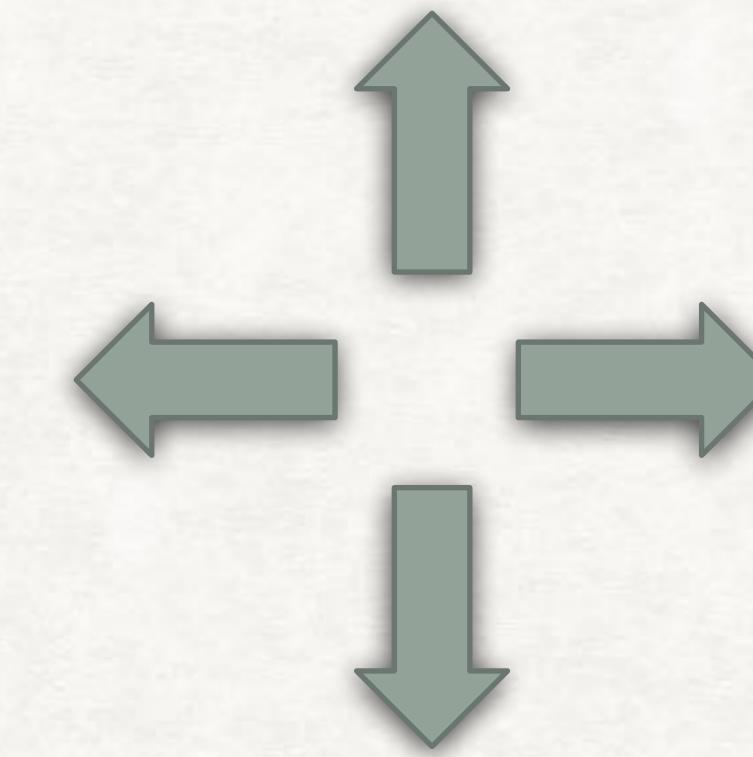
Q-VALUES IN TEMPORAL DIFFERENCE (REINFORCEMENT) LEARNING



Maximise expected return

Q (state, action)

We do not know the Q values



explore (randomness)

exploit (take "best" action)

SUPERHUMAN PERFORMANCE ON ATARI GAMES

BY GOOGLE DEEPMIND



<https://deepmind.com/research/dqn/>

ALPHAGO ZERO

BY GOOGLE DEEPMIND

► Starting from scratch



ALPHA ZERO

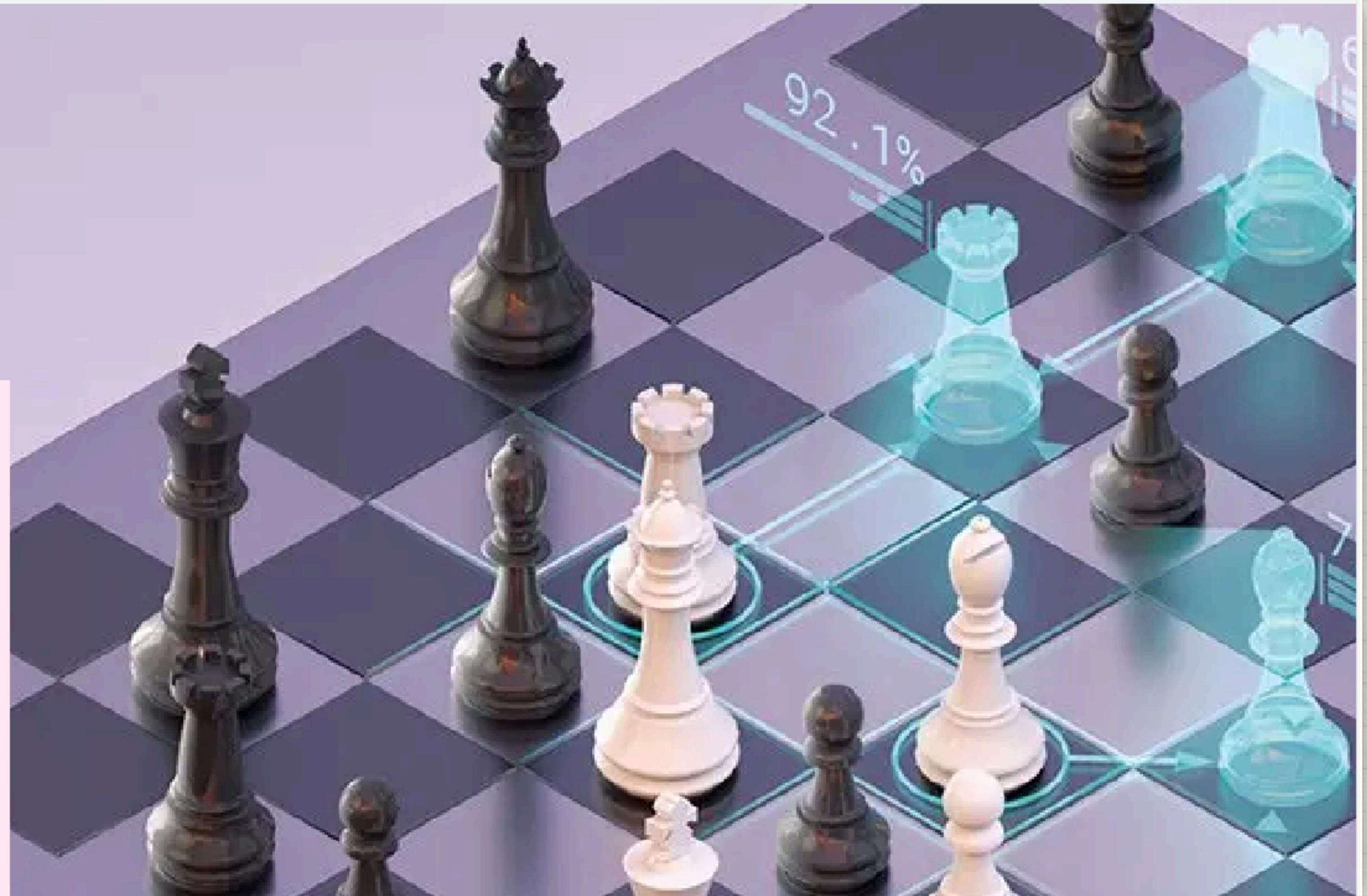
BY GOOGLE DEEPMIND



BLOG POST
RESEARCH

AlphaZero: Shedding new light on chess, shogi, and Go

06 DEC 2018









THE SCHOOL OF EPICURUS

AND HIS MOST PRAISED STUDENT

Themista of Lampsacus (Greek: Θεμίστη), the wife of Leonteus, was a student of Epicurus, early in the 3rd century BC.^[1] Epicurus' school was unusual in the 3rd century, in that it allowed women to attend, and we also hear of Leontion attending Epicurus' school around the same time. Cicero ridicules Epicurus for writing "countless volumes in praise of Themista," instead of more worthy men such as Miltiades, Themistocles or Epaminondas.^[2] Themista and Leonteus named their son Epicurus.^[3]

Girton College, Cambridge est. 1869



WOMEN'S RIGHT TO VOTE

SWITZERLAND, 1990

12 December 1971	Bern, Thurgau
23 January 1972	St. Gallen
30 January 1972	Uri
5 March 1972	Schwyz and Graubünden
30 April 1972	Nidwalden
24 September 1972	Obwalden
30 April 1989	Appenzell Ausserrhoden
27 November 1990	Appenzell Innerrhoden (by decision of the Federal Supreme Court of Switzerland)

The Jura, created by secession from Berne on 20 March 1977, has always had women's suffrage.

“
Pleasure is our first and kindred good.
It is the starting-point of every choice
and of every aversion

— Epicurus' *Letter to Menoeceus*
Diogenes Laertius, *Lives of Eminent Philosophers*

”

“
and to it we come back, inasmuch as we
make feeling the rule by which to judge of
every good thing.

— Epicurus' *Letter to Menoeceus*
Diogenes Laertius, *Lives of Eminent Philosophers*

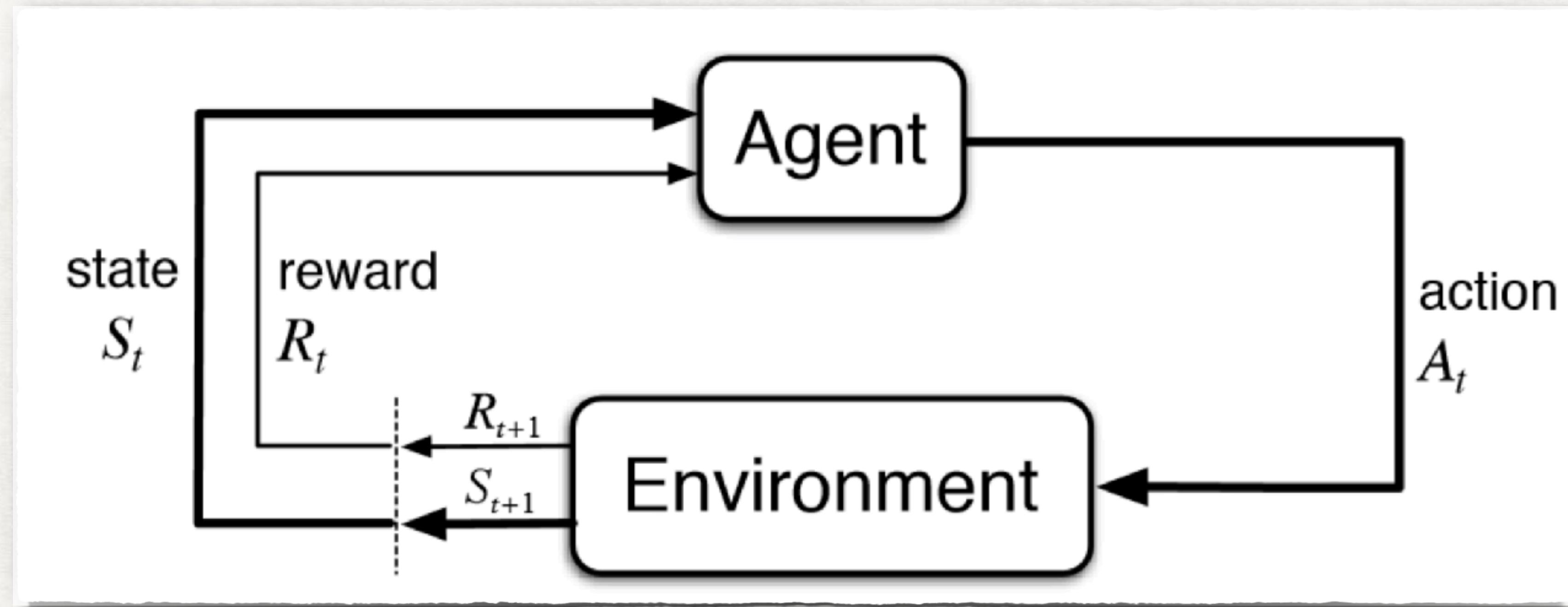
”

REWARDS AND PUNISHMENTS IN EPICUREAN PHILOSOPHY

Good=Pleasure

Evil=Pain

REINFORCEMENT LEARNING SCHEME COMPATIBLE WITH THE EPICUREAN VIEW OF THE WORLD



Sutton and Barto (2018), with permission

REWARDS AND PUNISHMENTS IN REINFORCEMENT LEARNING

Good=Reward

Evil=Negative Reward
(i.e. Punishment)

“

By pleasure we mean the absence of pain in the body and of trouble in the soul.

— Epicurus' Letter to Menoeceus

Diogenes Laertius, *Lives of Eminent Philosophers*

”

“
It is not an unbroken succession of drinking-bouts
and of merrymaking [...], which produce a
pleasant life; it is sober reasoning, searching out
the grounds of every choice and avoidance, [...]”

— Epicurus' *Letter to Menoeceus*
Diogenes Laertius, *Lives of Eminent Philosophers*

EXPECTED RETURN IN REINFORCEMENT LEARNING

$$G_t = R_{t+1} + R_{t+2} + R_{t+3} + \dots + R_{t+N}$$



Vasilaki (2017) arXiv:1710.04582

DISCOUNT FACTOR IN REINFORCEMENT LEARNING

$$G_t = R_{t+1} + R_{t+2} + R_{t+3} + \dots + R_{t+N}$$

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \dots$$

$$0 \leq \gamma < 1$$

This is why I am impatient

FRIENDSHIP VS ROMANCE

IN EPICUREAN PHILOSOPHY

$$Q(\textit{friendship}) > Q(\textit{romance})$$

FRIENDSHIP VS ROMANCE

IN EPICUREAN PHILOSOPHY

$$Q(state, friendship) > Q(state, romance)$$

REWARD PERCEPTION AND THE SARSA ALGORITHM

$$\Delta Q(S_t, A_t) = \alpha [(R_{t+1} + \gamma Q(S_{t+1}, A_{t+1})) - Q(S_t, A_t)]$$

What I “actually” get

Anticipated reward

A positive reward may feel like punishment

A negative reward may feel like reward

REWARD PERCEPTION AND THE SARSA ALGORITHM

$$\Delta Q(S_t, A_t) = \alpha (1 - 0)$$

What I “actually” get

Anticipated reward

Reward, Positive Change

REWARD PERCEPTION AND THE SARSA ALGORITHM

$$\Delta Q(S_t, A_t) = \alpha \quad ((-1) \quad - \quad 0)$$

What I “actually” get

Anticipated reward

Punishment, Negative Change

REWARD PERCEPTION AND THE SARSA ALGORITHM

$$\Delta Q(S_t, A_t) = \alpha (1 - 10)$$

What I “actually” get

Anticipated reward

Negative change:

A positive reward may feel like punishment

REWARD PERCEPTION AND THE SARSA ALGORITHM

$$\Delta Q(S_t, A_t) = \alpha \quad ((-1) - (-10))$$

What I “actually” get

Anticipated reward

Positive change:

A negative reward (punishment) may feel like reward

EFFORT AND THE SARSA ALGORITHM

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \dots$$

$$0 \leq \gamma < 1$$

The effort involved in taking an action
can be modelled as a negative reward

Not obvious to an external observer!

HAPPINESS AND THE SARSA ALGORITHM

- I am better off when I have low expectations.
- I only observe other people successes (rewards). Not their efforts!
- When considering achievements, is an inherent element of luck.

REINFORCEMENT LEARNING

PREREQUISITES

- Derivatives (& partial derivatives), the chain rule.
- Probabilities & Statistics.
- Matrix Algebra.
- You might need pen and paper (or stylus and tablet) for the lab and even for the lectures.

REINFORCEMENT LEARNING

KEY TOPICS

- Immediate Rewards (Bandit problems).
- Future Rewards (Q-Learning, SARSA, Deep Reinforcement Learning...).
 - For Deep RL we will be covering ANN and Deep Learning.
 - Bellman Equations and Reinforcement Learning.
- Research topics.

REINFORCEMENT LEARNING

GENERATIVE AI

- Is permitted in accordance to the school regulation (see student handbook).
- Organic to exploring the lab material.
- Has been used to support the material of this module.

REINFORCEMENT LEARNING

EVALUATION

- One assignment covering theory and practice (40%, pass/fail).
- A formal exam (60%, 0-100 scale).

REINFORCEMENT LEARNING

CONTACT

- Module discussion Board on Blackboard.
- During the lab sessions 3-5 on Thursdays.
- Office hours 5-6pm on Thursdays.

SUGGESTED READING

- Introduction to Reinforcement Learning, R.S. Sutton and A.G. Barto” <http://incompleteideas.net/book/the-book.html>

THANK YOU!