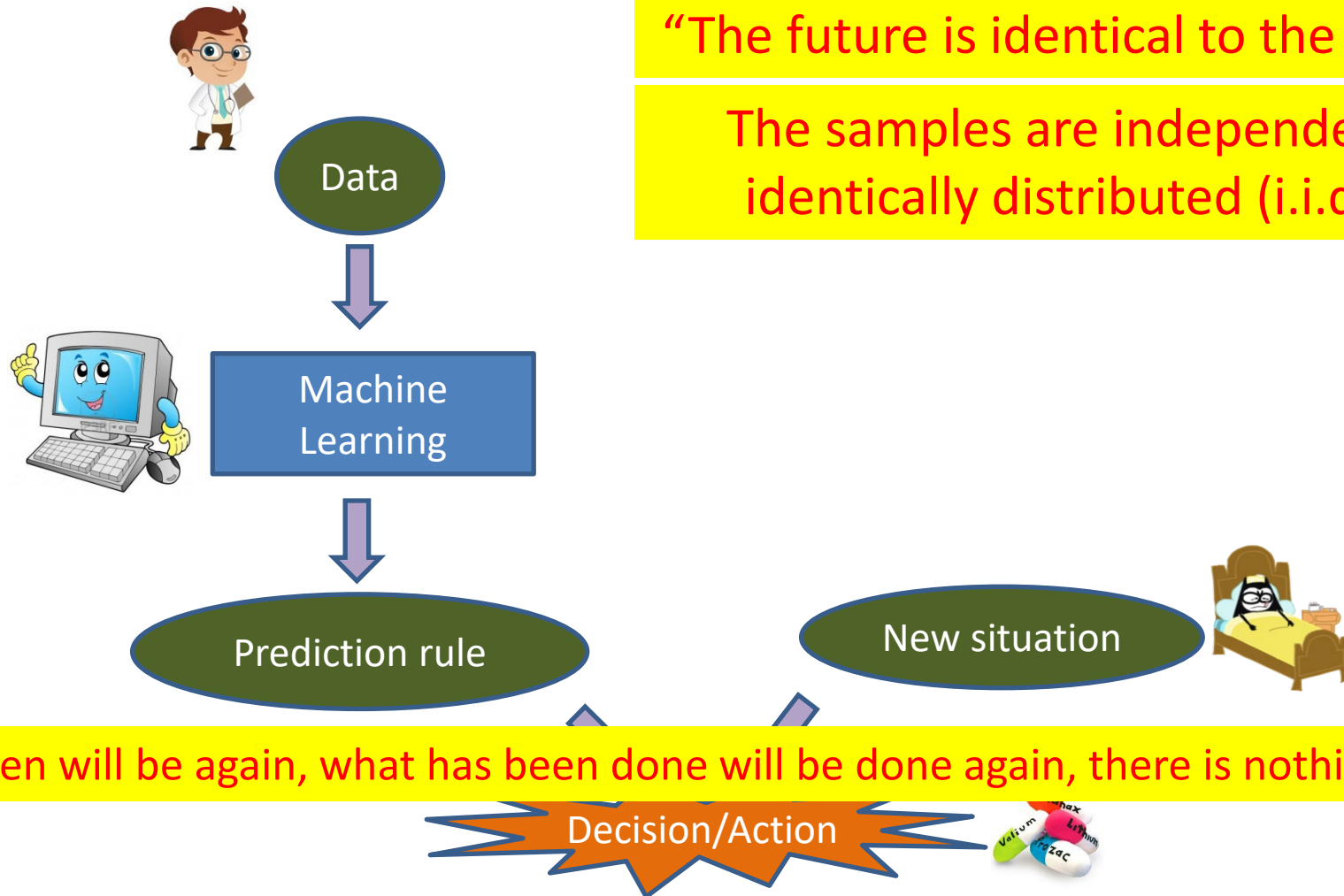# Online and Reinforcement Learning

Yevgeny Seldin

University of Copenhagen

"Classical" (Batch) Machine Learning

Data

Machine Learning

Prediction rule

New situation

Decision/Action

Assumption
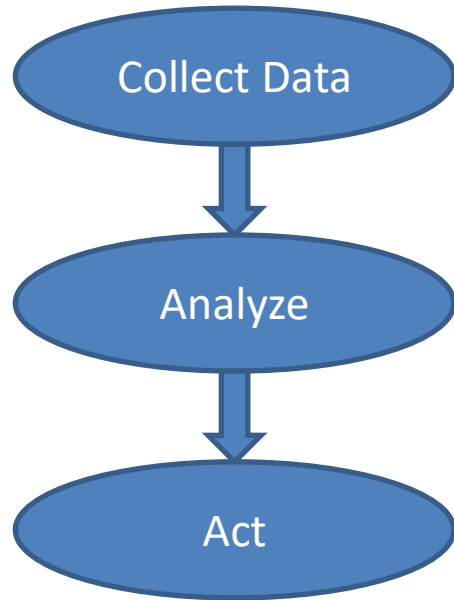"The future is identical to the past"

The samples are independent identically distributed (i.i.d.)

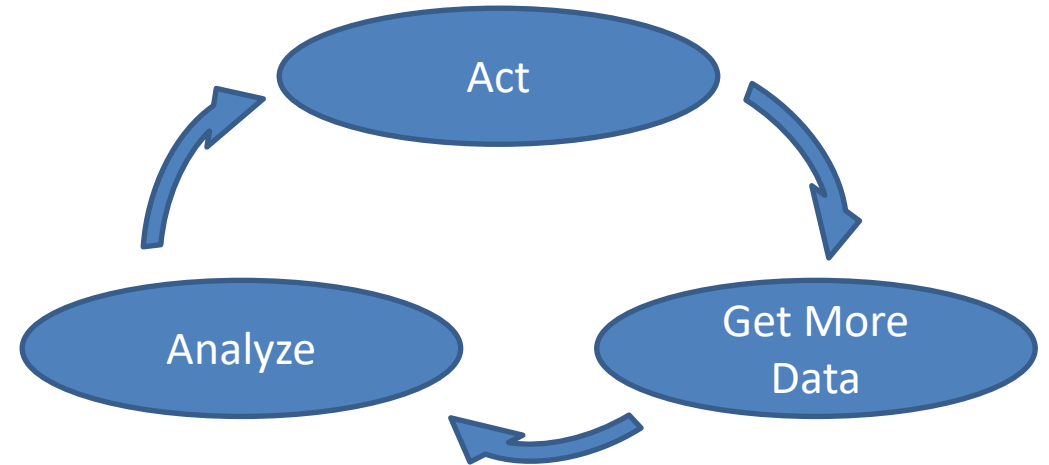"What has been will be again, what has been done will be done again, there is nothing new under the sun."

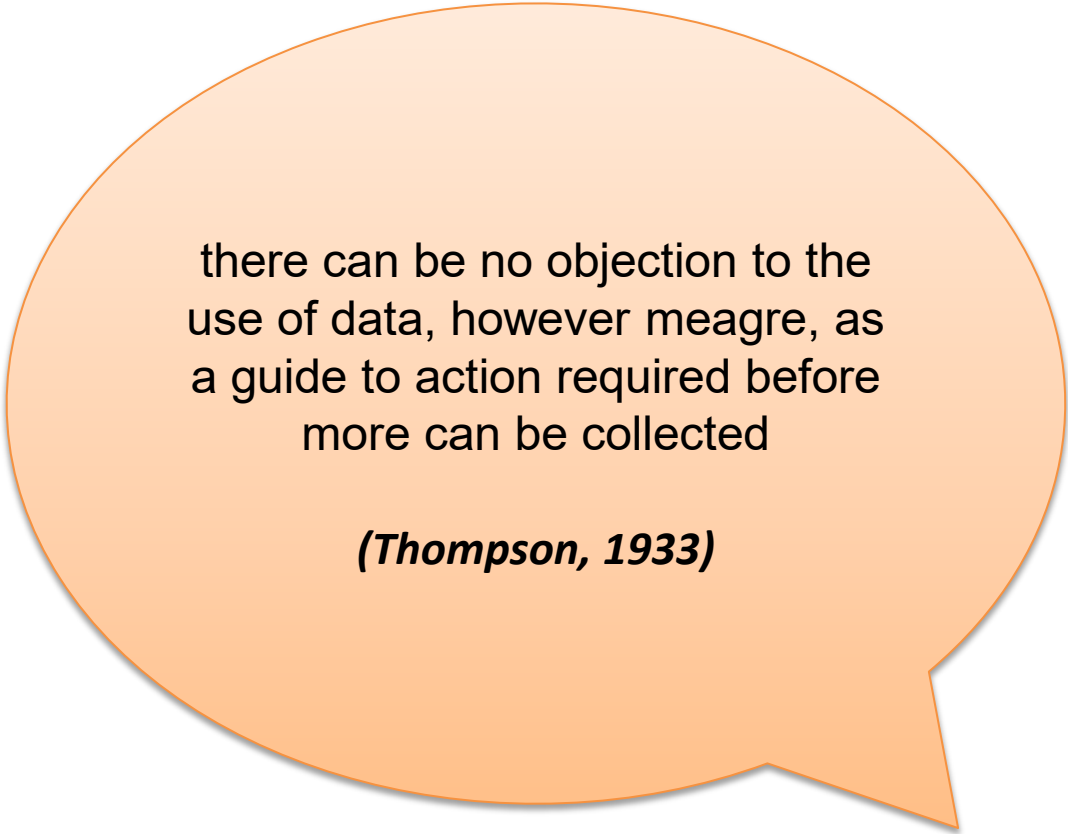# How Online is different from "batch"?

Batch Learning

Online Learning

Examples

- Investment in the stock market
- Online advertising/personalization
- Online routing
- Games
- Robotics
- …

When do we need Online Learning?

there can be no objection to the use of data, however meagre, as a guide to action required before more can be collected

(Thompson, 1933)

When do we need Online Learning?

Until recently, statistical theory has been restricted to the design and analysis of sampling experiments in which the size and composition of the samples are completely determined before the experimentation begins. The reasons for this are partly historical, dating back to the time when the statistician was consulted, if at all, only after the experiment was over, and partly intrinsic in the mathematical difficulty of working with anything but a fixed number of independent random variables. A major advance now appears to be in the making with the creation of a theory of the sequential design of experiments, in which the size and composition of the samples are not fixed in advance but are functions of the observations themselves.
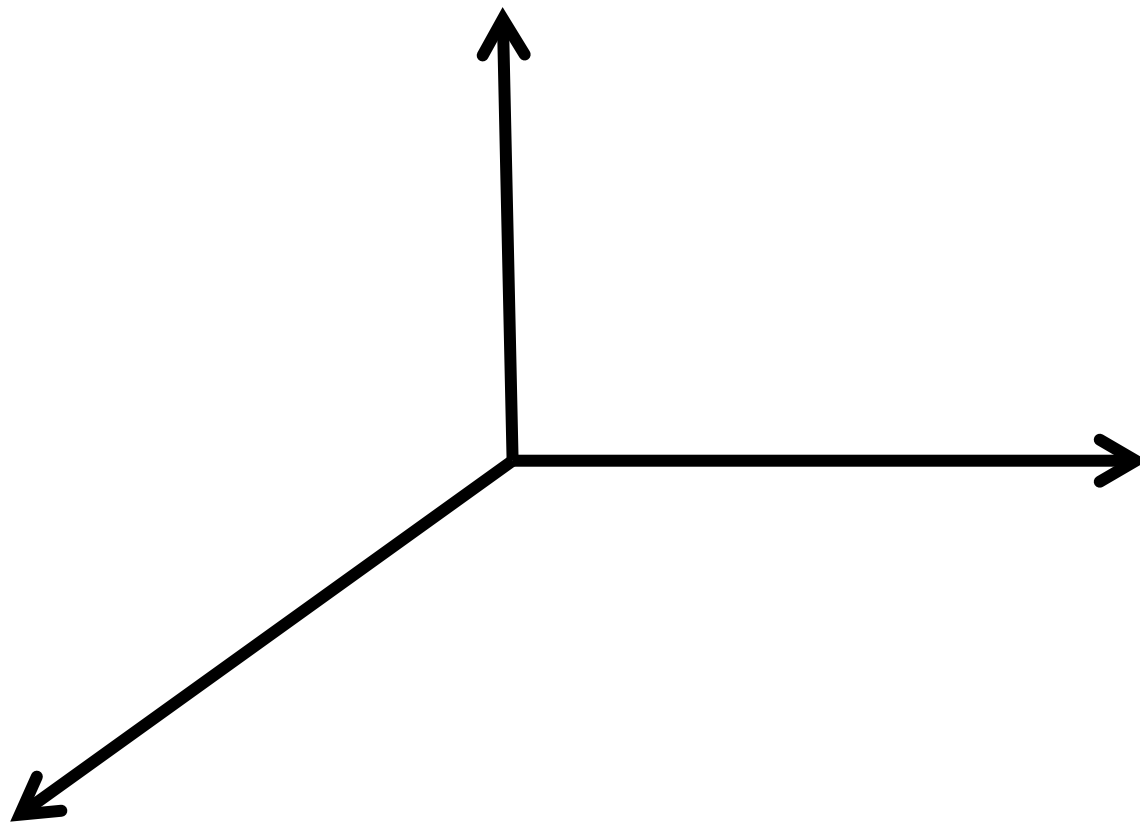
(Robbins, 1952)

When do we need Online Learning?

- Interactive learning

- "Adversarial" game-theoretic settings

  – No assumption on similarity of past and future

- Intelligent data collection

- Large-scale data analysis
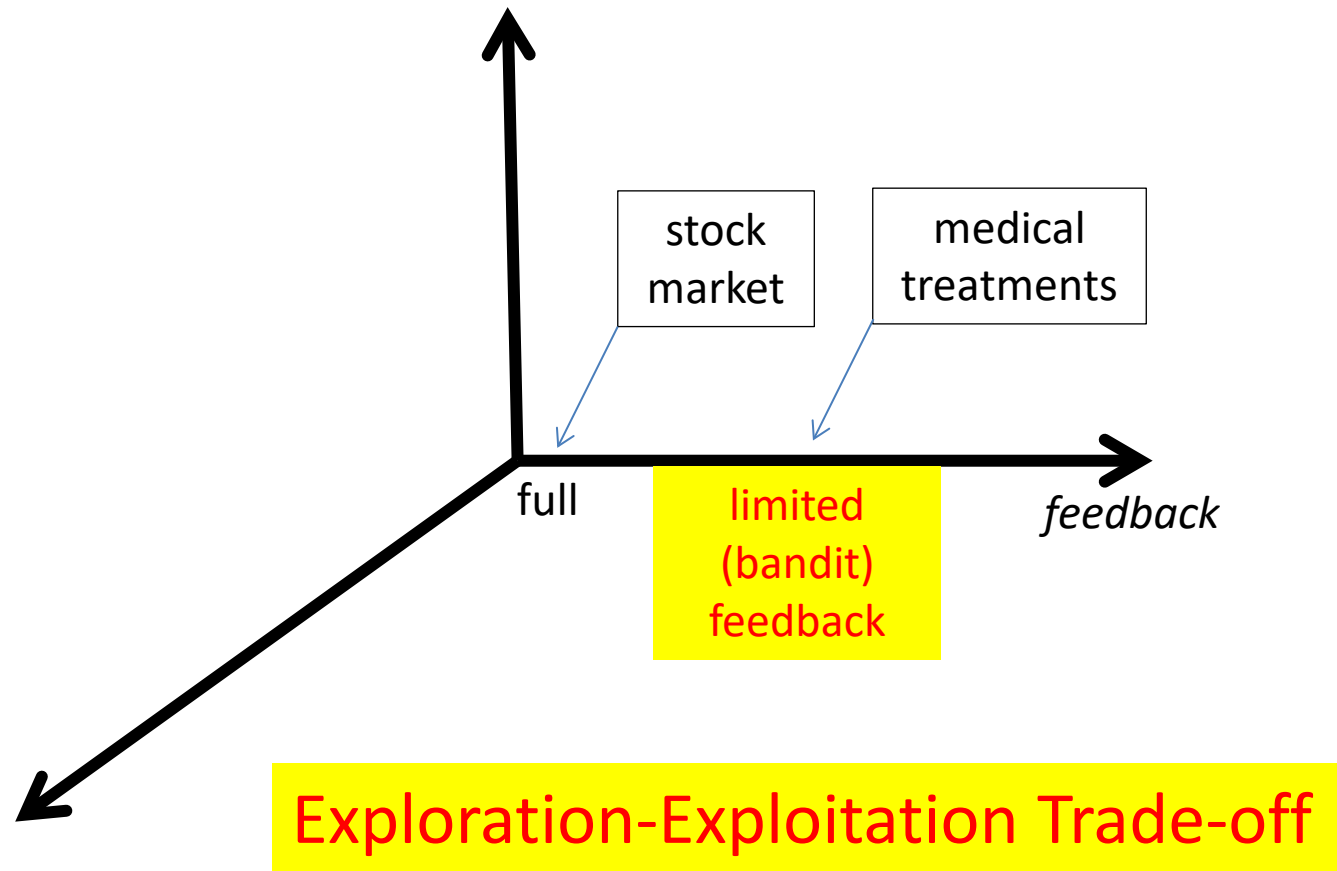
The Space of Online Learning Problems

The Space of Online Learning Problems
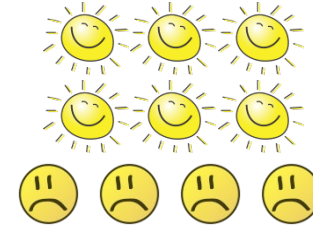
The Space of Online Learning Problems

stock
market

medical
treatments

full

limited
(bandit)
feedback

feedback
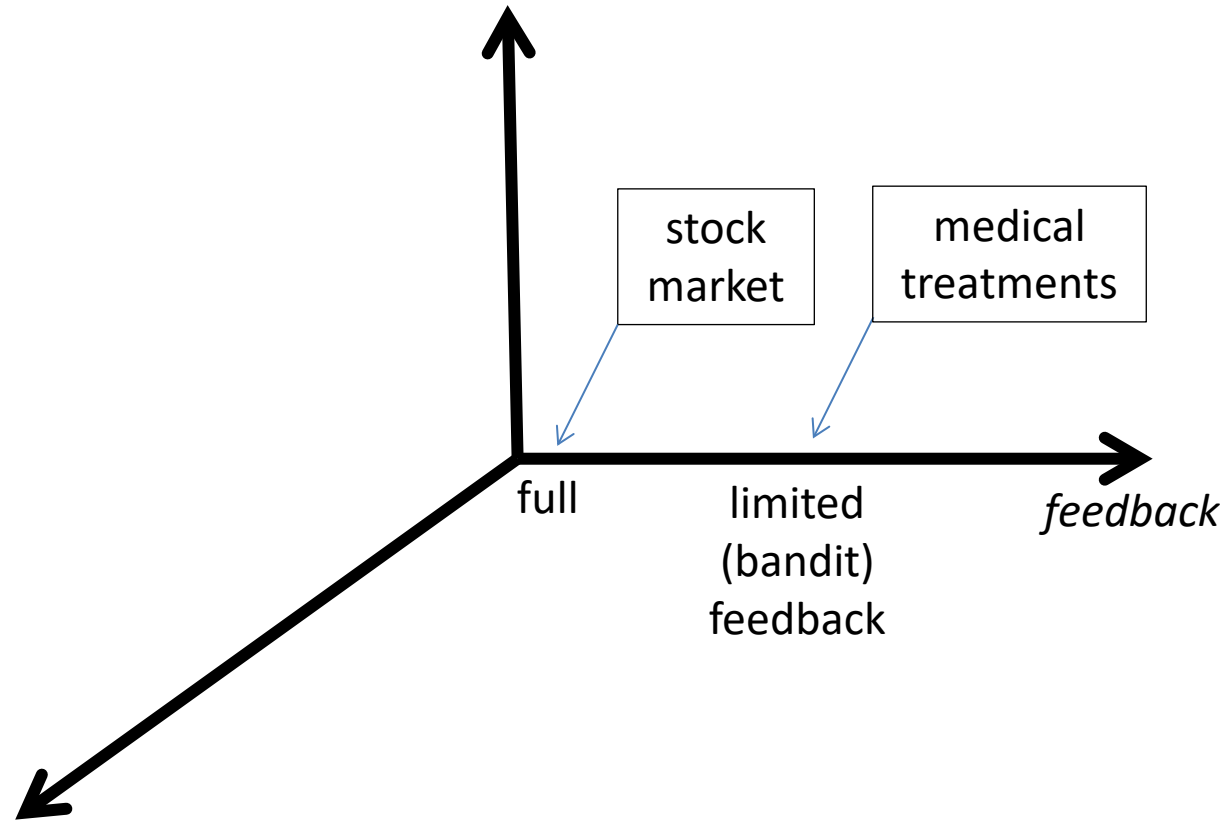
Exploration-Exploitation Trade-off

Exploration-Exploitation Trade-off

Never tried   0/2   6/10

What drug to give to a new patient

When there are more patients to come…

Act

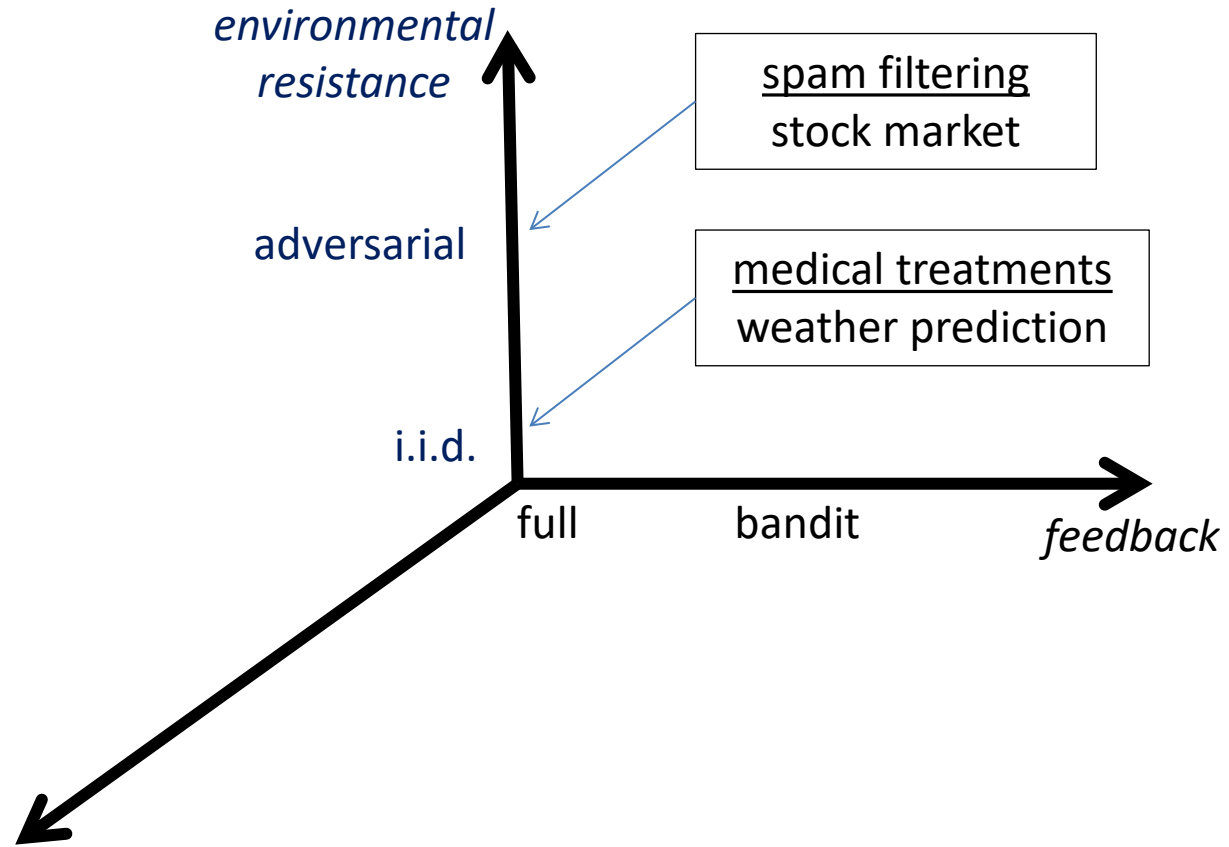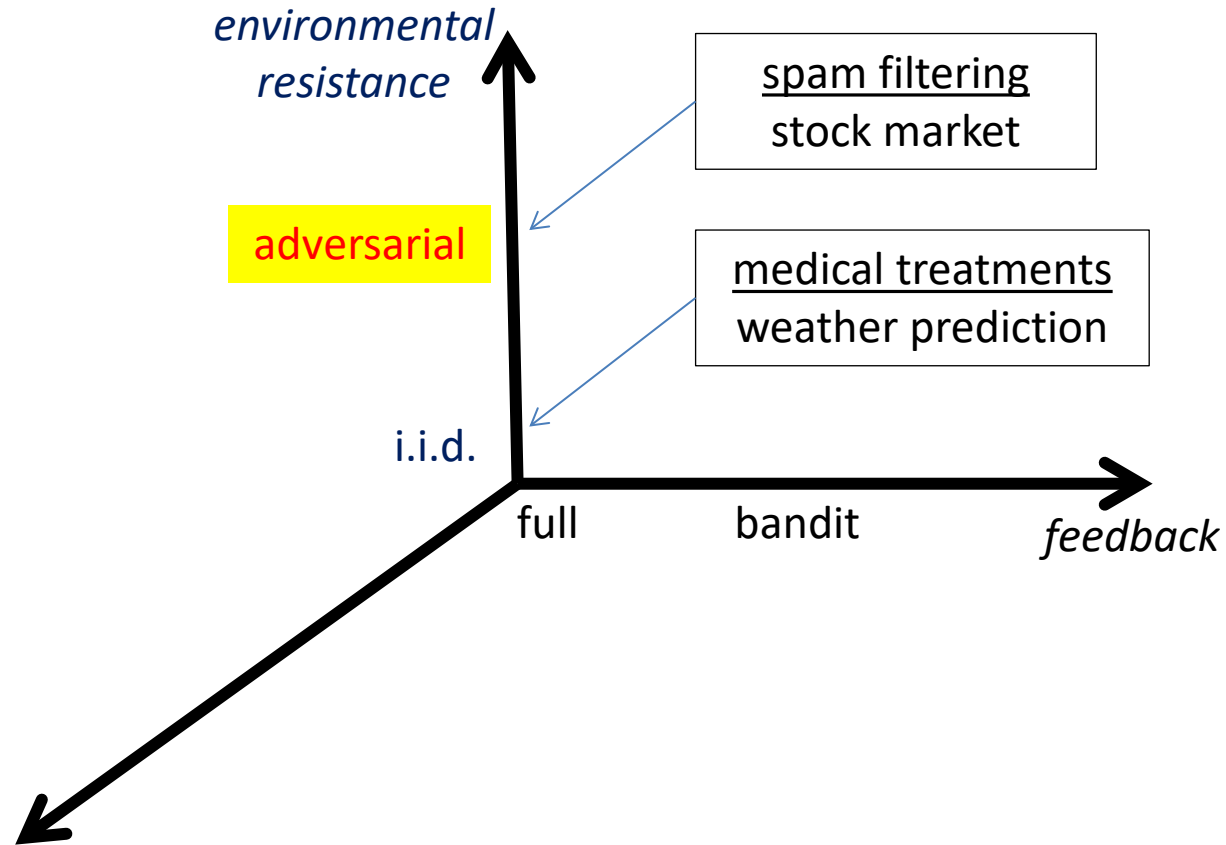Analyze   More Data
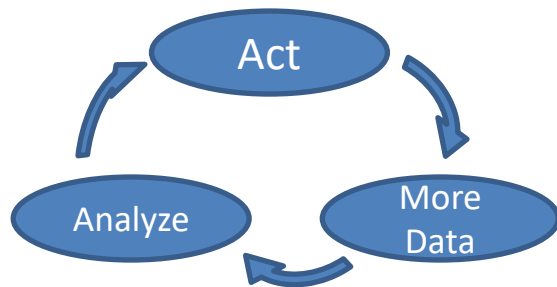
We are building the dataset for ourselves

The Space of Online Learning Problems

The Space of Online Learning Problems
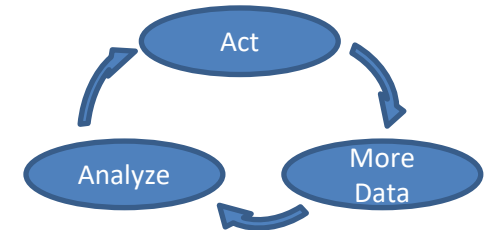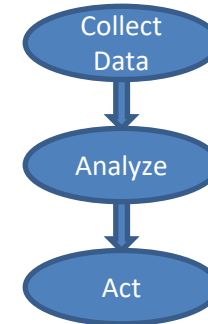
Learning in Adversarial Environments
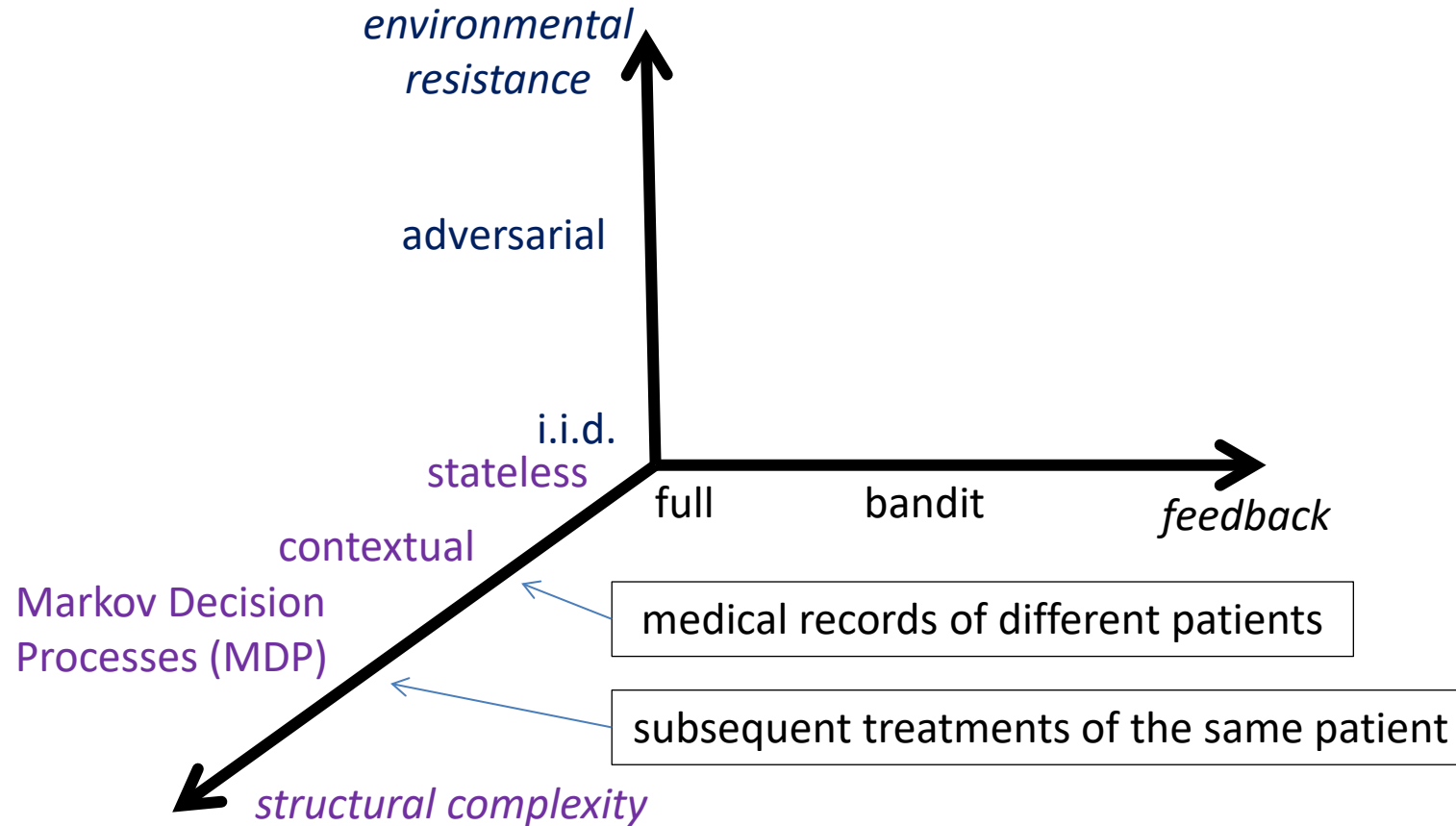
- Game theoretic setting

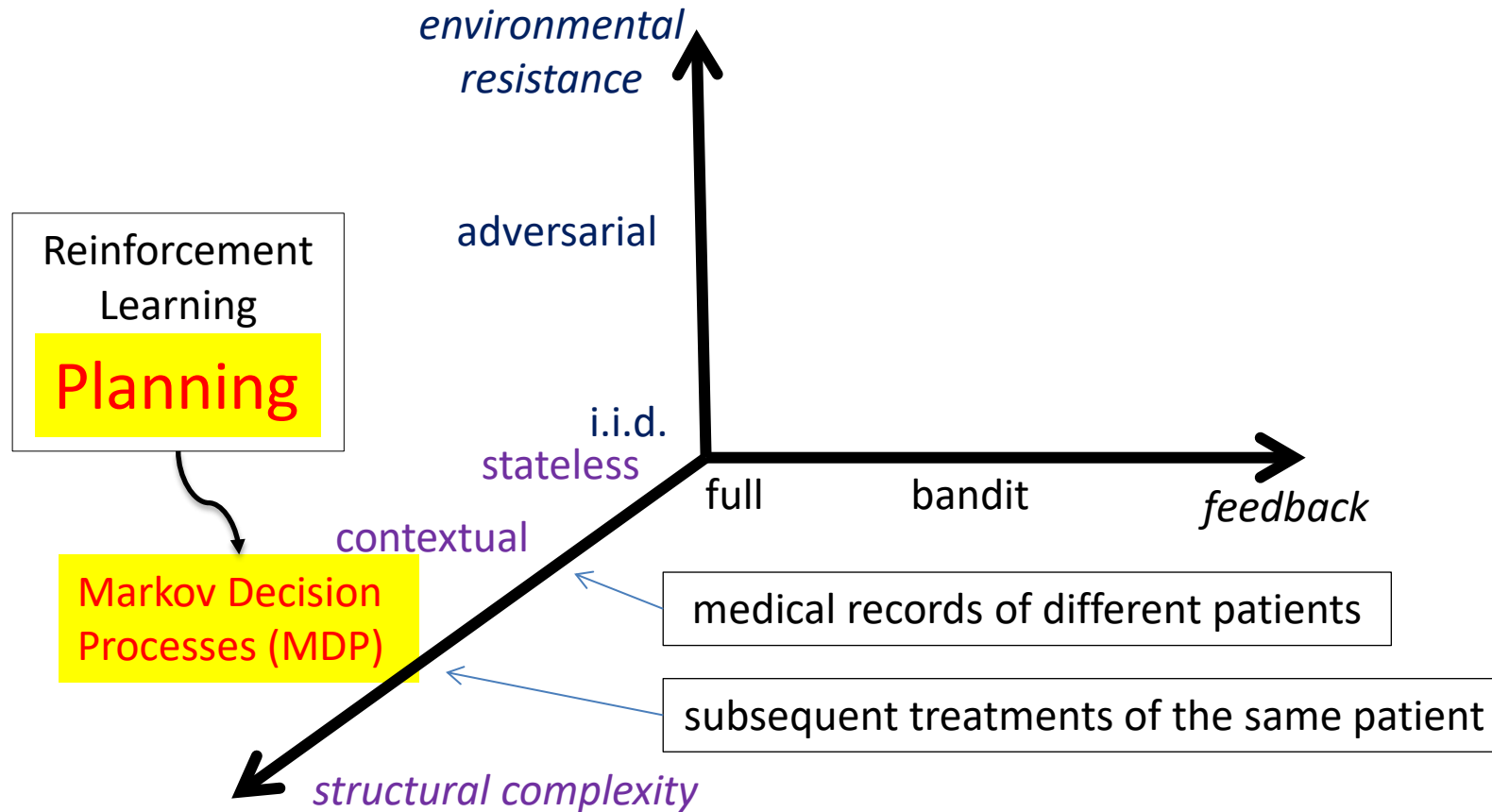- Cannot be treated in batch learning

- Evaluation measure: *regret*
  - Difference in performance compared to the best choice in hindsight (out of a limited set)
  - E.g. investment revenue vs. the best stock in hindsight

The Space of Online Learning Problems

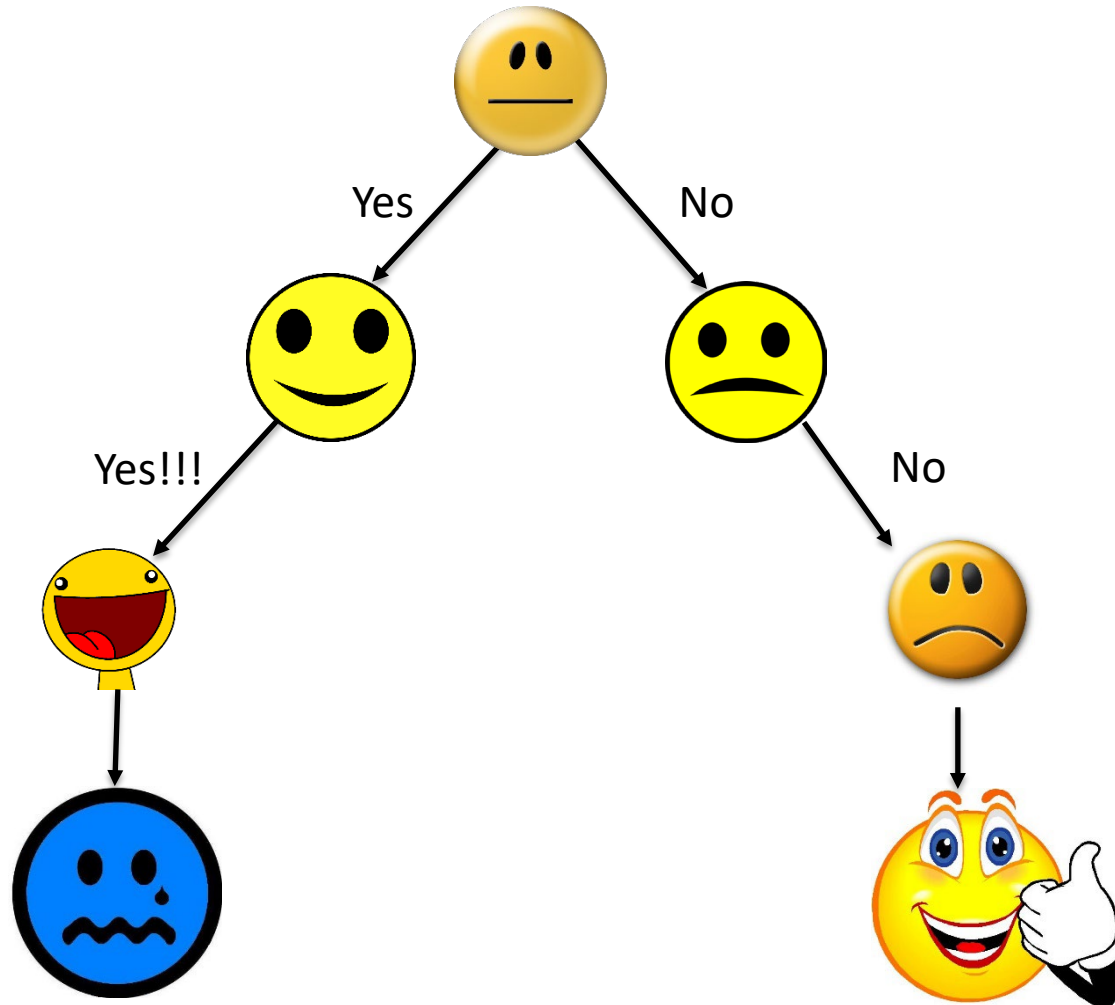# The Space of Online Learning Problems
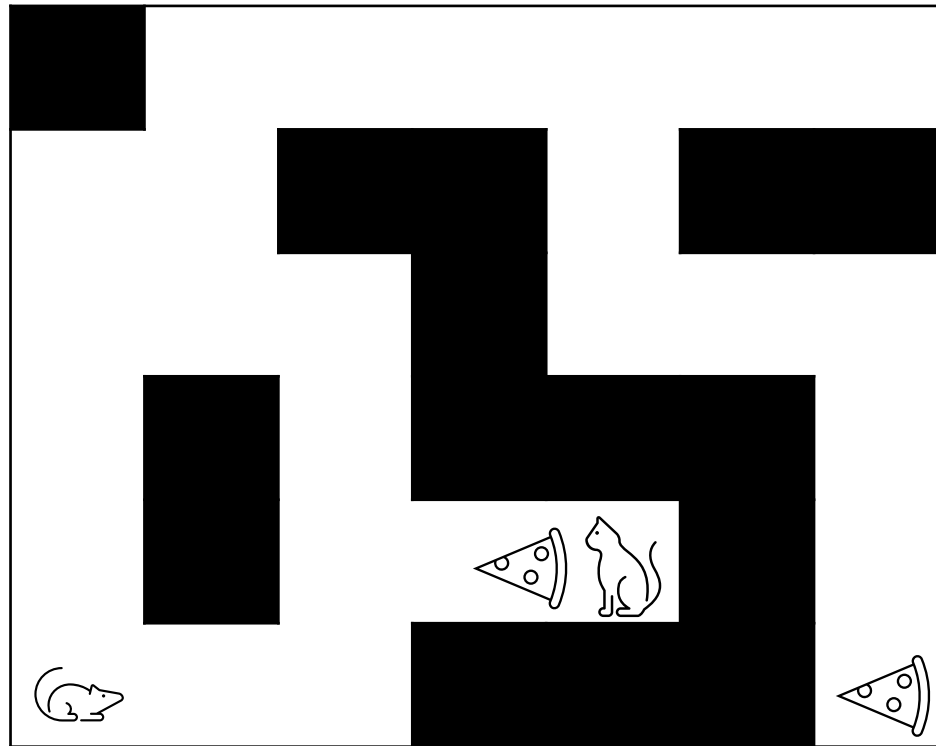
Planning

Would you like to have a drink?

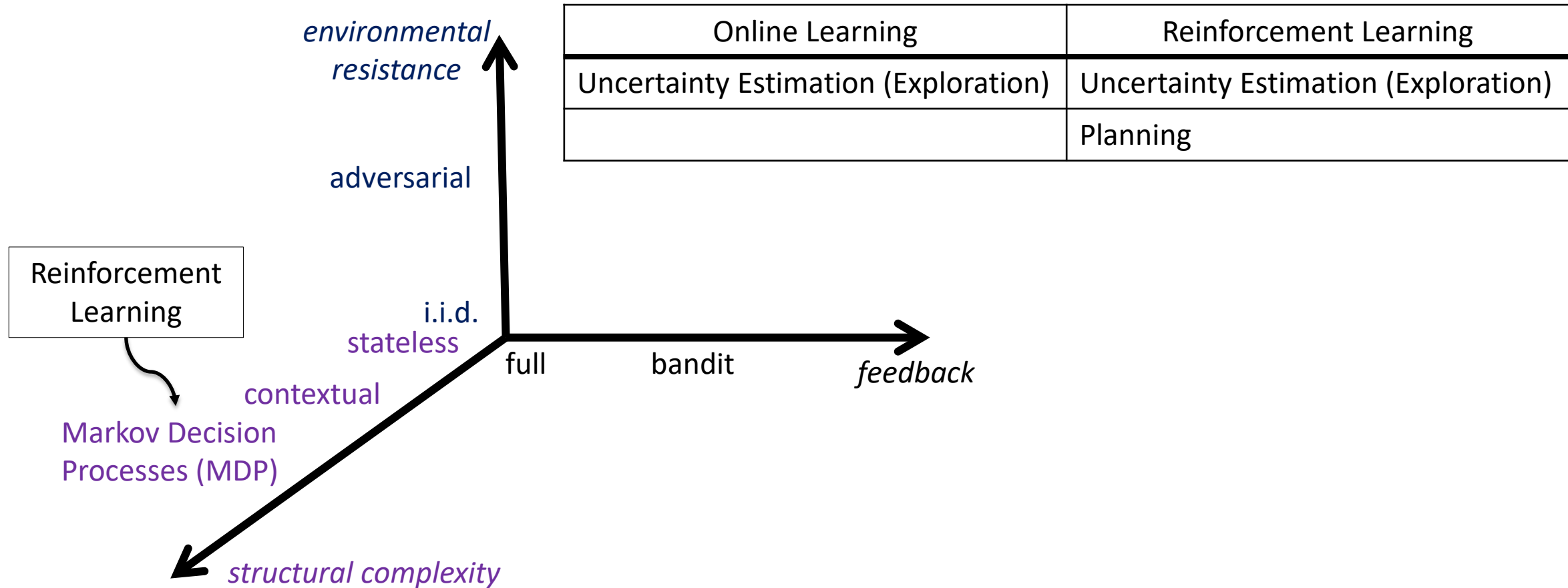Would you like to have another drink?

The next morning....

Yes

No

Yes!!!

No

Planning

- Even if the immediate outcomes are known, long-term goals require planning

The Space of Online Learning Problems

Online Learning	Reinforcement Learning
Uncertainty Estimation (Exploration)	Uncertainty Estimation (Exploration)
	Planning

environmental resistance

adversarial

Reinforcement Learning

i.i.d.
stateless

full   bandit   *feedback*

contextual

Markov Decision Processes (MDP)

structural complexity

The Space of Online Learning Problems