An Analytical Study of Utility Functions in Multi-Objective Reinforcement Learning

Manel Rodríguez-Soto,[1] Juan Antonio Rodríguez-Aguilar,[1] and Maite López-Sánchez,[2]

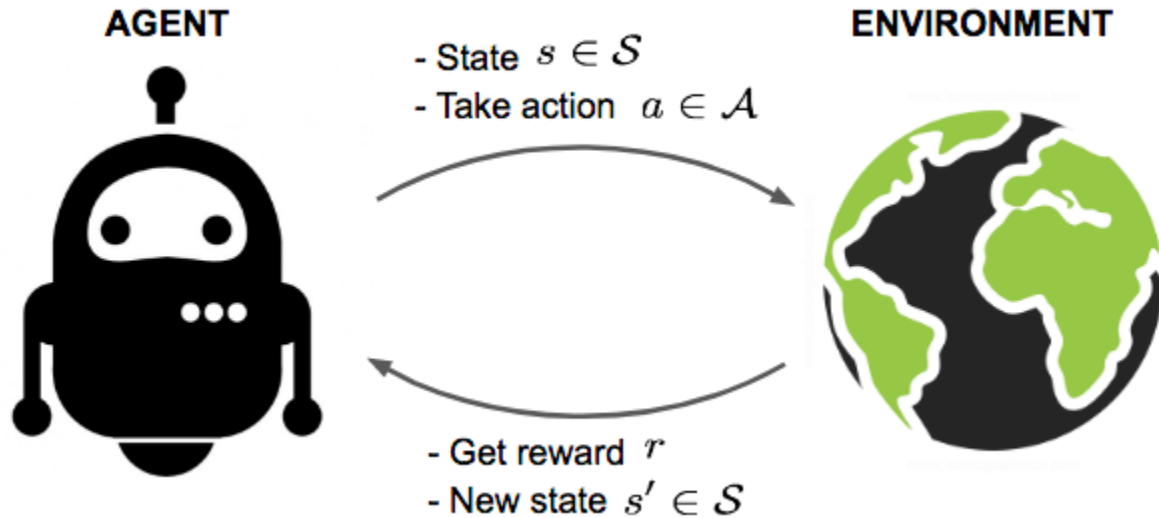[1] *Artificial Intelligence Research Institute (IIIA-CSIC).* [2] *Universitat de Barcelona (UB).*

Contact: manel.rodriguez@iiia.csic.es

Motivation

Multi-Objective reinforcement learning (MORL) is a *promising* research field.

But there are still *crucial theoretical analises missing*…

MORL background

- State $s \in \mathcal{S}$
- Take action $a \in \mathcal{A}$

- Get reward r
- New state $s' \in \mathcal{S}$

AGENT

ENVIRONMENT

**Multi-Objective
Markov Decision Process**

MORL background

AGENT

ENVIRONMENT

- State $s \in \mathcal{S}$
- Take action $a \in \mathcal{A}$

- Get reward r
- New state $s' \in \mathcal{S}$
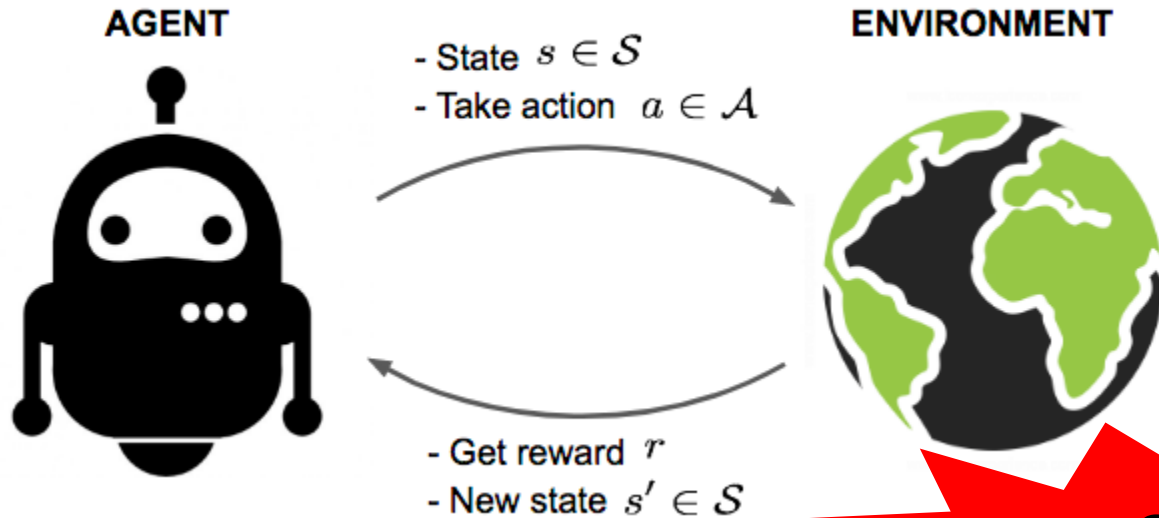
**Multi-Objective
Markov Decision Process**

REWARDS ARE VECTORIAL

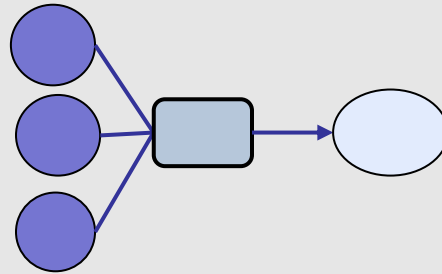# MORL background

Policy

The agent's behaviour

MORL background

Policy

The agent's behaviour
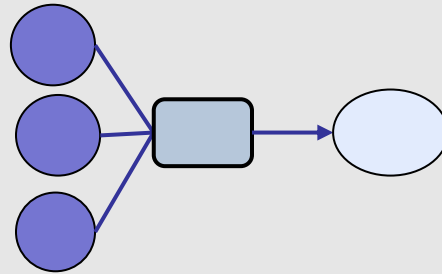
Utility function

Compacts all objectives into one

MORL background

Policy

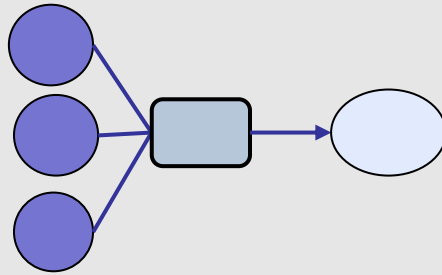The agent's behaviour

Utility function

Compacts all objectives into one

MORL background

Policy

The agent's behaviour

Utility function

Compacts all objectives into one

Preference relation

Expresses the preferred outcomes

Research questions

- Which **utility functions** have **policies** optimising them?

Research questions

- Which **utility functions** have **policies** optimising them?
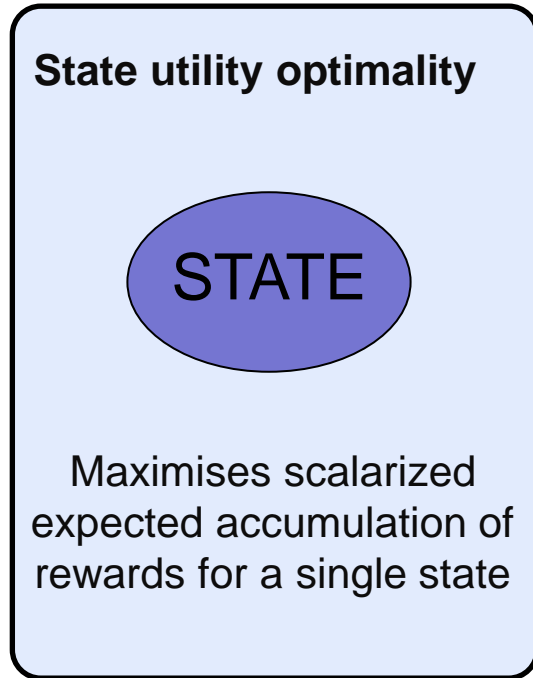
- Which **preferences** can be expressed as **utility functions**?
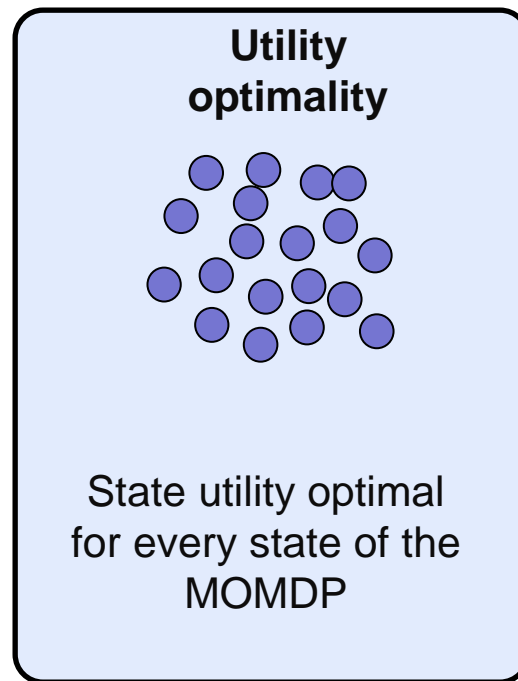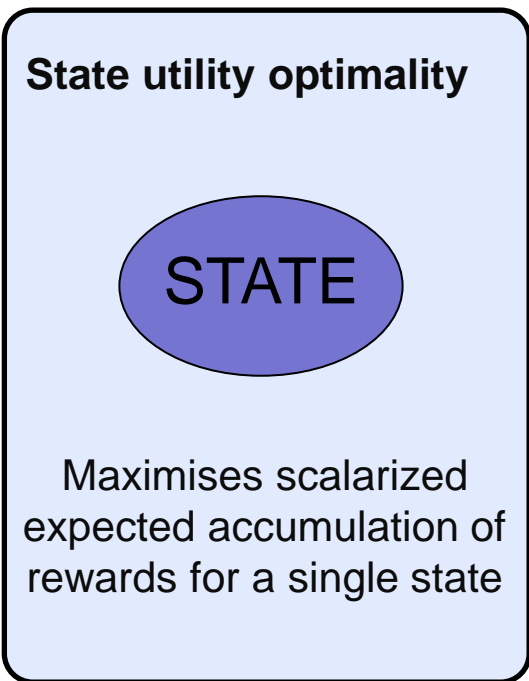
Research questions

- Which **utility functions** have **policies** optimising them?

- Which **preferences** can be expressed as **utility functions**?

Our novel MORL definitions

State utility optimality

STATE

Maximises scalarized expected accumulation of rewards for a single state

Our novel MORL definitions

State utility optimality

STATE

Maximises scalarized expected accumulation of rewards for a single state

Utility optimality

State utility optimal for every state of the MOMDP

Our theoretical results

Stationary state utility optimal policies are guaranteed to exist for…

- **Continuous utility functions**

Our theoretical results

Stationary state utility optimal policies are guaranteed to exist for…

- **Continuous utility functions**

Stationary utility optimal policies are guaranteed to exist for…

- **Composition of affine utility functions and strictly monotonically increasing functions**

Our theoretical results

Stationary state utility optimal policies are guaranteed to exist for…

- **Continuous utility functions**

Stationary utility optimal policies are guaranteed to exist for…

- **Composition of affine utility functions and strictly monotonically increasing functions**

In conclusion…

- **It may happen that the policy your algorithm learns is only state utility optimal at the initial state, but not at the rest!!!**

Take-home message

When using MORL, be wary of whether your algorithm is really learning **utility optimal** policies. Otherwise, you may learn a **contradictory policy**!

Take-home message

When using MORL, be wary of whether your algorithm is really learning **utility optimal** policies. Otherwise, you may learn a **contradictory policy**!

Contact:
manel.rodriguez@iiia.csic.es