

Multi-Objective Multi-Agent Learning: Evolutionary and Reinforcement Learning Perspectives

Nice to meet you

Gaurav Dixit (Oregon State University)

Roxana Rădulescu (Utrecht University)

Patrick Mannion (University of Galway)

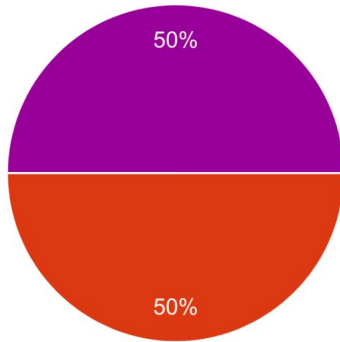


ECAI Tutorial, Santiago de Compostela, 2024

<https://moma-ecai24.github.io/>

Multi-Objective Multi-Agent Learning: Evolutionary and Reinforcement Learning Perspectives

What is your (primary) background discipline
8 responses



- Game Theory & Mechanism Design
- Reinforcement Learning
- Negotiation
- Social Choice
- Planning & Path Finding
- Human-AI interaction
- Robotics
- Social Networks

▲ 1/2 ▼

Nice to meet you



Hi, I'm Gaurav!

- Postdoc at Oregon State University and the AI Caring Institute
- Focus on cooperative decision making for asymmetric agents (agents with distinct capabilities and objectives)
- Research interests: multiagent systems, evolutionary and reinforcement learning, multi-objective decision making



Robust Multiagent Coordination



Socially Complex Situations and Care Coordination

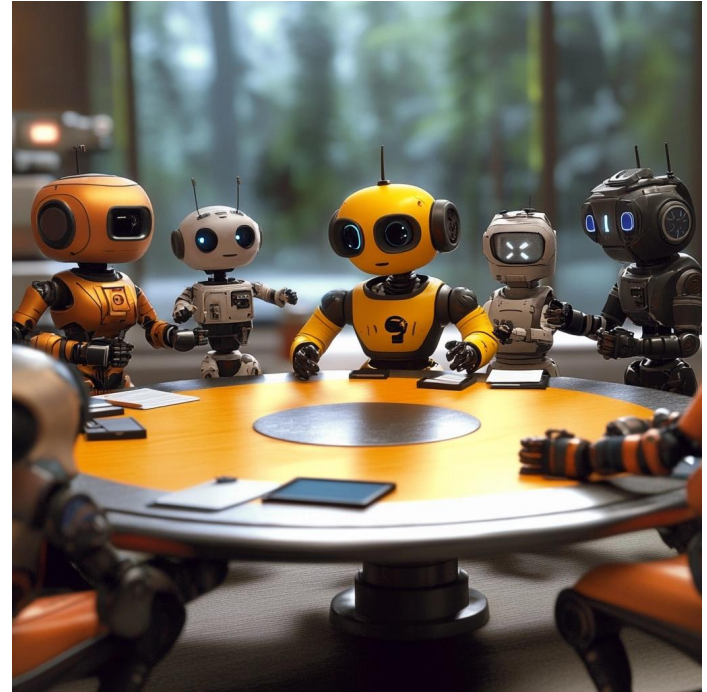



Ethics and Trust

<https://gdixit.com>

Hi, I'm Roxana!

- Assistant professor at Utrecht University, Netherlands
- Develop **multi-agent decision making systems** where **each agent** is driven by **different objectives** and goals, under the paradigm of **multi-objective multi-agent reinforcement learning**
- Keywords: multi-objective reinforcement learning, multi-agent systems, multi-objective game theory



 @rox_teo

<http://roxanaradulescu.com>

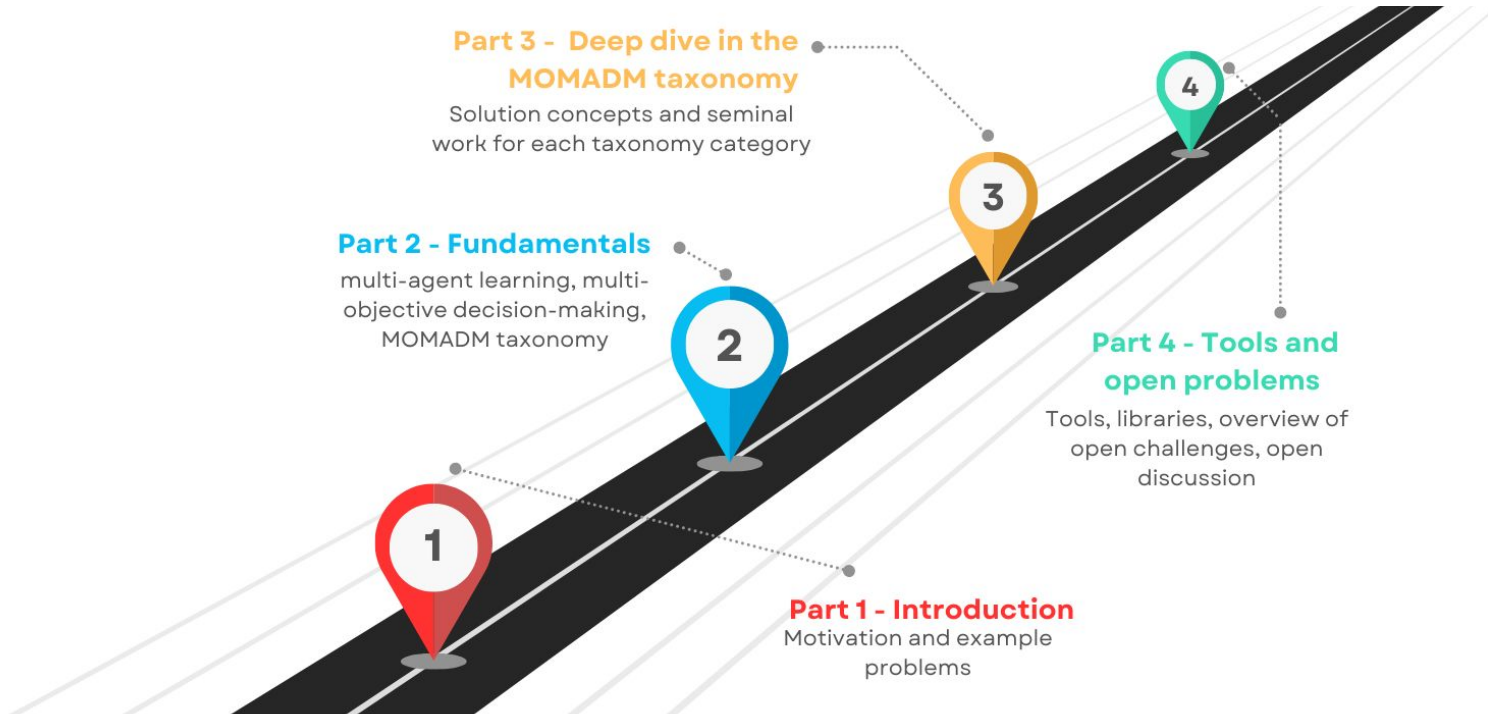
Hi, I'm Patrick!

- Lecturer Above the Bar (Assistant Professor) at University of Galway
- My view - many real world problems (like traffic!) are fundamentally **multi-objective** and **multi-agent**
- Research interests: multi-objective decision making, reinforcement learning, multi-agent systems, game theory, optimisation



The screenshot shows the top of a news article on the Connacht Tribune website. The header features the site's name 'CONNACHT TRIBUNE' in white on a dark blue background, with a search icon on the left and a menu icon on the right. Below the header, there are navigation buttons for 'Home' and 'City Tribune'. The main article title is 'Galway is seventh-worst city in Europe for car traffic congestion'. To the right of the title is a small circular profile picture of the author, Stephen Corrigan, and text indicating 'Author: Stephen Corrigan' and '~ 1 minutes read'. Below the title, it says 'Published: 27 January 2023' and 'From this week's Galway City Tribune'. The article's main image shows a long line of cars stuck in traffic on a city street, with a 'Galway City Tribune' logo overlaid in the bottom right corner.

Tutorial Roadmap



Part 1 - Introduction

Why multiple objectives? (Patrick)

Because life is not simple

- What are your objectives for your current research project?
 - Publishing asap?
 - Quality of conference/journal?
 - Collaboration potential?
 - Flag-posting?
 - Increasing funding potential?
 - Finishing your PhD?



Because life is not simple

- What are your objectives for your current research project?
 - Publishing asap?
 - Quality of conference/journal?
 - Collaboration potential?
 - Flag-posting?
 - Increasing funding potential?
 - Finishing your PhD?
- How about your co-authors?



Scalar Reward Design Process

- Design/tweak scalar reward function
- (Re-)Train RL agent using new/updated reward function (may take hours or days)
- Evaluate the outcome (and try to figure out what went wrong!)

Scalar Reward Design Process

- Design/tweak scalar reward function
- (Re-)Train RL agent using new/updated reward function (may take hours or days)
- Evaluate the outcome (and try to figure out what went wrong!)
- Repeat until the desired agent behaviour is (finally) learned

- **Wasteful and time consuming** process - each trained agent must be discarded if the reward function changes

- Designers implicitly bake in trade-offs between different behaviours
 - **Should AI engineers make the decisions about these trade-offs?**

Scalar Reward Design Process

- Design/tweak scalar reward function
- (Re-)Train RL agent using new/updated reward function (may take hours or days)
- Evaluate the outcome (and try to figure out what went wrong!)



- GT Sophy - Super Human Racing AI Agent, Sony AI
- Objectives: high precision race car control, efficient racing tactics and manoeuvres, while respecting an imprecisely defined racing etiquette
- With enough time and computation, good results can be achieved

The reward hypothesis

That all of what we mean by goals and purposes can be well thought of as the maximization of the expected value of the cumulative sum of a received scalar signal (called reward)'



Reward is enough

David Silver*, Satinder Singh, Doina Precup, Richard S. Sutton



ARTICLE INFO

Article history:

Received 12 November 2020
Received in revised form 28 April 2021
Accepted 12 May 2021
Available online 24 May 2021

Keywords:

Artificial intelligence
Artificial general intelligence
Reinforcement learning
Reward

ABSTRACT

In this article we hypothesise that intelligence, and its associated abilities, can be understood as subserving the maximisation of reward. Accordingly, reward is enough to drive behaviour that exhibits abilities studied in natural and artificial intelligence, including knowledge, learning, perception, social intelligence, language, generalisation and imitation. This is in contrast to the view that specialised problem formulations are needed for each ability, based on other signals or objectives. Furthermore, we suggest that agents that learn through trial and error experience to maximise reward could learn behaviour that exhibits most if not all of these abilities, and therefore that powerful reinforcement learning agents could constitute a solution to artificial general intelligence.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Autonomous Agents and Multi-Agent Systems (2022) 36:41
<https://doi.org/10.1007/s10458-022-09575-5>



Scalar reward is not enough: a response to Silver, Singh, Precup and Sutton (2021)

Peter Vamplew¹ · Benjamin J. Smith² · Johan Källström³ · Gabriel Ramos⁴ · Roxana Rădulescu⁵ · Diederik M. Roijers^{6,7} · Conor F. Hayes⁸ · Fredrik Heintz³ · Patrick Mannion⁸ · Pieter J. K. Libin^{6,9,10} · Richard Dazeley¹¹ · Cameron Foale¹

Accepted: 2 July 2022 / Published online: 16 July 2022
© The Author(s) 2022

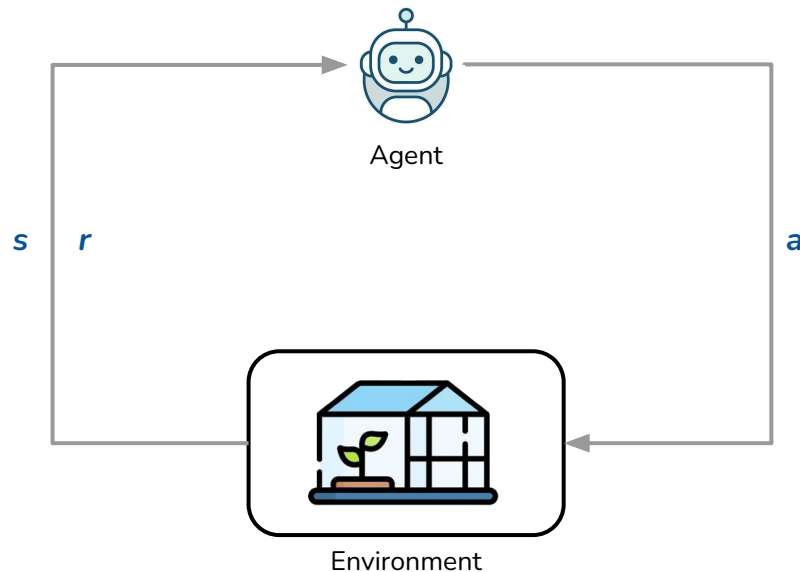
Abstract

The recent paper “Reward is Enough” by Silver, Singh, Precup and Sutton posits that the concept of reward maximisation is sufficient to underpin all intelligence, both natural and artificial, and provides a suitable basis for the creation of artificial general intelligence. We contest the underlying assumption of Silver et al. that such reward can be scalar-valued. In this paper we explain why scalar rewards are insufficient to account for some aspects of both biological and computational intelligence, and argue in favour of explicitly multi-objective models of reward maximisation. Furthermore, we contend that even if scalar reward functions can trigger intelligent behaviour in specific cases, this type of reward is insufficient for the development of human-aligned artificial general intelligence due to unacceptable risks of unsafe or unethical behaviour.

Part 2 - MOMADM Fundamentals

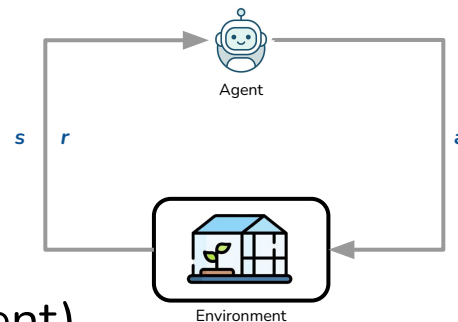
2.1 Multi-Agent Learning (Gaurav)

Markov Decision Process (MDP)



Markov Decision Process (MDP)

- MDP = $\langle S, A, T, \gamma, R \rangle$
 - Set of **s**tates
 - Set of **a**ctions
 - The **t**ransition function (dynamics of the environment)
 - A **r**eward function $R : S \times A \times S \rightarrow \mathbb{R}$



Reinforcement Learning

State

s

Action

a

Reward

r

Policy (Agent)

$\pi(a | s)$

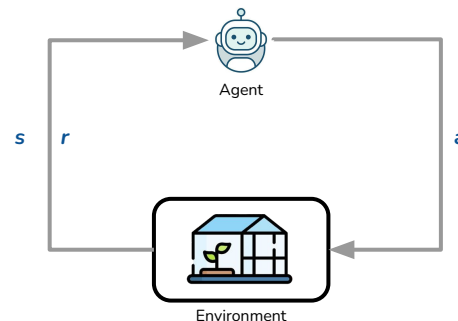
$\pi : S \times A \rightarrow [0, 1]$

Episode Length

T

Transitions

(s, a, s', r)



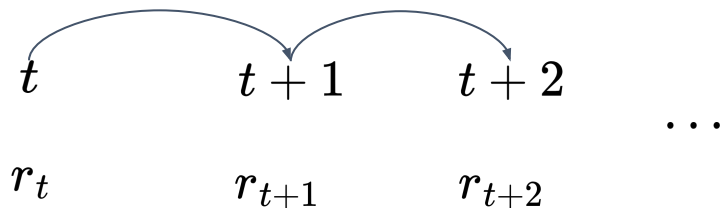
Objective

$$\max \mathbb{E} [\sum_t r_t]$$

Discount Factor



Agent



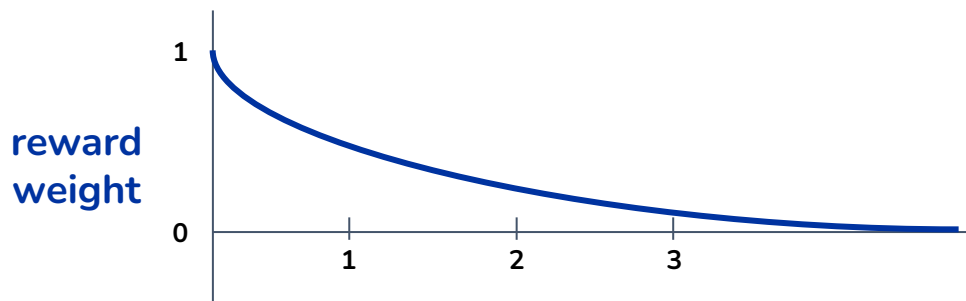
Objective

$$\max \mathbb{E} [\sum_t r_t]$$

Maximized Objective

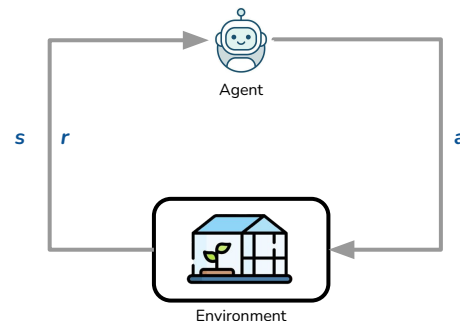
$$\max \mathbb{E} [\sum_t \gamma^t r_t]$$

Discount Factor

$$\gamma$$


Reinforcement Learning

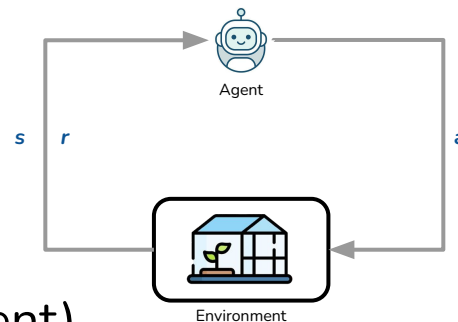
State	s
Action	a
Reward	r
Policy (Agent)	$\pi(a s) \quad \pi : S \times A \rightarrow [0, 1]$
Episode Length	T
Transitions	(s, a, s', r)
Discounted Return	$G_t = \sum_{i=t}^T \gamma^i r_{t+i}$



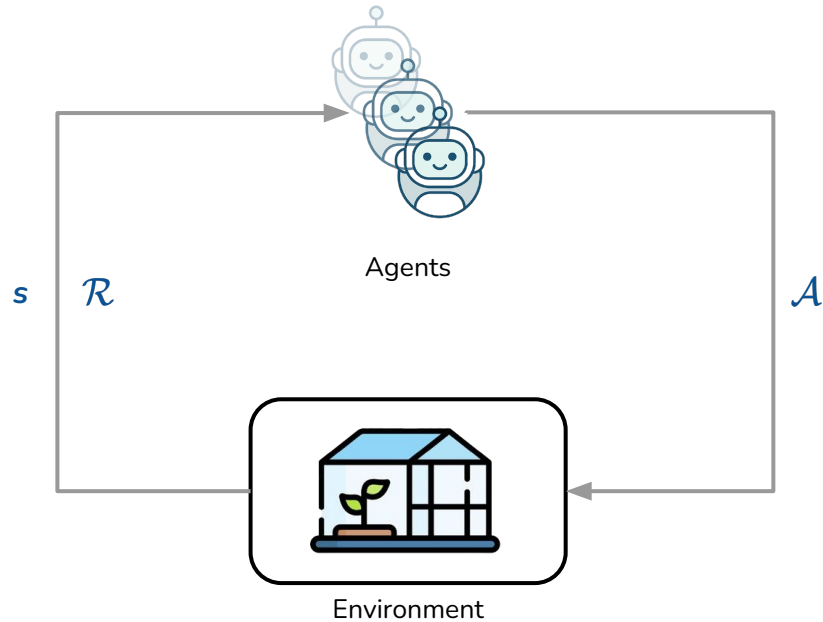
Objective
 $\max \mathbb{E} [\sum_t r_t]$

Markov Decision Process (MDP)

- MDP = $\langle \mathcal{S}, \mathcal{A}, T, \gamma, R \rangle$
 - Set of **s**tates
 - Set of **a**ctions
 - The **t**ransition function (dynamics of the environment)
 - A **r**eward function $R : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$
 - Discount factor $\gamma \in [0, 1]$



Multiagent MDP



Multiagent MDP

• MDP = $\langle S, A, T, \gamma, \mathcal{R} \rangle$

• Set of **s**tates

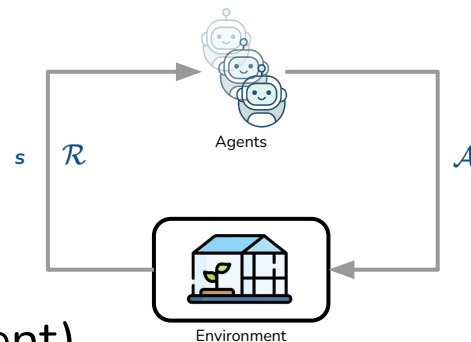
• Set of joint **a**ctions $\mathcal{A} = A_1 \times \dots \times A_n$

• The **t**ransition function (dynamics of the environment)

• **R**eward function

$$R : S \times A \times S \rightarrow \mathbb{R}$$

• Discount factor $\gamma \in [0, 1]$



Multiagent MDP: Stochastic Game

• MDP = $\langle S, A, T, \gamma, \mathcal{R} \rangle$

• Set of **s**tates

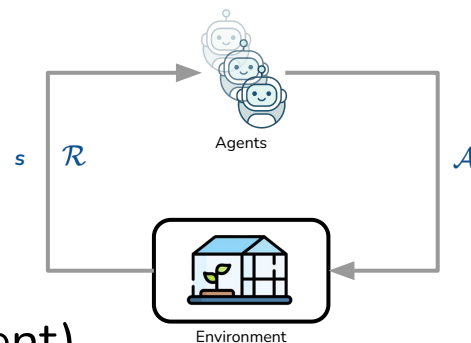
• Set of joint **a**ctions $\mathcal{A} = A_1 \times \dots \times A_n$

• The **t**ransition function (dynamics of the environment)

• **R**eward functions $\mathcal{R} = R_1 \times \dots \times R_n$

$$R_i : S \times A \times S \rightarrow \mathbb{R}$$

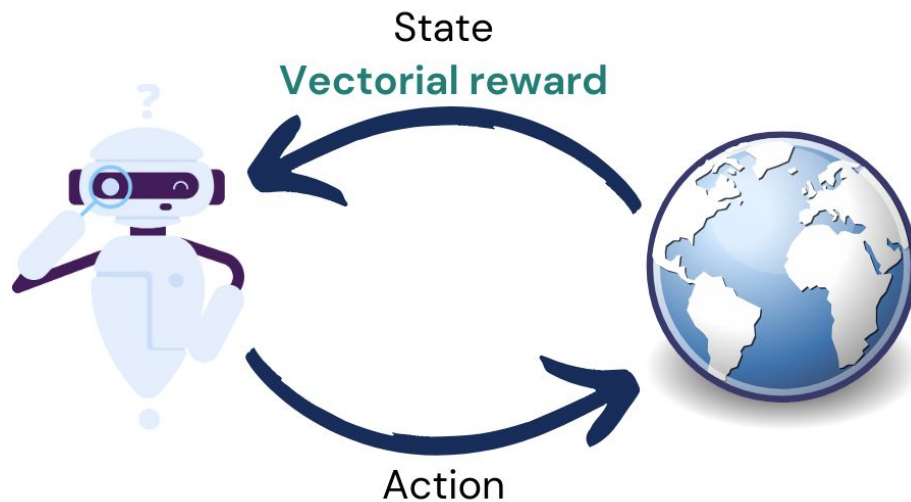
• Discount factor $\gamma \in [0, 1]$



2.2 Multi-Objective Decision-Making (Roxana)

Multi-Objective Reinforcement Learning

- Vector-valued reward function
 - $r = [r_{\text{objective1}}, r_{\text{objective2}}, \dots]$
- Length of the reward vector = number of objectives



Multi-Objective Reinforcement Learning

- Multi-Objective MDP $\langle S, A, T, \gamma, \mathbf{R} \rangle$
 - Set of states
 - Set of actions
 - A vectorial reward function $\mathbf{R}: S \times A \times S \rightarrow \mathbb{R}^d$
 $d \geq 2$ objectives
 - Transition function (dynamics of the environment)
 - Discount factor $\gamma \in [0, 1]$

Value Functions and Policies

The agent behaves according to a policy:

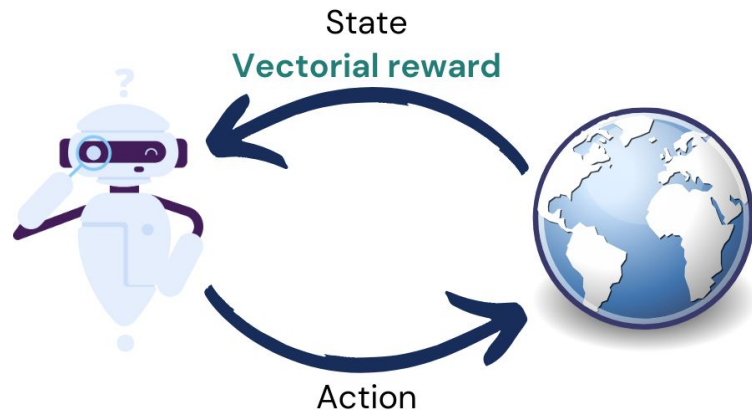
$$\pi : S \times A \rightarrow [0, 1]$$

The value function of a policy in a MOMDP:

$$\mathbf{V}^\pi = \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k \mathbf{r}_{k+1} \mid \pi, \mu \right]$$

where

$$\mathbf{r}_{k+1} = \mathbf{R}(s_k, a_k, s_{k+1})$$



Value Functions and Policies

- Vectorial value functions now supply only a partial ordering, even for a given state:

$$V_i^\pi(s) > V_i^{\pi'}(s) \text{ but } V_j^\pi(s) < V_j^{\pi'}(s)$$

- We can no longer determine which values are optimal without additional information about how to prioritize the objectives

Utility Functions

A utility function, \mathcal{U} , is used to represent a user's preferences over objectives

Utility function maps a vector reward to a scalar utility:

$$u : \mathbb{R}^d \rightarrow \mathbb{R}$$

\mathcal{U} is generally assumed to be monotonically increasing:

$$\left(\forall o, V_o^\pi > V_o^{\pi'} \right) \implies u(\mathbf{V}^\pi) \geq u(\mathbf{V}^{\pi'})$$

Utility Functions - Examples

- Linear utility function:

$$u(\mathbf{V}^\pi) = \mathbf{w}^\top \mathbf{V}^\pi$$

- Each element \mathbf{w} specifies how much one unit of value for the corresponding objective contributes to the scalarised value
- The elements of the weight vector are all positive real numbers and sum to 1

Utility Functions - Examples

- The **product utility function** - seeks to make the objective values as balanced as possible

$$u(\mathbf{V}^\pi) = \prod_{o=1}^d V_o^\pi$$

2.3 Solution Sets

Solution Sets

- The utility function is often **unknown**
- In multi-objective settings there can now be multiple possibly optimal value vectors \mathbf{V}
- We need to reason about **sets of possibly optimal value vectors and policies** when thinking about solutions to MORL problems

Undominated Set

- The most general set of solutions: the undominated set
- The undominated set, U , is the subset of all possible policies Π for which there exists a possible utility function u with a maximal scalarised value:

$$U(\Pi) = \left\{ \pi \in \Pi \mid \exists u, \forall \pi' \in \Pi : u(\mathbf{V}^\pi) \geq u(\mathbf{V}^{\pi'}) \right\}$$

Pareto Coverage Set

If the utility function u is any monotonically increasing function, then the **Pareto Coverage Set (PCS)** coincides with the undominated set:

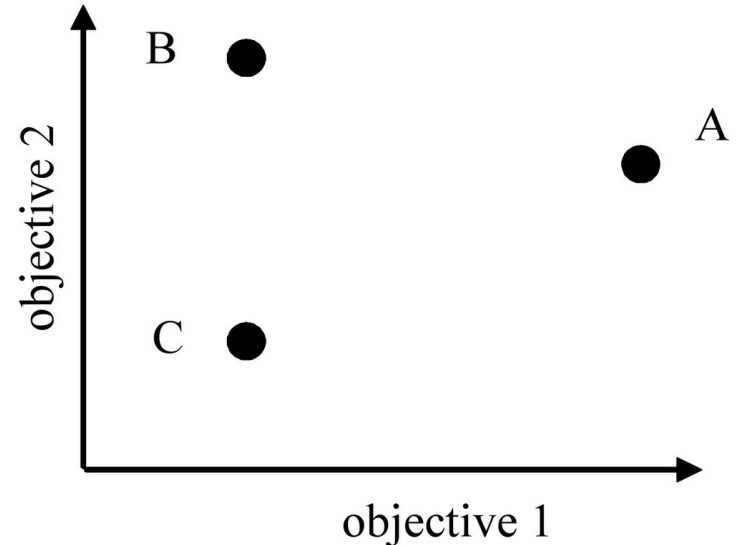
$$PCS(\Pi) = \{\pi \in \Pi \mid \nexists \pi' \in \Pi : \mathbf{V}^{\pi'} \succ_P \mathbf{V}^{\pi}\}$$

where \succ_P is the Pareto dominance relationship:

$$\mathbf{V}^{\pi} \succ_P \mathbf{V}^{\pi'} \iff (\forall i : \mathbf{V}_i^{\pi} \geq \mathbf{V}_i^{\pi'}) \wedge (\exists i : \mathbf{V}_i^{\pi} > \mathbf{V}_i^{\pi'})$$

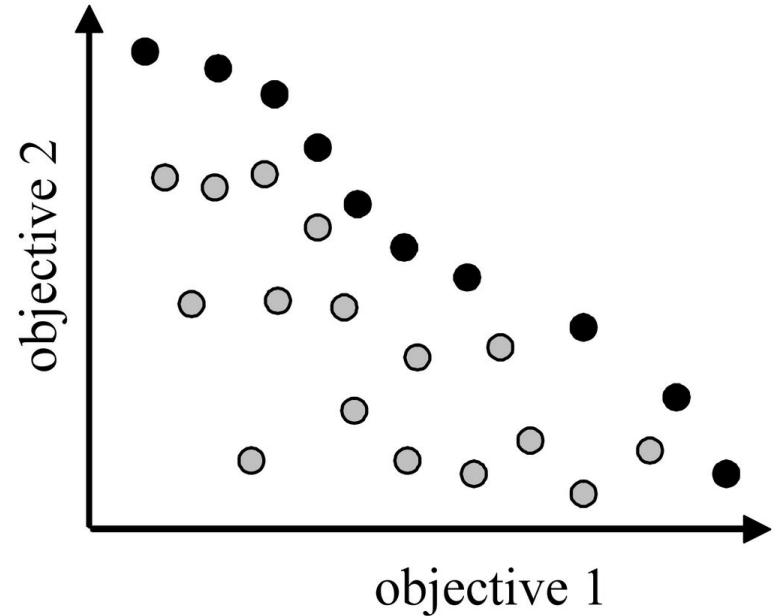
Pareto Front

- The Pareto front (PF) is the set containing the value functions corresponding to the PCS policies
- Pareto dominance illustration, maximising objectives
- Solution A **strongly dominates** solution C
- Solution B **weakly dominates** solution C
- A and B are **incomparable**



Pareto Front

- Black points indicate solutions which form the Pareto front
- Grey solutions are dominated by at least one member of the Pareto front

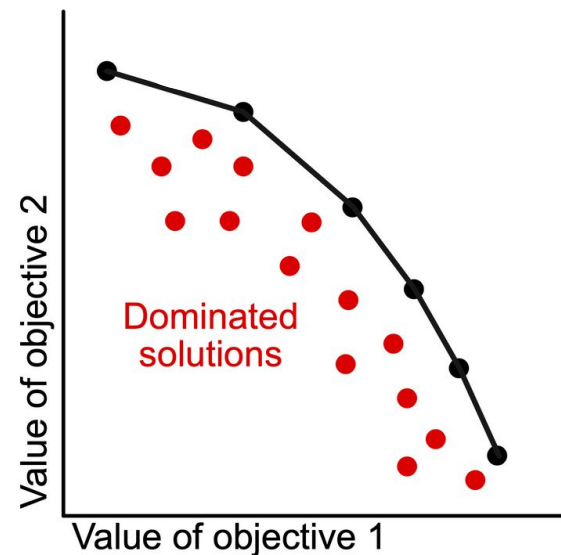


Convex Coverage Set (Convex Hull)

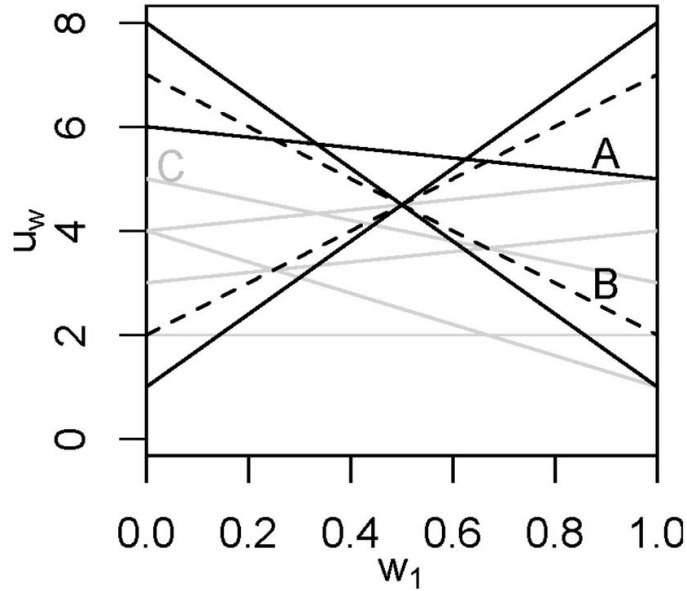
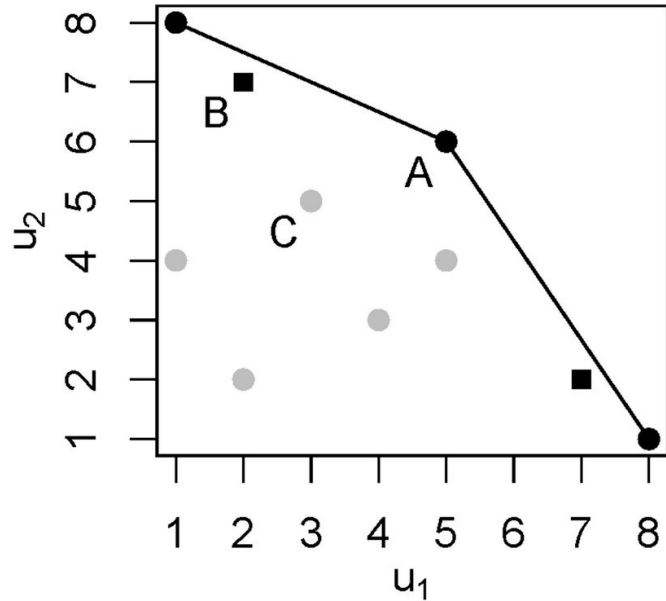
- The convex coverage set is the undominated set for non-decreasing linear utility functions

$$CCS(\Pi) = \{ \pi \in \Pi \mid \exists \mathbf{w}, \forall \pi' \in \Pi : \mathbf{w}^\top \mathbf{V}^\pi \geq \mathbf{w}^\top \mathbf{V}^{\pi'} \}$$

- The Convex Hull (CH) contains the value functions corresponding to the CCS policies

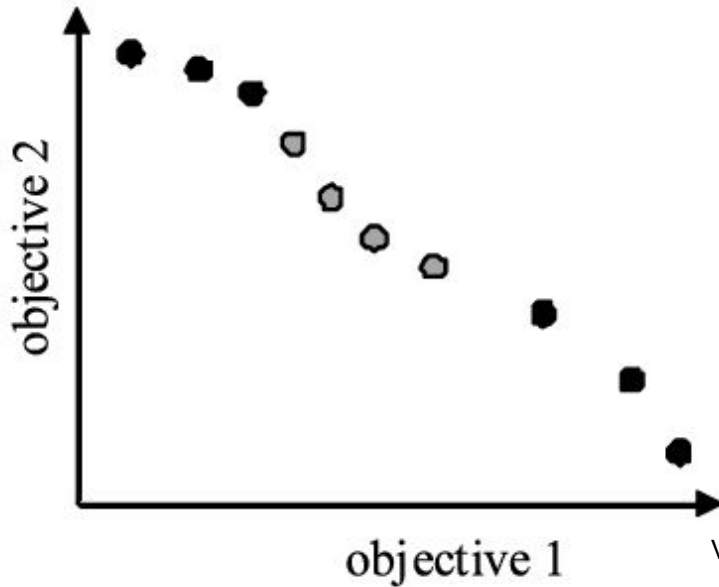


Convex Hull versus Pareto Front



● CH
● & ■ PF

Limitations of linear utilities



- Pareto front containing a concave region, indicated by the grey points
- Fundamental limitation of linear scalarisation: it cannot find policies which lie in non-convex regions of the Pareto front

Vamplew, P., Dazeley, R., Berry, A., Issabekov, R., & Dekker, E. (2011). Empirical evaluation methods for multiobjective reinforcement learning algorithms. *Machine learning*, 84, 51-80.

Axiomatic approach

- Defines the optimal solution set to be the Pareto front
- **Advantage:** retrieve a solution set containing an optimal policy for any possible monotonically increasing utility function, without any need to explicitly consider the details of those potential utility functions
- **Disadvantages:**
 - the set is typically large, and may be prohibitively expensive to retrieve
 - cannot exploit existing domain knowledge (e.g., in practical settings)

Utility-based approach

- Puts the user's utility at the center of the learning process
- Solution should be derived from utility (not axiomatically assumed)



- The properties of the user's utility function may drastically alter the desired solution, and what methods are available

Utility-based approach

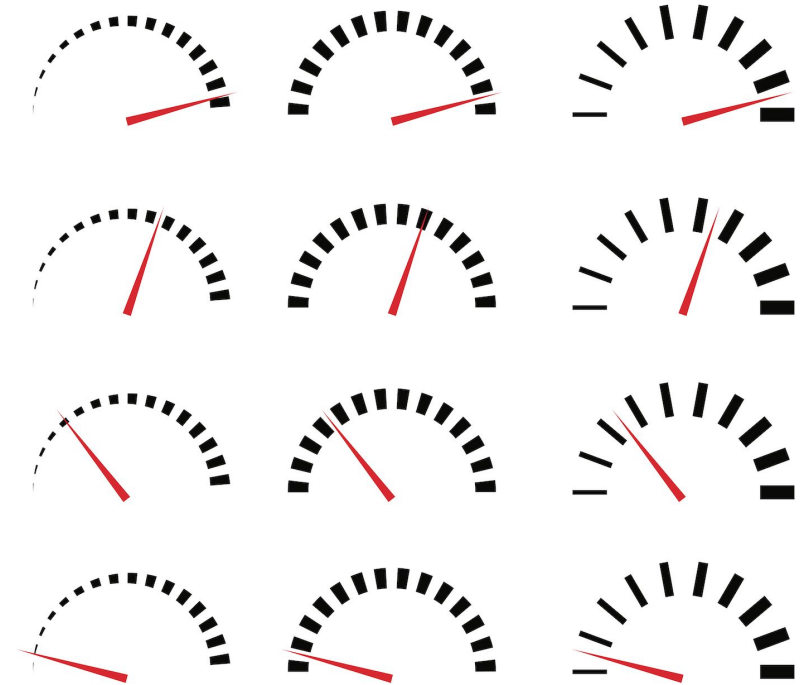


- Does not exclude or contradict the use of axiomatic methods
- When it is not possible to establish any constraints on the user's utility function, or other characteristics of the solution, prior to learning, we should still resort to axiomatic methods

Optimisation criteria

- Vectorial reward function
- Utility-based perspective

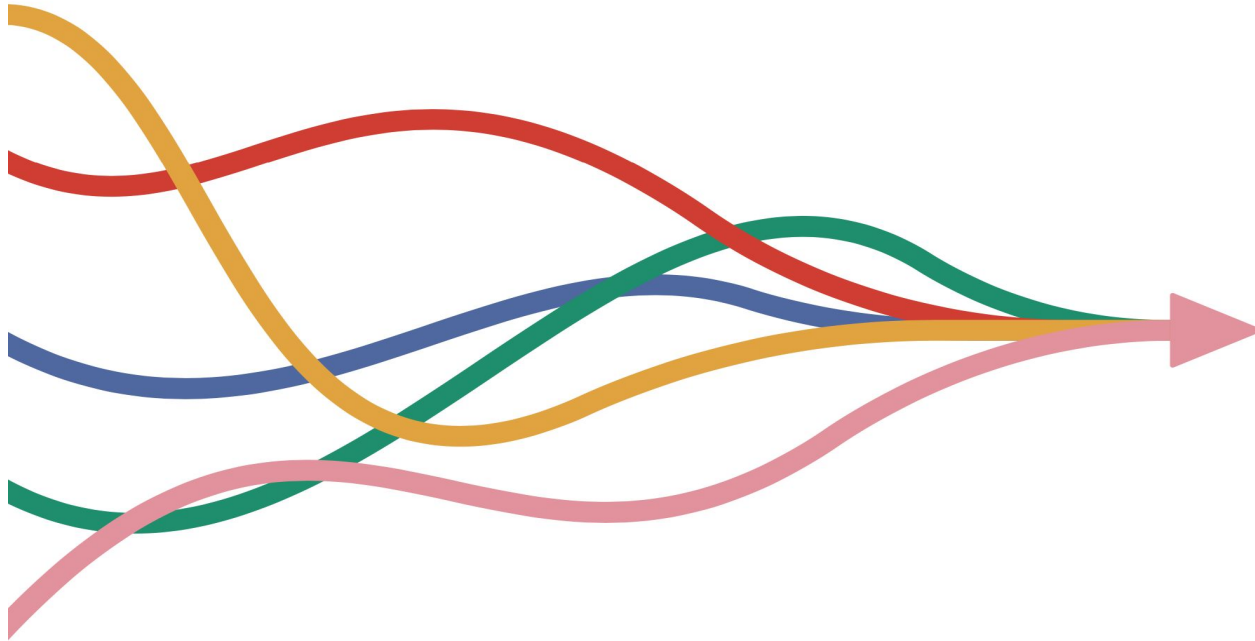
$$u_i: \mathbb{R}^d \rightarrow \mathbb{R}$$



Optimisation criteria



Optimisation criteria



Optimisation criteria



- Expected Scalarised Returns (ESR)
 - Calculate the expectation of the utility from the payoffs
 - Utility of an individual policy execution



Optimisation criteria



- Expected Scalarised Returns (ESR)
 - Calculate the expectation of the utility from the payoffs
 - Utility of an individual policy execution



- Scalarised Expected Returns (SER)
 - Calculate the utility of the expected payoff
 - Utility of the average payoff from several executions of the policy



Optimisation criteria



- Expected Scalarised Returns (ESR)

$$V_u^\pi = \mathbb{E} \left[u \left(\sum_{t=0}^{\infty} \gamma^t \mathbf{r}_t \right) \mid \pi, \mu_0 \right]$$



- Scalarised Expected Returns (SER)

$$V_u^\pi = u \left(\mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t \mathbf{r}_t \mid \pi, \mu_0 \right] \right)$$





- The utility of a user is derived from a single execution of a policy
- Understudied in the RL literature
- ESR set is defined as the set of optimal solutions



Hayes, C. F., Verstraeten, T., Roijers, D. M., Howley, E., & Mannion, P. (2022). Expected scalarised returns dominance: a new solution concept for multi-objective decision making. *Neural Computing and Applications*, 1-21.



- The utility of a user is derived from multiple executions of a policy (i.e., user is concerned about achieving an optimal utility over multiple policy executions)
- Most commonly used optimisation criterion in multi-objective RL and planning
- Coverage set is defined as a set of optimal solutions for all possible utility functions



Optimisation criteria



- Note that:
 - $SER = ESR$ under linear scalarisation
- Which criterion should be chosen for optimisation depends on how the policies are used in practice



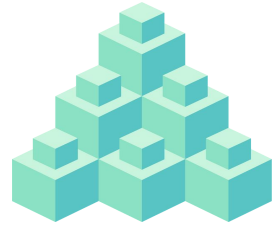
Evaluation Metrics

- Axiomatic-based evaluation metrics
 - The hypervolume metric
 - The ϵ -metric
 - Cardinality
- Utility-based evaluation metrics
 - Expected utility metric (EUM)
 - Maximal utility loss (MUL)

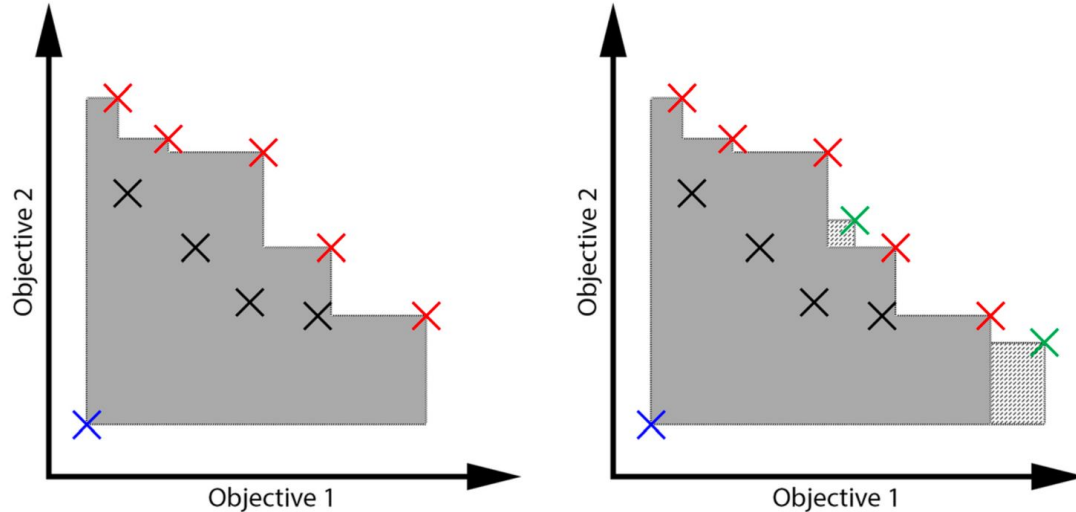


Hypervolume

- Measures the (hyper-)volume in value-space
Pareto-dominated by the set of policies in an approximate coverage set
- Correlates with (but is not equal to) the spread of a set of undominated solutions over the possible multi-objective solution space



Hypervolume



- Left: The hypervolume for a 2-objective maximisation problem. Solutions in red form the undominated set, solutions in black are said to be dominated. The shaded area denotes the hypervolume of the undominated set with respect to the reference point (shown in blue).
- Right: The effect of adding two new points (shown in green) to the undominated set

Hypervolume

- Hypervolume values are difficult to interpret
- The benefit of a certain increase or decrease in hypervolume is not readily apparent to the end user:
 - adding a non-dominated solution at the extreme ends could lead to a large increase in the hypervolume, even if this additional solution is of little interest to the end user
 - adding a new solution close to other solutions in the non-dominated set can result in a minimal increase in hypervolume, even if the new solution is valuable to the end user

Expected utility metric

- Aims to directly evaluating an agent's ability to maximize user utility
- Defined as the expected utility for a user from this solution set, under some prior distribution over user utility functions
- Under the SER optimality criterion:

$$\text{EUM} = \mathbb{E}_{P_u} \left[\max_{\pi \in \mathcal{S}} u(\mathbf{V}^\pi) \right]$$

- Useful when we care about the agent's ability to do well across many different utility functions

Maximal utility loss

- Measures the maximal loss in utility that occurs when taking a policy from a given solution set, instead of the full set of possibly optimal solutions
- Under the SER optimality criterion:

$$\text{MUL} = \max_{u \in \mathcal{U}} \left(\max_{\pi^* \in \mathcal{S}^*} u(\mathbf{V}^{\pi^*}) - \max_{\pi \in \mathcal{S}} u(\mathbf{V}^{\pi}) \right)$$

Connections to other problems - POMDPs

- If one assumes **linear utility** functions, **POMDPs are a superclass of MOMDPs**
- Intuitively, imagine there would be a "true objective" and the linear weights of the utility function would form a "belief" over what the true objective would be
- MOMDPs and POMDPs have different interpretations
- Theoretical results and methods from POMDPs can be transferred/adapted for MOMDPs (e.g., Optimistic Linear Support - approximate the CCS)

Multi-objective as multi-agent problems

- **Objectives are not agents**
 - Cast single-agent multi-objective problems as multi-agent problems, with each agent representing a competing objective
 - Either through voting rules or Nash equilibria a policy is selected

Q: What do you think is the potential issue with such approaches?

Multi-objective as multi-agent problems

- Such mechanisms offer no guarantees with respect to the user's utility and it is unclear if these "compromise solutions" represents desired trade-offs or not
- Note that objectives can be more or less important and may have non-linear interactions in the utility function
- This is different than determining trade-offs between the individual payoffs of agents

Multi-objective as multi-agent problems

- But **altruistic agents can see other agents as objectives**
 - An altruistic agent could see the utility of other agents as objectives
 - Modelling other agents as objectives enables explicitly imposing fairness between these objectives, i.e., the utilities of the agents
- Lorenz dominance ordering - a refinement of Pareto dominance, adding a predilection towards a more balanced distribution of values over the objectives

Human-aligned agents

- How to ensure that the decisions and behaviour of autonomous agents are **safe, trustworthy, aligned, interpretable, fair, and unbiased**
- Since these are additional considerations beyond maximising the agent's primary reward, there is a clear link to multi-objective approaches

2.4 Multi-objective multi-agent decision making (Gaurav)

Because life really is not simple

- What are your objectives for your current research project?
 - Publishing asap?
 - Quality of conference/journal?
 - Collaboration potential?
 - Flag-posting?
 - Increasing funding potential?
 - Finishing your PhD?
- How about your co-authors?



Life is still not simple

- What are your objectives for your current research project?
 - Publishing asap?
 - Quality of conference/journal?
 - Collaboration potential?
 - Flag-posting?
 - Increasing funding potential?
 - Finishing your PhD?
- Setting?

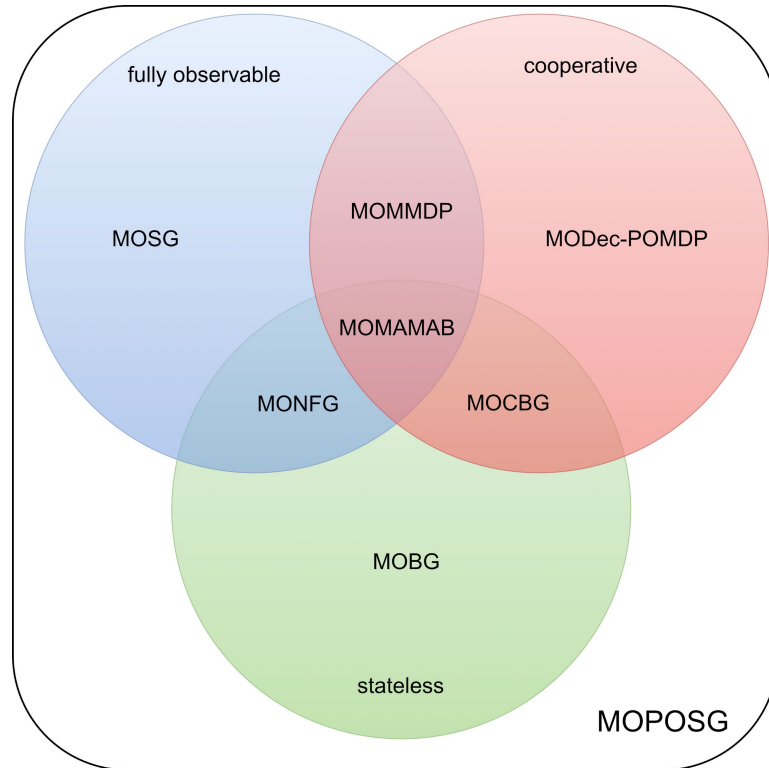


Life is still not simple at all?

- What are your objectives for your current research project?
 - Publishing asap?
 - Quality of conference/journal?
 - Collaboration potential?
 - Flag-posting?
 - Increasing funding potential?
 - Finishing your PhD?
- Truly cooperative though?

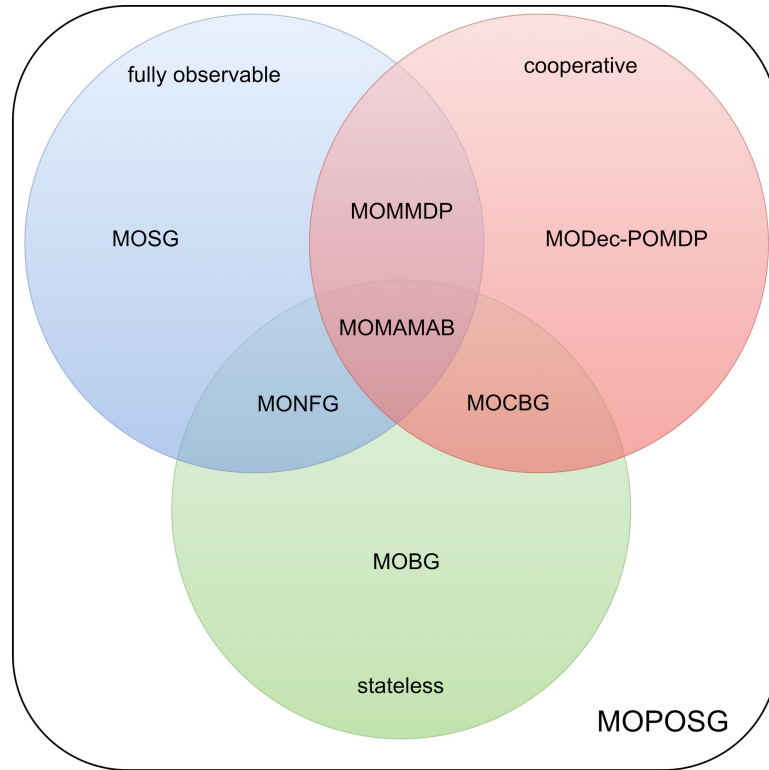


Mathematical models



Models:
On the basis of rewards (in objectives) and observations (about states).

Mathematical models



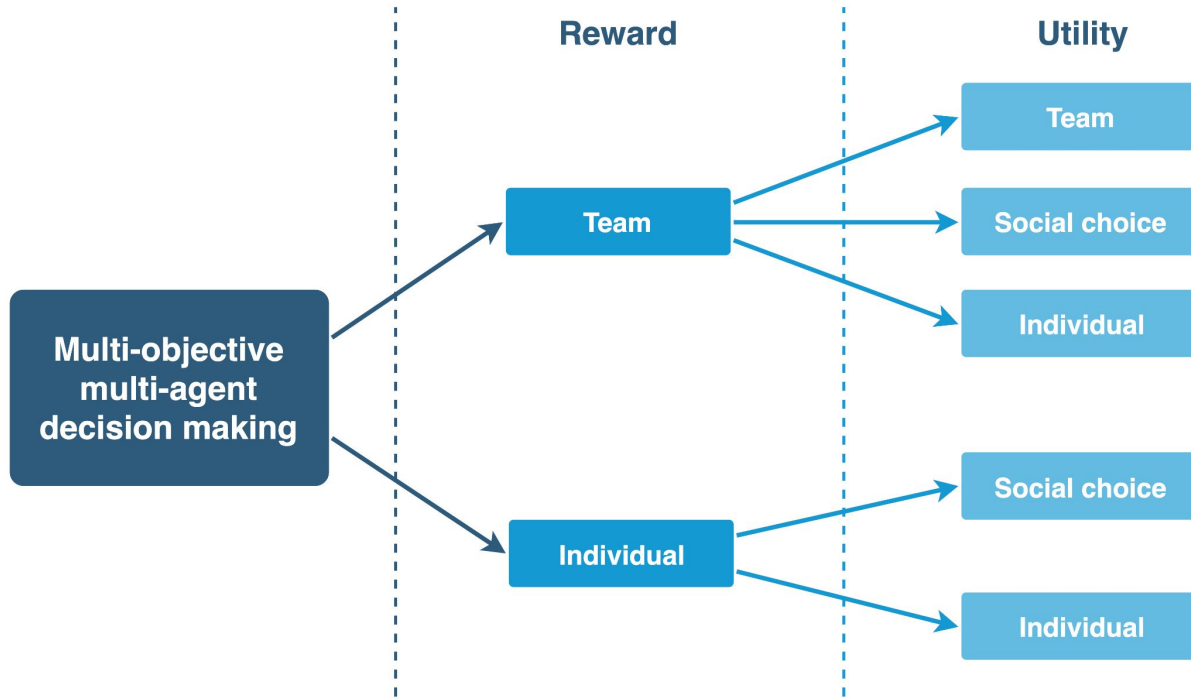
Models:

On the basis of rewards (in objectives) and observations (about states).

But utility is not yet modelled!

MOMADM taxonomy based on rewards and utilities (Patrick)

Taxonomy



Rădulescu, R., Mannion, P., Roijers, D. M., & Nowé, A. (2020). Multi-objective multi-agent decision making: a utility-based analysis and survey. *Autonomous Agents and Multi-Agent Systems*, 34(1), 1-52.

Taxonomy



Solution concepts

		UTILITY		
		TEAM	SOCIAL CHOICE	INDIVIDUAL
REWARD	TEAM	Coverage sets	Mechanism design	Coverage sets (+ Negotiation) Equilibria and stability concepts
	INDIVIDUAL		Mechanism design	Equilibria and stability concepts Coverage Sets as best responses

Rădulescu, R., Mannion, P., Roijers, D. M., & Nowé, A. (2020). Multi-objective multi-agent decision making: a utility-based analysis and survey. *Autonomous Agents and Multi-Agent Systems*, 34(1), 1-52.

Coverage sets

- Contain at least one optimal policy for each possible utility function
- **TRTU**: rewards and derived utility is shared between agents, with one utility function selected during execution
- **TRIU**: agent can (contractually) agree which policy to execute
- **IRIU**: set of possible best responses to the behaviour of other agents

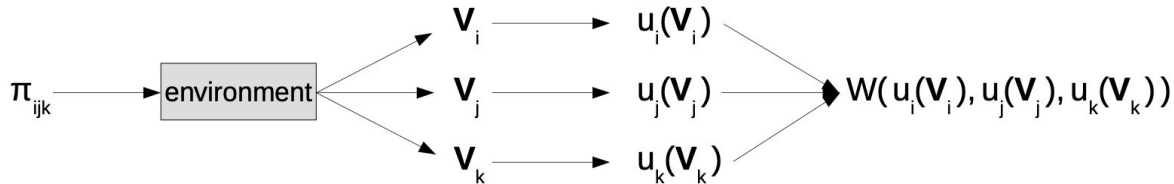
		UTILITY		
		TEAM	SOCIAL CHOICE	INDIVIDUAL
REWARD	TEAM	Coverage sets	Mechanism design	Coverage sets (+ Negotiation) Equilibria and stability concepts
	INDIVIDUAL		Mechanism design	Equilibria and stability concepts Coverage Sets as best responses

Coverage sets: negotiation

- Automated negotiation
 - Autonomous negotiating agents, representing their user's interests/preferences
 - Reach a compromise that satisfies all the involved parties
 - Pursue equity (i.e., fairness and justice)
- Baarslag, T., Kaisers, M., Gerding, E., Jonker, C. M., & Gratch, J. (2017). When will negotiation agents be able to represent us? The challenges and opportunities for autonomous negotiators. International Joint Conferences on Artificial Intelligence.
- Aydoğan, R., & Jonker, C. M. (2023). A Survey of Decision Support Mechanisms for Negotiation. In Recent Advances in Agent-Based Negotiation: Applications and Competition Challenges (pp. 30-51). Singapore: Springer Nature Singapore.

Social Welfare and Mechanism Design

- System perspective: what is a socially desirable outcome



		UTILITY		
		TEAM	SOCIAL CHOICE	INDIVIDUAL
REWARD	TEAM	Coverage sets	Mechanism design	Coverage sets (+ Negotiation) Equilibria and stability concepts
	INDIVIDUAL		Mechanism design	Equilibria and stability concepts Coverage Sets as best responses

Design a system that forces agents to be truthful about their utilities and leads to optimal solution under W

Equilibria and stability concepts

- Stable outcomes from which self-interested agents have no incentive to deviate
- Nash equilibria, correlated equilibria, cyclic equilibria, coalition formation

		UTILITY		
		TEAM	SOCIAL CHOICE	INDIVIDUAL
REWARD	TEAM	Coverage sets	Mechanism design	Coverage sets (+ Negotiation) Equilibria and stability concepts
	INDIVIDUAL		Mechanism design	Equilibria and stability concepts Coverage Sets as best responses



Nash Equilibrium

- No agent can improve their utility by unilaterally deviating from the joint strategy π^{NE}

- Nash equilibrium under SER:

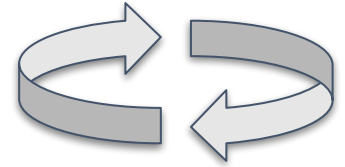
$$\mathbb{E}u_i [\mathbf{p}_i(\pi_i^{\text{NE}}, \pi_{-i}^{\text{NE}})] \geq \mathbb{E}u_i [\mathbf{p}_i(\pi_i, \pi_{-i}^{\text{NE}})]$$

- Nash equilibrium under ESR:

$$u_i [\mathbb{E}\mathbf{p}_i(\pi_i^{\text{NE}}, \pi_{-i}^{\text{NE}})] \geq u_i [\mathbb{E}\mathbf{p}_i(\pi_i, \pi_{-i}^{\text{NE}})]$$

Other solution concepts

- Cyclic Nash equilibria
 - No agent can improve their utility by unilaterally deviating from a joint cyclic strategy
- Correlated equilibria
 - No agent can improve their utility by unilaterally deviating from the recommendation of the correlated signal, given by an external mechanism



Part 3 - Deep dive in the MOMADM taxonomy

3.1 Team Reward - Team Utility



Team Reward Team Utility

Yliniemi et al. (2016). Multi-Objective Multiagent Credit Assignment in reinforcement learning and NSGA-II

Rojers et al. (2013). Multi-objective variable elimination for collaborative graphical games

Rojers et al. (2014). Linear support for multi-objective coordination graph

Brys et al. (2014). Distributed learning and multi-objectivity in traffic light control

Agrawal et al. (2015). Non-additive multi-objective robot coalition formation

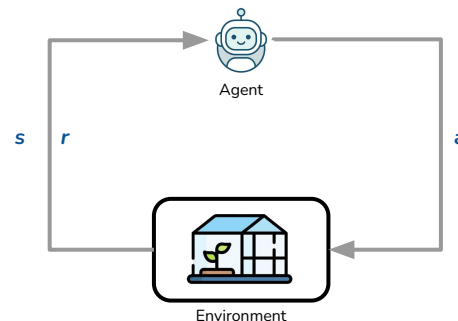
Credit assignment in EA

Joint actions with
coordination graphs

Scalarization and Coalitions
in RL + EAs

Reinforcement Learning

State	s
Action	a
Reward	r
Policy (Agent)	$\pi(a s) \quad \pi : S \times A \rightarrow [0, 1]$
Episode Length	T
Transitions	(s, a, s', r)
Discounted Return	$G_t = \sum_{i=t}^T \gamma^i r_{t+i}$

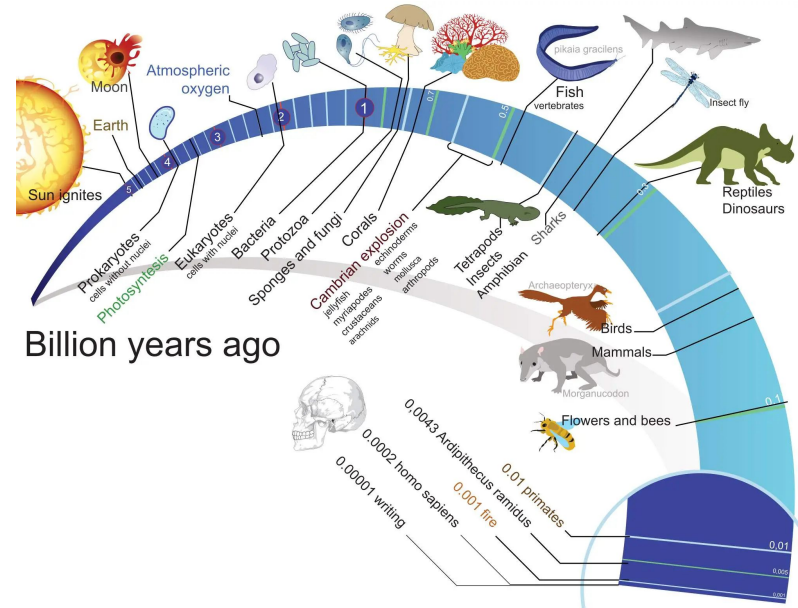


Objective
 $\max \mathbb{E} [\sum_t r_t]$

Why Evolutionary Algorithms?

How do we optimize without gradients?

- Creative diverse solutions
- Good at escaping local minima
- Can find partial solutions
- Good at combining partial solutions

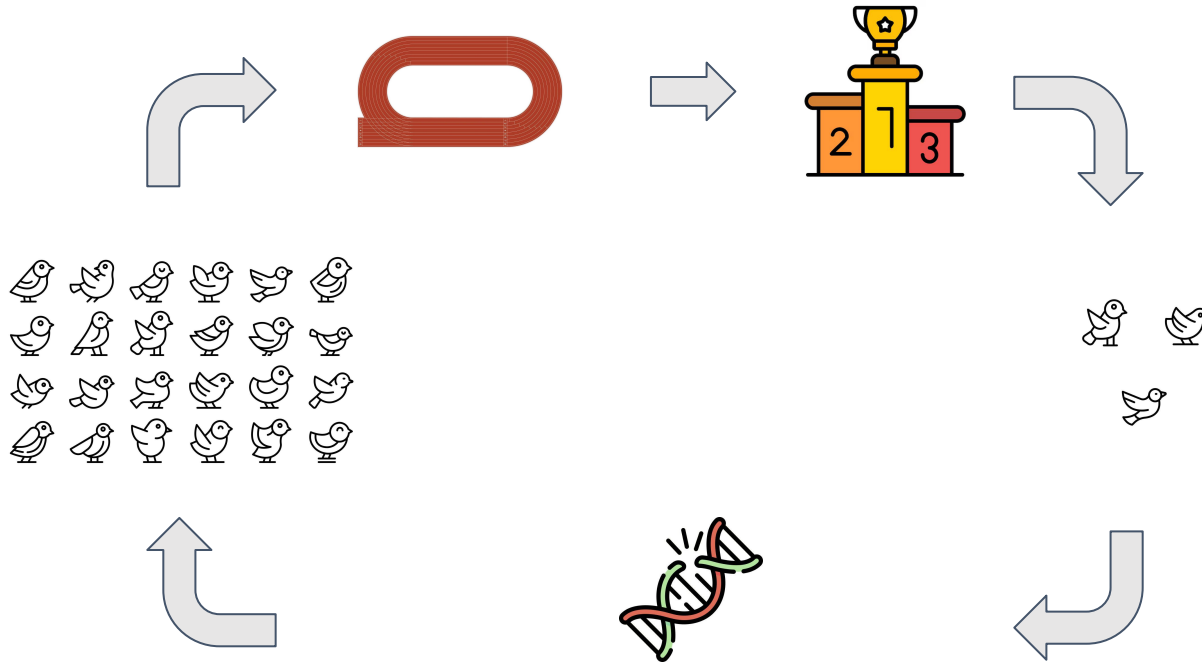


Cully et al. (2015). Robots that can adapt like animals

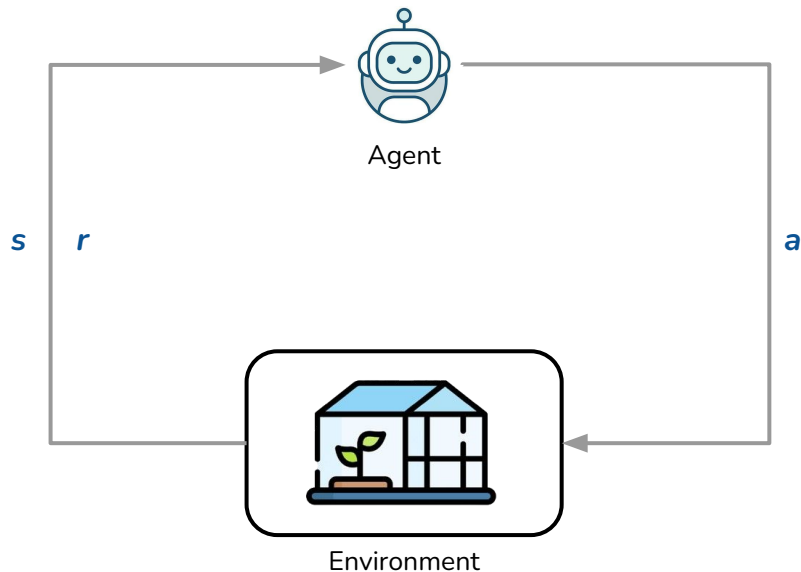
Such et al. (2017). Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning

Chatzilygeroudis et al. (2021). Quality-diversity optimization: a novel branch of stochastic optimization

Evolutionary Algorithms



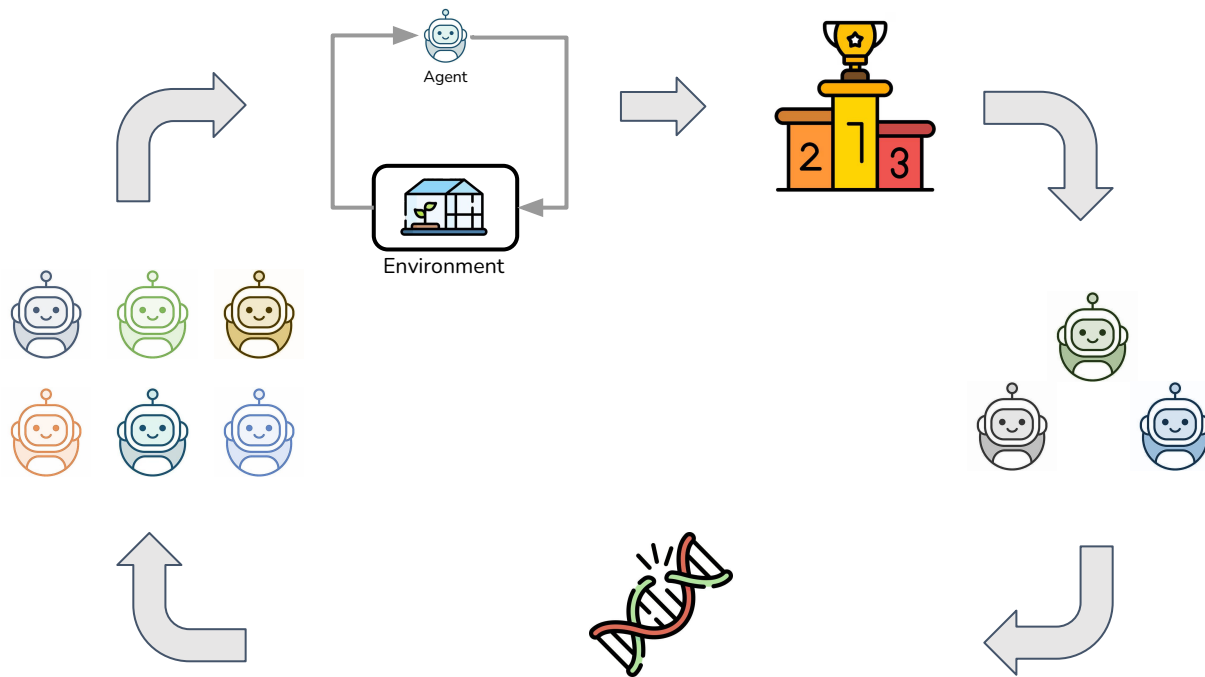
How to solve RL Problems?



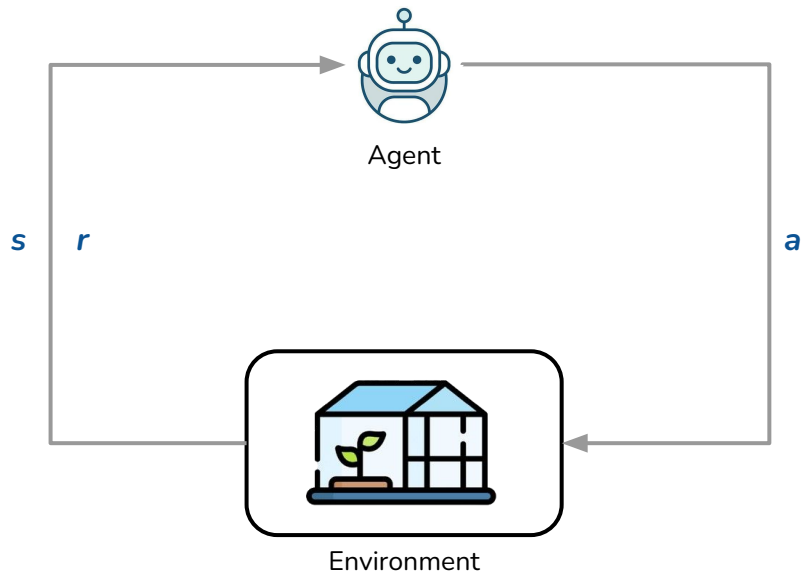
Objective

$$J = \max E [R (\tau)]$$

Evolutionary Perspective for RL

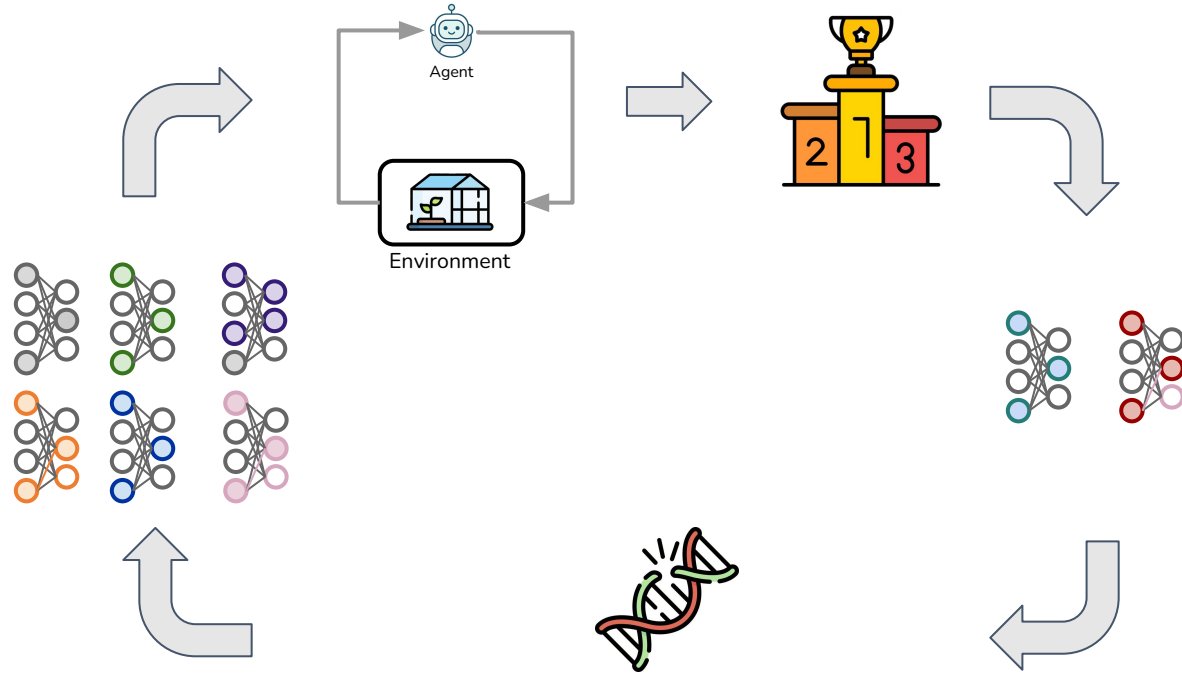


Evolutionary Perspective for RL

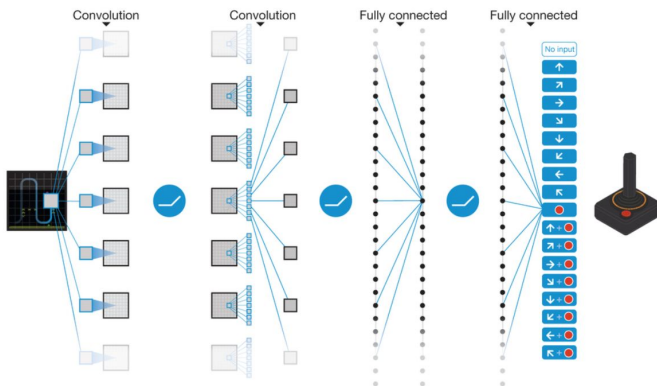


$$F(\pi_\theta) = R(\tau) = r_1 + r_2 + \dots + r_T = \sum_T r_t$$

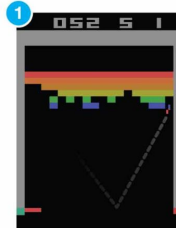
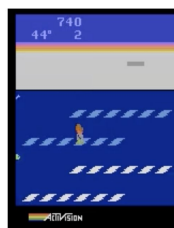
Neuroevolution for RL



Neuroevolution for RL

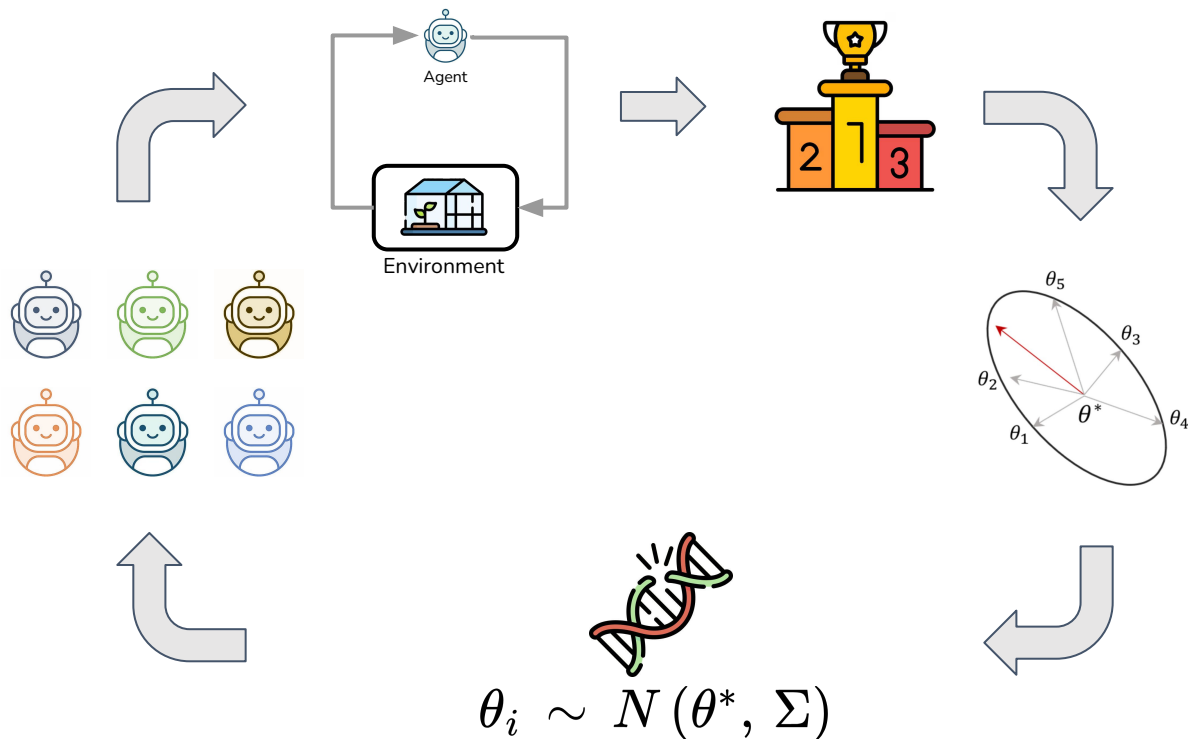


Hyperparameter	Humanoid Locomotion	Image Hard Maze	Atari
Population Size (N)	12,500+1	20,000+1	1,000+1
Mutation power (σ)	0.00224	0.005	0.002
Truncation Size (T)	625	61	20
Number of Trials	5	1	1
Archive Probability		0.01	

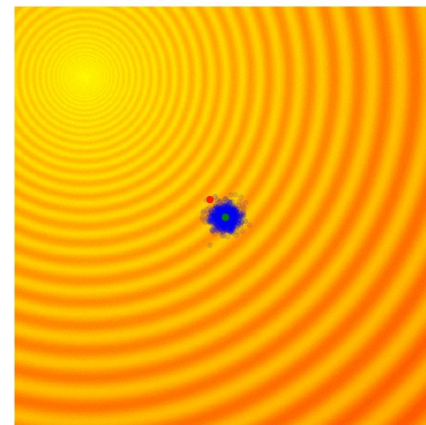
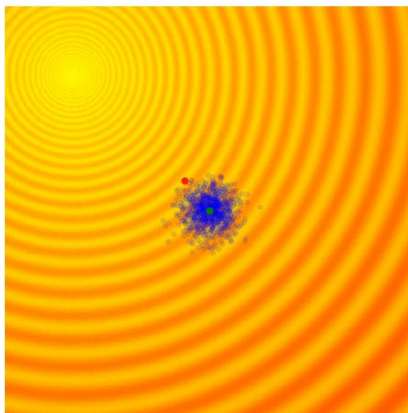
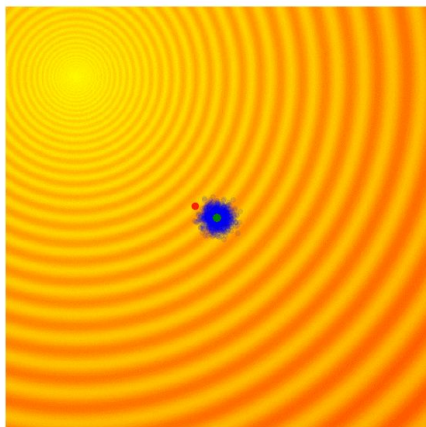


Such et al. (2018). Deep Neuroevolution: Genetic Algorithms Are a Competitive Alternative for Training Deep Neural Networks for Reinforcement Learning

Evolution Strategies for RL



Evolution Strategies for RL



David Ha (2017). Visual Evolution Strategies <https://blog.otoro.net/2017/10/29/visual-evolution-strategies>

Salmans et al. (2017). Evolution Strategies as a Scalable Alternative to Reinforcement Learning

Chrabaszcz et al. (2018). Back to Basics: Benchmarking Canonical Evolution Strategies for Playing Atari

Team Reward Team Utility

Yliniemi et al. (2016). Multi-Objective Multiagent Credit Assignment in reinforcement learning and NSGA-II

Credit assignment in EA

Rojers et al. (2013). Multi-objective variable elimination for collaborative graphical games

Rojers et al. (2014). Linear support for multi-objective coordination graph

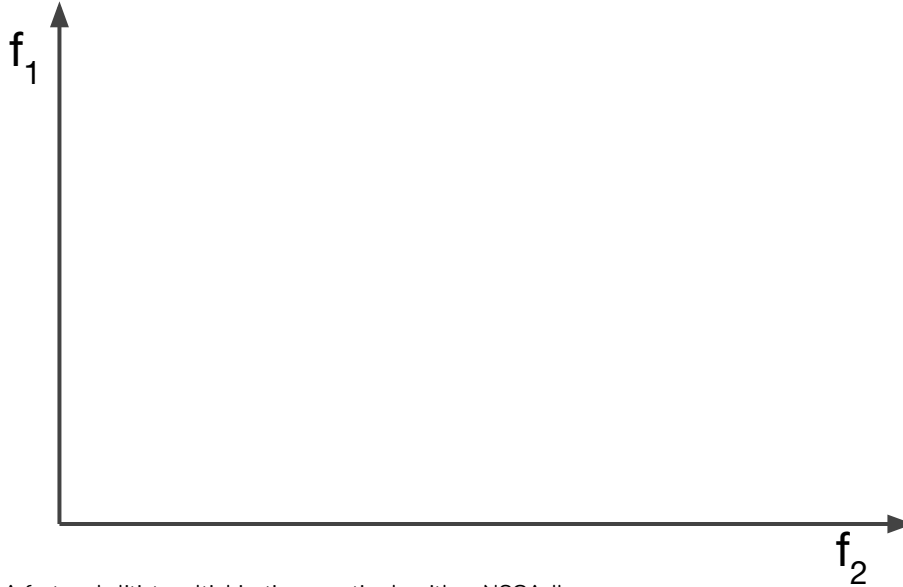
Joint actions with
coordination graphs

Brys et al. (2014). Distributed learning and multi-objectivity in traffic light control

Agrawal et al. (2015). Non-additive multi-objective robot coalition formation

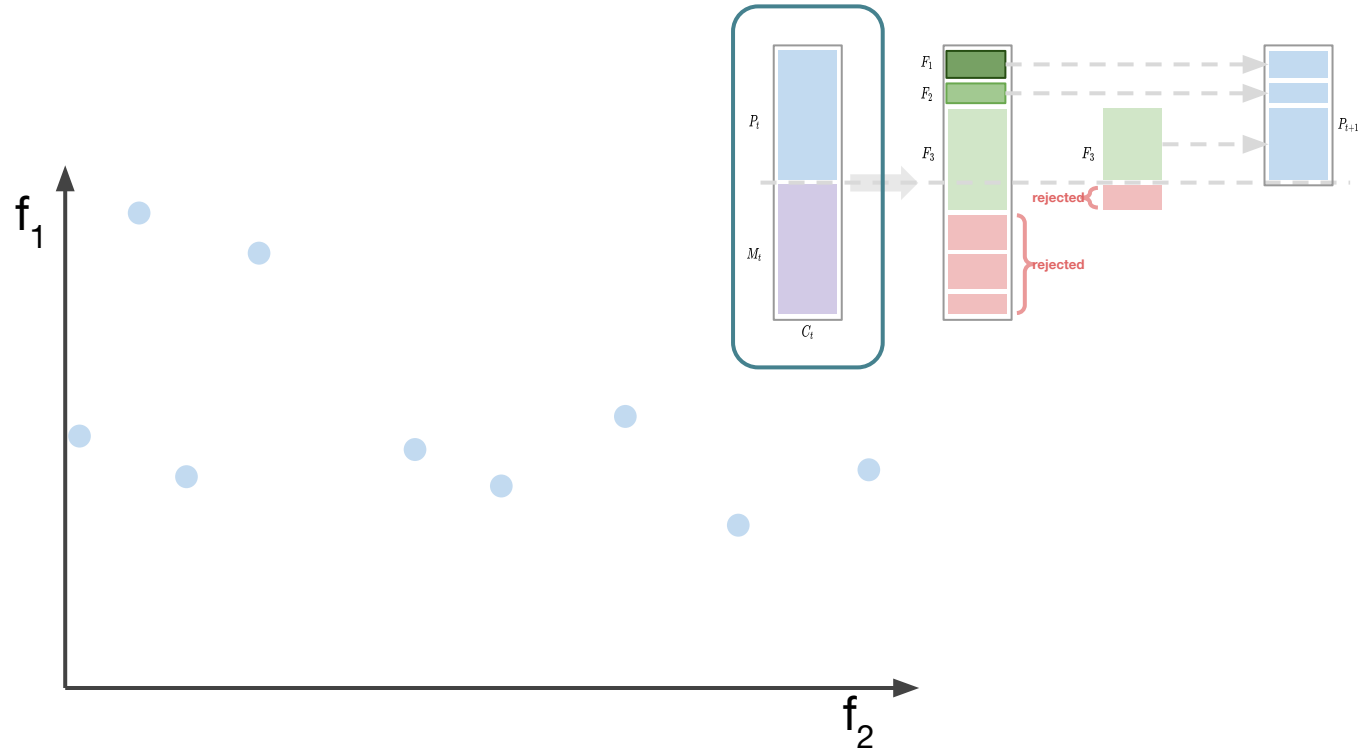
Scalarization and Coalitions
in RL + EAs

Non-dominated Sorting Genetic Algorithm (NSGA-II)

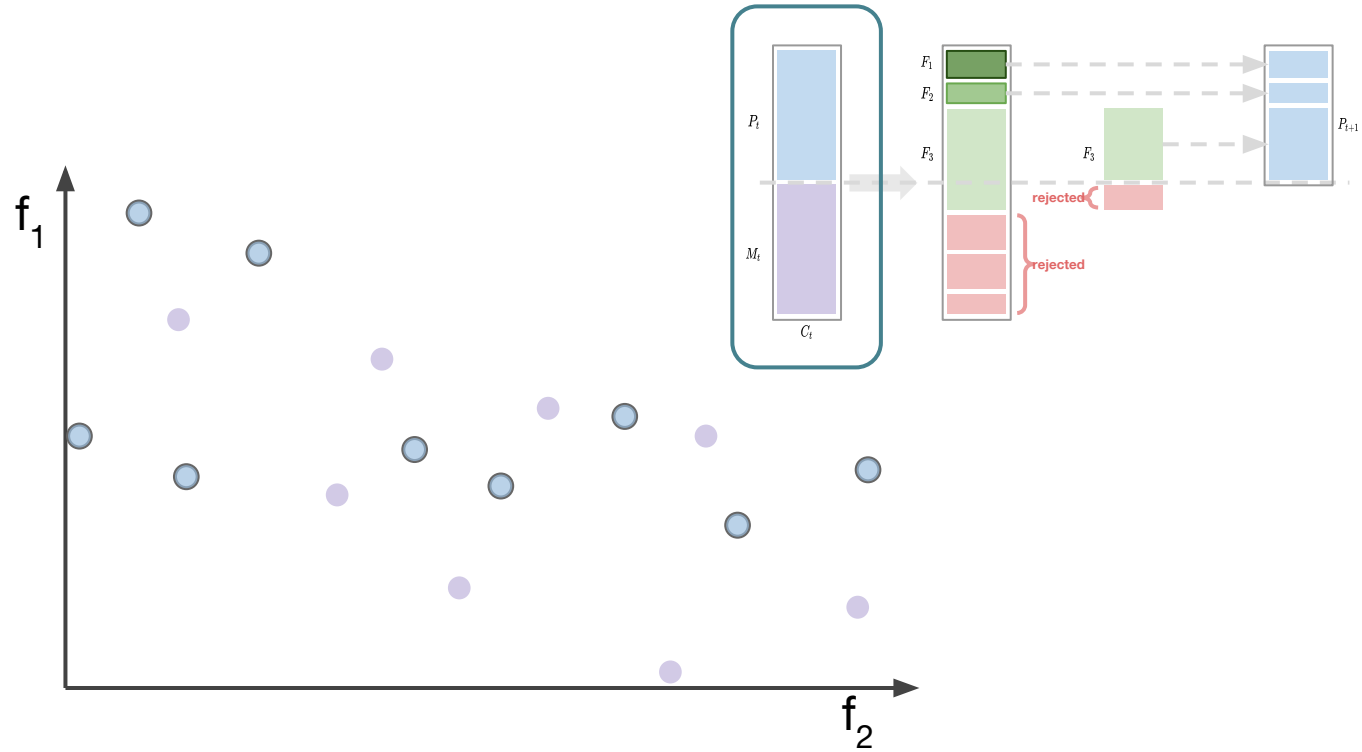


Deb et al. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II

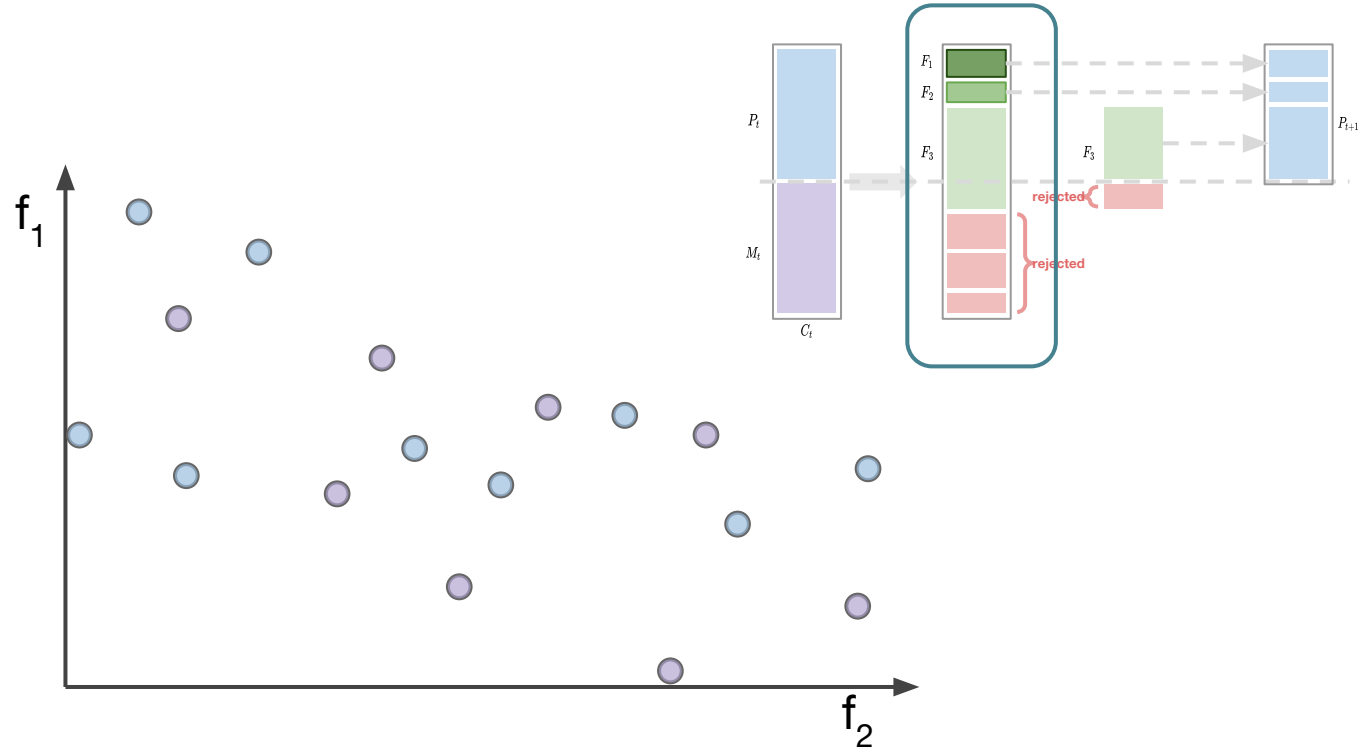
NSGA-II



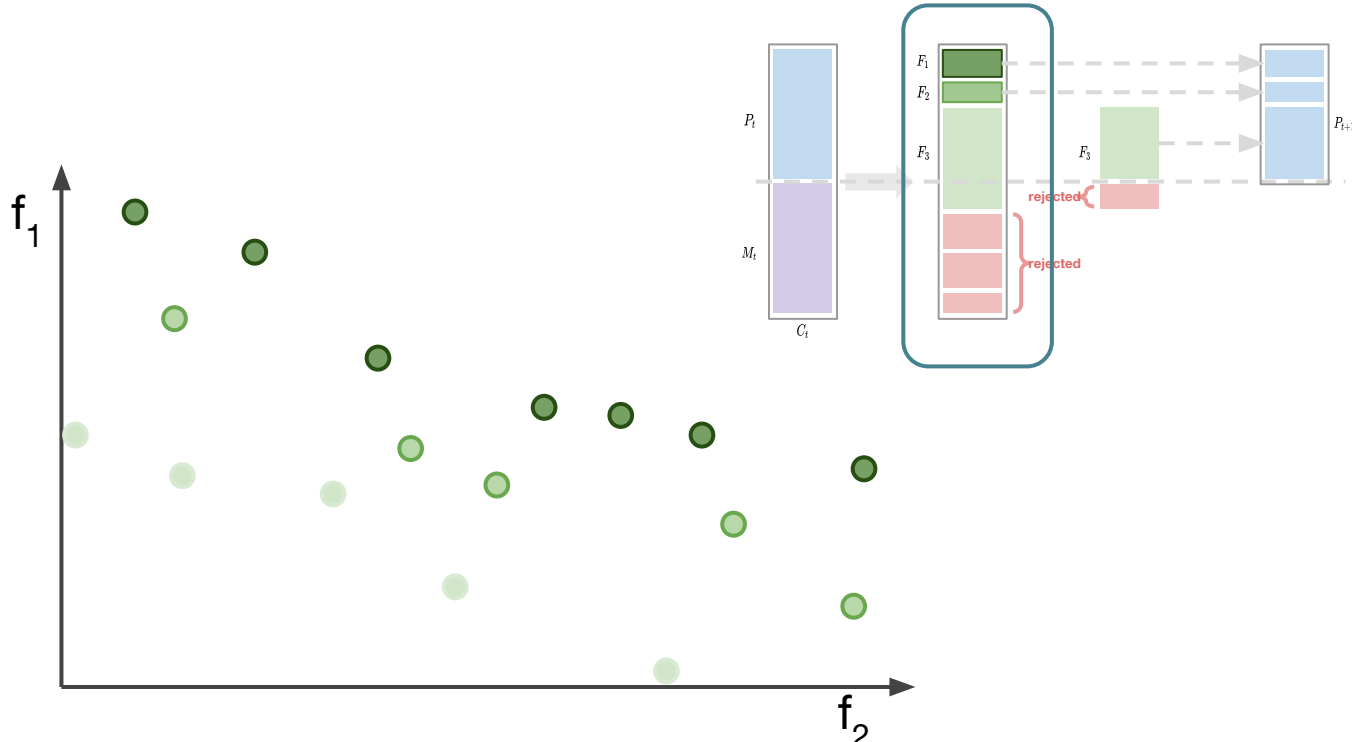
NSGA-II



NSGA-II

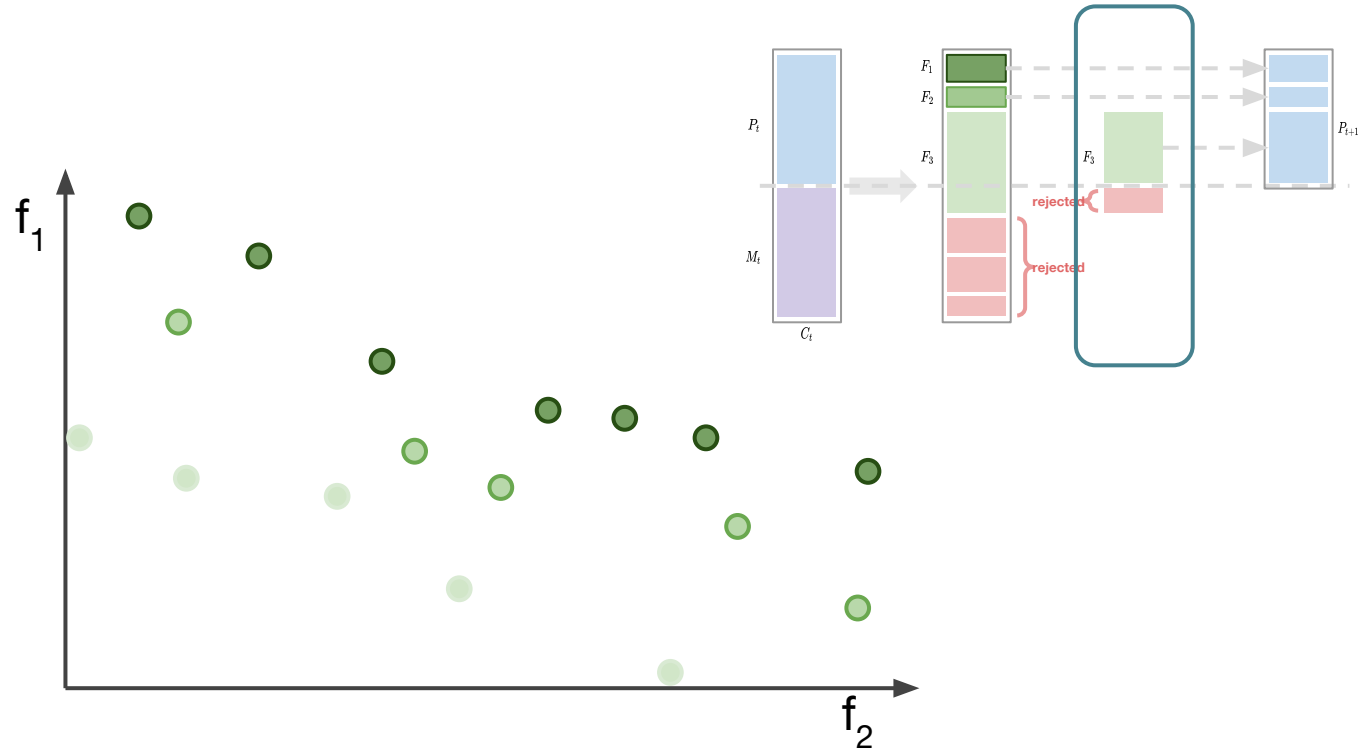


NSGA-II

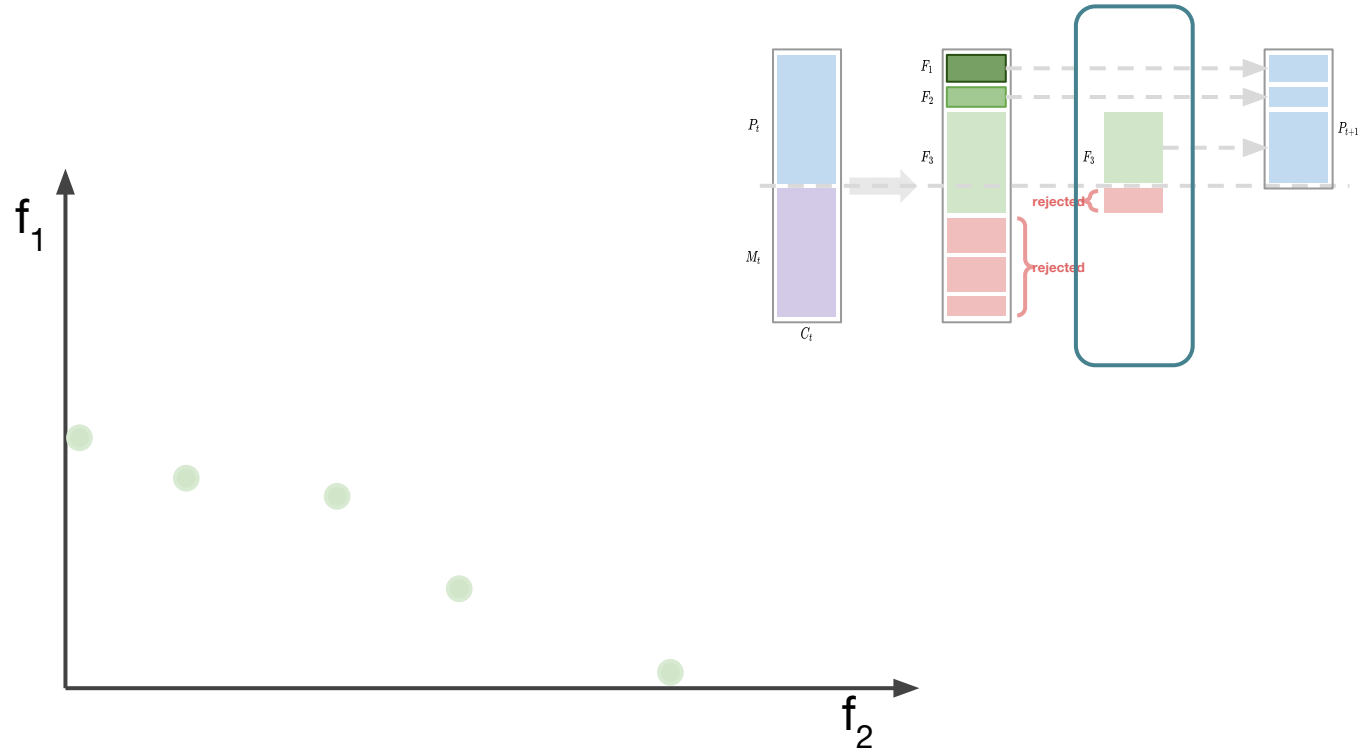


$A(x_1, y_1)$ is dominated by $B(x_2, y_2)$ when: $(x_1 \leq x_2 \text{ and } y_1 \leq y_2)$ and $(x_1 < x_2 \text{ or } y_1 < y_2)$

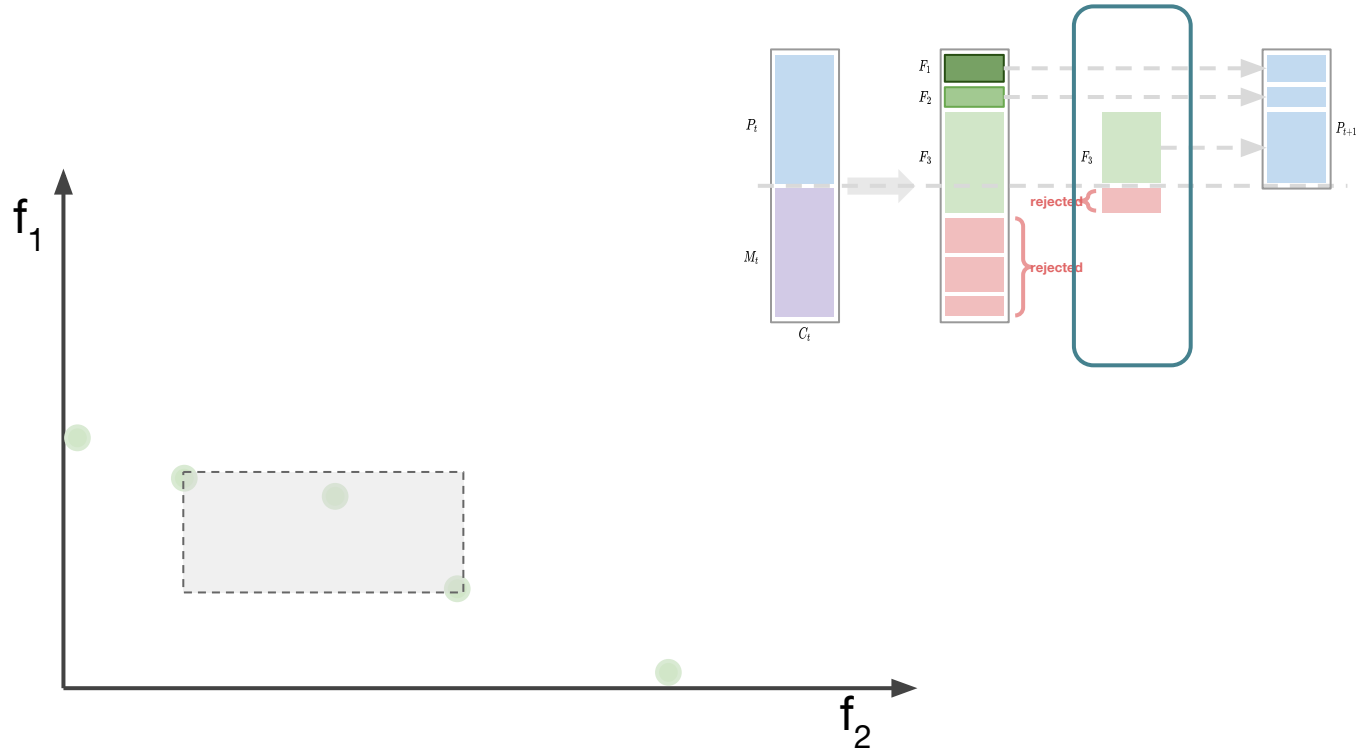
NSGA-II



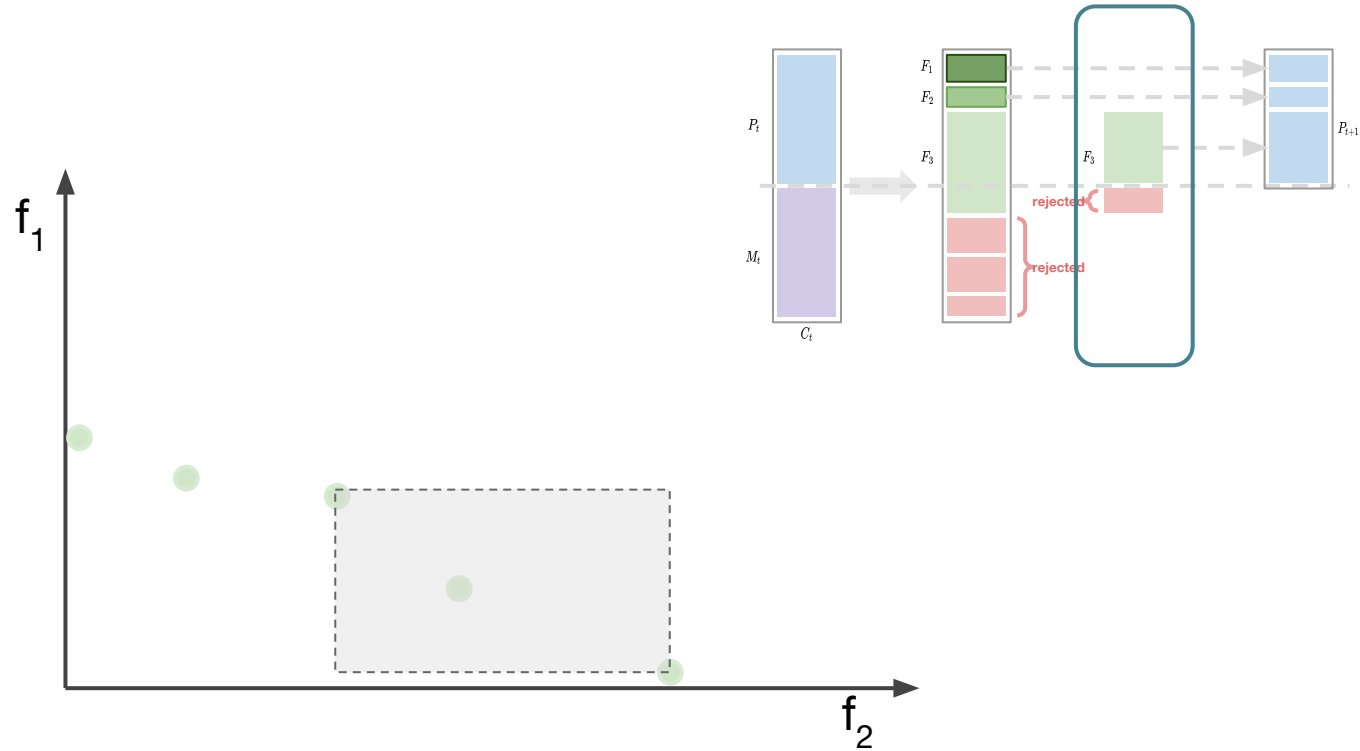
NSGA-II



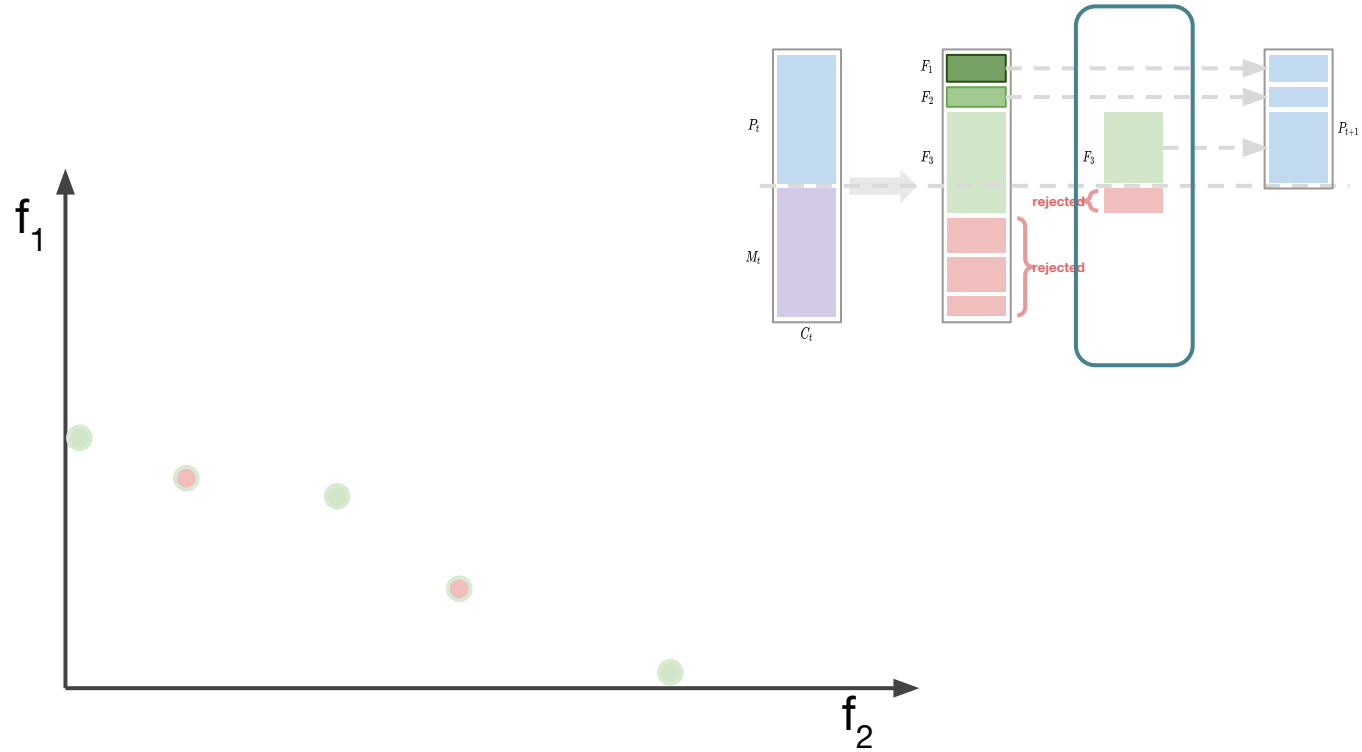
NSGA-II



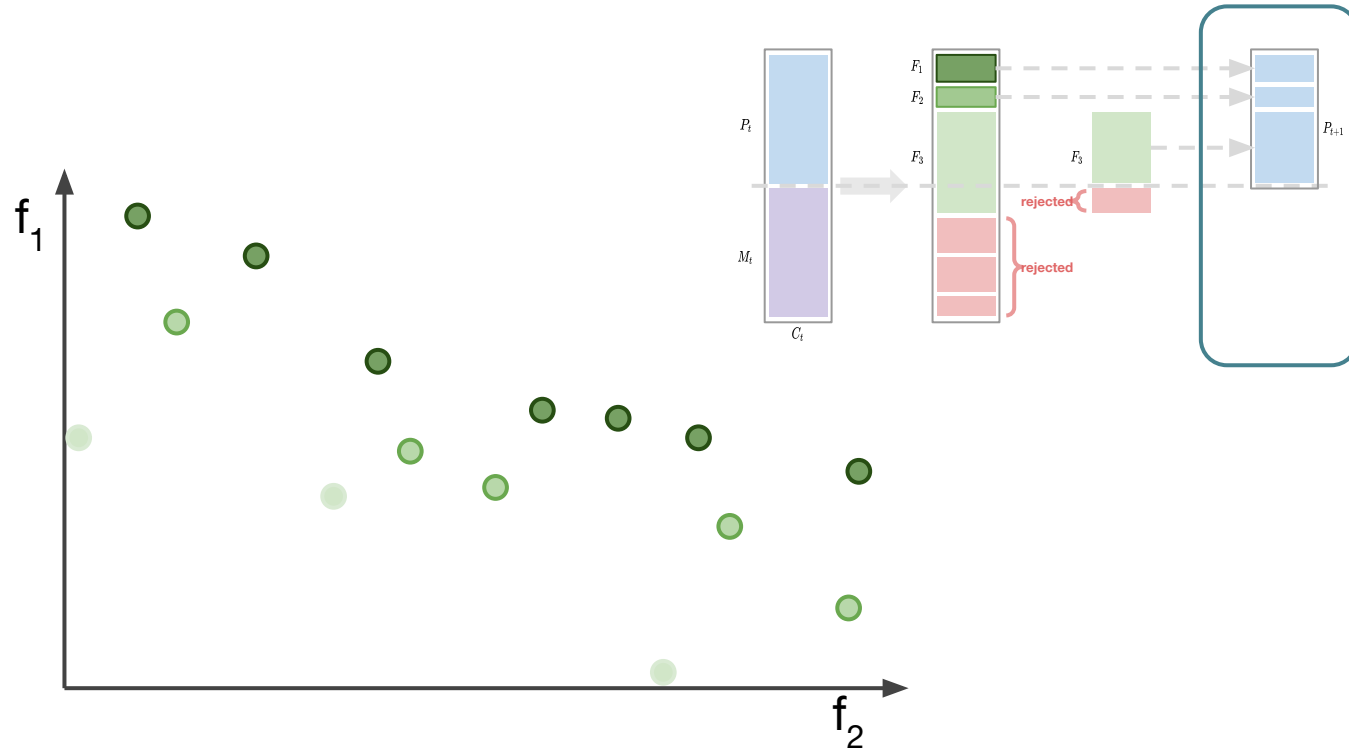
NSGA-II



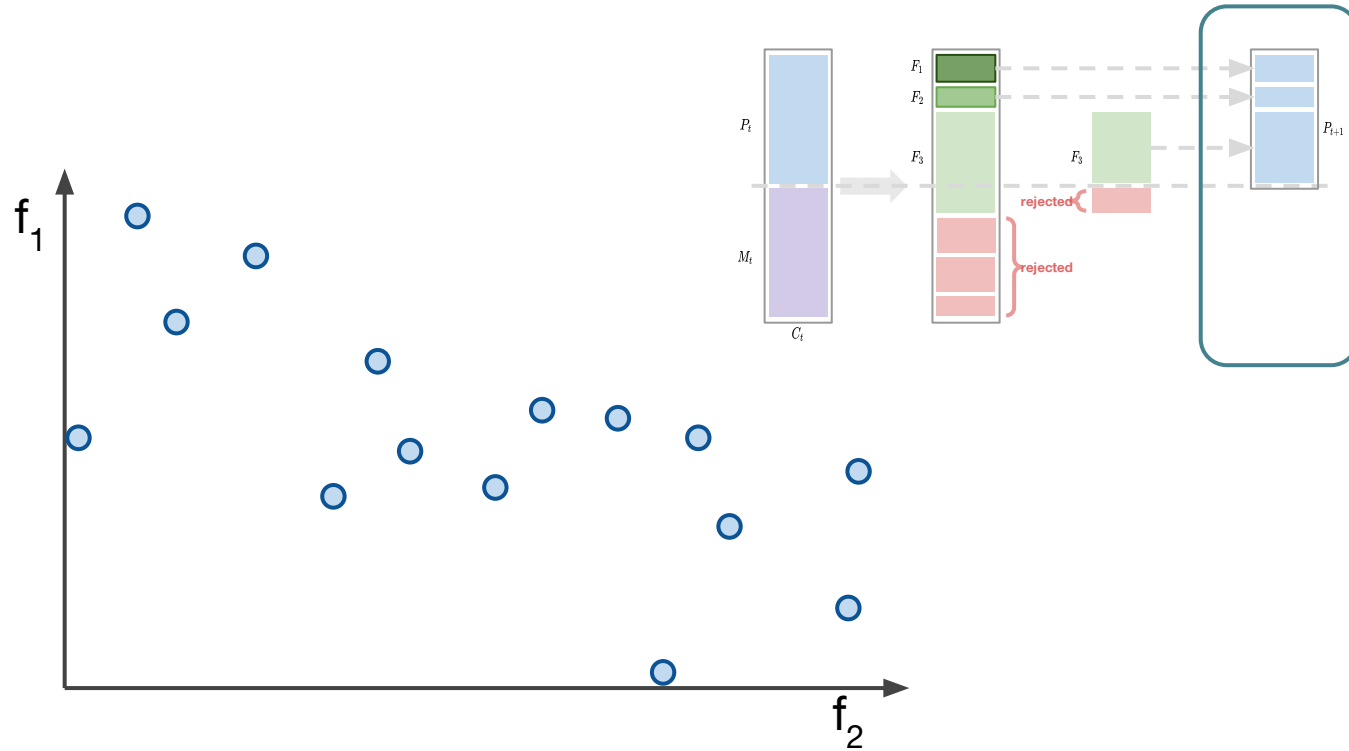
NSGA-II



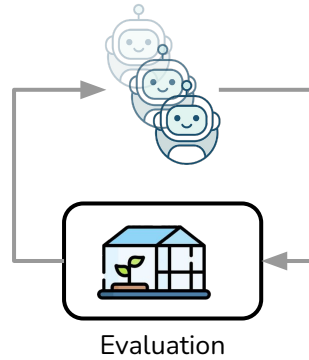
NSGA-II



NSGA-II

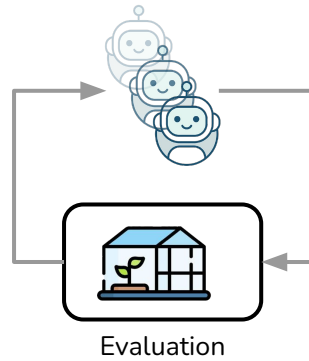


Multiagent Credit Assignment

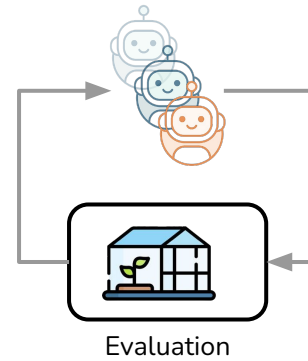


R

Difference Reward

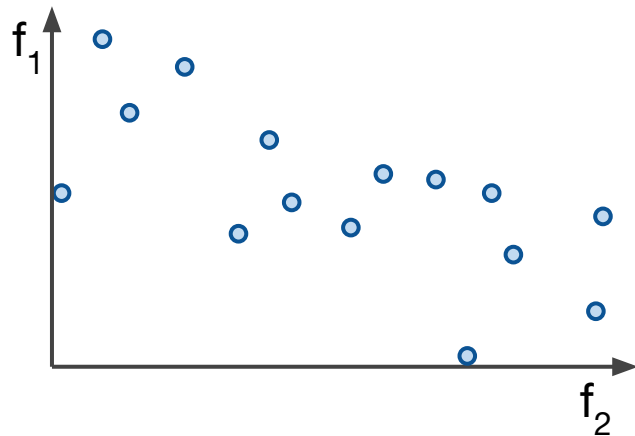


R

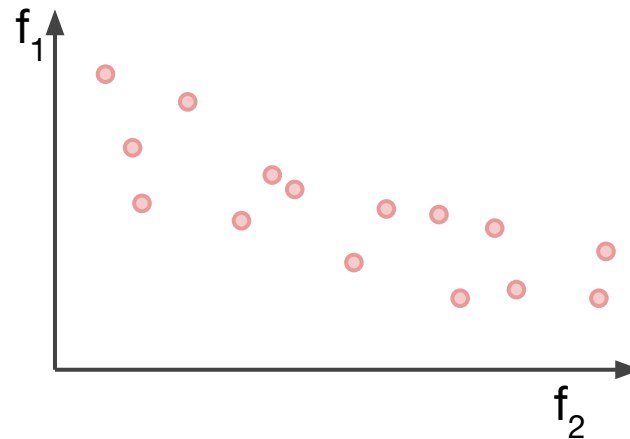


$$D = R - R_{-i}$$

NSGA-II - Multiagent Credit Assignment

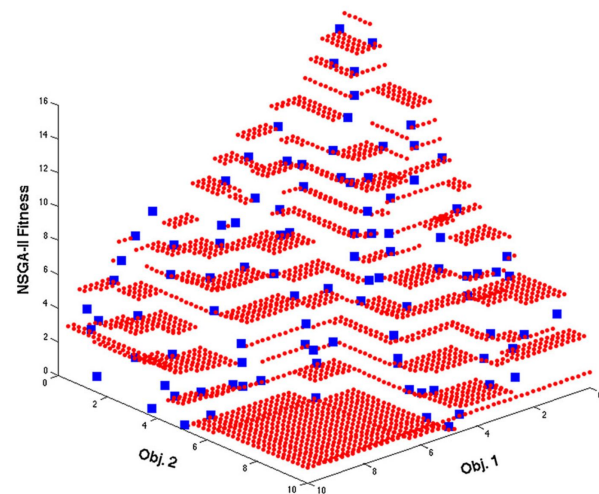
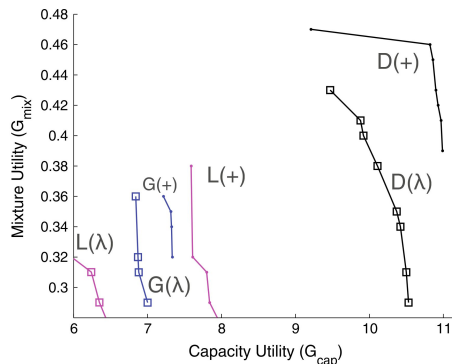
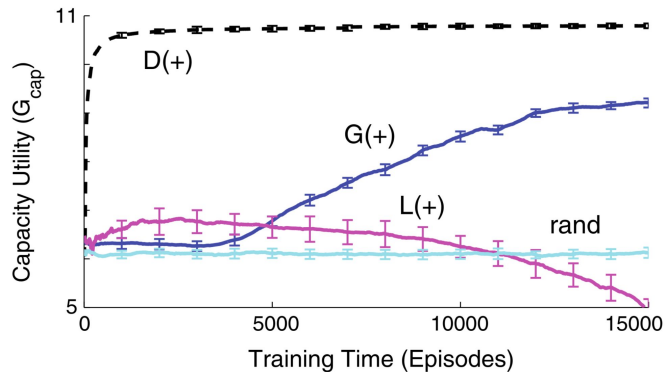


\mathcal{R}



$\mathcal{D} = \mathcal{R} - \mathcal{R}_{-i}$

NSGA-II - Multiagent Credit Assignment



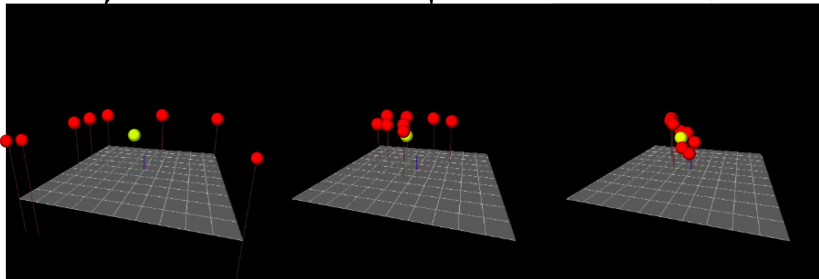
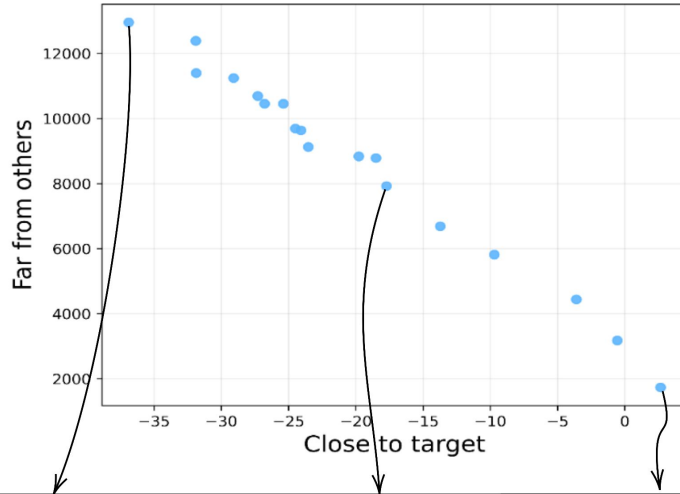
Yliniemi et al. (2016). Multi-Objective Multiagent Credit Assignment in reinforcement learning and NSGA-II

MOMAPPO (Roxana)

- Extension of the MAPPO algorithm to return a Pareto set of multi-agent policies in cooperative problems
- Employs decomposition to divide the MO problem into a collection of single-objective problems solved by a multi-agent RL algorithm

Felten, F., Ucak, U., Azmani, H., Peng, G., Röpke, W., Baier, H., ... & Rădulescu, R. (2024, August). MOMAland: A Set of Benchmarks for Multi-Objective Multi-Agent Reinforcement Learning. In *Multi-objective Decision Making Workshop at ECAI 2024*.

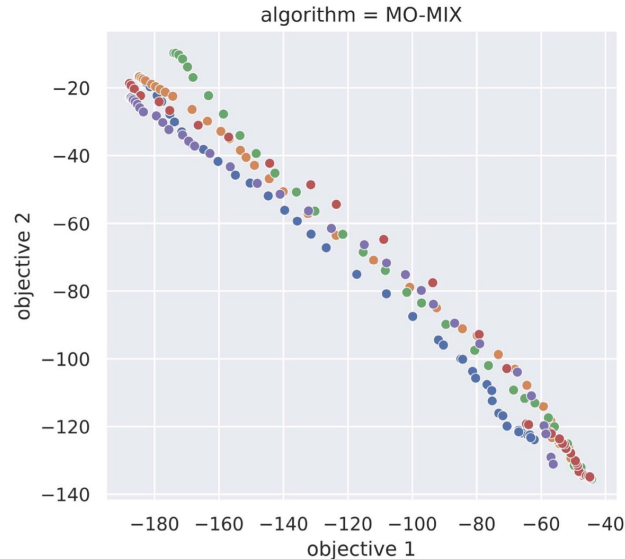
MOMAPPO



Felten, F., Ucak, U., Azmani, H., Peng, G., Röpke, W., Baier, H., ... & Rădulescu, R. (2024, August). MOMAland: A Set of Benchmarks for Multi-Objective Multi-Agent Reinforcement Learning. In *Multi-objective Decision Making Workshop at ECAI 2024*.

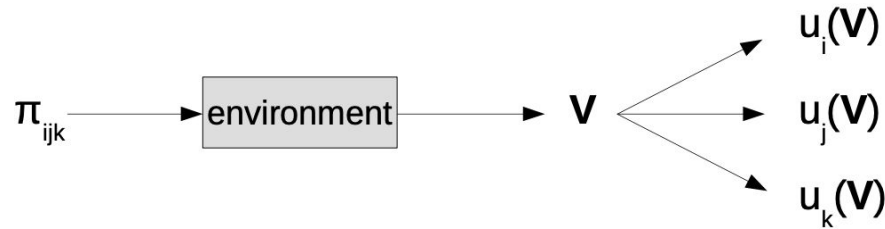
MO-MIX

- Conditions the value function network on the preferences
- Uses a Multi-Objective Mixing Network to concatenate the agent's values

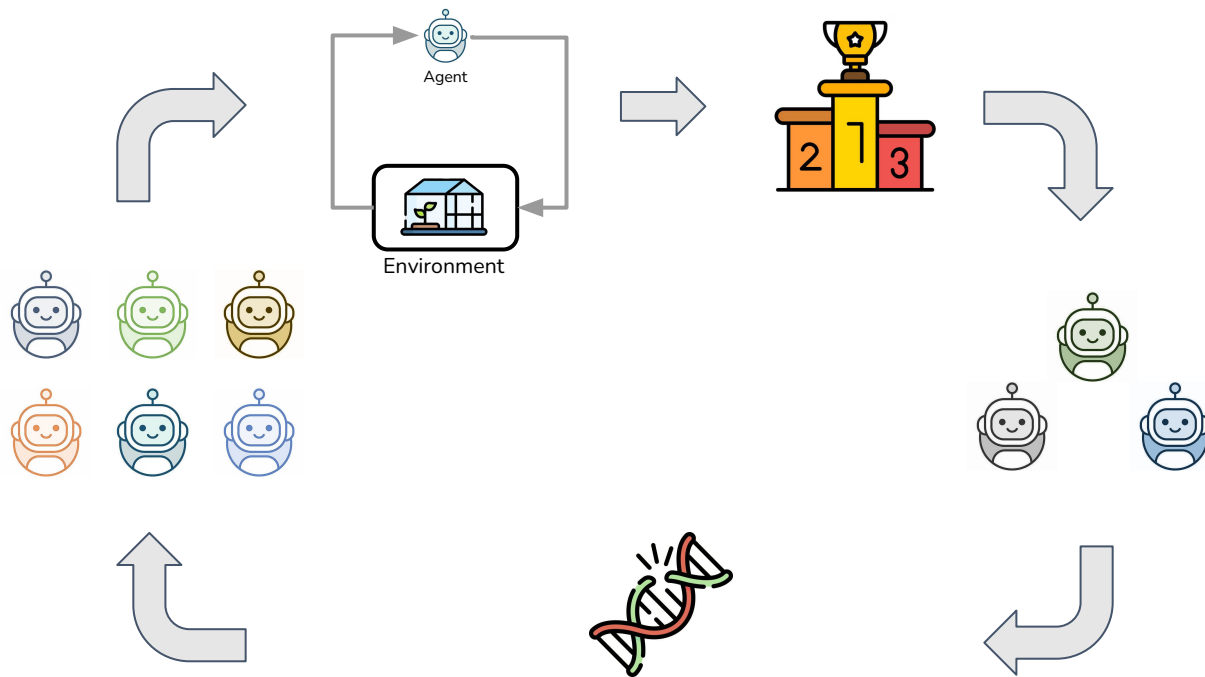


Hu, T., Luo, B., Yang, C., & Huang, T. (2023). MO-MIX: Multi-objective multi-agent cooperative decision-making with deep reinforcement learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10), 12098-12112.

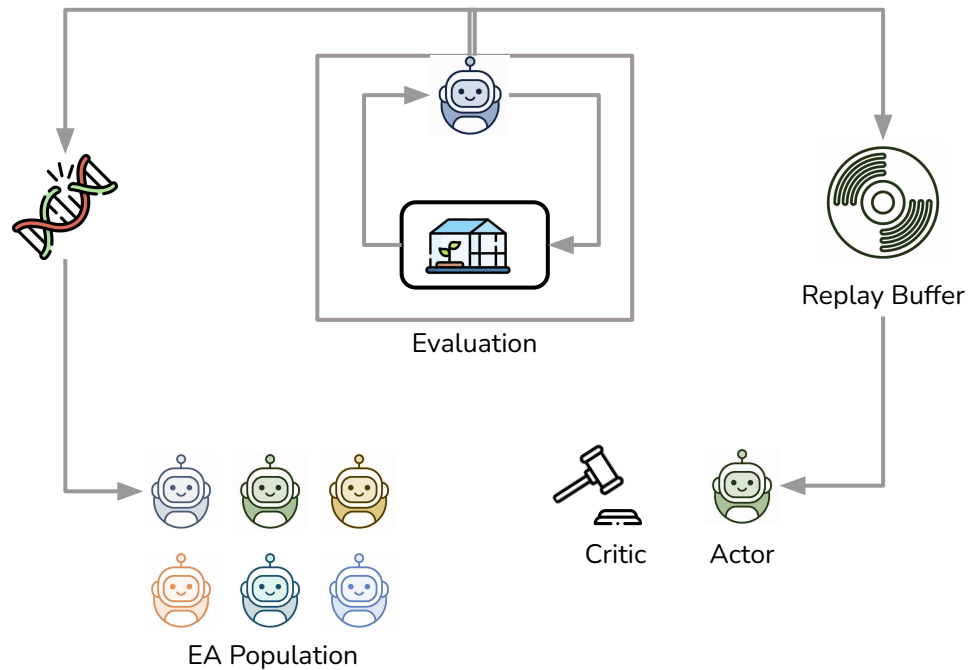
3.2 Team Reward - Individual utility



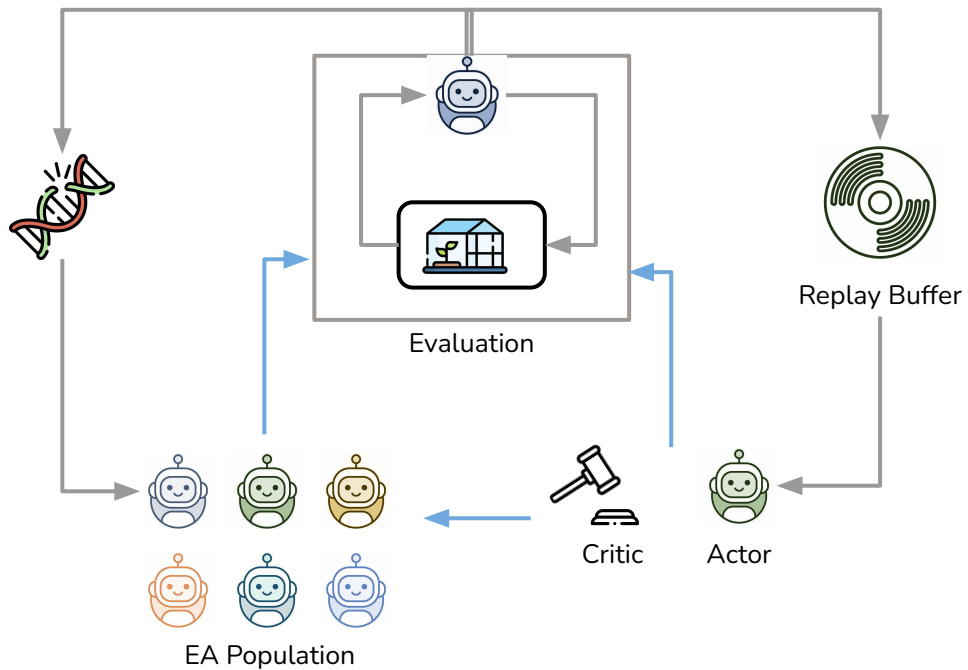
Evolutionary Perspective for RL



Evolutionary and Reinforcement Learning Loop



Evolutionary Reinforcement Learning (ERL)



EA population

=> Parameter-space exploration

RL actor

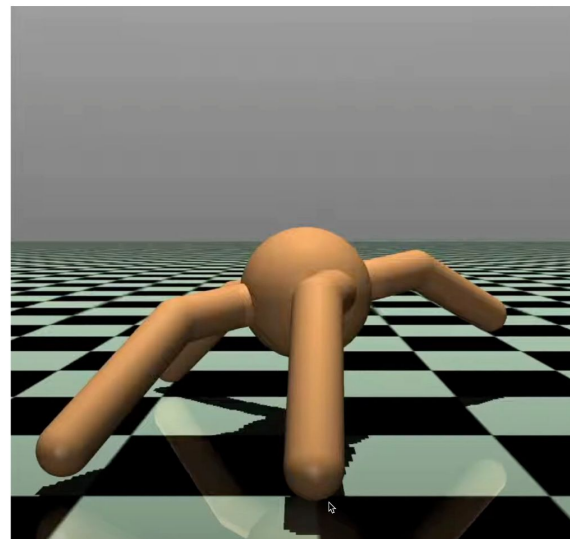
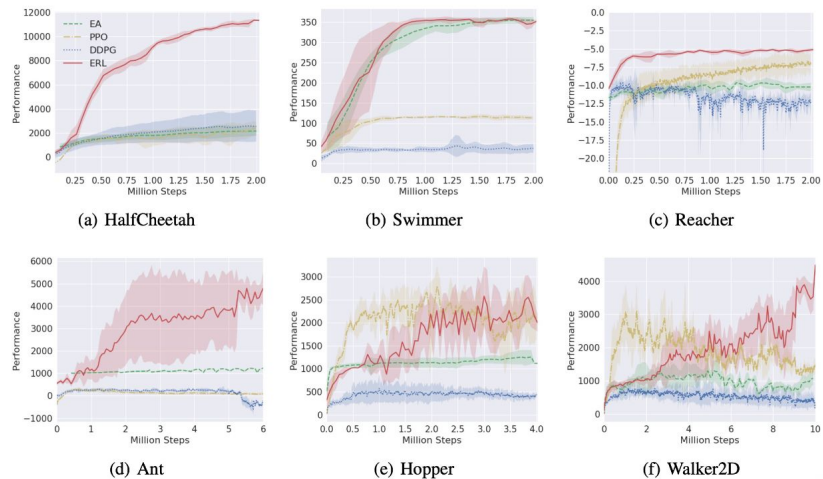
=> action-space exploration

Actor is periodically injected in the EA population

Khadka & Tumer (2018). Evolution-guided policy gradient in reinforcement learning

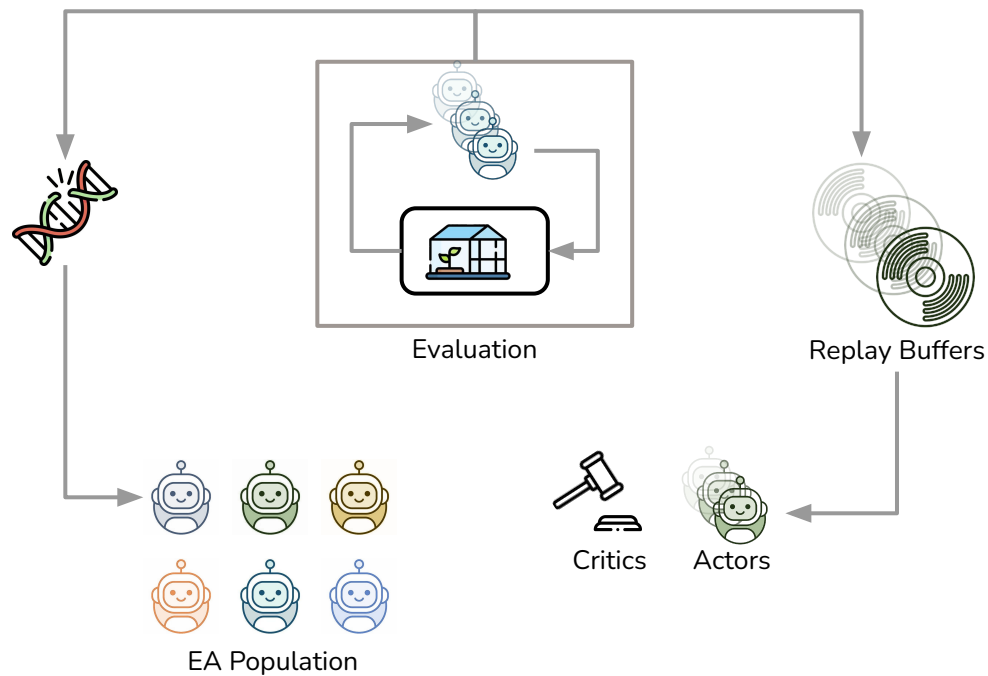
Evolutionary Reinforcement Learning (ERL)

Outperforms EA and RL baselines

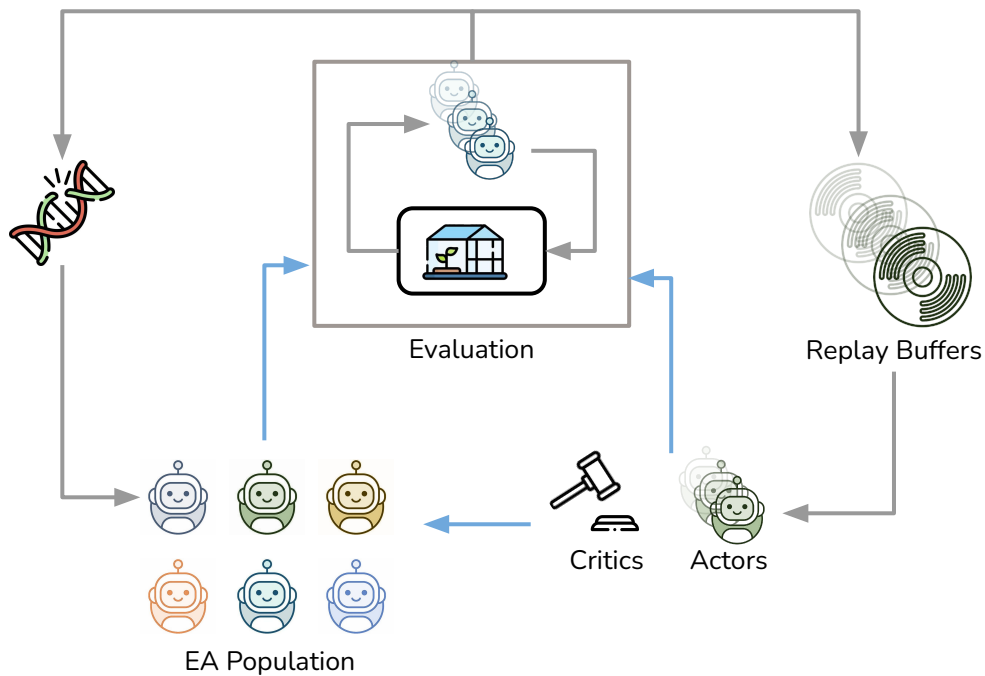


Khadka & Tumer (2018). Evolution-guided policy gradient in reinforcement learning

Multi-Objective Multiagent ERL



Multi-Objective Multiagent ERL



EA population

=> Improve team trade-offs

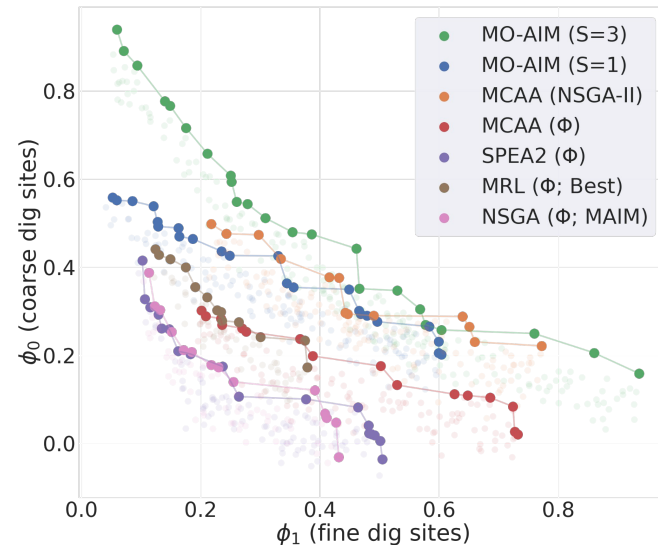
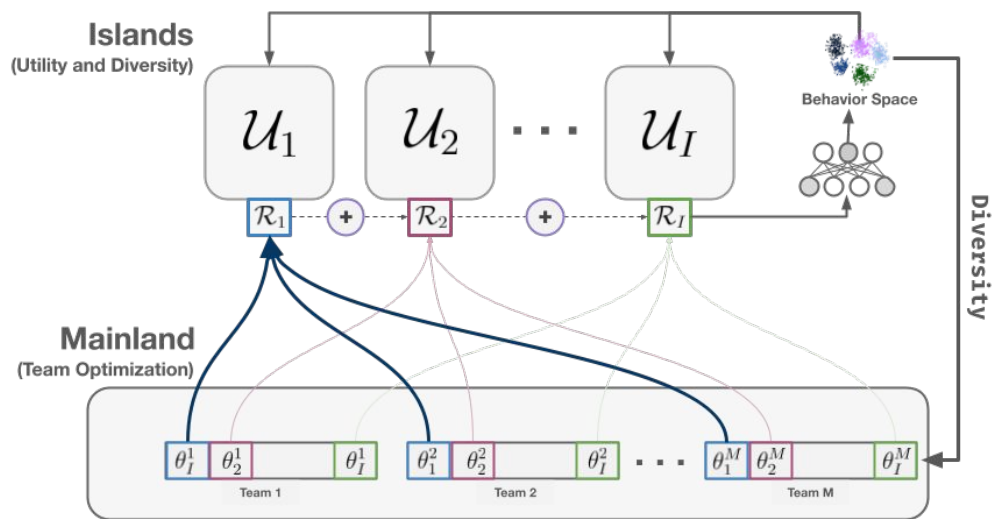
RL actors

=> maximize individual utilities

Actors are periodically injected in the EA population

Dixit et al. (2023). Learning synergies for multi-objective optimization in asymmetric multiagent systems

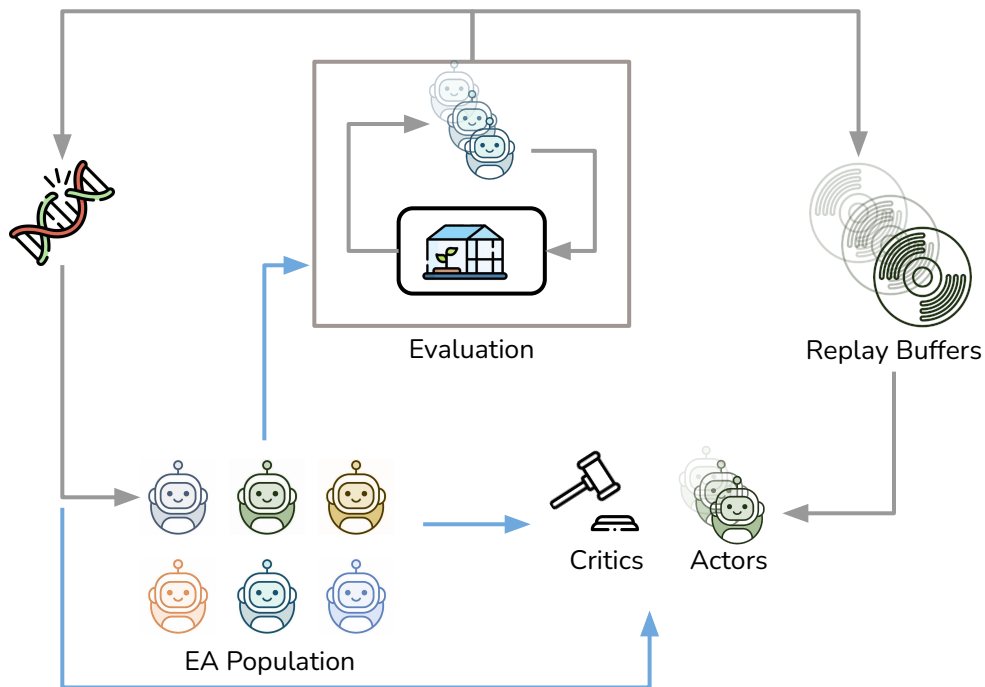
Multi-Objective Multiagent ERL



Outperforms MOEA and MORL baselines

Dixit et al. (2023). Learning synergies for multi-objective optimization in asymmetric multiagent systems

Add Diversity for Improved Coverage



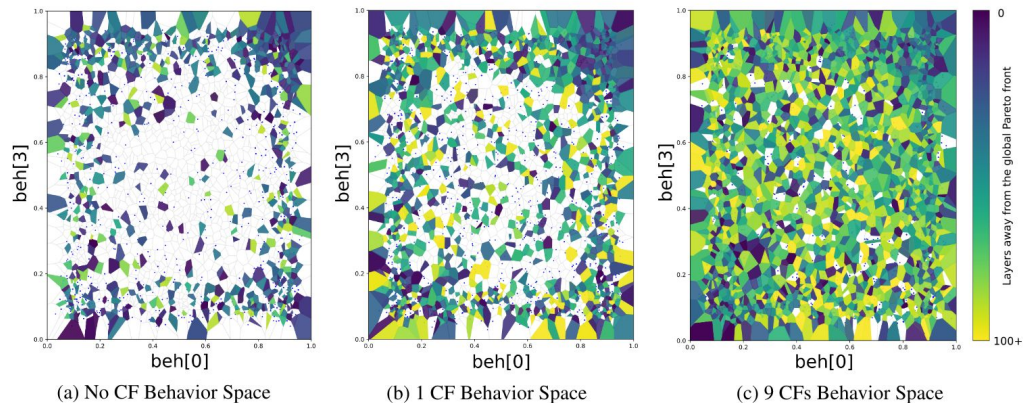
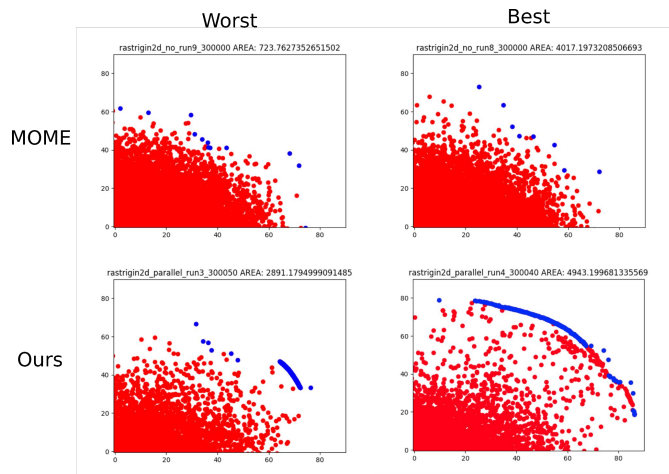
Generate EA population by sampling
from RL actors

Update RL actors to maximize coverage
in the behavior space

Diverse policies are periodically
injected in the EA population

Dixit et al. (2023). Learning synergies for multi-objective optimization in asymmetric multiagent systems

Quality-Diversity for MOMA Problems



Outperforms MOEA and Diversity-Search baselines

Nickelson et al. (2023). Shaping the Behavior Space with Counterfactual Agents in Multi-Objective Map Elites

MOMA: EA and RL

Yliniemi et al. (2016). Multi-Objective Multiagent Credit Assignment in reinforcement learning and NSGA-II

Khadka & Tumer (2018). Evolution-guided policy gradient in reinforcement learning

Leibo et al. (2019). Malthusian Multi-Objective Reinforcement Learning

Dixit et al. (2023). Learning synergies for multi-objective optimization in asymmetric multiagent systems

Nickelson et al. (2023). Shaping the Behavior Space with Counterfactual Agents in Multi-Objective Map Elites

Credit assignment in RL and EA

Evolutionary + Reinforcement
Learning

Diversity in utilities, trade-offs +
better coverage

Multi-Objective Normal Form Games (Patrick)

- Introduced by Blackwell in 1956
- MONFG - tuple (N, A, \mathbf{p}) , with $n \geq 2$ and $C \geq 2$ objectives, where:
 - $N = \{1, \dots, n\}$ – set of players
 - $A = A_1 \times \dots \times A_n$ – set of actions
 - $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_n)$ – vectorial payoffs

Example - SER

$$u(p_1, p_2) = p_1 \cdot p_2$$

	A	B
A	(10, 2); (10, 2)	(0, 0); (0, 0)
B	(0, 0); (0, 0)	(2, 10); (2, 10)

Example - Nash equilibrium

$$u(p_1, p_2) = p_1 \cdot p_2$$

$$u(10, 2) = 10 \cdot 2 = 20$$

	A	B
A	(10, 2); (10, 2)	(0, 0); (0, 0)
B	(0, 0); (0, 0)	(2, 10); (2, 10)

Example - Cyclic Nash equilibrium

$$u(p_1, p_2) = p_1 \cdot p_2$$

- Joint cyclic strategy
 - Player 1: {A, B}
 - Player 2: {A, B}

$$u\left(\frac{10+2}{2}, \frac{2+10}{2}\right) = u(6, 6) = 36$$

	A	B
A	(10, 2); (10, 2)	(0, 0); (0, 0)
B	(0, 0); (0, 0)	(2, 10); (2, 10)

Example - Correlated equilibrium

- Correlated strategy σ
 - 50% (A, A)
 - 50% (B, B)

$$u(p_1, p_2) = p_1 \cdot p_2$$

$$u\left(\frac{10+2}{2}, \frac{2+10}{2}\right) = u(6, 6) = 36$$

	A	B
A	(10, 2); (10, 2)	(0, 0); (0, 0)
B	(0, 0); (0, 0)	(2, 10); (2, 10)

(Im)balancing Act Game

- 2 players, 2 objective
- Same payoff vector for both players

$$u_1([p_1, p_2]) = p_1^2 + p_2^2$$
$$u_2([p_1, p_2]) = p_1 \cdot p_2$$

	L	M	R
L	[4,0]	[3,1]	[2,2]
M	[3,1]	[2,2]	[1,3]
R	[2,2]	[1,3]	[0,4]

ESR Equilibrium

- equilibrium 1: (0.75, 0, 0.25) and (0, 1, 0)
expected utilities: 10 and 3
- equilibrium 2: (0.25, 0, 0.75) and (0, 1, 0)
expected utilities: 10 and 3

ESR	L	M	R
L	16,0	10,3	8,4
M	10,3	8,4	10,3
R	8,4	10,3	16,0

	L	M	R
L	[4,0]	[3,1]	[2,2]
M	[3,1]	[2,2]	[1,3]
R	[2,2]	[1,3]	[0,4]

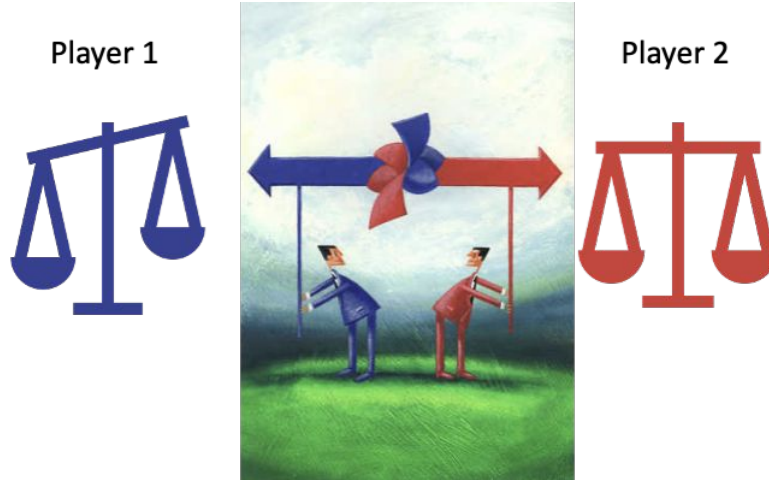
$$u_1([p_1, p_2]) = p_1^2 + p_2^2$$

$$u_2([p_1, p_2]) = p_1 \cdot p_2$$

Rădulescu, R., Mannion, P., Zhang, Y., Roijers, D. M., & Nowé, A. (2020). A utility-based analysis of equilibria in multi-objective normal-form games. *The Knowledge Engineering Review*, 35.

SER Equilibrium?

- In finite MONFGs, where each agent seeks to maximise the utility under **SER**, **Nash equilibria need not exist.**



	L	M	R
L	[4,0]	[3,1]	[2,2]
M	[3,1]	[2,2]	[1,3]
R	[2,2]	[1,3]	[0,4]

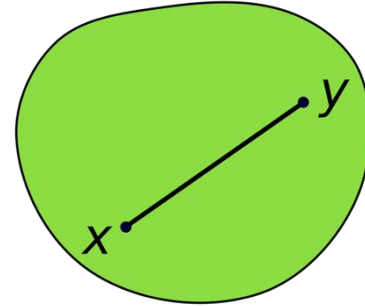
$$u_1([p_1, p_2]) = p_1^2 + p_2^2$$

$$u_2([p_1, p_2]) = p_1 \cdot p_2$$

Rădulescu, R., Mannion, P., Zhang, Y., Roijers, D. M., & Nowé, A. (2020). A utility-based analysis of equilibria in multi-objective normal-form games. *The Knowledge Engineering Review*, 35.

Bridging continuous games and MONFGs

- Continuous games:
 - Single objective
 - Infinite number of pure strategies
 - Continuous payoff functions
 - Benefit from many theoretical results
 - Algorithmically challenging

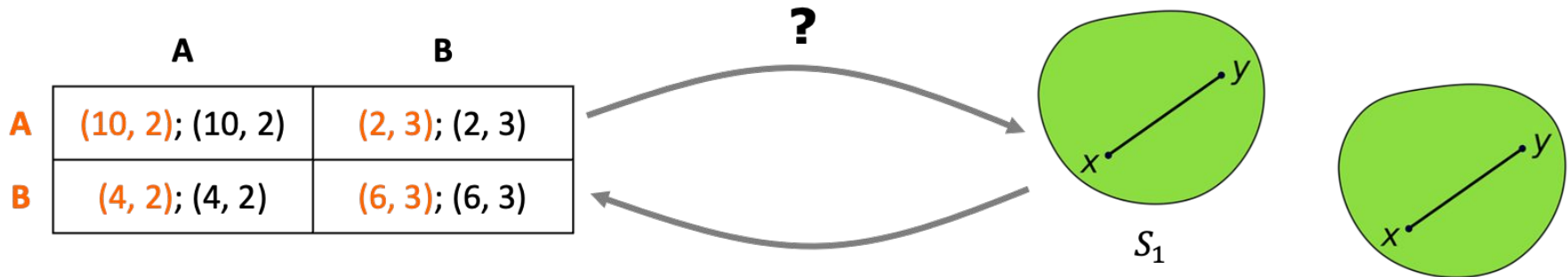


Assumption: convex strategy set

Röpke, W., Groenland, C., Rădulescu, R., Nowé, A., & Roijers, D. M. (2023). Bridging the Gap Between Single and Multi Objective Games. *AAMAS 2023*.

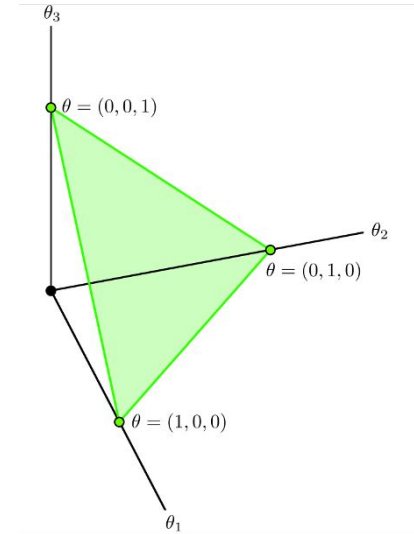
Bridging continuous games and MONFGs

- Build mapping between MONFGs and continuous games
- Ensure that it preserves key dynamics
- Leverage the link for theoretical and algorithmic improvements



Bridging continuous games and MONFGs

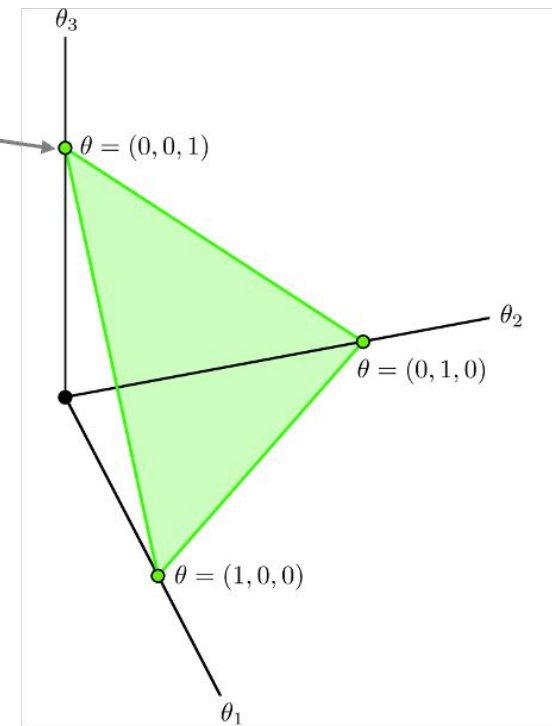
- Every mixed strategy in the MONFG becomes a pure strategy in the continuous game



Röpke, W., Groenland, C., Rădulescu, R., Nowé, A., & Roijers, D. M. (2023). Bridging the Gap Between Single and Multi Objective Games. *AAMAS 2023*.

Bridging continuous games and MONFGs

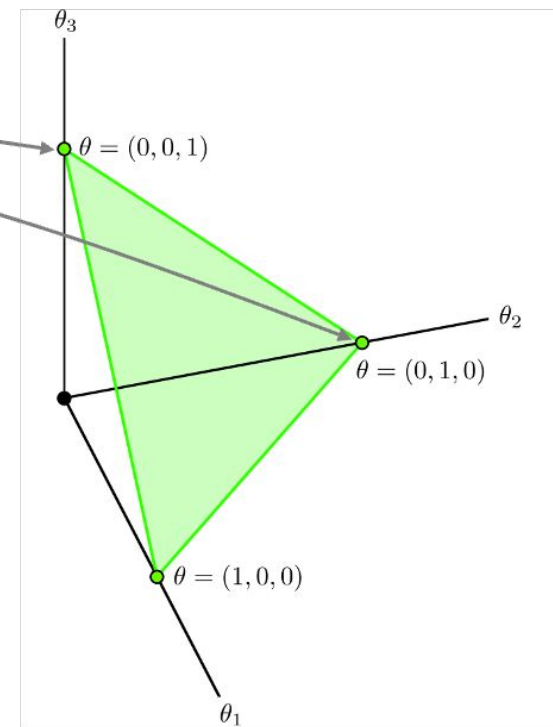
	A	B	C
A	(4, 1); (4, 1)	(1, 2); (4, 2)	(2, 1); (1, 2)
B	(3, 1); (2, 3)	(3, 2); (6, 3)	(1, 2); (2, 1)
C	(1, 2); (2, 1)	(2, 1); (1, 2)	(1, 3); (1, 3)



Bridging continuous games and MONFGs

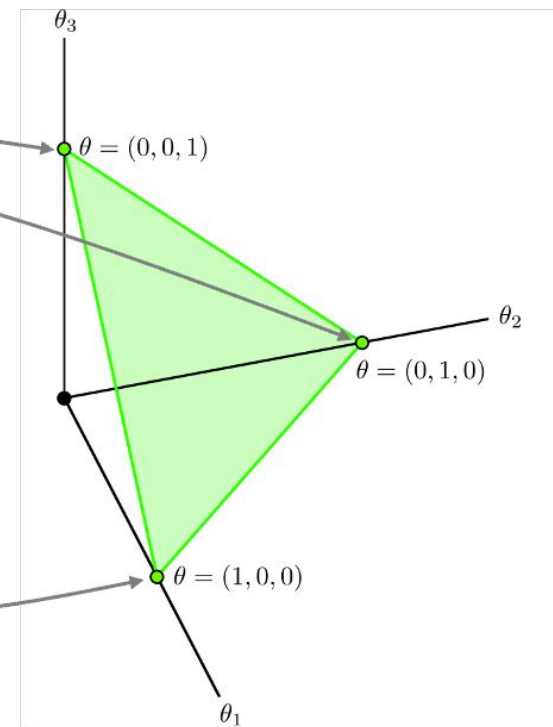
A B C

A	(4, 1); (4, 1)	(1, 2); (4, 2)	(2, 1); (1, 2)
B	(3, 1); (2, 3)	(3, 2); (6, 3)	(1, 2); (2, 1)
C	(1, 2); (2, 1)	(2, 1); (1, 2)	(1, 3); (1, 3)



Bridging continuous games and MONFGs

	A	B	C
A	(4, 1); (4, 1)	(1, 2); (4, 2)	(2, 1); (1, 2)
B	(3, 1); (2, 3)	(3, 2); (6, 3)	(1, 2); (2, 1)
C	(1, 2); (2, 1)	(2, 1); (1, 2)	(1, 3); (1, 3)

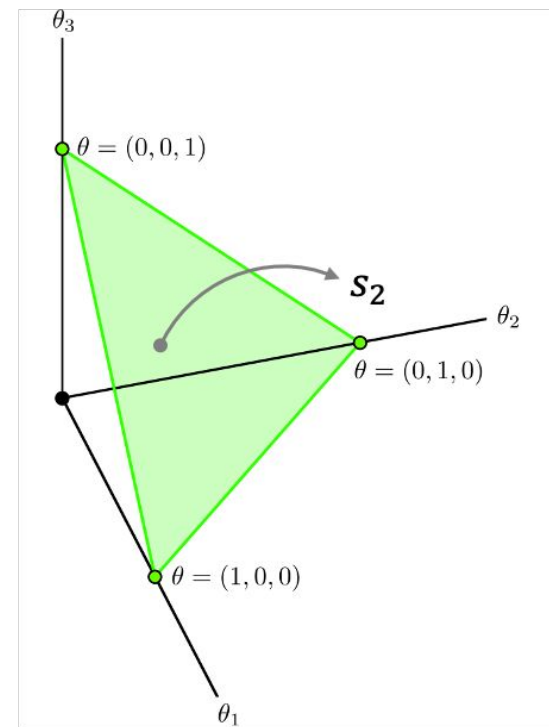


Bridging continuous games and MONFGs

$\frac{1}{4}$ $\frac{1}{4}$ $\frac{2}{4}$

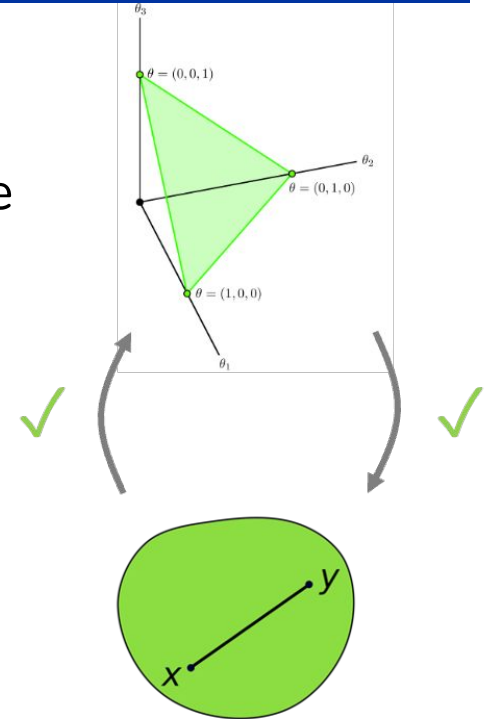
A **B** **C**

A	(4, 1); (4, 1)	(1, 2); (4, 2)	(2, 1); (1, 2)
B	(3, 1); (2, 3)	(3, 2); (6, 3)	(1, 2); (2, 1)
C	(1, 2); (2, 1)	(2, 1); (1, 2)	(1, 3); (1, 3)



Theoretical insights

- Mixed strategy equilibria in the MONFG are pure strategy equilibria in the continuous game
- Continuous games are not guaranteed to have a pure strategy Nash equilibrium
 - ▶ Nash equilibria are not guaranteed in MONFGs



Relations between optimisation criteria

- **Mixed strategies**

- **No relation** between both optimisation criteria **in general**

$$u(x, y) = 0.1 * x + \max(0, x) * \max(0, y)$$

	A	B
A	(1, 0); (1, 0)	(0, 1); (0, 1)
B	(0, 1); (0, 1)	(-10, 0); (-10, 0)

Multi-objective reward vectors

	A	B
A	0.1; 0.1	0; 0
B	0; 0	-0.1; -0.1

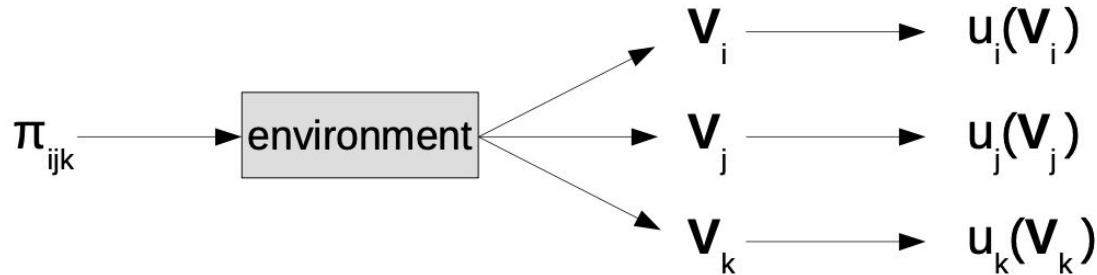
Scalarised utility for both agents

No sharing of number of equilibria or equilibria themselves

Relations between optimisation criteria

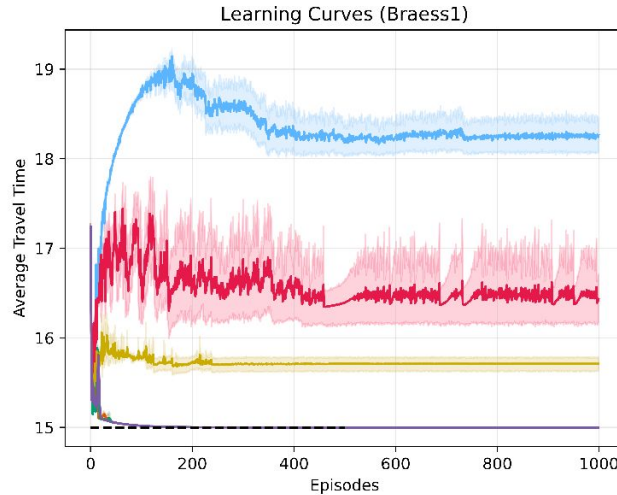
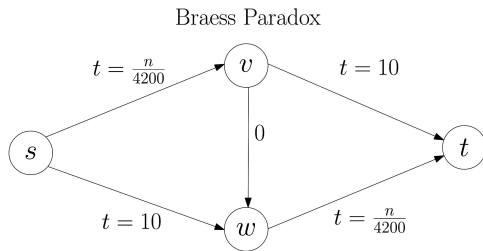
- **Pure strategies**
 - Pure strategy equilibrium under SER is also one under ESR
 - Bidirectional when assuming (quasi)convex utility functions
- We can extend this to **blended settings**
 - Pure strategy equilibrium under SER is also one in any blended setting
 - Bidirectional when assuming (quasi)convex utility functions

3.3 Individual Reward - Individual utility



Scalarised Individual Q-learning

- Assume a known linear utility function for each agent



Mannion, P., Devlin, S., Duggan, J., & Howley, E. (2018). Reward shaping for knowledge-based multi-objective multi-agent reinforcement learning. *The Knowledge Engineering Review*, 33, e23.

Felten, F., Ucak, U., Azmani, H., Peng, G., Röpke, W., Baier, H., Mannion, P., Roijers, D.M., Terry, J.K., Talbi, E.G. and Danoy, G., 2024. MOMAland: A Set of Benchmarks for Multi-Objective Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:2407.16312*.

RIU with dynamic preferences

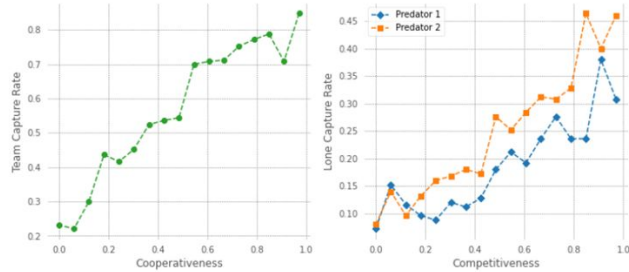


Figure 6: Tuning performance for two predator agents with matched preferences

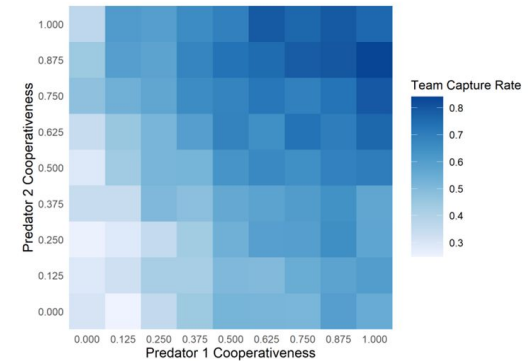
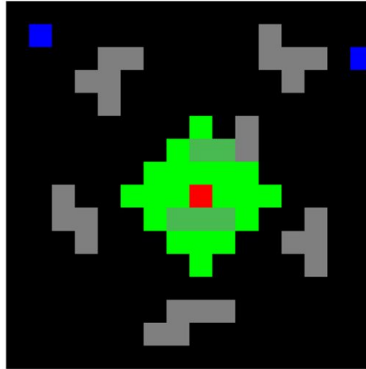
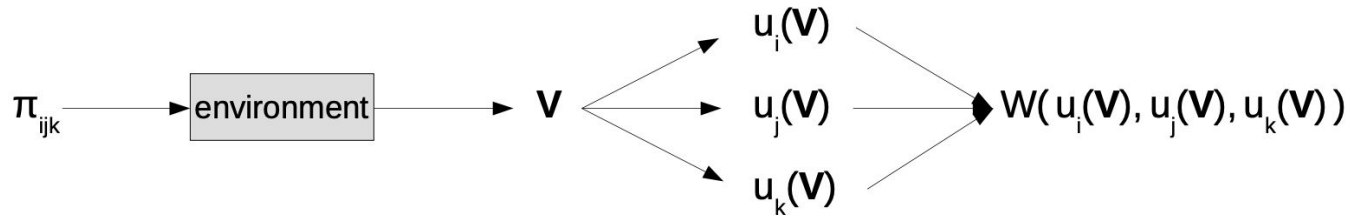


Figure 7: Tuning performance for two predator agents with varied preferences

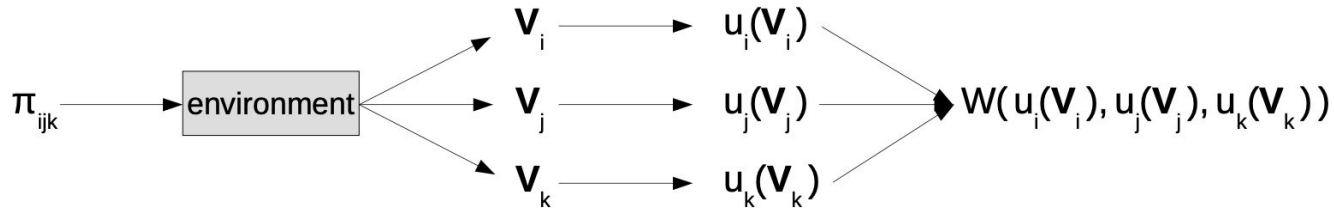
- Multi-objective Wolfpack environment – two predators (blue) must catch a prey (red)
- Separate rewards for team capture and lone capture
- Changing the agents' preferences for team/lone captures influences the team capture rate

O'Callaghan, D. and Mannion, P., 2021, May. Tunable behaviours in sequential social dilemmas using multi-objective reinforcement learning. In *Proceedings of the 20th international conference on autonomous agents and multiagent systems* (pp. 1610-1612).

3.4 Team Reward - Social Choice (Roxana)



3.5 Individual Reward - Social Choice



Generalised Toll-based Q-learning

- Considers the toll-based route choice problem, where self-interested agents repeatedly commute attempting to minimise their travel times and costs
- Realigns agents' heterogeneous preferences over travel time and monetary expenses to obtain a system-efficient equilibrium

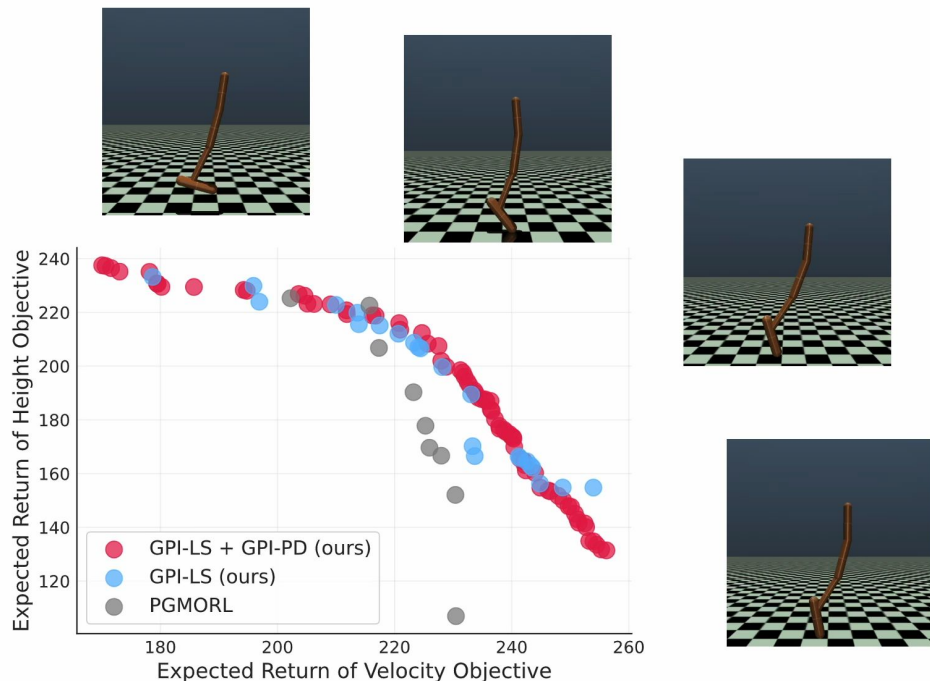
Ramos, G. D. O., Rădulescu, R., Nowé, A., & Tavares, A. R. (2020, May). Toll-based learning for minimising congestion under heterogeneous preferences. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems* (pp. 1098-1106).

Additional work

- Conor F. Hayes, Timothy Verstraeten, Diederik M. Roijers, Enda Howley, Patrick Mannion - Multi-Objective Coordination Graphs for the Expected Scalarised Returns with Generative Flow Models. *In European Workshop on Reinforcement Learning (EWRL 2022)*, Milan, 2022
- Rădulescu, R., Verstraeten, T., Zhang, Y., Mannion, P., Roijers, D. M., & Nowé, A. (2022). Opponent learning awareness and modelling in multi-objective normal form games. *Neural Computing and Applications*, 1-23.
- Röpke, W., Roijers, D. M., Nowé, A., & Rădulescu, R. (2022). Preference communication in multi-objective normal-form games. *Neural Computing and Applications*, 1-26.

Part 4 Tools and open problems

Benchmarks and Tools - MO-Gymnasium






Alegre et al. 2022. MO-Gym: A Library of Multi-Objective Reinforcement Learning Environments. In Proceedings of the 34th Benelux Conference on Artificial Intelligence BNAIC/Benelearn 2022.

Benchmarks and Tools - MO-Gymnasium



Environments

Env	Obs/Action spaces	Objectives	Description
<code>deep-sea-treasure-v0</code> 	Discrete / Discrete	<code>[treasure, time_penalty]</code>	Agent is a submarine that must collect a treasure while taking into account a time penalty. Treasures values taken from Yang et al. 2019 .
<code>resource-gathering-v0</code> 	Discrete / Discrete	<code>[enemy, gold, gem]</code>	Agent must collect gold or gem. Enemies have a 10% chance of killing the agent. From Barret & Narayanan 2008 .
<code>fruit-tree-v0</code> 	Discrete / Discrete	<code>[nutri1, ..., nutri6]</code>	Full binary tree of depth $d=5,6$ or 7 . Every leaf contains a fruit with a value for the nutrients Protein, Carbs, Fats, Vitamins, Minerals and

Benchmarks and Tools - MORL-baselines



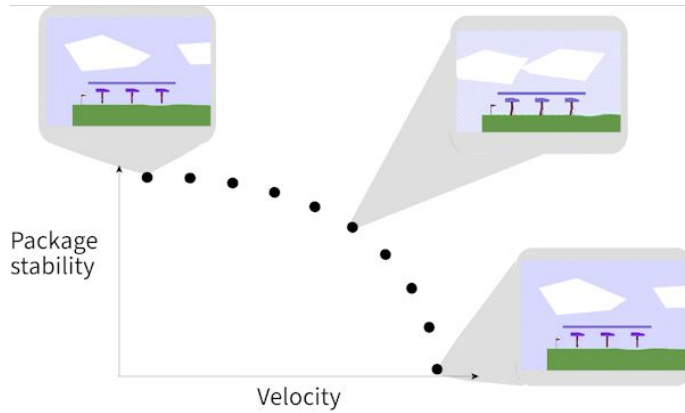
Implemented Algorithms

Name	Single/Multi-policy	ESR/SER	Observation space	Action space	Paper
GPI-LS + GPI-PD	Multi	SER	Continuous	Discrete / Continuous	Paper and Supplementary Materials
MORL/D	Multi	/	/	/	Paper
Envelope Q-Learning	Multi	SER	Continuous	Discrete	Paper
CAPQL	Multi	SER	Continuous	Continuous	Paper
PGMORL ¹	Multi	SER	Continuous	Continuous	Paper / Supplementary Materials
Pareto Conditioned Networks (PCN)	Multi	SER/ESR ²	Continuous	Discrete / Continuous	Paper
Pareto Q-Learning	Multi	SER	Discrete	Discrete	Paper
MO Q Learning	Single	SER	Discrete	Discrete	Paper
MPMOQLearning (outer loop MOQL)	Multi	SER	Discrete	Discrete	Paper
Optimistic Linear Support (OLS)	Multi	SER	/	/	Section 3.3 of the thesis

Felten, F., Alegre, L. N., Nowe, A., Bazzan, A., Talbi, E. G., Danoy, G., & C da Silva, B. (2024). A toolkit for reliable benchmarking and research in multi-objective reinforcement learning. *Advances in Neural Information Processing Systems*, 36.

MOMAland – MOMADM benchmarks

- Open source Python library for developing and comparing multi-objective multi-agent reinforcement learning algorithms



 MOMAland

<https://momaland.farama.org/>

MOMAland – MOMADM benchmarks

Domain	# of agents	# of objectives	stochastic transitions?	full observability?	partial observability possible?	team rewards possible?	individual rewards possible?	discrete/continuous state (d/c)	discrete/continuous actions (d/c)
MO-BPD	2-n	2	✗ ³	✗	✓	✓	✓	d	d
MO-ItemGathering	2-n	2-d	✗ ³	✓	✗	✗	✓	d	d
MO-GemMining	2-n	2-d	✗ ³	-	-	✗	✓	-	d
MO-RouteChoice	2-n	2	✗ ³	-	-	✗	✓	-	d
MO-PistonBall	2-n	3	✗ ³	✗	✓	✗	✓	c	d/c
MO-MW-Stability	2-n	2	✗	✓	✓	✓	✓	c	c
CrazyRL/Surround	2-n	2	✗	✓	✗	✓	✓	c	c
CrazyRL/Escort	2-n	2	✗	✓	✗	✓	✓	c	c
CrazyRL/Catch	2-n	2	✓	✓	✗	✓	✓	c	c
MO-Breakthrough	2	1-4	✗	✓	✗	✗	✓	d	d
MO-Connect4	2	2-20	✗	✓	✗	✗	✓	d	d
MO-Ingenious	2-6	2-6	✗	✓	✓	✓	✓	d	d
MO-SameGame	1-5	2-10	✗ ³	✓	✗	✓	✓	d	d



MOMAland

<https://momaland.farama.org/>

Open questions

- Solution concepts for the axiomatic approach to MOMA problems
- Results for more complex (e.g., sequential, partially observable) settings
- Integrated pipelines for planning -> negotiation -> execution
- Utility modelling, e.g., inferring preferences of other agents via demonstrations or opponent modelling
- Strategic disclosure of utility information to the other agents

Multi-Objective Decision Making Workshop

- Tomorrow
- Full day workshop
- Location: V1 Ferrol at the School of Communication Sciences



<https://modem2024.vub.ac.be/>

Thank you for listening

- Questions?
- Any additional questions drop us a message at:
 - dixitg@oregonstate.edu
 - r.t.radulescu@uu.nl
 - patrick.mannion@universityofgalway.ie

This tutorial was based (primarily) on

- Hayes, C. F., Rădulescu, R., Bargiacchi, E., Källström, J., Macfarlane, M., Reymond, M., ... & Roijers, D. M. (2022). A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems*, 36(1), 26.
- Rădulescu, R., Mannion, P., Roijers, D. M., & Nowé, A. (2020). Multi-objective multi-agent decision making: a utility-based analysis and survey. *Autonomous Agents and Multi-Agent Systems*, 34(1), 1-52.
- Rădulescu, R., Mannion, P., Zhang, Y., Roijers, D. M., & Nowé, A. (2020). A utility-based analysis of equilibria in multi-objective normal-form games. *The Knowledge Engineering Review*, 35.
- Rădulescu, R. (2021). *Decision Making in Multi-Objective Multi-Agent Systems: A Utility-Based Perspective*. Brussels: Crazy Copy Center Productions.
- Roijers, D. M., Vamplew, P., Whiteson, S., & Dazeley, R. (2013). A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research*, 48, 67-113.
- Röpke, W., Roijers, D. M., Nowé, A., & Rădulescu, R. (2022). On Nash equilibria in normal-form games with vectorial payoffs. *Autonomous Agents and Multi-Agent Systems*, 36(2), 53.
- Röpke, W., Groenland, C., Rădulescu, R., Nowé, A., & Roijers, D. M. (2023). Bridging the Gap Between Single and Multi Objective Games. *AAMAS 2023*.