DIEDERIK M. ROIJERS

DIEDERIK.YAMAMOTO-ROIJERS@HU.NL, HTTP://ROIJERS.INFO

Multiple objectives: why, how and what now?

ABOUT ME

HU University of Applied Sciences Utrecht

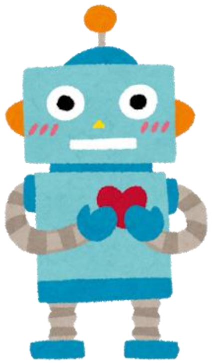
Microsystems Technology

Institute of ICT

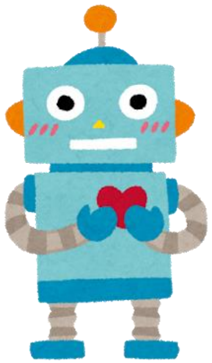
Vrije Universiteit Brussel

AI Research Group

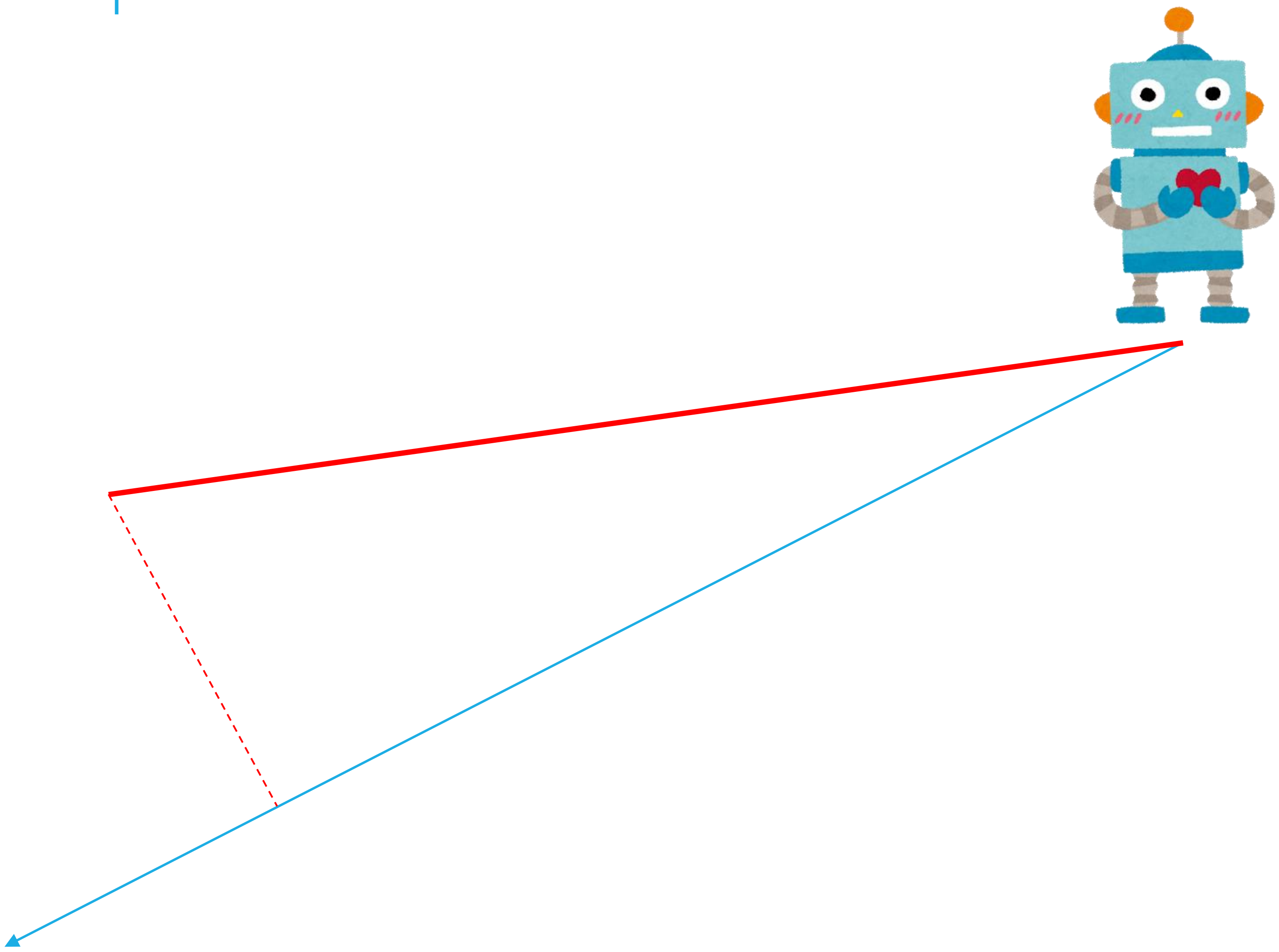
ROBOT



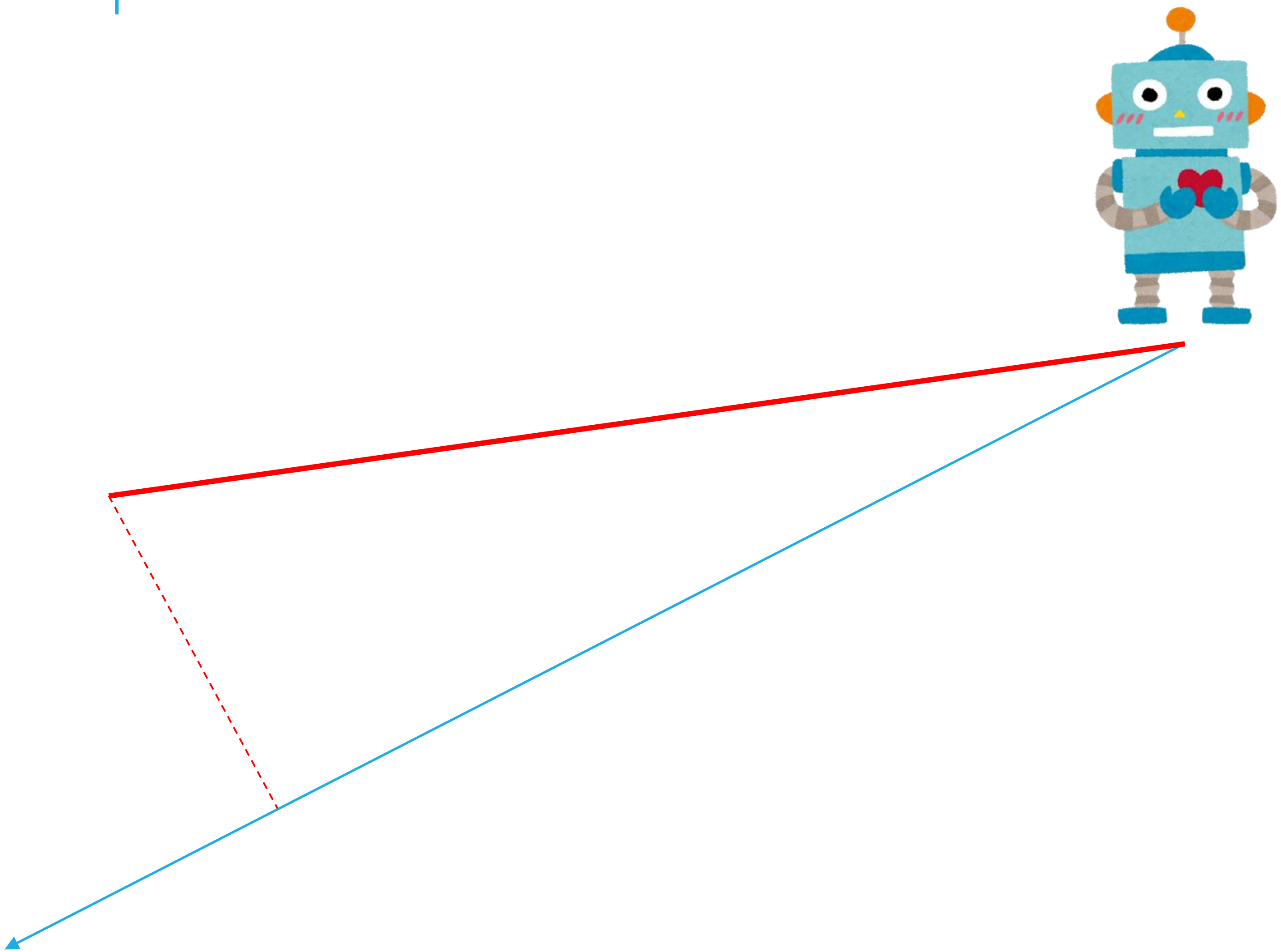
ROBOT



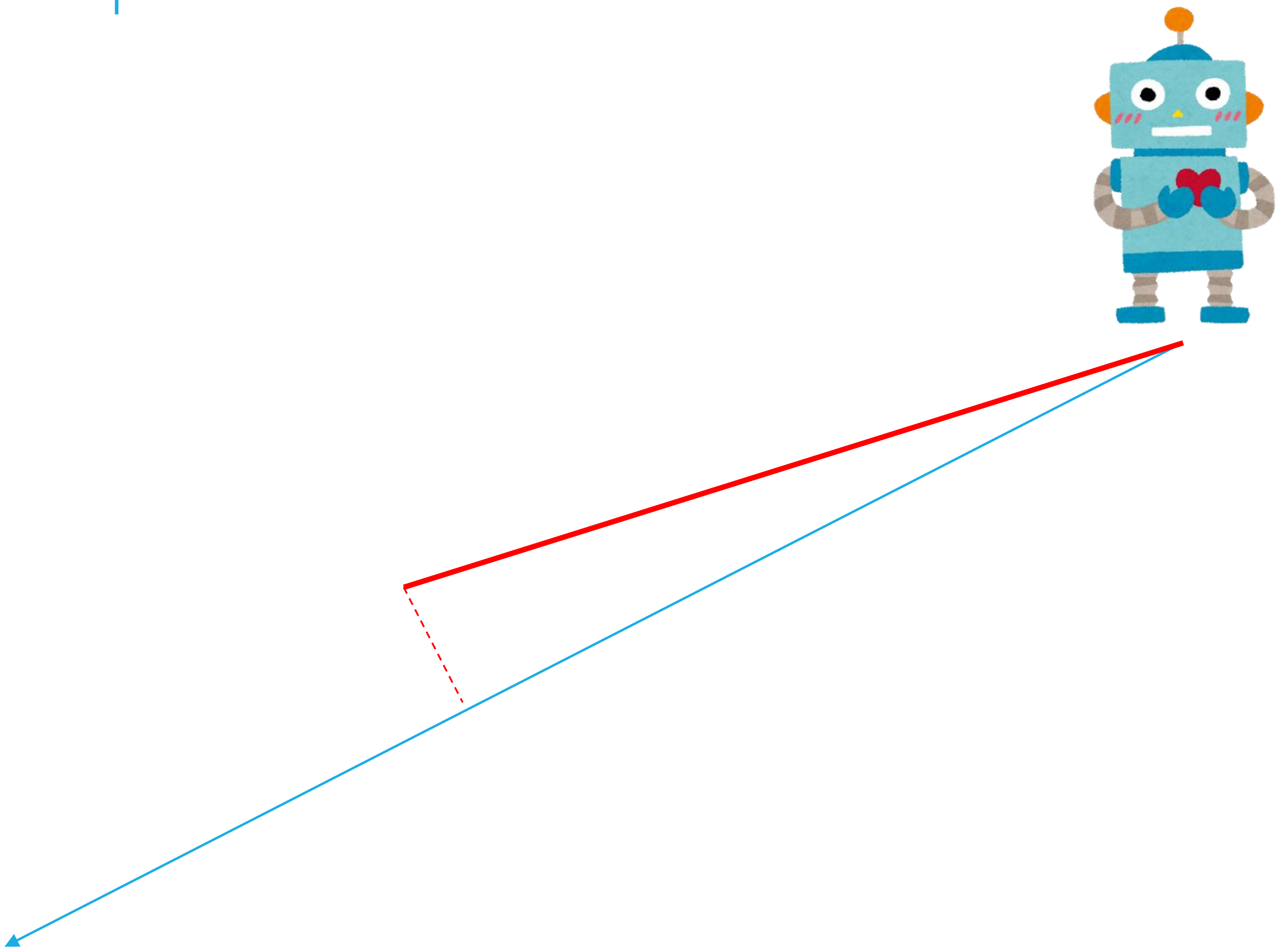
ROBOT



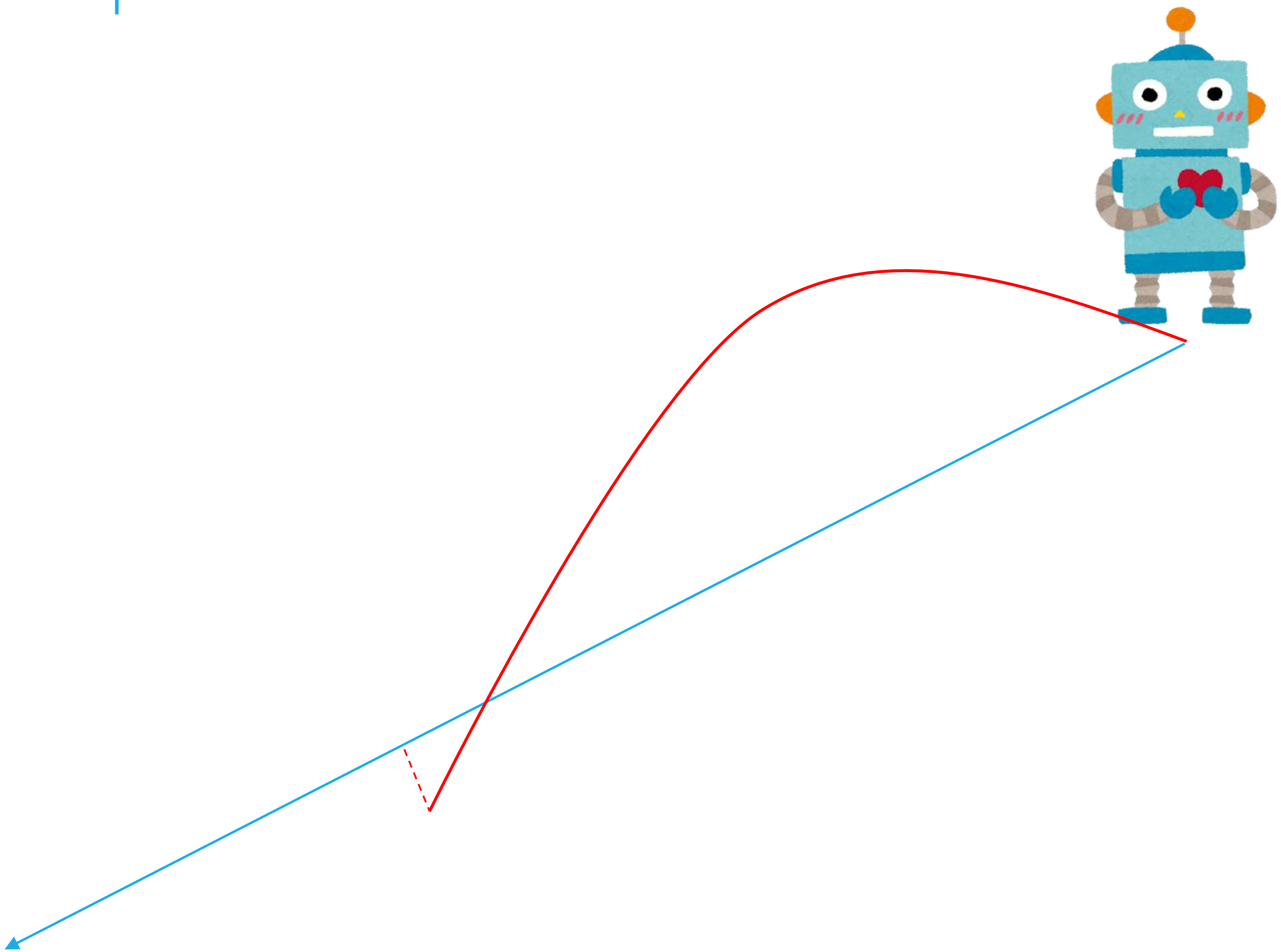
ROBOT



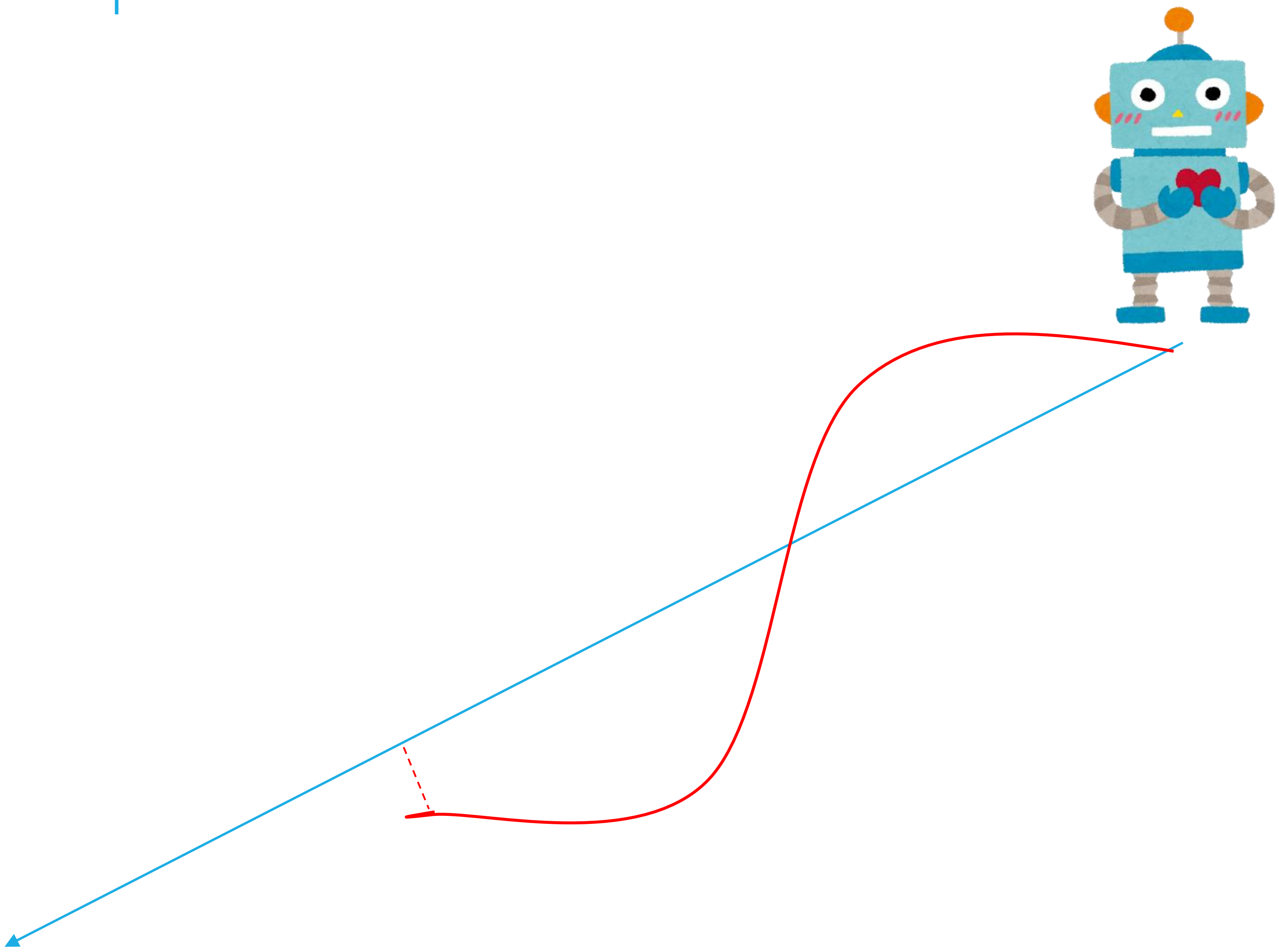
ROBOT



ROBOT



ROBOT



ROBOT

It's never easy!

Move 30 sec in direction x: three objectives?

1. max projected length
2. min angle end point
3. min path length to get to end point

Engineering the a reward function until…

ROBOT

It’s never easy!

Move 30 sec in direction x: three objectives?

1. max projected length
2. min angle end point
3. min path length to get to end point

Engineering the a reward function until… it works…

ROBOT

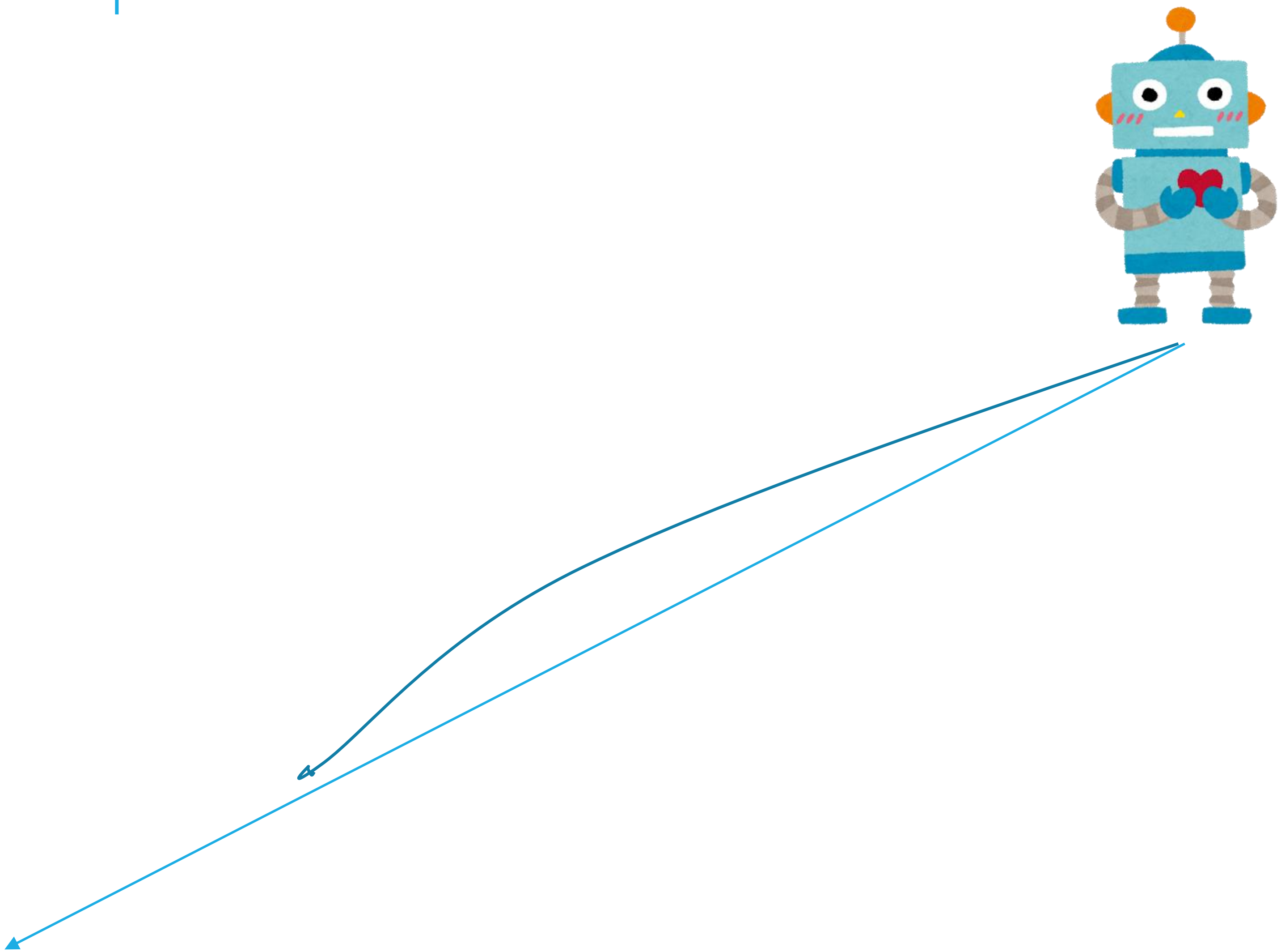
It's never easy!

Move 30 sec in direction x: three objectives?

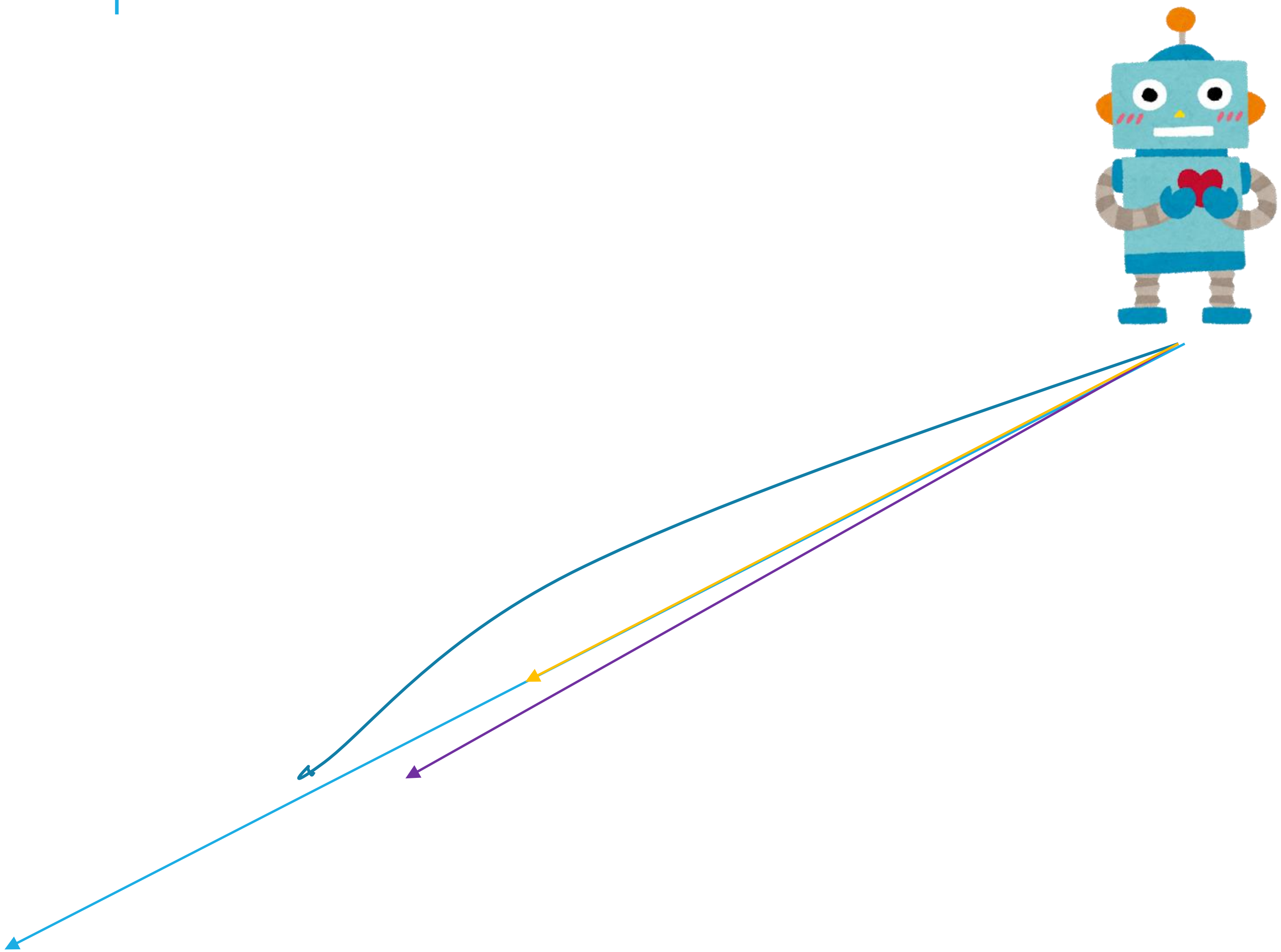
1. max projected length
2. min angle end point
3. min path length to get to end point

Engineering the a reward function until… it works… sort of…

ROBOT



WAS THIS REALLY THE BEST?!



ROBOT STORY MORAL

Even simple problems have multiple objectives

- Bryce et al 2007: probabilistic planning *is* multi-objective

Engineering single-objective reward function is a semi-blind process

Single-objective reward functions make implicit decisions about what is optimal (without explicitly reasoning about it)

ROBOT STORY MORAL

Even simple problems have multiple objectives

- Bryce et al 2007: probabilistic planning *is* multi-objective

Engineering single-objective reward function is a semi-blind process

Single-objective reward functions make implicit decisions about what is optimal (without explicitly reasoning about it)

… might be okay, but we don’t know?

MORAL IMPLICATIONS

Self-driving cars? Robots in human environments? Insurance intake?

Is it even ethical to take a single-objective approach?

Human-aligned AI is a multi-objective problem (Vamplew et al., 2018)

WHEN THE STAKES ARE HIGH

We really need to see the alternatives

We really don’t want the designers/engineers of algorithms deciding what the (ethically / socially) optimal thing to do is

We need to be able to adjust in the face of new situations

The responsible people need to take the shots, not the AI (researchers)

(I’ll get back to this.)

MULTI-OBJECTIVE PROBLEMS

Formalisation

Utility-based approach

FROM THE MORALS TO MORL

- Vector-valued reward/value functions
- Meaningful objectives:
  - easy to define
  - easy to interpret the results

$$\mathbf{V} : \Pi \to \mathbb{R}^n$$

MULTI-OBJECTIVE MARKOV DECISION PROCESS

MOMDP
$$\mathbf{R}_t = \sum_{k=0}^{\infty} \gamma^k \mathbf{r}_{t+k+1}$$ $$\mathbf{V}^{\pi}(s) = E[\mathbf{R}_t \mid \pi, s_t = s]$$ $$\mathbf{V}^{\pi}(s) = \sum_a \pi(s,a) \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma \mathbf{V}^{\pi}(s')]$$

DECISION MAKERS

"Owners" of the utility

Utility-based approach $u :: \mathbb{R}^n \to \mathbb{R}$

Utility function can be implicit or explicit

Monotonically increasing in all objectives

MULTI-OBJECTIVE DECISION MAKING

Necessary when scalarising the problem with the utility function a priori is impossible, infeasible, or undesirable

- unknown / uncertain
- not explicit
- changeable / subject to adjustments
- subject to review

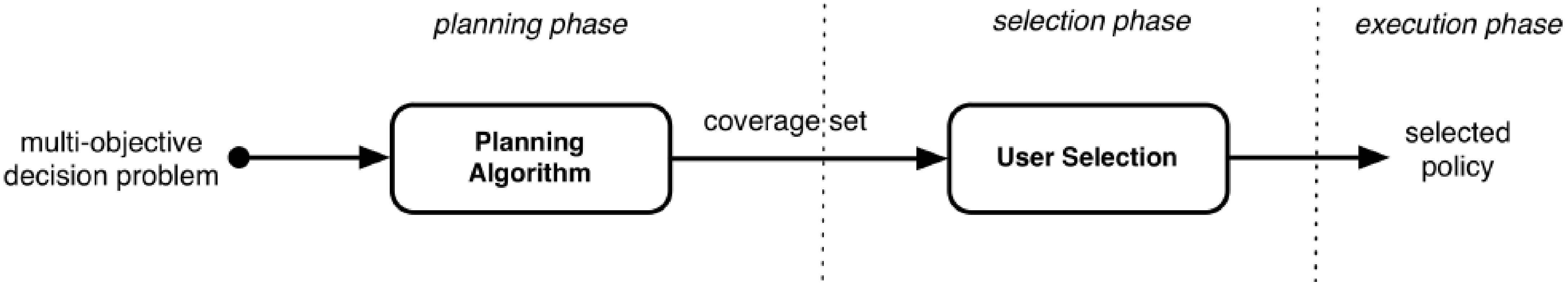
MULTI-OBJECTIVE DECISION MAKING

Necessary when scalarising the problem with the utility function a priori is impossible, infeasible, or undesirable

- unknown / uncertain
- not explicit
- changeable / subject to adjustments
- subject to review

so basically always

DECISION SUPPORT SCENARIO



WHAT IS OPTIMAL?

Decision support scenario

We don’t fully know $u :: \mathbb{R}^n \rightarrow \mathbb{R}$

At least one optimal solution for all possible u

within the *allowed set of policies* (search/policy space)

Coverage set

DERIVE COVERAGE SET

1. Multi-objective scenario

- Known utility function: single policy
- (Partially) unknown utility function / decision support: multiple policies

2. Properties of utility function

- Linear
- Monotonically increasing

3. Allowable policies

- Deterministic
- Stochastic

TAXONOMY AND LESSONS LEARNT

Optimal solution sets

Assumptions

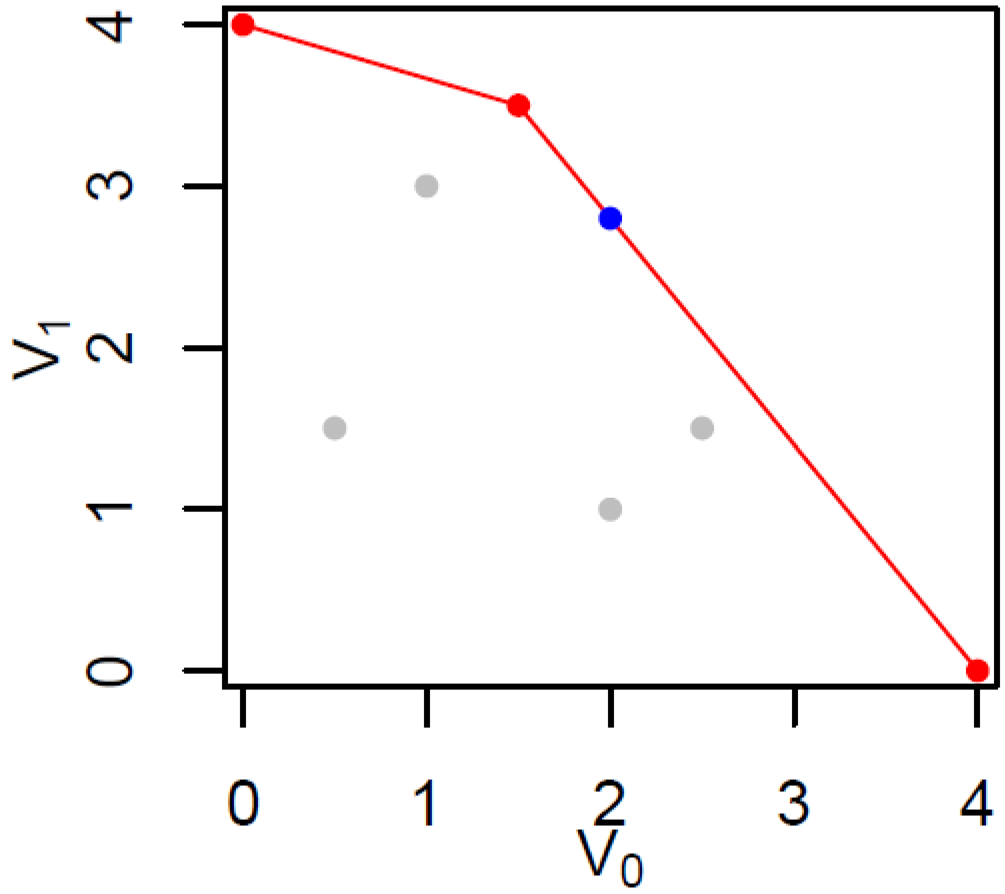
Settings

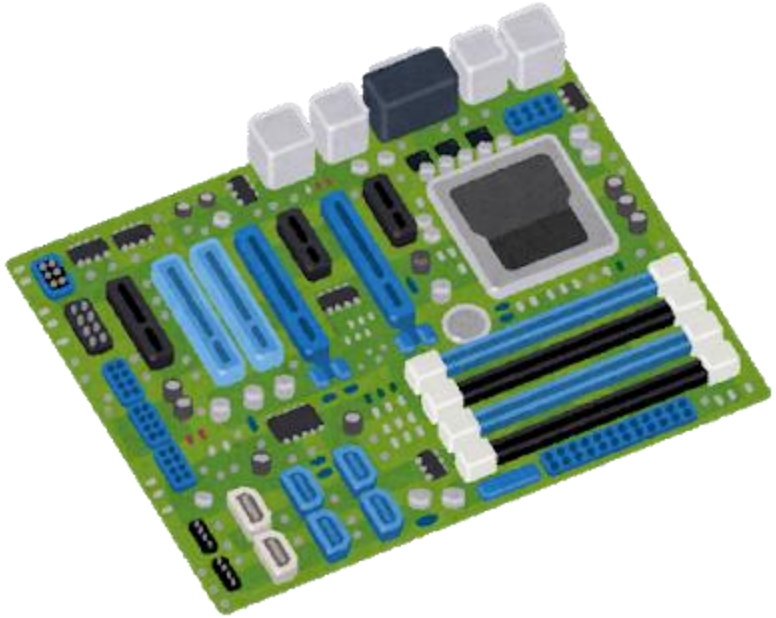
Positioning, positioning, positioning

TAXONOMY

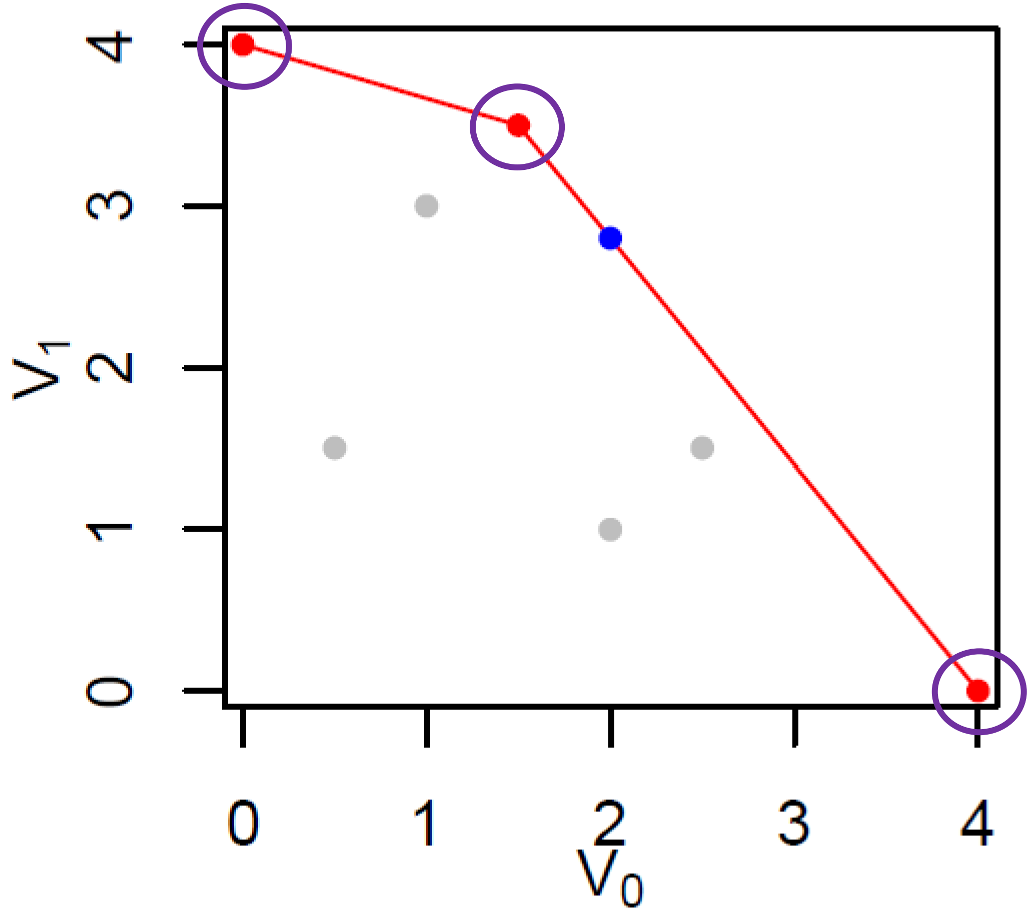
<table>
<tr><th></th><th colspan="2">single policy</th><th colspan="2">multiple policies</th></tr>
<tr><th></th><th>deterministic</th><th>stochastic</th><th>deterministic</th><th>stochastic</th></tr>
<tr><td>linear u</td><td colspan="2">one deterministic stationary policy</td><td colspan="2">convex coverage set of deterministic stationary policies</td></tr>
<tr><td>possibly non-linear u (monotonically increasing)</td><td>one deterministic non-stationary policy</td><td>one mixture policy of two or more deterministic stationary policies</td><td>Pareto coverage set of deterministic non-stationary policies</td><td>convex coverage set of deterministic stationary policies</td></tr>
</table>

CCS VERSUS PCS (PARETO FRONT)



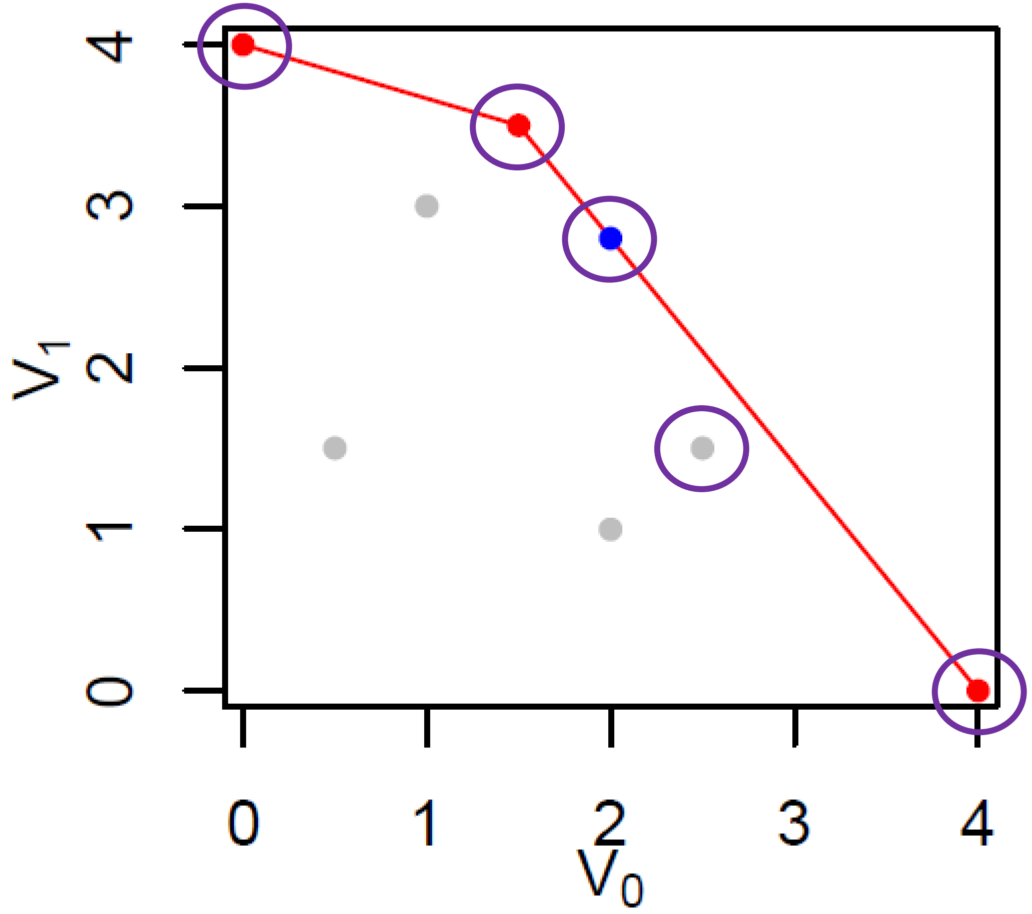


CCS VERSUS PCS (PARETO FRONT)



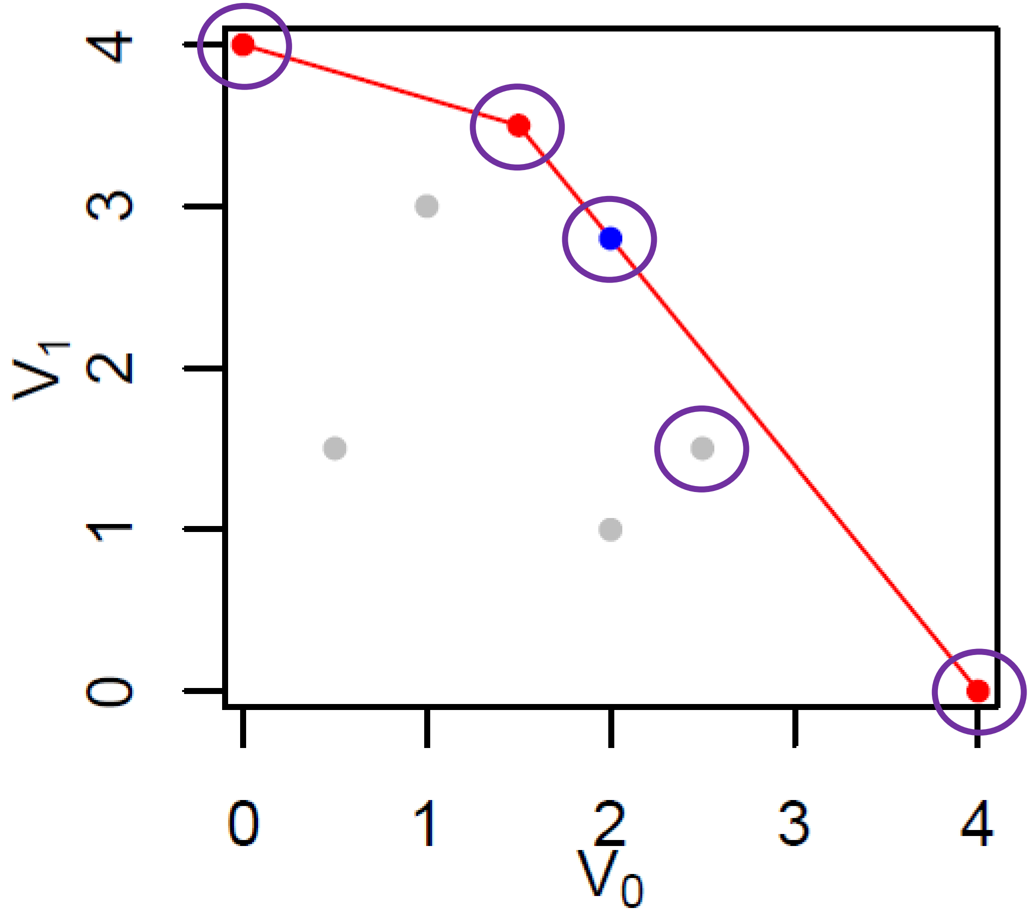
$$u_{\boldsymbol{w}}(\boldsymbol{V}) = \boldsymbol{w} \cdot \boldsymbol{V}$$

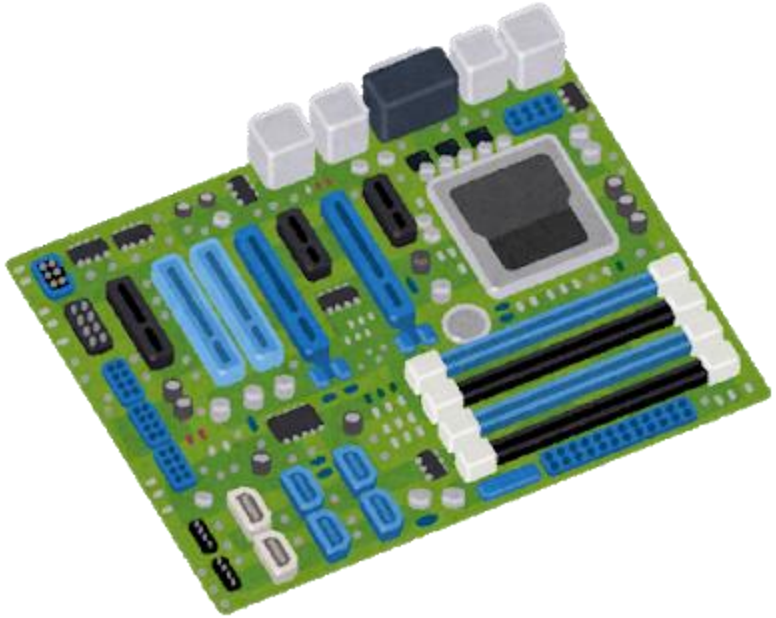
CCS VERSUS PCS (PARETO FRONT)



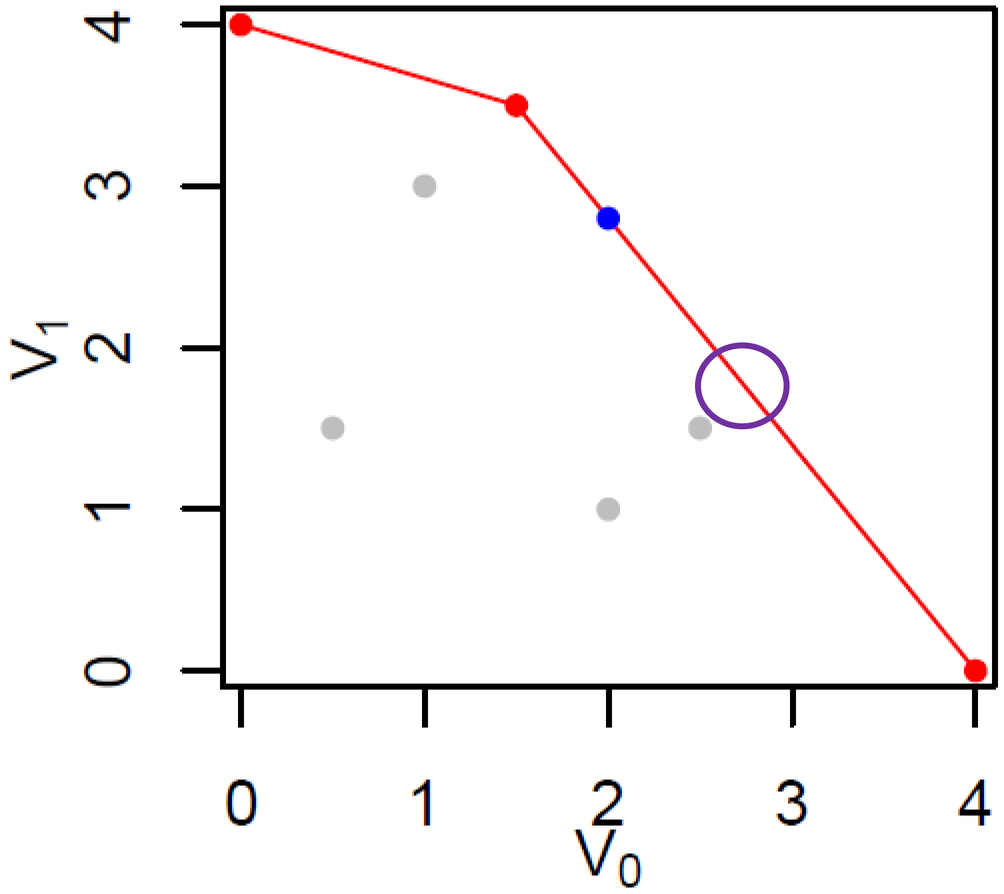
Non-linear u,
deterministic policies

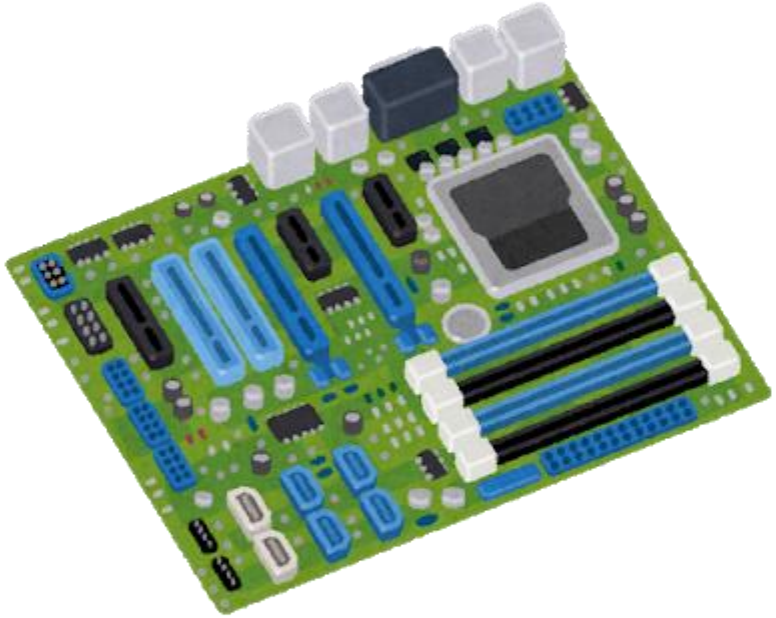
CCS VERSUS PCS (PARETO FRONT)





CCS VERSUS PCS (PARETO FRONT)

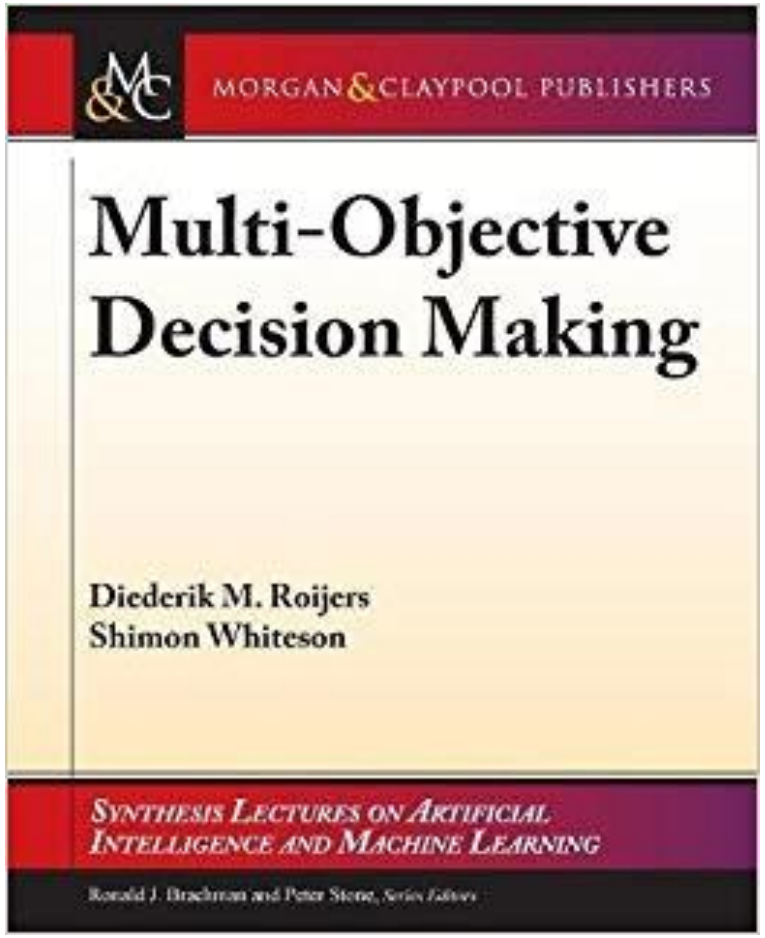




Stochastic policies are often OK

No Pareto front needed

FOR PRECISE DEFINITIONS SEE



Diederik M. Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley - A Survey of Multi-Objective Sequential Decision-Making. *Journal of Artificial Intelligence Research*, 48:67–113, 2013.

TAXONOMY

	single policy		multiple policies	
	deterministic	stochastic	deterministic	stochastic
linear u	one deterministic stationary policy		convex coverage set of deterministic stationary policies	
possibly non-linear u (monotonically increasing)	one deterministic non-stationary policy	one mixture policy of two or more deterministic stationary policies	Pareto coverage set of deterministic non-stationary policies	convex coverage set of deterministic stationary policies

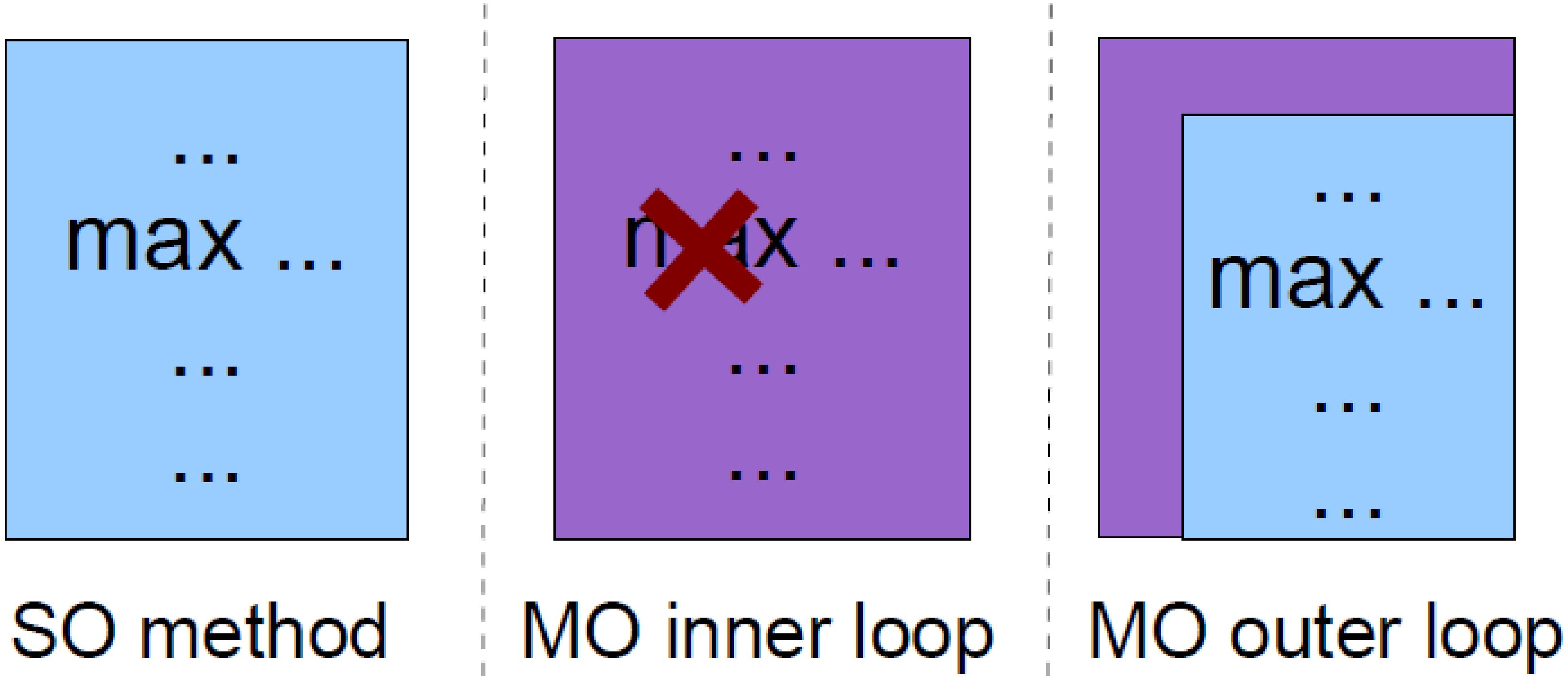
CONVEX COVERAGE SETS

Viable in a lot of problems if stochastic policies are allowed

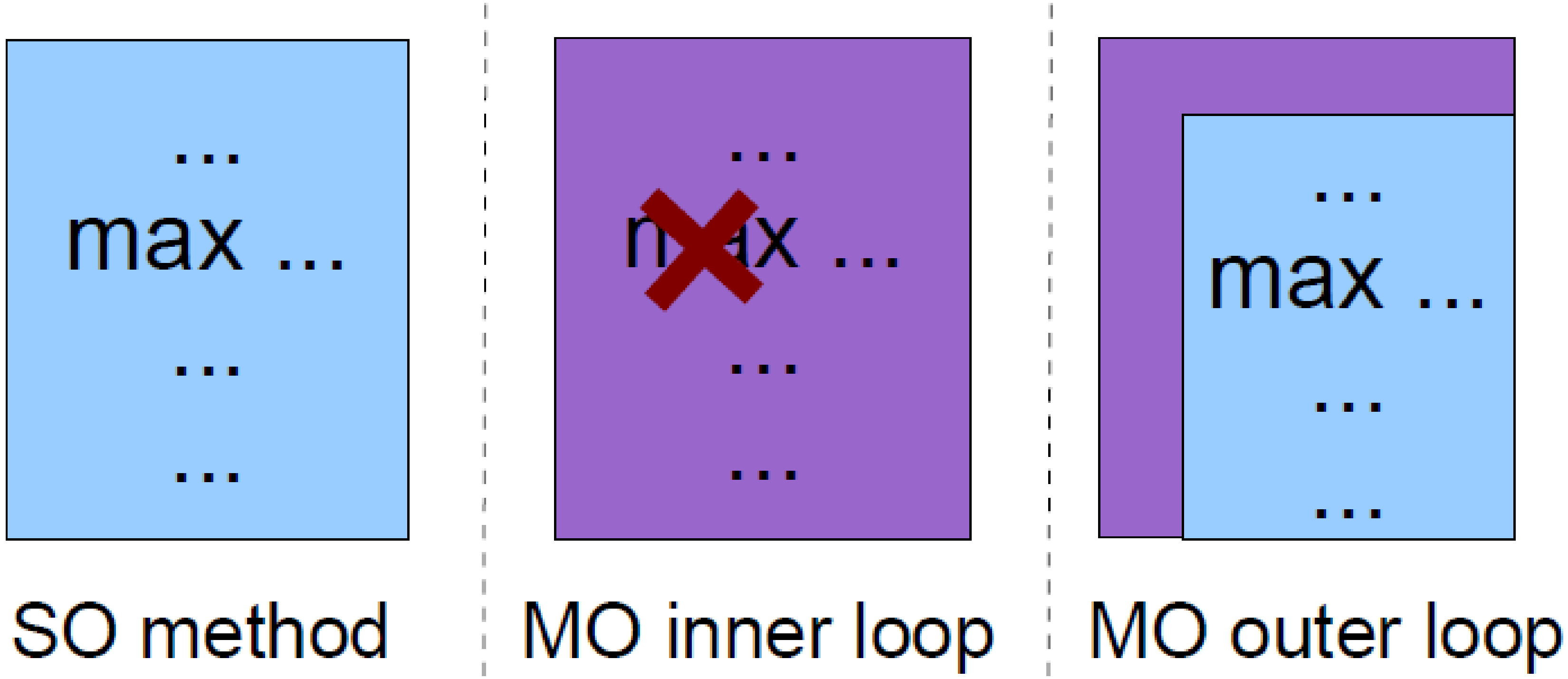
Linear utility functions distribute over expectations: for known weights single-objective methods still work. Very convenient!

Under linear utility functions, POMDPs are a mathematically equivalent superclass.
No need to prove much (!) (convergence, etc.) Can take inspiration from POMDP methods.

INNER LOOP VERSUS OUTER LOOP



INNER LOOP VERSUS OUTER LOOP



Outer loop methods (CCS) are easier, and faster for 2, 3 objectives

Inner loop methods scale better in the numbers of objectives

BACK UP: LESSONS LEARNT

Utility-based approach: derive your optimal set

Helps to position the paper

Positioning is important; useful methods, theory, and tricks can be used depending on it.

WHERE NEED WE GO FROM HERE

Particularism

We need to change what we think is optimal

Non-static AI is multi-objective

WHAT DOES MO ENABLE US TO DO

Reason about problems in a natural way (in meaningful statistics)

Inform human decision makers about viable alternatives

Adjust to changes in utility judgements

Helps us engineer AI solutions

Helps us make application of AI viable

Helps us make AI long-lived

NECESSITY

AI has an ever stronger impact

NECESSITY

AI has an ever stronger impact

So I don't trust researchers and engineers to make the trade-offs between important objectives

NECESSITY

AI has an ever stronger impact

So I don’t trust researchers and engineers to make the trade-offs between important objectives → ethical perspective

And I don’t trust anybody to get it right in one go

SELF-DRIVING CAR: ACCIDENT AVOIDANCE

AI takes risks with driver's life to save the life of a child running onto the street, and may cause damage to parked vehicles

What is fair?

How much risk is acceptable?

SELF-DRIVING CAR: ACCIDENT AVOIDANCE

AI takes risks with driver's life to save the life of a child running onto the street, and may cause damage to parked vehicles

What is fair?

How much risk is acceptable?

I don't know!

SELF-DRIVING CAR: ACCIDENT AVOIDANCE

This is the domain of decision-makers that are typically not the people that design the algorithms.

But algorithms do need to take immediate action

It will make trade-offs between objectives

Were those okay?

Review and adjust

PARTICULARIST ETHICS AND MO

What the ethically optimal course of action is, is determined the particular relevant factors in each situation. It is always possible to add factors that change the optimal action.

PARTICULARIST ETHICS AND MO

What the ethically optimal course of action is, is determined the particular relevant factors in each situation. It is always possible to add factors that change the optimal action.

“Pessimistic” (safe) view: the utility function depends on the domain and situations in which we apply the AI

PARTICULARIST ETHICS AND MO

What the ethically optimal course of action is, is determined the particular relevant factors in each situation. It is always possible to add factors that change the optimal action.

"Pessimistic" (safe) view: the utility function depends on the domain and situations in which we apply the AI

New objectives may arise!

MO CHALLENGES

We need:

- systems that model objectives explicitly

- that can interact with decision makers

- who may change the definition, and even the number of objectives

MO CHALLENGES

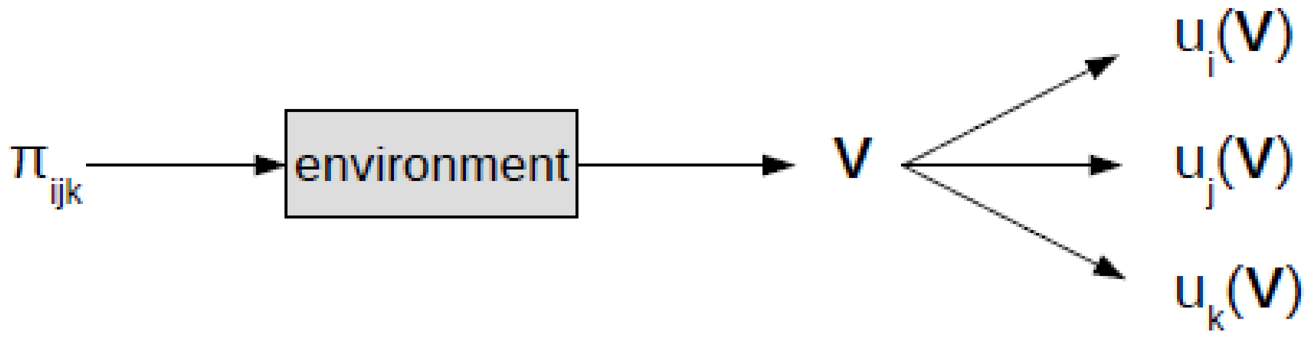
We need:

- systems that model objectives explicitly
- that can interact with decision makers
- who may change the definition, and even the number of objectives
- we cannot currently do this… at all
- we need to extend our test horizons, long-term utility

FINAL REMARKS

Multi-agent settings
Acknowledgements
SER vs ESR
Interactive settings

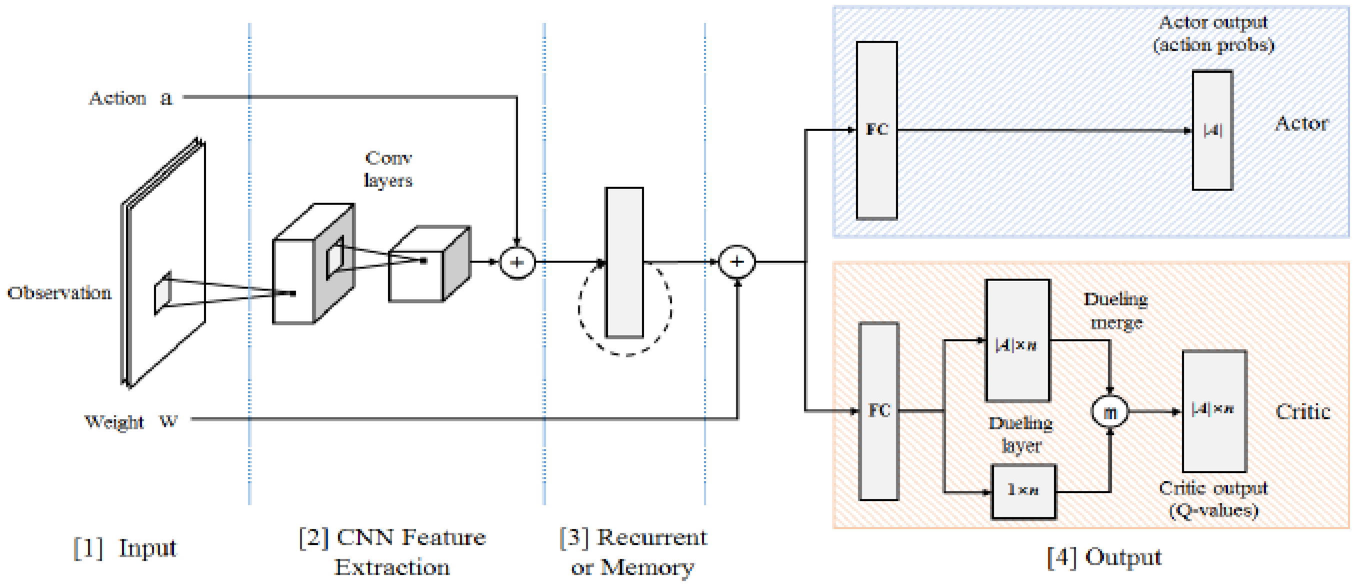
MULTI-AGENT SETTINGS



Check out: Roxana Rădulescu's talks at ALA and AAMAS

- Roxana Rădulescu, Patrick Mannion, Diederik M. Roijers, Ann Nowé - Multi-Objective Multi-Agent Decision Making: A Utility-based Analysis and Survey. *Autonomous Agents and Multi-Agent Systems* (JAAMAS), **34**, 10 (2020). Special issue on New Horizons in Multiagent Learning.
- Yijie Zhang, Roxana Rădulescu, Patrick Mannion, Diederik M. Roijers, Ann Nowé - Opponent Modelling for Reinforcement Learning in Multi-Objective Normal Form Games, In *Proceedings of the Nineteenth International Joint Conference on Autonomous Agents and Multiagent Systems*, May 2020

DEEP PARTIALLY OBSERVABLE MORL



Check out:

- Xiaodong Nian, Athirai A. Irissappane, Diederik M. Roijers - DCRAC: Deep Conditioned Recurrent Actor-Critic for Multi-Objective Partially Observable Environments. In: *AAMAS 2020: Proceedings of the Nineteenth International Joint Conference on Autonomous Agents and Multi-Agent Systems*, May 2020

MO PAPERS AT ALA

#31 Conor F Hayes, Enda Howley and Patrick Mannion - *Dynamic Thresholded Lexicographic Ordering*

#28 Peter Vamplew, Cameron Foale and Richard Dazeley – *A Demonstration of Issues with Value-Based Multiobjective Reinforcement Learning Under Stochastic State Transitions*

MANY THANKS

Roxana Rădulescu, Zoltan Istvan Zardai, Patrick Mannion, Ann Nowé, Peter Vamplew, Richard Dazeley, Luisa M. Zintgraf, Frans Oliehoek, Shimon Whiteson, Denis Steckelmacher, Eugenio Bargiacchi, Hélène Plisnier, Pieter Libin, Timothy Verstraeten, Matthieu Reymond, Matthijs T.J. Spaan, Mathijs de Weerdt, Joris Scharpff, Dirk Sierag, Maarten Inja, Chiel Kooijman, Maarten de Waard, Joost van Doorn, Daan Odijk, Maarten de Rijke, Gongjin Lan, Axel Abels, Tom Lennaerts, Felipe Leno Da Silva, Cyntia E.H. Nishida, Anna H. Reali Costa, Xiaodong Nian, Athirai A. Irissappane, Ayumi Igarashi, Yijie Zhang, Dean Webb, Hossam Mossalam, Yannis Assael, Roberta Piscitelli, …

MEDICAL: SER?

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

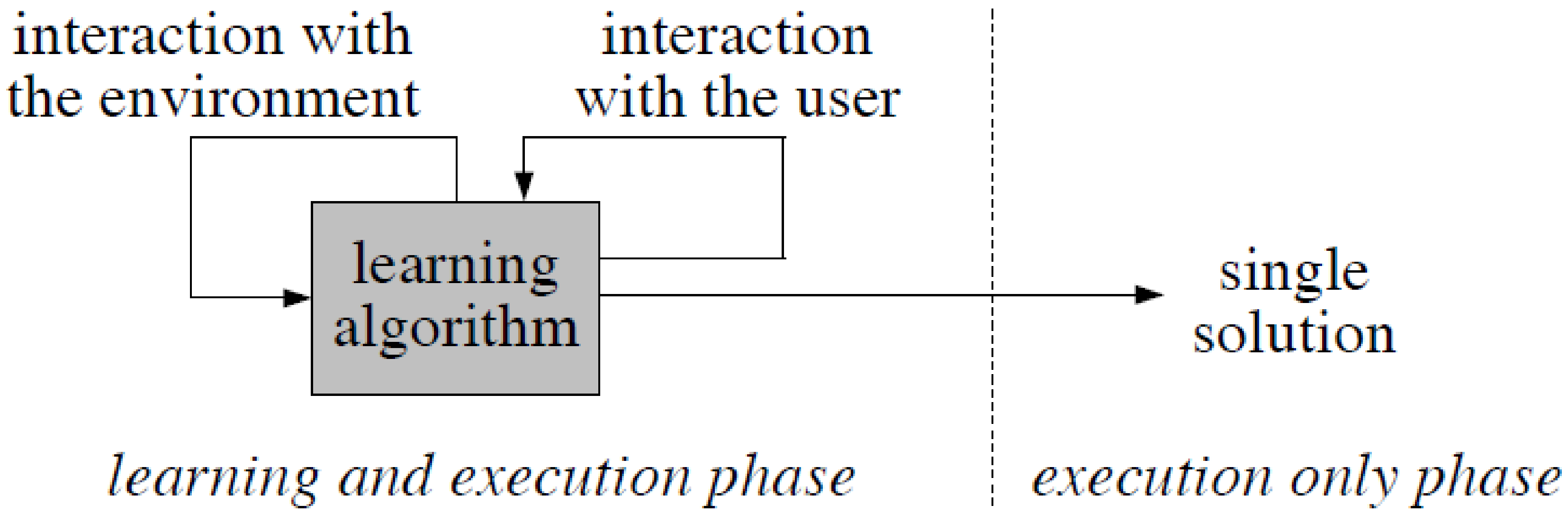
MEDICAL: ESR!

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

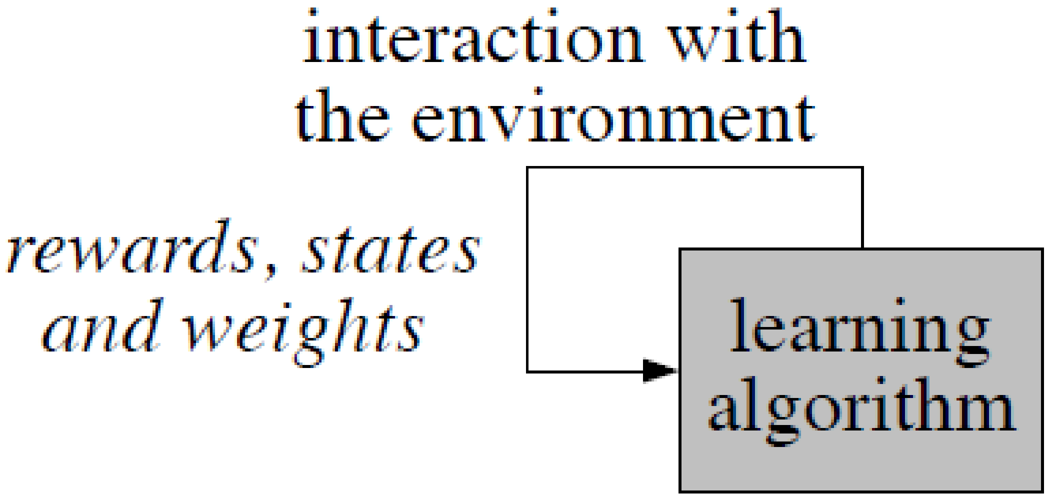
MEDICAL: ESR!

Setting can fundamentally change optimality (again)

INTERACTIVE DECISION SUPPORT



DYNAMIC WEIGHTS



learning and execution phase