Guaranteeing the Learning of Ethical Behaviour through Multi-Objective Reinforcement Learning

ALA 2021 - AAMAS

MANEL RODRÍGUEZ-SOTO

MAITE LÓPEZ-SÁNCHEZ

JUAN ANTONIO RODRÍGUEZ-AGUILAR

Motivation

Current advancements in AI show very impressive results.

Motivation

Current advancements in AI show very impressive results.

We need to guarantee that agents learn to behave ethically, namely in **alignment** with moral **values**.

Motivation

Current advancements in AI show very impressive results.

We need to guarantee that agents learn to behave ethically, namely in **alignment** with **moral values**.

This problem is called the ***value alignment*** problem**.**

Related Work

Common approach: apply Reinforcement Learning (RL) so the agent learns what is ethical by trial and error.

Related Work

Common approach: apply Reinforcement Learning (RL) so the agent learns what is ethical by trial and error.

How: designing an environment that incentivises ethical behaviours (or penalises unethical ones) by means of some exogenous reward function.

Related Work

Common approach: apply Reinforcement Learning (RL) so the agent learns what is ethical by trial and error.

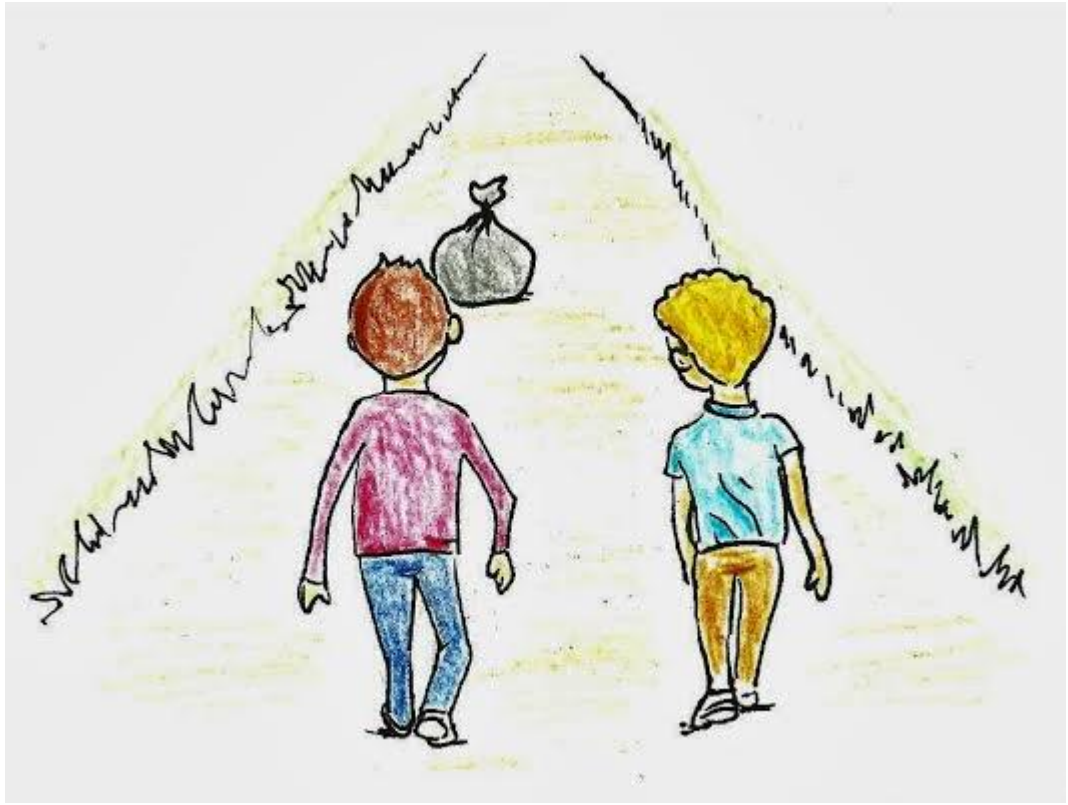
How: designing an environment that incentivises ethical behaviours (or penalises unethical ones) by means of some exogenous reward function.

Problem: lack of theoretical guarantees, we do not know under which circumstances the agent will learn an ethical behaviour or not.

Research Question

How to guarantee that the agent learns to behave ethically while still pursuing its individual objective?

Example: Public Civility Game



Individual objective:

get to the goal as fast as possible.

Example: Public Civility Game

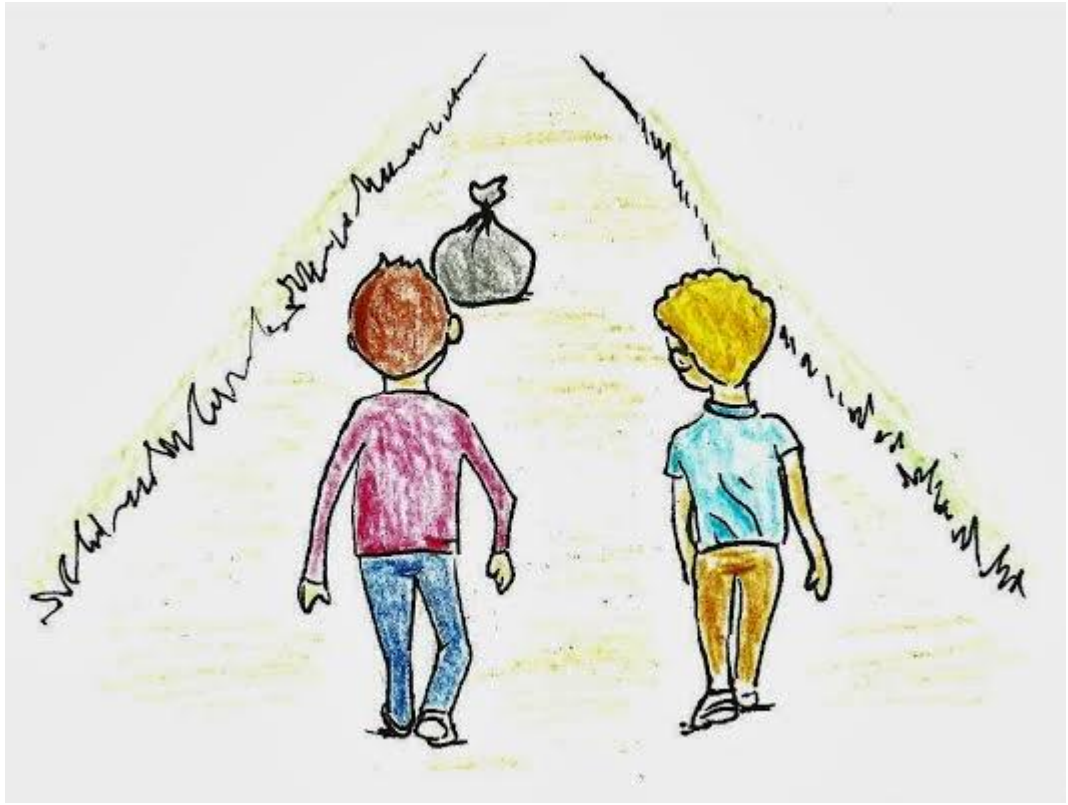
Individual objective:

get to the goal as fast as possible.

Ethical objective:

civility (bring the garbage to the bin).

Example: Public Civility Game



Individual objective:

get to the goal as fast as possible.

Ethical objective:

civility (bring the garbage to the bin).

Actions:

1. Move towards goal.
2. Throw garbage aside.
3. Bring garbage to bin.

Example: Public Civility Game

Each action has different consequences with respect to the two objectives:

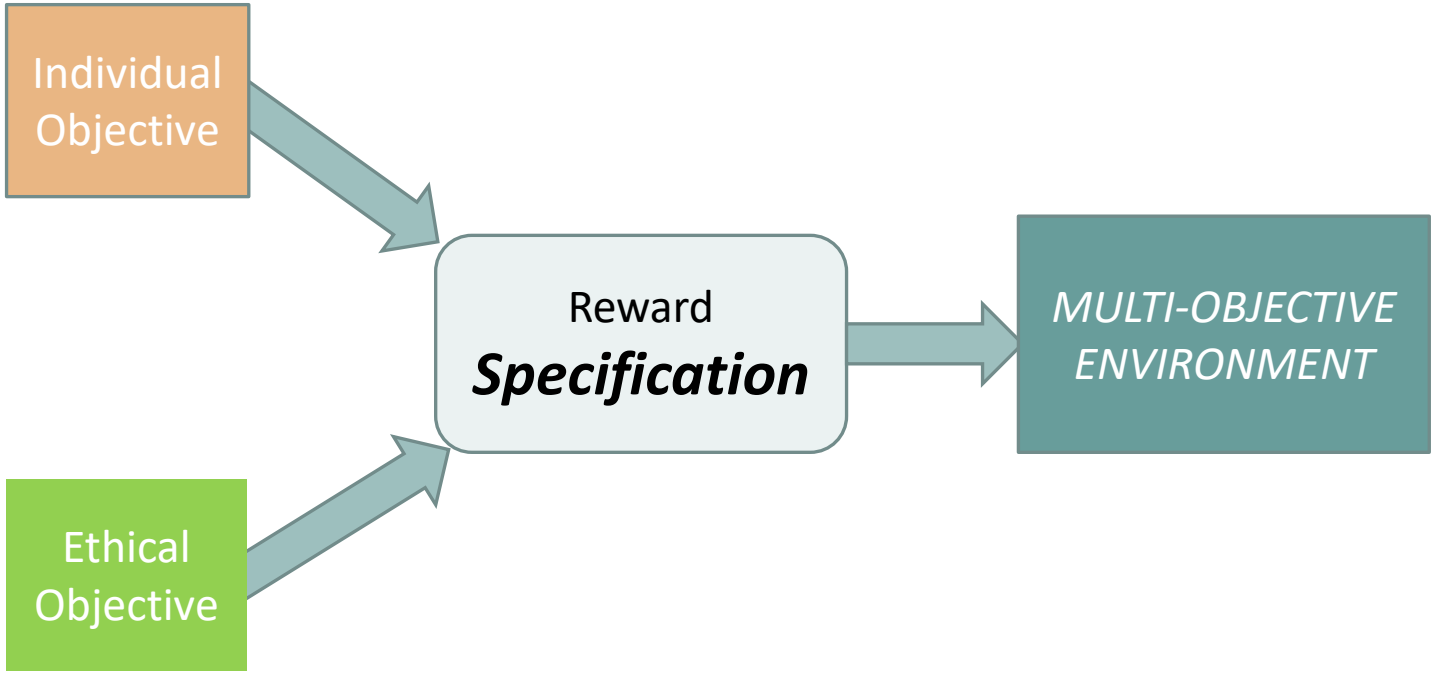
Objectives / Actions	Reach goal	Civility
Throw garbage aside		
Bring garbage to bin		

Example: Public Civility Game

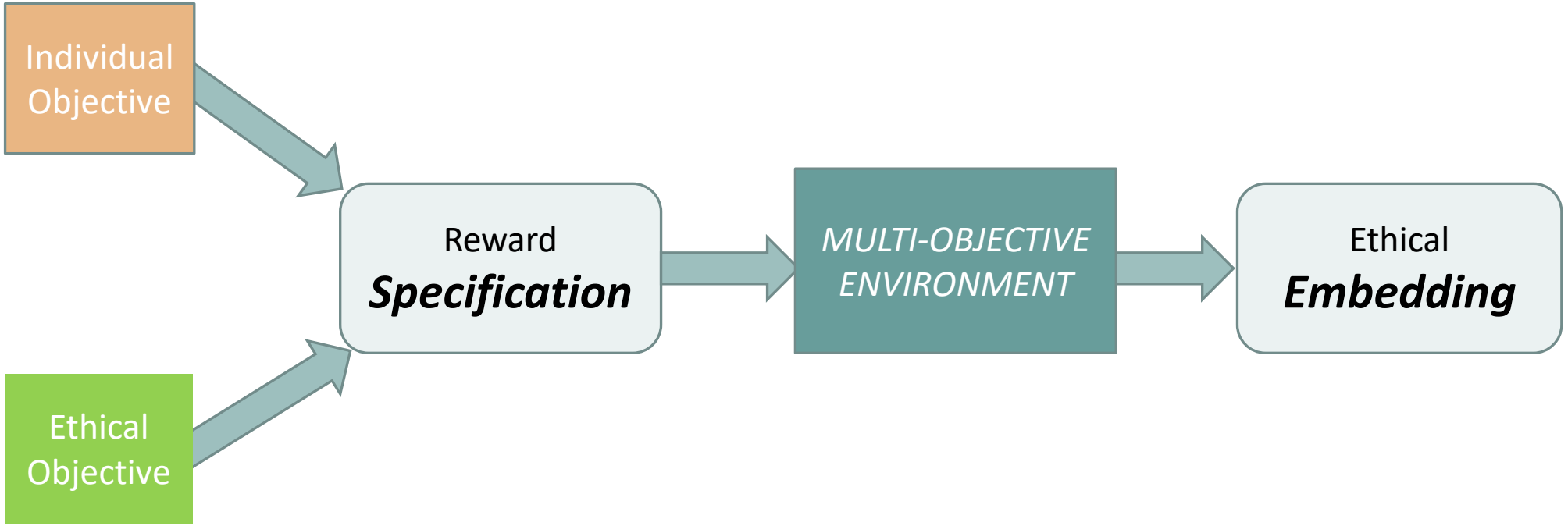
The public civility game is a natural instance of

MULTI-OBJECTIVE DECISION MAKING PROBLEM

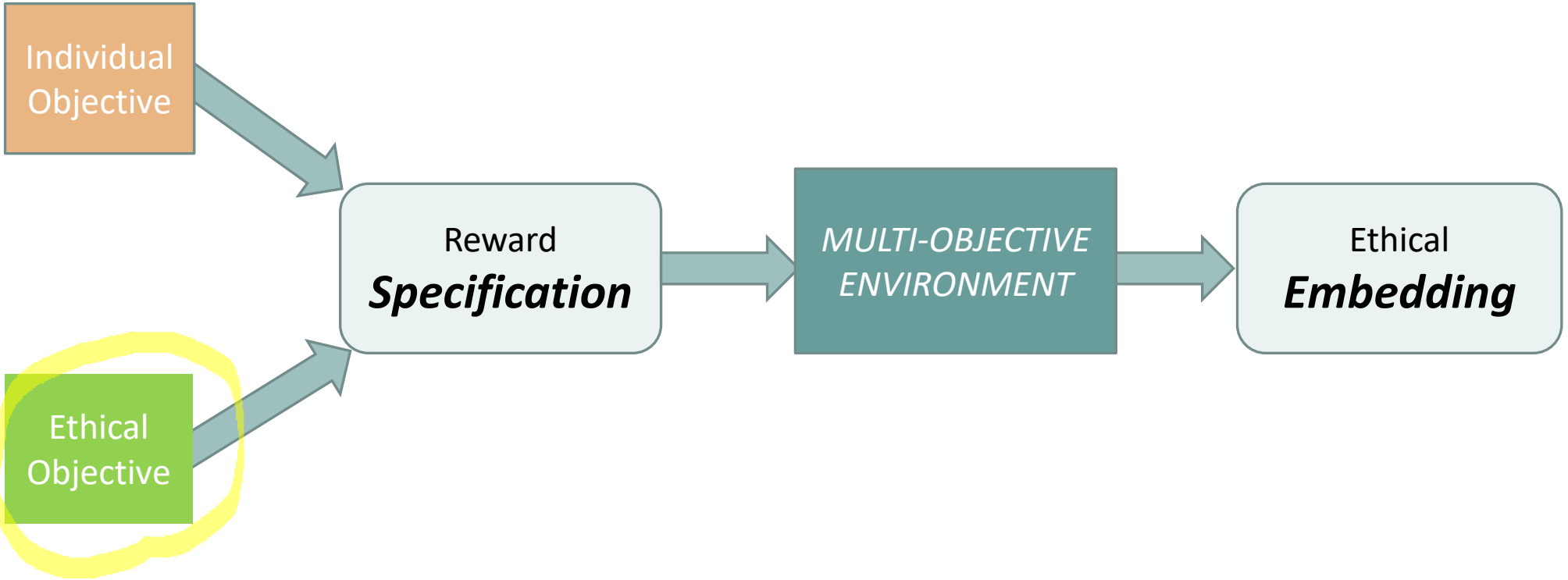
Our approach



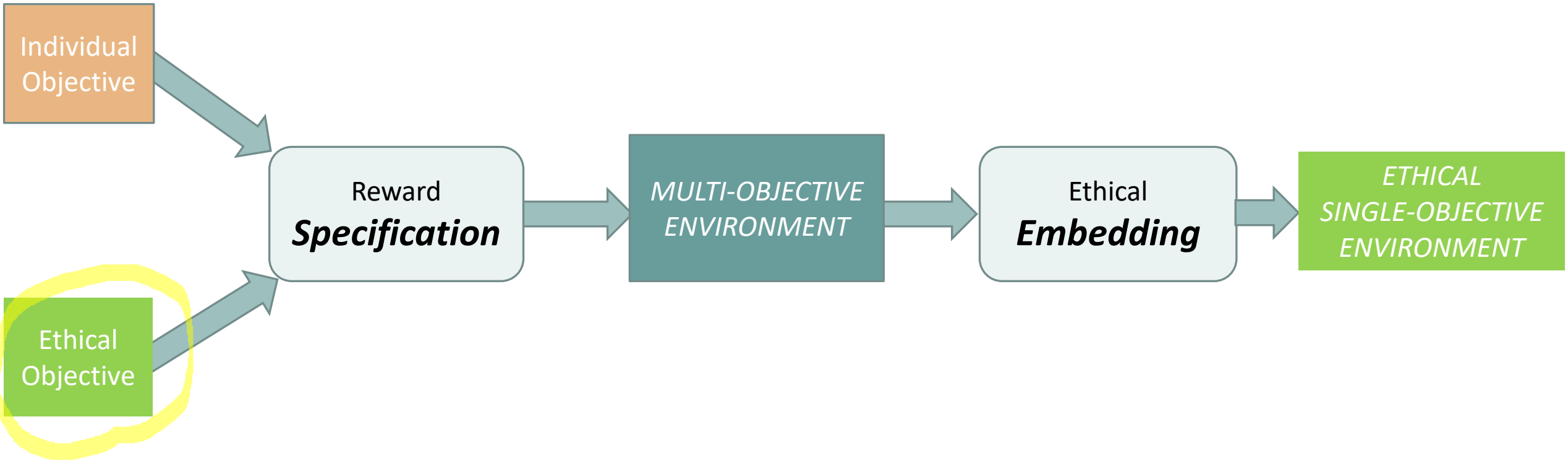
Our approach



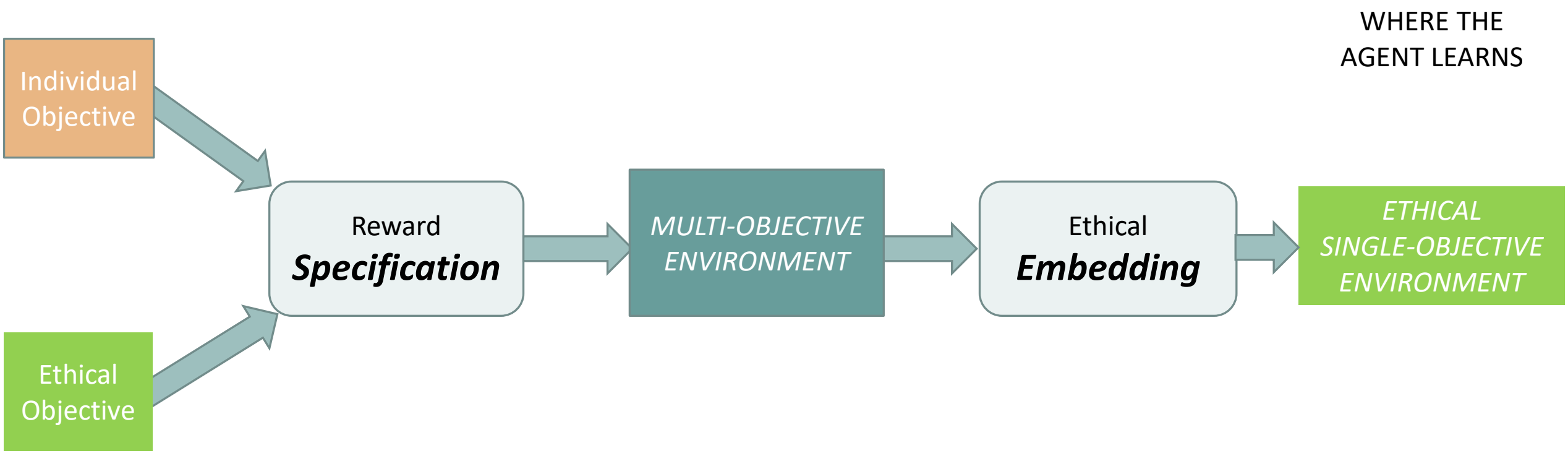
Our approach



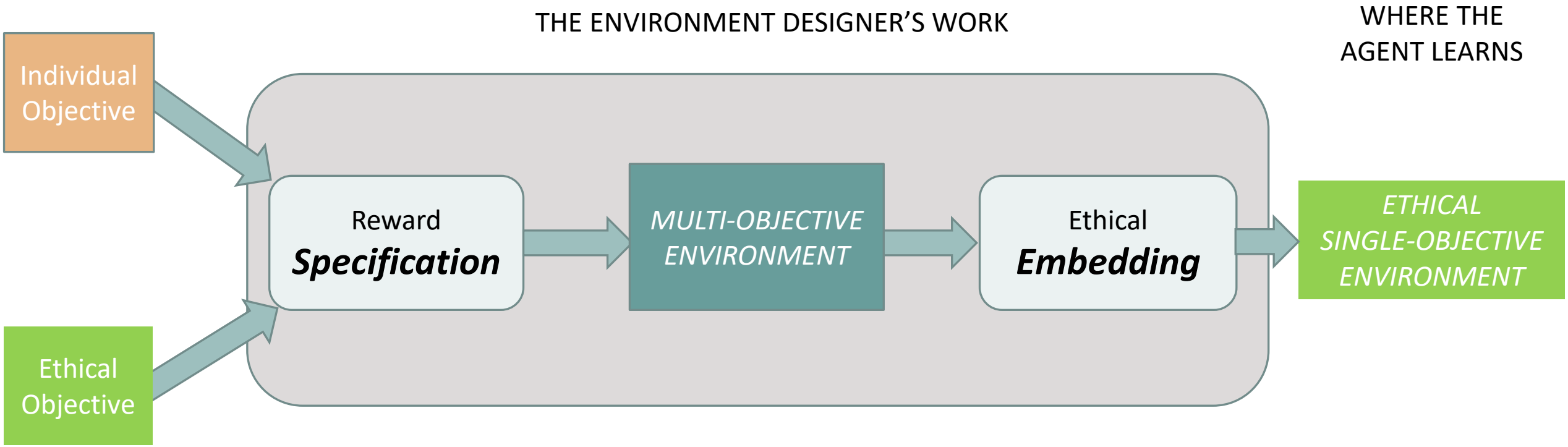
Our approach



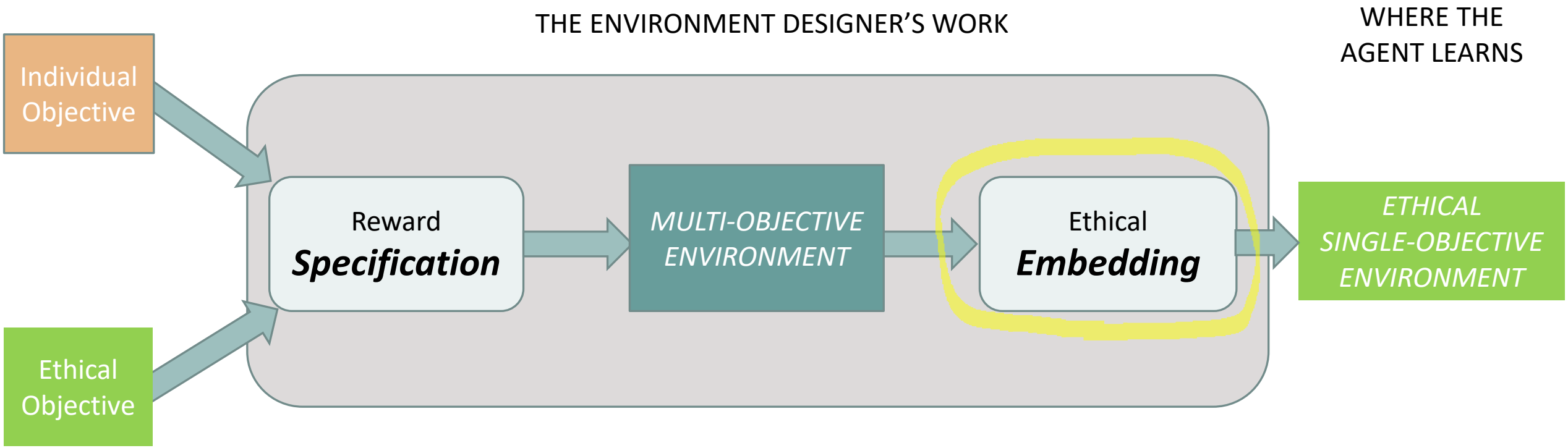
Our approach



Our approach



Our approach



Our approach

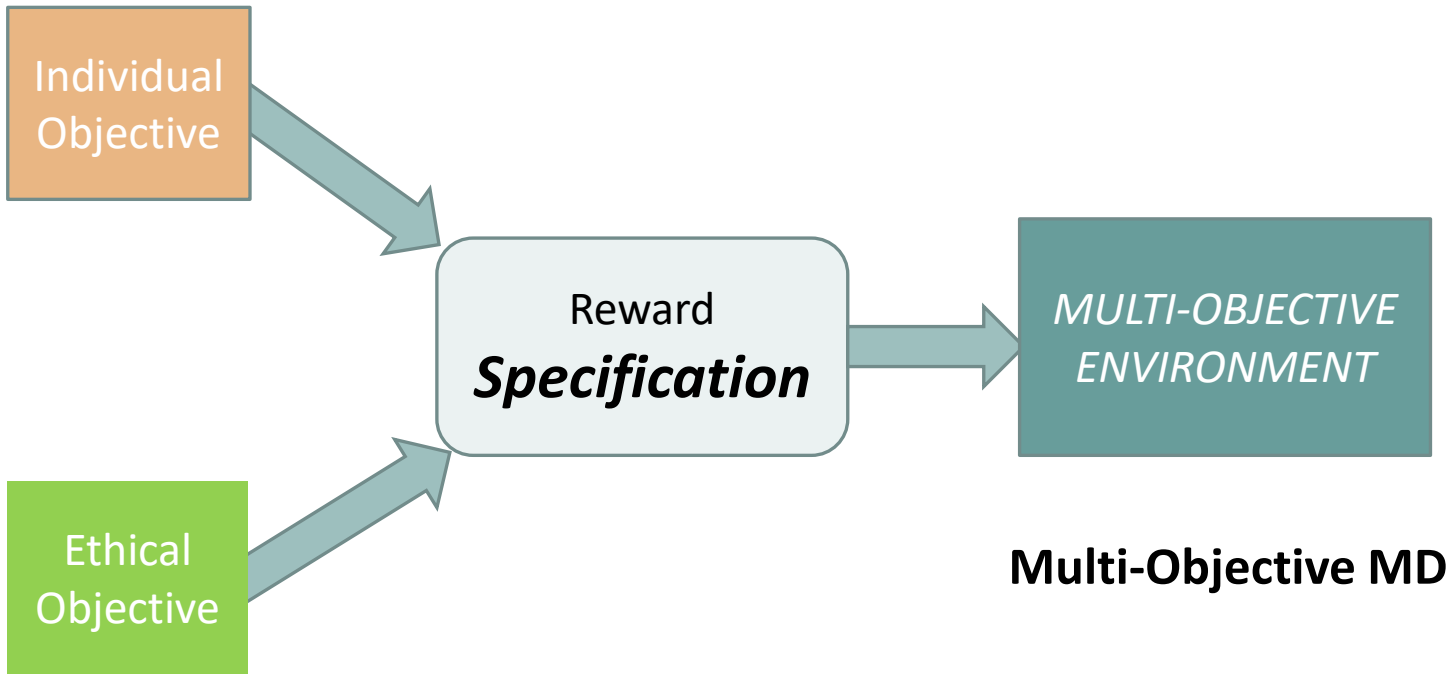
Design an ethical environment that **guarantees** that the agent learns to behave **ethically** while pursuing its individual objective.

Contributions

Theoretical means to design a learning environment wherein ethical behaviour is guaranteed.

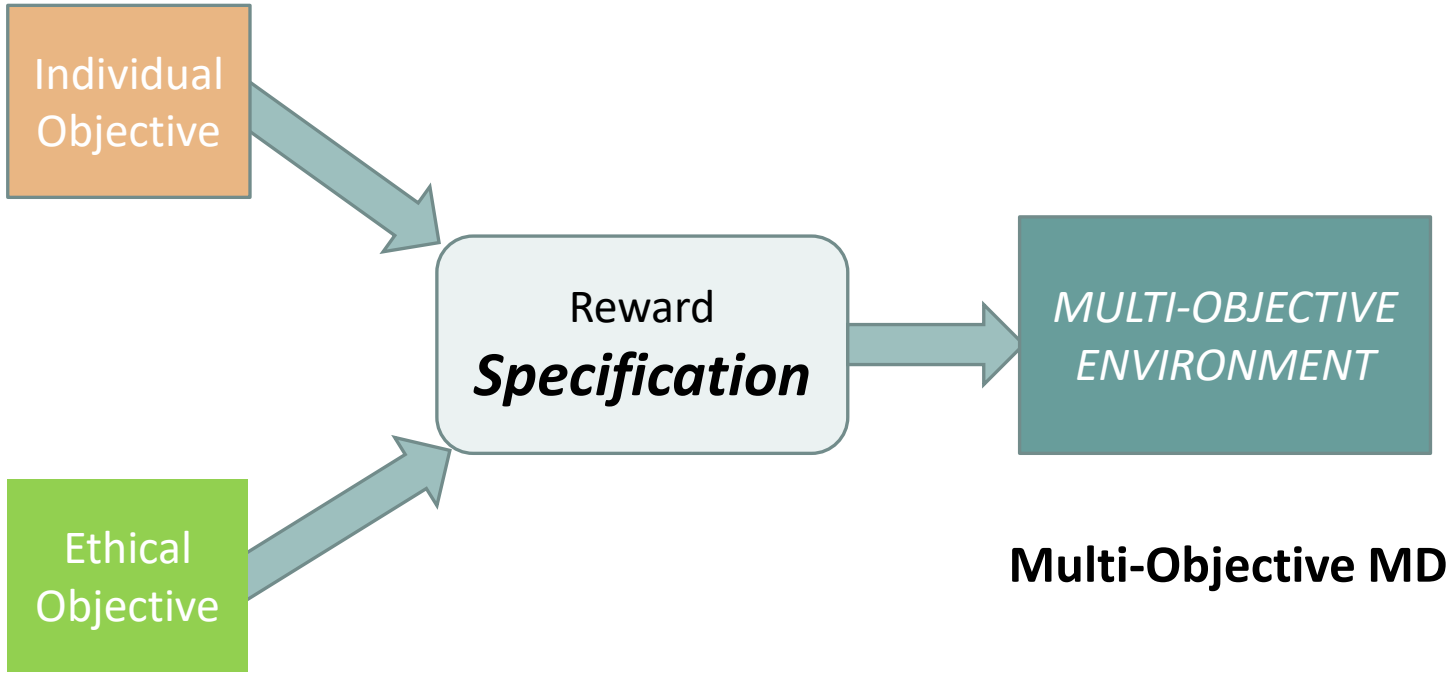
An algorithm for automating the design of the learning environment.

Formalising the ethical embedding problem



Formalising the ethical embedding problem

We define this kind of MOMDP as an **Ethical** MOMDP



Formalising the ethical embedding problem

We divide the ethical reward function in two equally important components:

- ***N**ormative component:* it penalises the violation of ethical requirements
- ***E**valuative component:* it rewards the performance of praiseworthy actions.

Formalising the ethical embedding problem

We divide the ethical reward function in two equally important components:

- ***N**ormative component:* it penalises the violation of ethical requirements
- ***E**valuative component:* it rewards the performance of praiseworthy actions.

MULTI-OBJECTIVE ENVIRONMENT

Multi-Objective MDP
with rewards $(R_0, R_N + R_E)$

Formalising the ethical embedding problem

Multi-Objective MDP:

A **Multi-Objective** Markov Decision Process (MOMDP) is an MDP with a vectorial reward function:

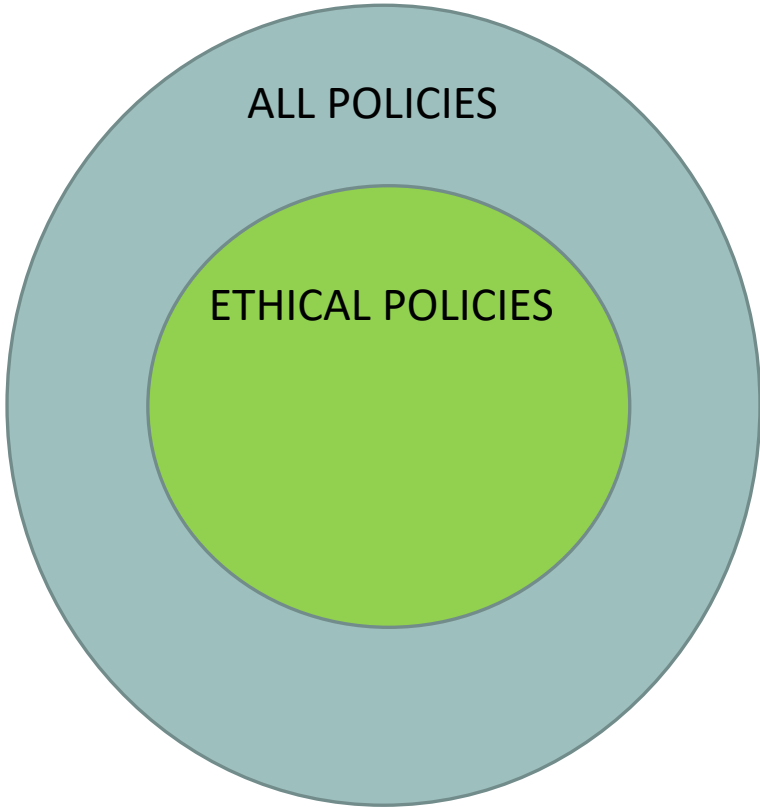
$$\vec{R} = (R_0, \dots, R_n).$$

Thus, the value function of an MOMDP is also vectorial: $\vec{V} = (V_0, \dots, V_n)$.

Formalising the ethical embedding problem

Ethical policies

An **ethical** policy is a policy with the maximum accumulation of the two components of ethical rewards (V_N and V_E).



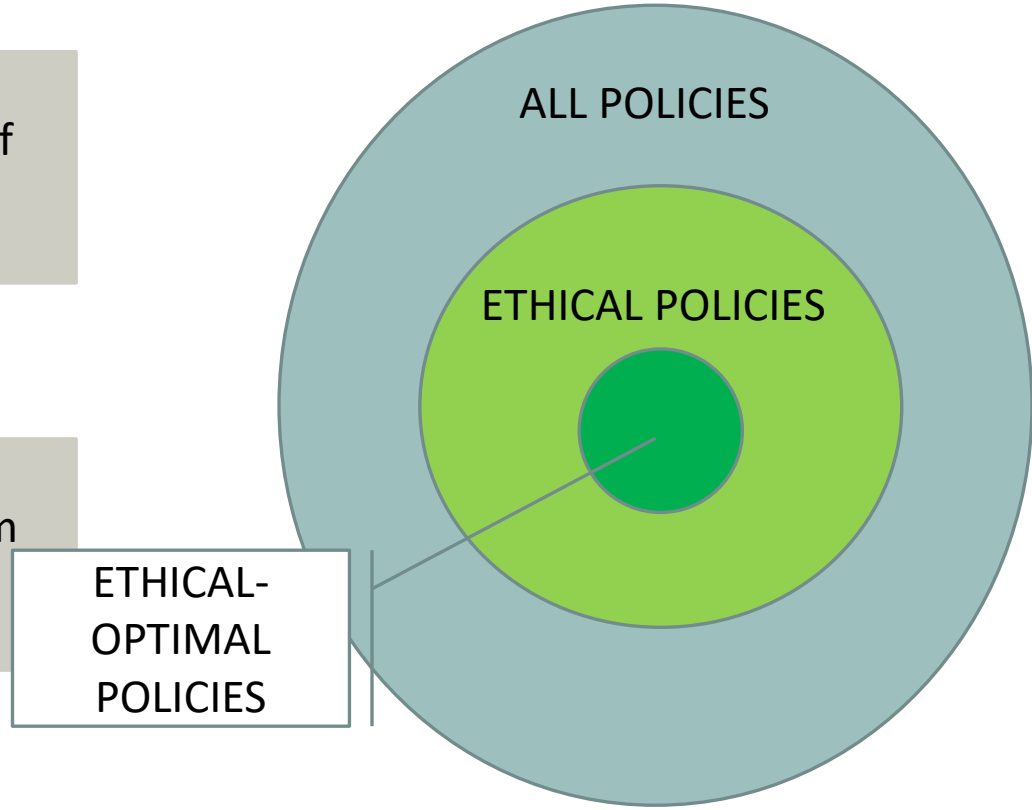
Formalising the ethical embedding problem

Ethical policies

An **ethical** policy is a policy with the maximum accumulation of the two components of ethical rewards (V_N and V_E).

Ethical-optimal policies

An **ethical-optimal** policy is an ethical policy with the maximum accumulation of individual rewards (V_0).



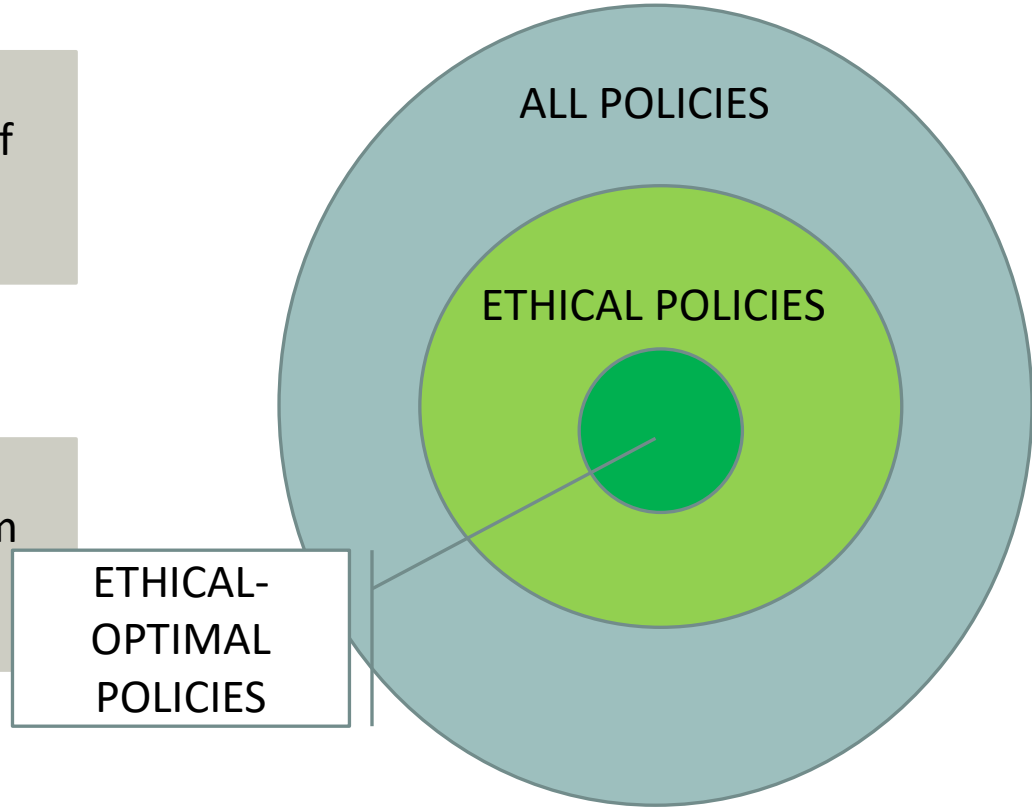
Formalising the ethical embedding problem

Ethical policies

An **ethical** policy is a policy with the maximum accumulation of the two components of ethical rewards (V_N and V_E).

Ethical-optimal policies

An **ethical-optimal** policy is an ethical policy with the maximum accumulation of individual rewards (V_0).



How to create the ethical environment in which the agent learns an ethical-optimal policy?

Formalising the ethical embedding problem

Scalarisation function: transforms a multi-objective environment into a single-objective environment.

Formalising the ethical embedding problem

Scalarisation function: transforms a multi-objective environment into a single-objective environment.

Typically a linear combination of the different objectives.

Formalising the ethical embedding problem

Scalarisation function: transforms a multi-objective environment into a single-objective environment.

Typically a linear combination of the different objectives. For Ethical MOMDPs:

$$f(V_0, V_N + V_E) = w_0 \cdot V_0 + w_e \cdot (V_N + V_E)$$

where w_0 is the individual weight, and w_e is the ethical weight.

Formalising the ethical embedding problem

Scalarisation function for an Ethical MOMDP.

$$f(V_0, V_N + V_E) = w_0 \cdot V_0 + w_e \cdot (V_N + V_E)$$

Formalising the ethical embedding problem

Scalarisation function for an Ethical MOMDP.

$$f(V_0, V_N + V_E) = w_0 \cdot V_0 + \boldsymbol{w_e} \cdot (V_N + V_E)$$

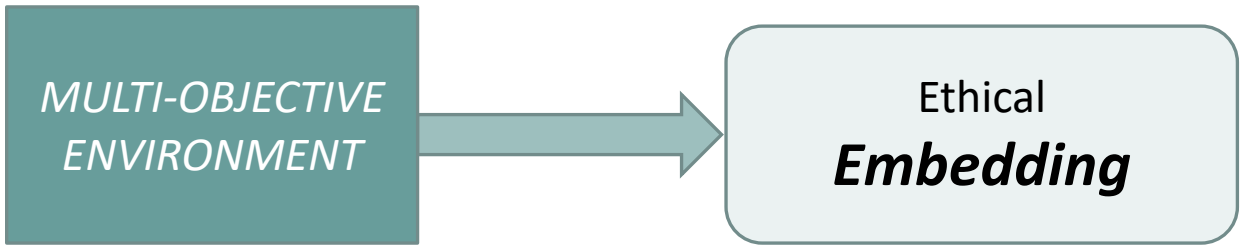
The environment designer's work: to set the appropriate **ethical weight**.

The ethical embedding problem

MULTI-OBJECTIVE ENVIRONMENT

Ethical MOMDP

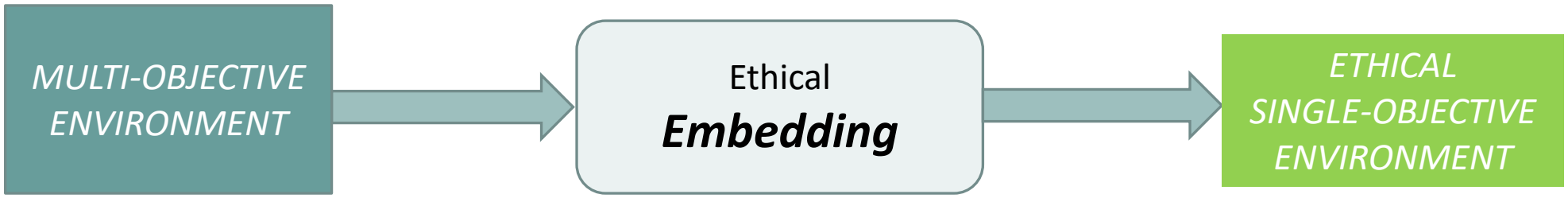
The ethical embedding problem



Ethical MOMDP

How to set the **ethical weight**?

The ethical embedding problem

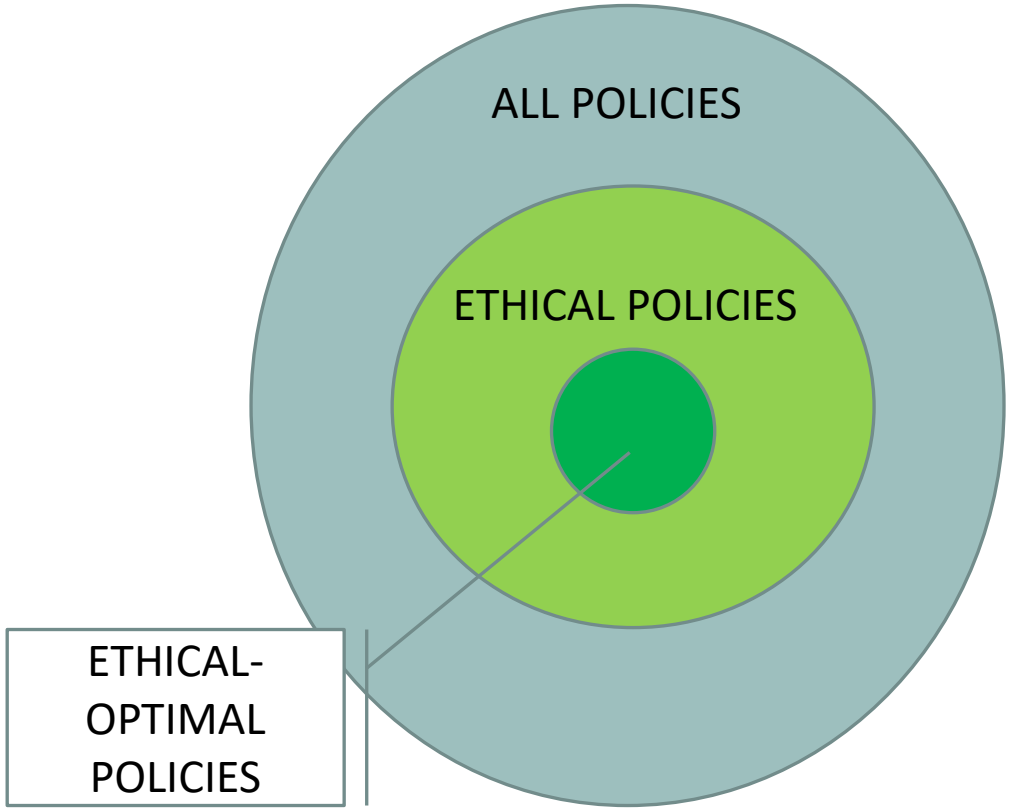


Ethical MOMDP

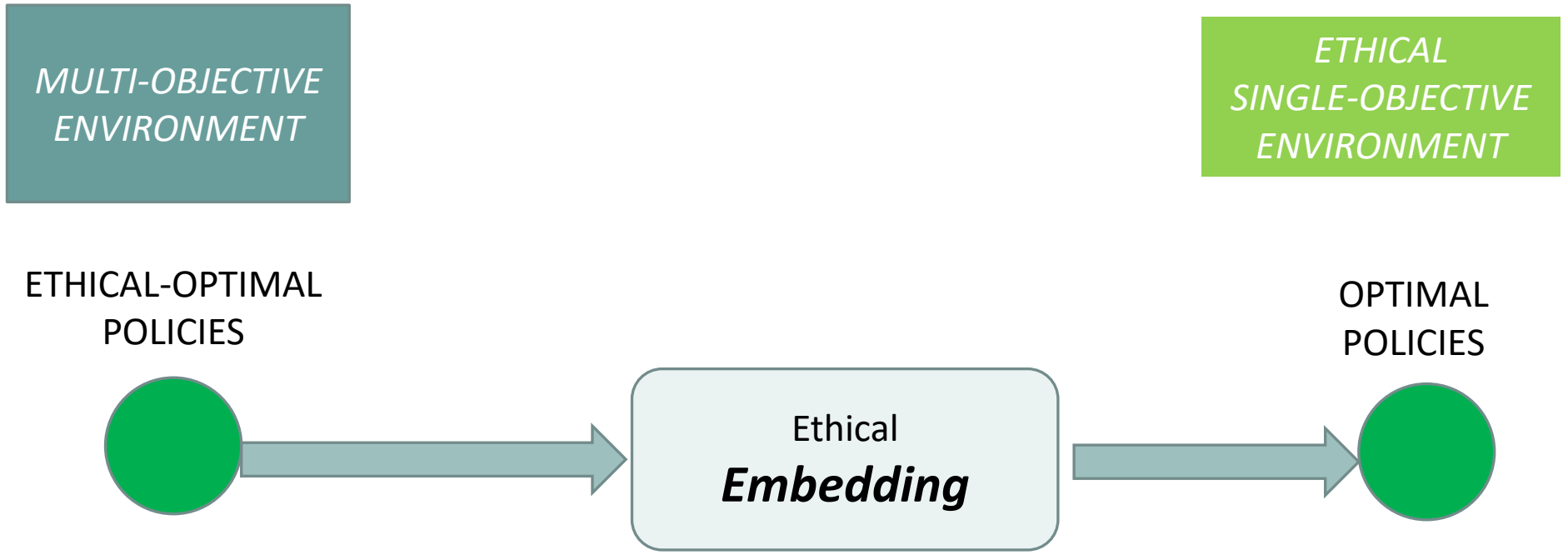
How to set the **ethical weight**?

Single-objective MDP where **all optimal policies** are **ethical-optimal**

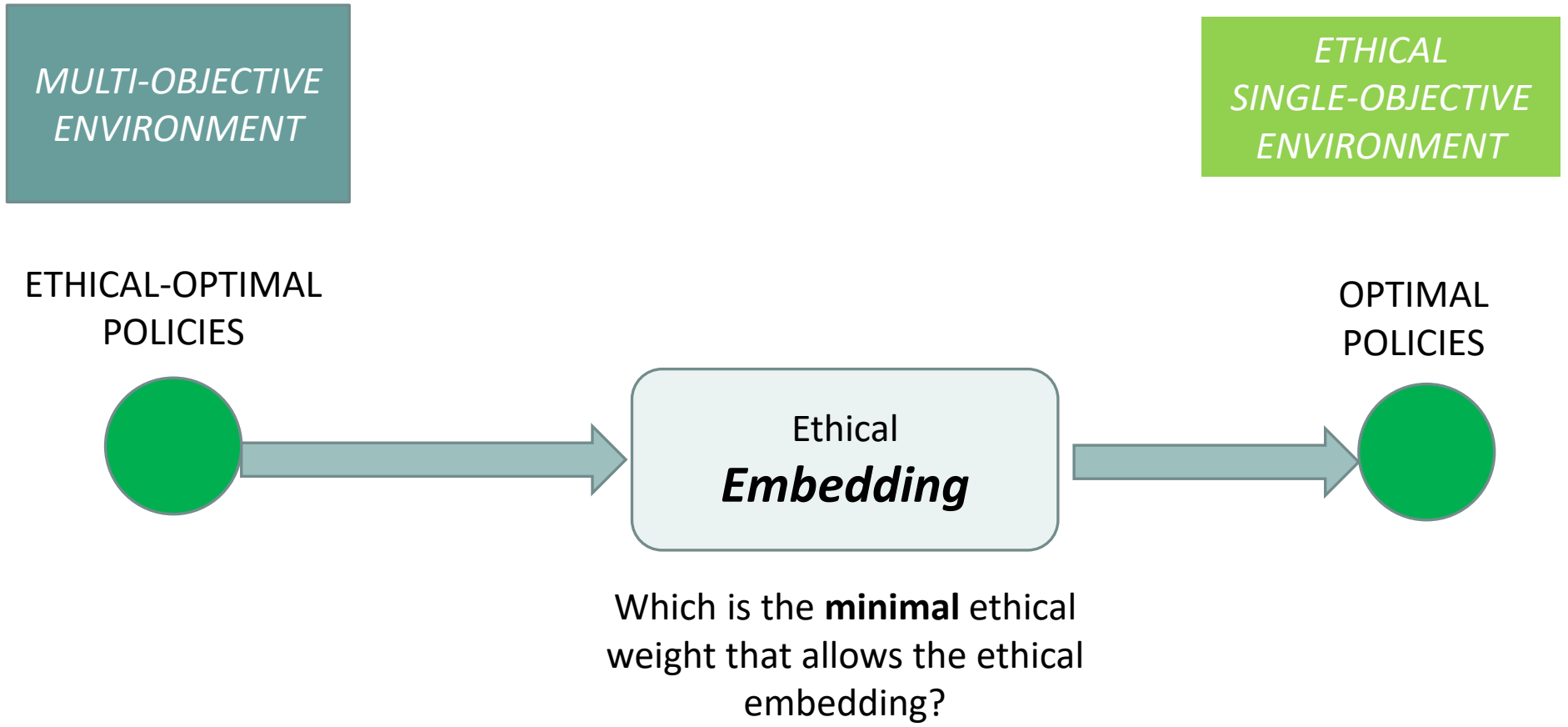
The ethical embedding problem



The ethical embedding problem



The ethical embedding problem



Solvability of the ethical embedding problem

Existence Condition

Given an Ethical MOMDP, there **exists** at least one ethical policy.

Solvability of the ethical embedding problem

Existence Condition

Given an Ethical MOMDP, there **exists** at least one ethical policy.

In plain words, if we want the agent to behave ethically,
it must be possible for it to behave ethically.

Solvability of the ethical embedding problem

Solution Existence Theorem

Given an ethical MOMDP, there **exists** an ethical weight
for which all optimal policies are ethical-optimal.

Solvability of the ethical embedding problem

Solution Existence Theorem

Given an ethical MOMDP, there **exists** an ethical weight
for which all optimal policies are ethical-optimal.

In plain words, **every** ethical embedding problem is solvable.

Solving the ethical embedding problem

- Our solution to the ethical embedding problem is a 3-step algorithm that obtains the desired ethical environment in which the agent learns to behave ethically.

Solving the ethical embedding problem

- Our solution to the ethical embedding problem is a 3-step algorithm that obtains the desired ethical environment in which the agent learns to behave ethically.

- We design this ethical environment directly from computing the appropriate minimal ethical weight.

Solving the ethical embedding problem

- Our solution to the ethical embedding problem is a 3-step algorithm that obtains the desired ethical environment in which the agent learns to behave ethically.

- We design this ethical environment directly from computing the appropriate ethical weight.

- The key concept for our solution is the computation of the **convex hull** of an MOMDP.

The **convex hull** (CH) of an MOMDP is the set of policies (and their associated value vectors) that are optimal for some weight vector.

Solving the ethical embedding problem

We compute a partial subset *P*
of the convex hull of an Ethical MOMDP.

Solving the ethical embedding problem

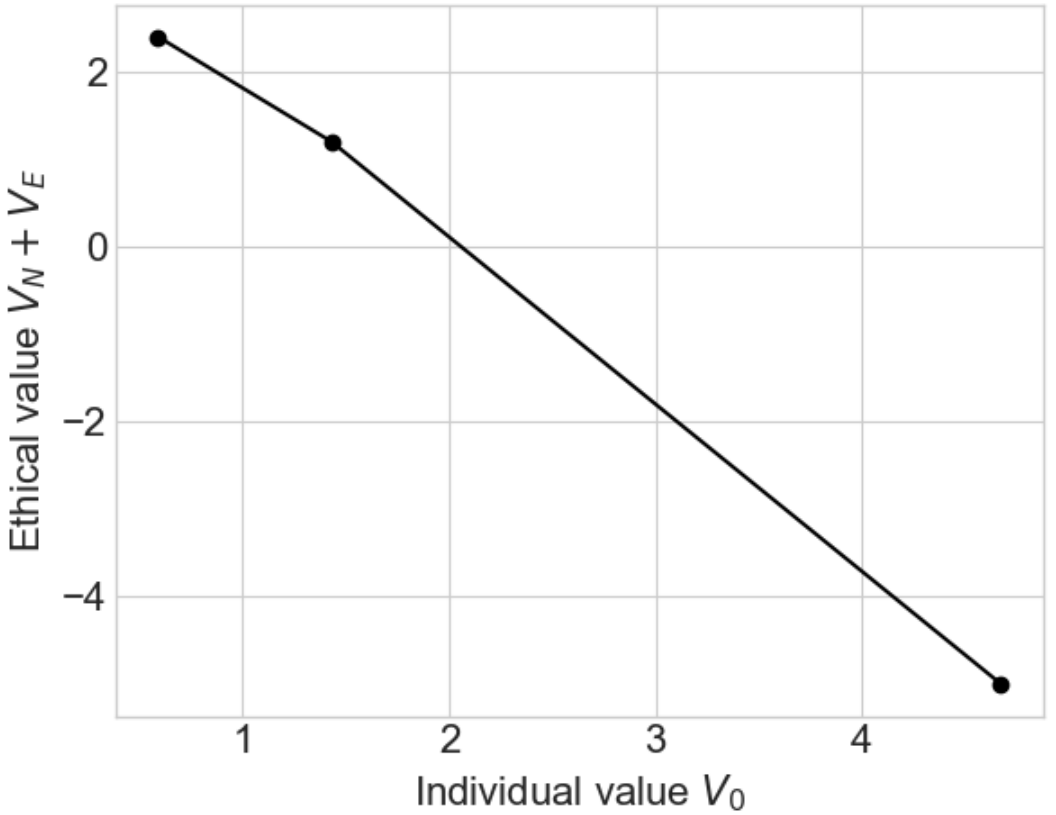
We compute a partial subset *P*
of the convex hull of an Ethical MOMDP.

This partial subset must contain the ethical-optimal value vector.

Solving the ethical embedding problem

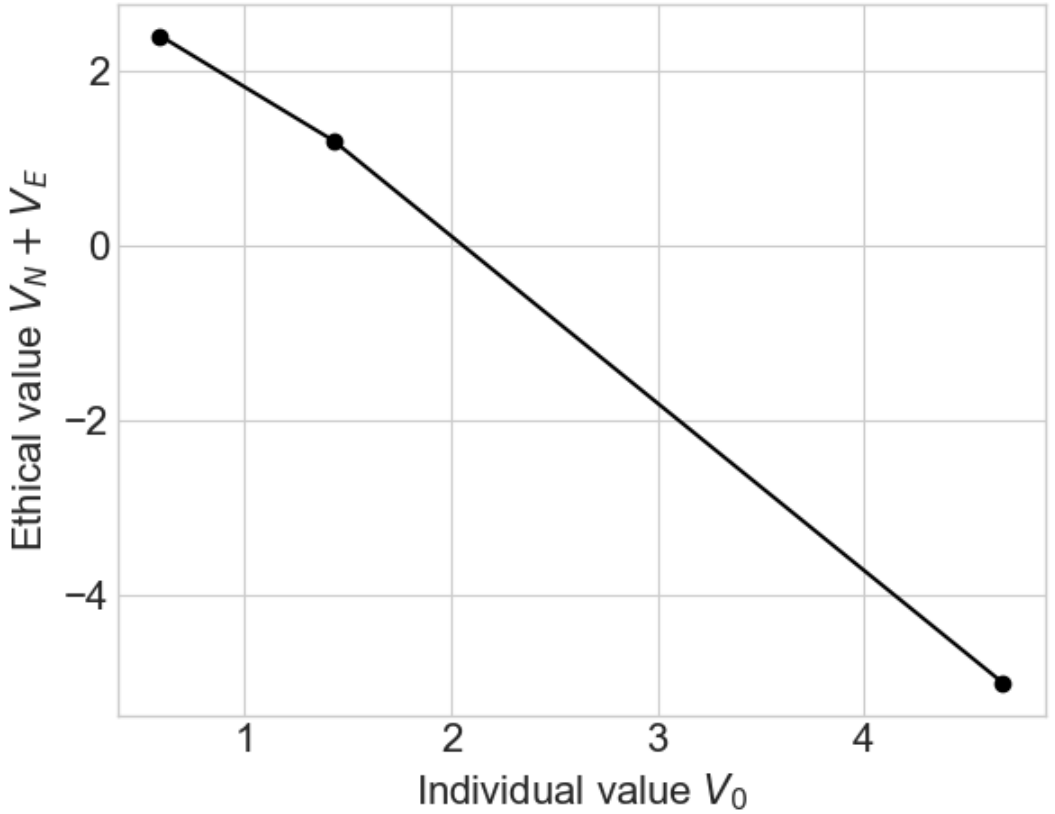
We compute a partial subset *P*
of the convex hull of an Ethical MOMDP.

This partial subset must contain the ethical-optimal value vector.



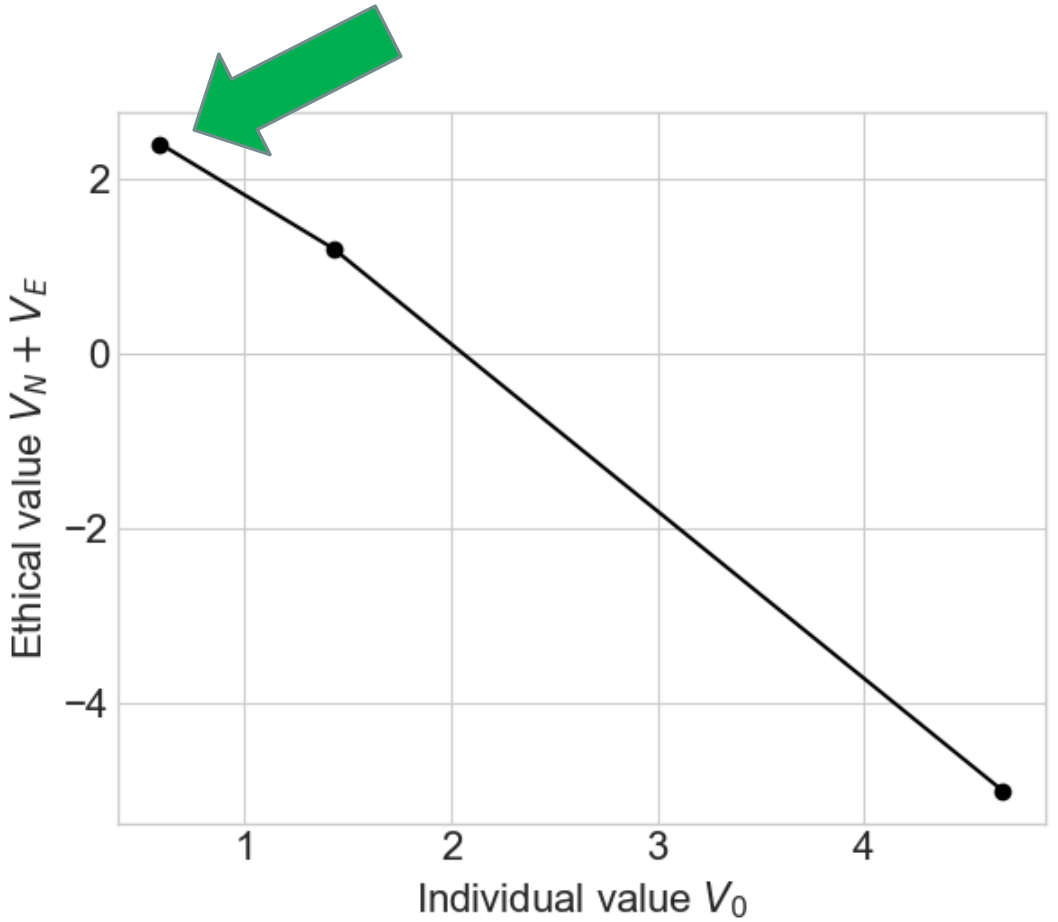
Solving the ethical embedding problem

In the computed partial convex hull we identify the **ethical-optimal** value vector.



Solving the ethical embedding problem

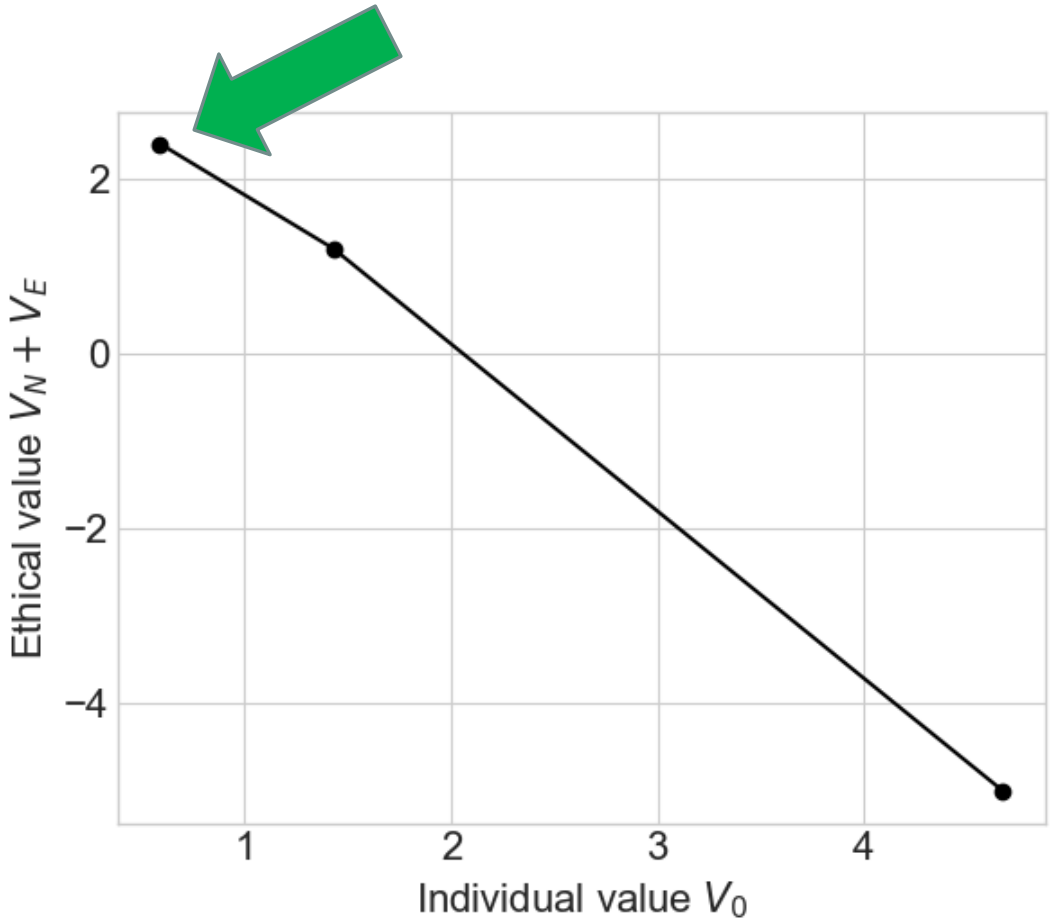
In the computed partial convex hull we identify the **ethical-optimal** value vector.



Solving the ethical embedding problem

In the computed partial convex hull we identify the **ethical-optimal** value vector.

We identify this value vector to find the minimal ethical weight solution.

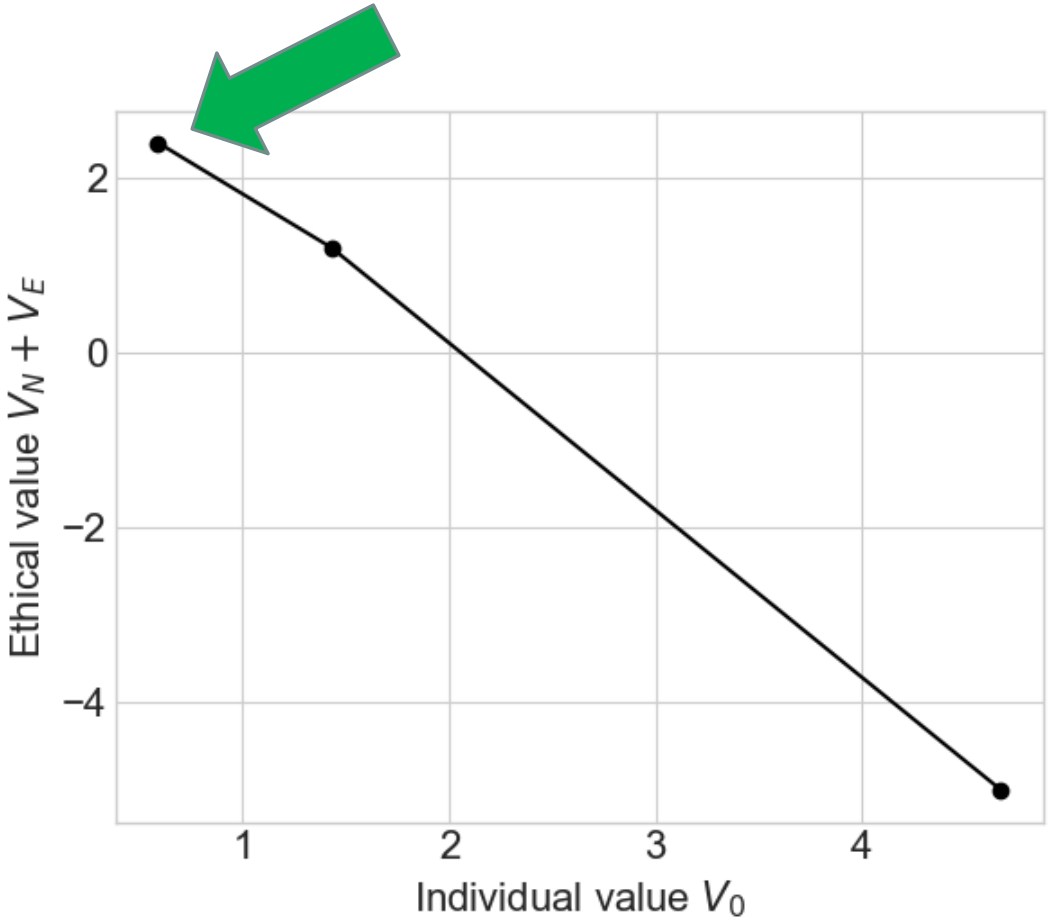


Solving the ethical embedding problem

In the computed partial convex hull we identify the **ethical-optimal** value vector.

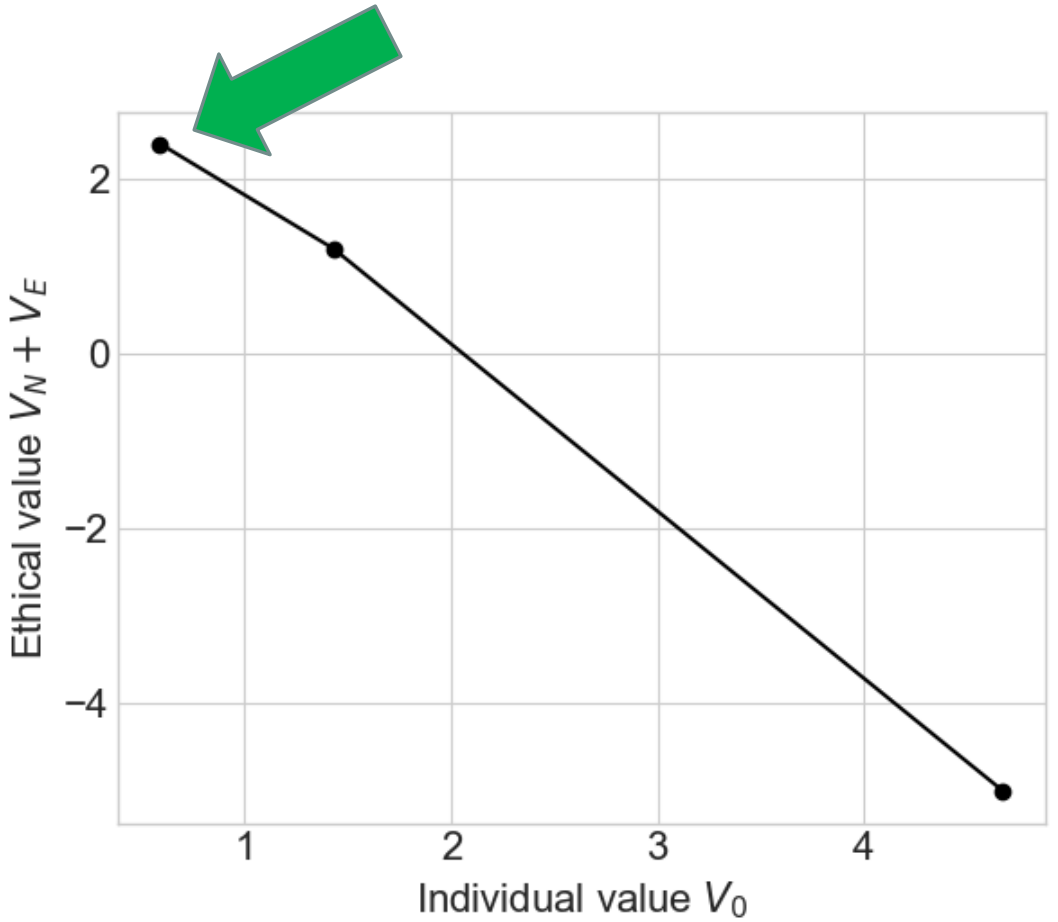
We identify this value vector to find the minimal ethical weight solution.

For such ethical weight, the ethical-optimal value vector will be the only optimal one.



Solving the ethical embedding problem

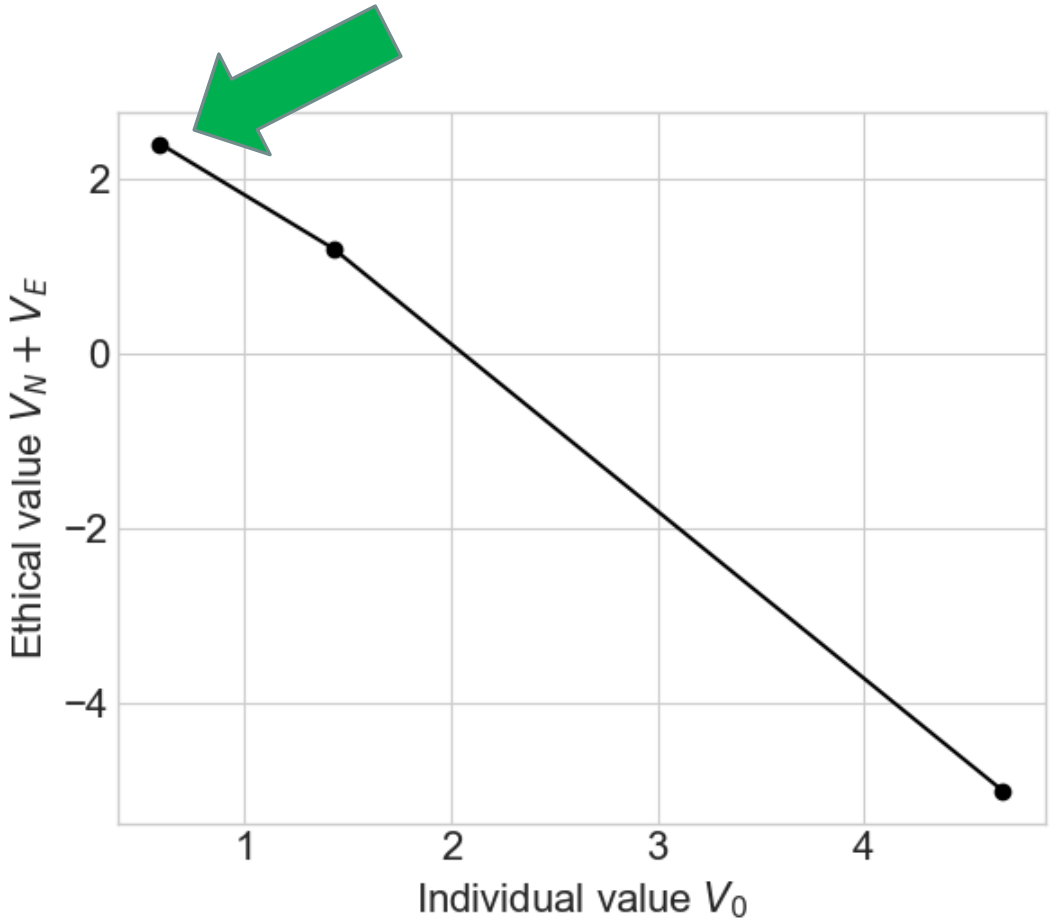
Computing the minimal ethical weight does not require to consider all value vectors on the partial convex hull.



Solving the ethical embedding problem

Computing the minimal ethical weight does not require to consider all value vectors on the partial convex hull.

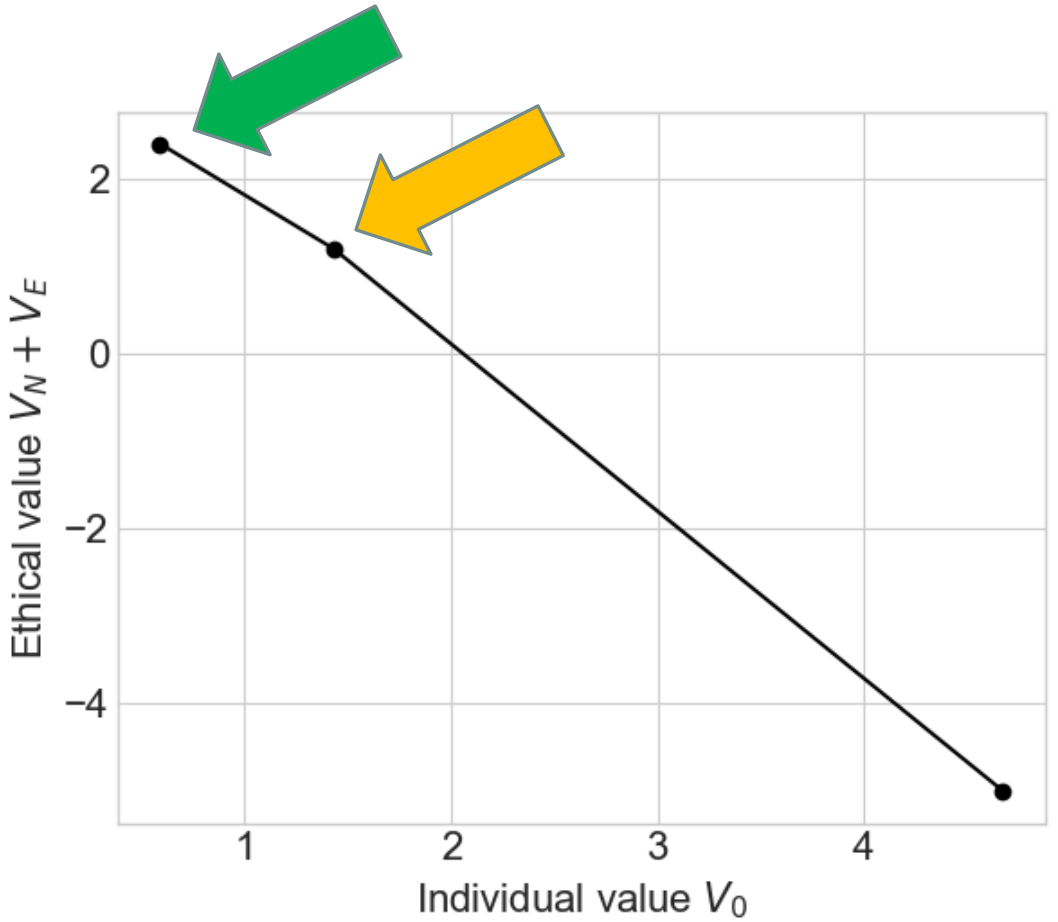
We only need to also identify the **second-best** value vector (in terms of ethical rewards).



Solving the ethical embedding problem

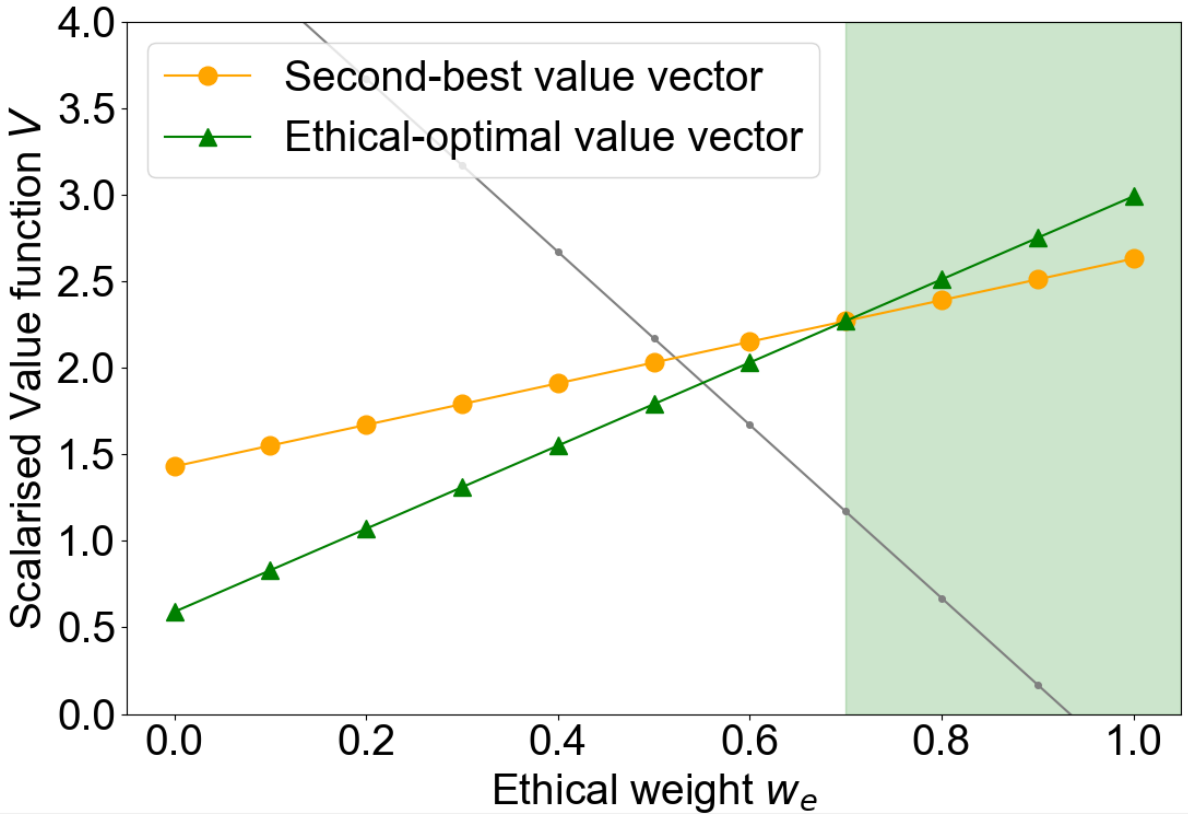
Computing the minimal ethical weight does not require to consider all value vectors on the partial convex hull.

We only need to also identify the **second-best** value vector (in terms of ethical rewards).



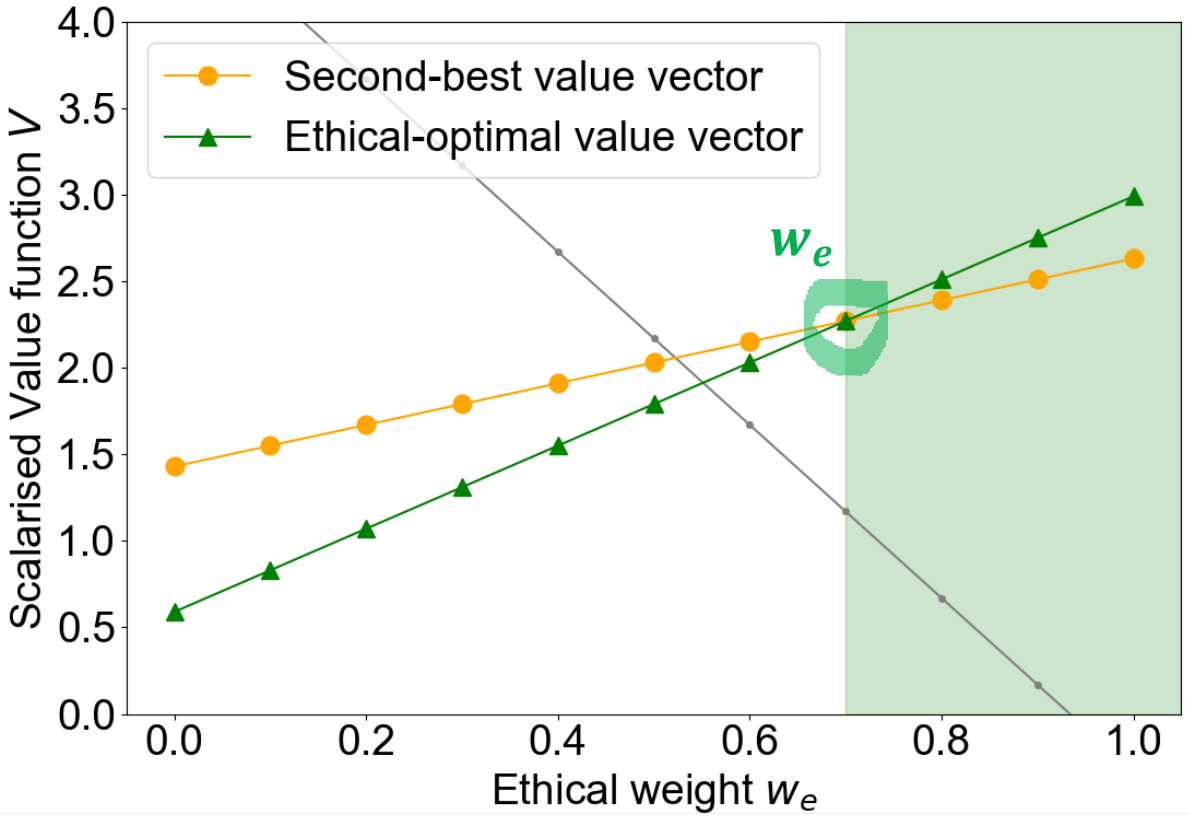
Solving the ethical embedding problem

After intersecting the two value vectors, the ethical-optimal becomes the only optimal one



Solving the ethical embedding problem

After intersecting the two value vectors, the ethical-optimal becomes the only optimal one

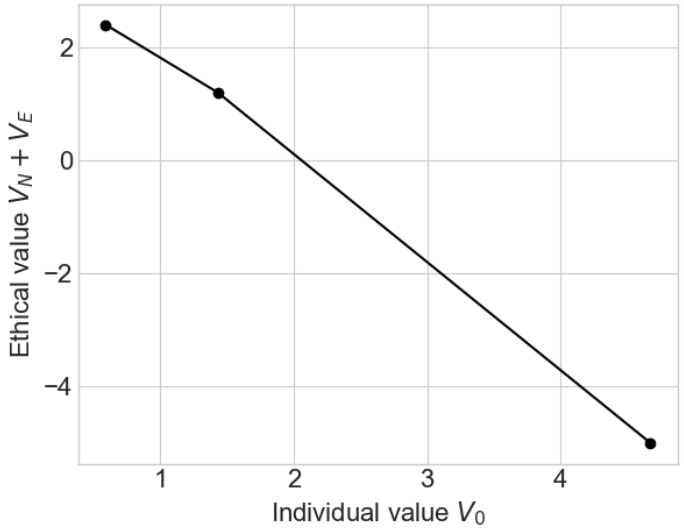


Solving the ethical embedding problem

In summary, our 3-step algorithm consists in:

Solving the ethical embedding problem

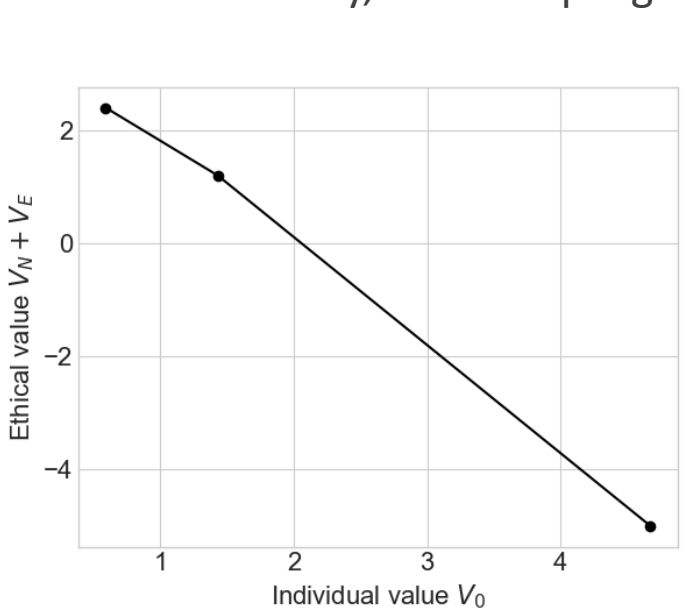
In summary, our 3-step algorithm consists in:

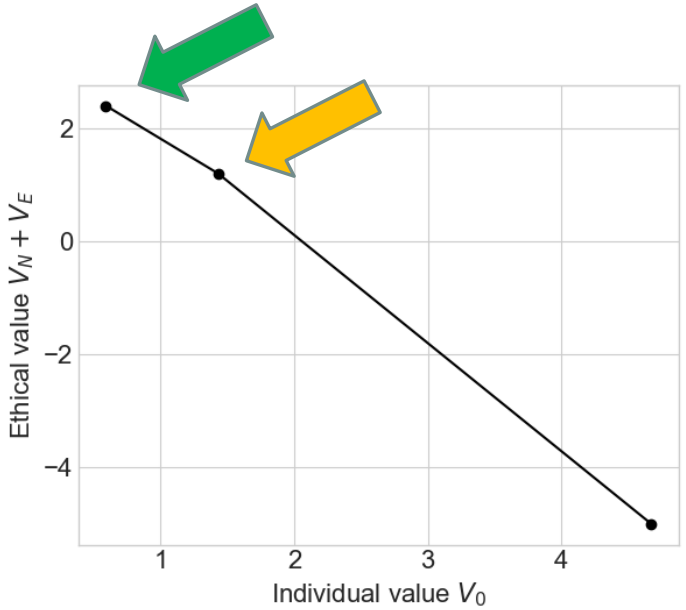


Compute the **partial convex hull** *P* of an Ethical MOMDP

Solving the ethical embedding problem

In summary, our 3-step algorithm consists in:



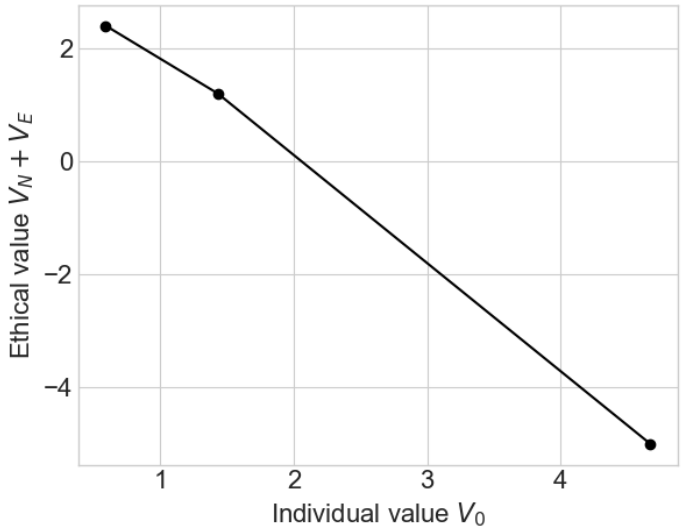


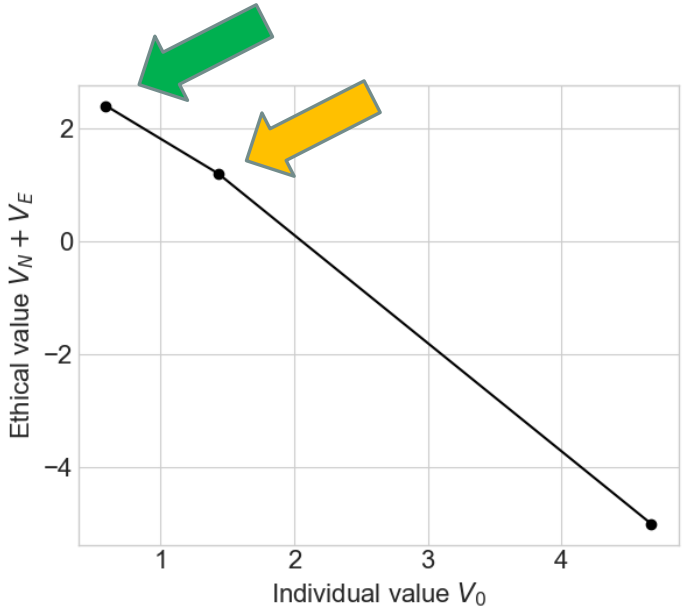
Compute the partial convex hull *P* of an Ethical MOMDP

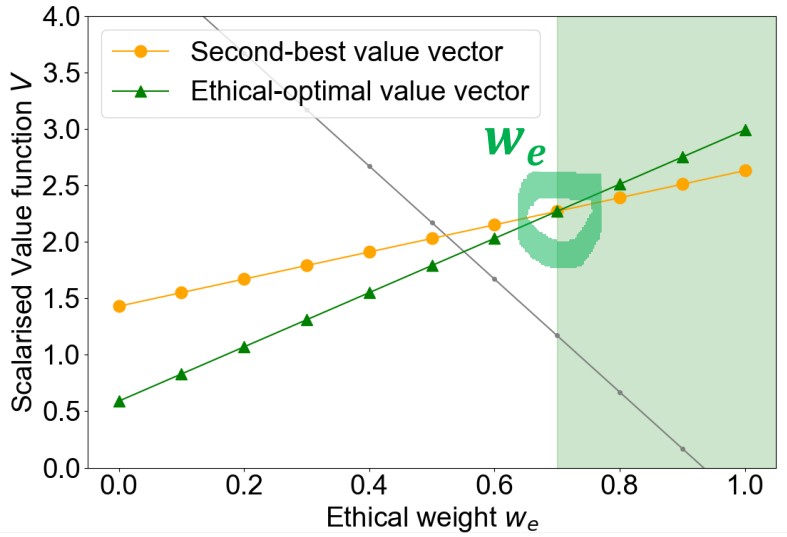
Extract the **Ethical-Optimal** and **Second-best** value vectors from *P*.

Solving the ethical embedding problem

In summary, our 3-step algorithm consists in:



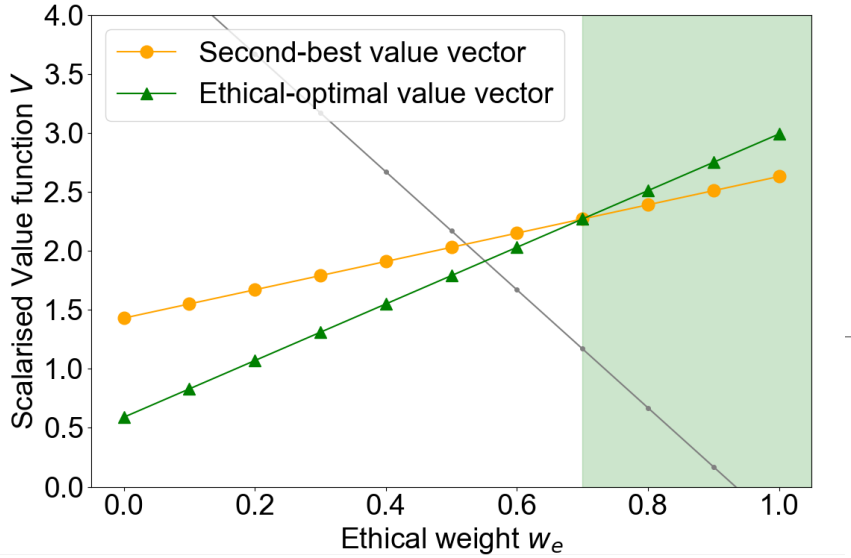


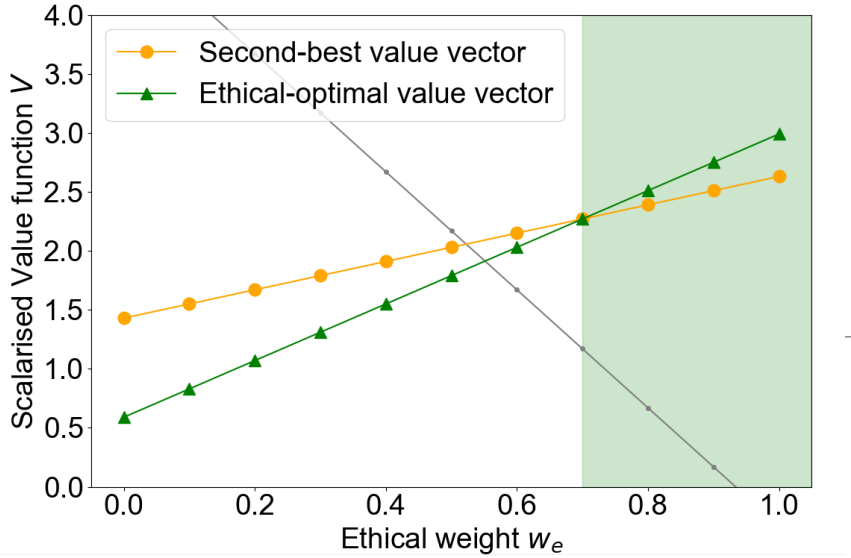


Compute the partial convex hull *P* of an Ethical MOMDP

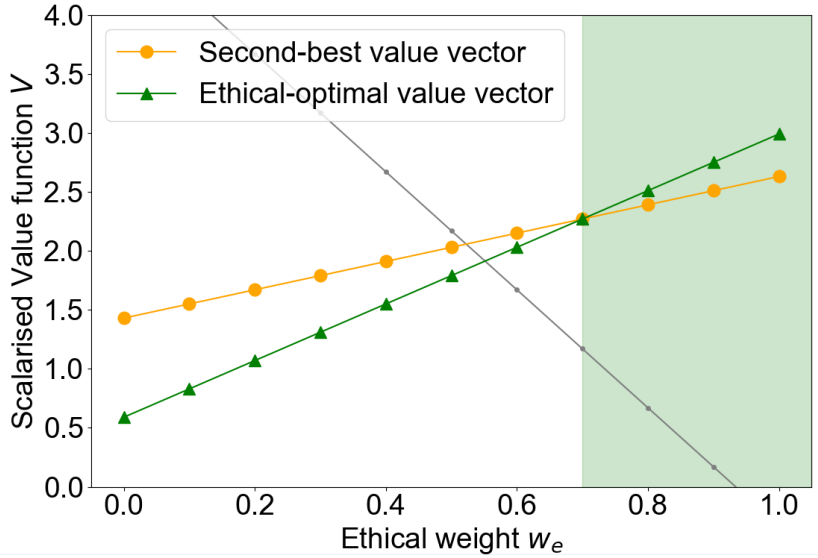
Extract the Ethical-Optimal and Second-best value vectors from *P*.

Obtain the minimal **ethical weight** w_e by computing where the two value vectors intersect.


Second-best value vector
Ethical-optimal value vector
Scalarised Value function V
Ethical weight w_e
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0



$$f(V_0, V_N + V_E) = w_0 \cdot V_0 + \boldsymbol{w_e} \cdot (V_N + V_E)$$



$$f(V_0, V_N + V_E) = w_0 \cdot V_0 + w_e \cdot (V_N + V_E)$$

MULTI-OBJECTIVE ENVIRONMENT

Ethical
Embedding

ETHICAL SINGLE-OBJECTIVE ENVIRONMENT

Solving the ethical embedding problem

- The computational cost of the algorithm resides in computing the partial convex hull of an MOMDP.
- CHVI algorithm requires $O(n \cdot log\ n)$ times what its single-objective version takes, where n is the number of points in the partial hull.

Conclusions

We make headway in tackling the problem of **designing ethical environments** by providing novel formal and algorithmic tools that build upon **Multi-Objective Reinforcement Learning**.

Conclusions

We make headway in tackling the problem of designing ethical environments by providing novel formal and algorithmic tools that build upon Multi-Objective Reinforcement Learning.

We characterise the obtention of an ethical environment as a **two-step process** that first specifies rewards and second performs an ethical embedding.

Conclusions

We make headway in tackling the problem of designing ethical environments by providing novel formal and algorithmic tools that build upon Multi-Objective Reinforcement Learning.

We characterise the obtention of an ethical environment as a two-step process that first specifies rewards and second performs an ethical embedding.

We formalise this last step as the ethical embedding problem and **theoretically prove** that it is always **solvable**.

Conclusions

We make headway in tackling the problem of designing ethical environments by providing novel formal and algorithmic tools that build upon Multi-Objective Reinforcement Learning.

We characterise the obtention of an ethical environment as a two-step process that first specifies rewards and second performs an ethical embedding.

We formalise this last step as the ethical embedding problem and **theoretically prove** that it is always **solvable**.

Our findings lead to an **algorithm for automating** the design of an **ethical environment** in which it will be in the best interest of the agent to behave ethically while still pursuing its individual objective.

Future work

We would like to further examine empirically our algorithm in more complex environments.

Any questions? Ask me at manel.rodriguez@iiia.csic.es or at the ALA-2021 Slack Channel 5.