TU Delft

Delft University of Technology

Document Version
Final published version

Citation (APA)
Biswas, P., Osika, Z., Tamassia, I., Whorra, A., Zatarain-Salazar, J., Kwakkel, J., Oliehoek, F. A., & Murukannaiah, P. K. (2025). Exploring Equity of Climate Policies using Multi-Agent Multi-Objective Reinforcement Learning. In J. Kwok (Ed.), *Proceedings of the 34th International Joint Conference on Artificial Intelligence, IJCAI 2025* (pp. 9573-9581). (IJCAI International Joint Conference on Artificial Intelligence). International Joint Conferences on Artificial Intelligence. https://doi.org/10.24963/ijcai.2025/1064

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)
Special Track on AI and Social Good

Exploring Equity of Climate Policies using Multi-Agent Multi-Objective Reinforcement Learning

Palok Biswas, **Zuzanna Osika**, **Isidoro Tamassia**, **Adit Whorra**, **Jazmin Zatarain-Salazar**, **Jan Kwakkel**, **Frans A. Oliehoek** and **Pradeep K. Murukannaiah**
Delft University of Technology, The Netherlands
{p.biswas, z.osika, j.zatarainsalazar, j.h.kwakkel, f.a.oliehoek, p.k.murukannaiah}@tudelft.nl
{isidorotamassia, aditwhorra}@gmail.com

Abstract

Addressing climate change requires coordinated policy efforts of nations worldwide. These efforts are informed by scientific reports, which rely in part on Integrated Assessment Models (IAMs), prominent tools used to assess the economic impacts of climate policies. However, traditional IAMs optimize policies based on a single objective, limiting their ability to capture the trade-offs among economic growth, temperature goals, and climate justice. As a result, policy recommendations have been criticized for perpetuating inequalities, fueling disagreements during policy negotiations. We introduce JUSTICE, the first framework integrating IAM with Multi-Objective Multi-Agent Reinforcement Learning (MOMARL). By incorporating multiple objectives, JUSTICE generates policy recommendations that shed light on equity while balancing climate and economic goals. Further, using multiple agents can provide a realistic representation of the interactions among the diverse policy actors. We identify equitable Pareto-optimal policies using our framework, which facilitates deliberative decision-making by presenting policymakers with the inherent trade-offs in climate and economic policy.

1 Introduction

Climate change poses a significant global threat, disproportionately affecting marginalized communities [Faus Onbargi, 2022; Rising *et al.*, 2022]. The extent of this threat remains uncertain due to the complex interplay between climate and socioeconomic systems [Burke *et al.*, 2018; van der Wijst *et al.*, 2021]. This uncertainty underscores a central challenge of climate justice: the fair allocation of burdens and benefits through the implementation of climate policies [Pozo *et al.*, 2020]. The complexity in assessing the implications of policy measures leads to contentious international negotiations, as seen in the Conference of the Parties (COP). Asymmetric impacts, divergent responsibilities, and differing national priorities often result in impasses and heated disagreements among policymakers [Wei *et al.*, 2013]. Climate change is thus characterized as a wicked problem in the policy domain [Lönngren and Svanström, 2016], necessitating interventions that seek to balance the diverse ethical preferences of stakeholders.

The Intergovernmental Panel on Climate Change (IPCC) is a key player in the global climate change discourse, and its assessment reports serve as the primary source of scientific information for international climate negotiations, such as the COP [Asayama, 2024]. These technical reports, crafted by experts from various disciplines, including climate science and economics, provide evidence-based guidance to support effective climate action through the use of Integrated Assessment Models (IAMs) [Cointe *et al.*, 2019]. IAMs integrate socioeconomic, climate and technological processes into a single framework to assess global mitigation pathways and simulate future socio-economic and climate scenarios.

Although IAMs are influential tools, they are criticized for suggesting inequitable mitigation policies that disadvantage developing countries [Rivadeneira and Carton, 2022; Gambhir *et al.*, 2022]. These models often simplify the complexity of climate change policies to a single objective, which is ethically problematic, as it favours dominant stakeholders and neglects non-economic metrics, such as temperature, biodiversity, and human mortality [Bromley and Beattie, 1973; Kasprzyk *et al.*, 2016]. Recent studies seek to address these shortcomings by enhancing IAMs with multiple objectives [Marangoni *et al.*, 2021; Ferrari *et al.*, 2022], or by incorporating multi-agent approaches, such as multi-regional studies [Zhang *et al.*, 2022]. However, the multi-objective studies treat the world as a single agent, which overlooks the interests of various stakeholders. Conversely, the multi-agent approach focuses on a single objective, neglecting the diverse preferences of agents during negotiations. Currently, no single modeling framework simultaneously addresses both the multiplicity of objectives and agents, thus overlooking the complexities of the real-world. The exclusion of either multiple objectives or multiple agents limits the applicability of IAMs. Our contribution fills this gap by developing JUSTICE, a Multi-Objective Multi-Agent Reinforcement Learning (MOMARL) IAM framework that emulates real-world negotiations and discovers equitable policy options. [1]

MOMARL is a powerful framework for complex decision-making that requires balancing conflicting objectives and co-

[1] Code: https://github.com/pollockDeVis/JUSTICE

ordinating independent decisions over time. In particular, it is a crucial tool for addressing problems that involve sequential decisions [Radulescu, 2024]. By extending reinforcement learning (RL) to accommodate multiple agents and multiple objectives, MOMARL enables trade-off management through vectorial rewards, where each component reflects performance on a specific objective. Despite the prevalence of societal problems involving multiple stakeholders with diverse preferences, the field of MOMARL remains underexplored. Simplifications such as hard-coding trade-offs or centralizing decisions limit practical applicability. Importantly, existing works are limited to theoretical examples or simple (grid-like) environments [Felten *et al.*, 2024b; Rădulescu *et al.*, 2020].

Our contribution is twofold. (1) For the IAM community, JUSTICE introduces a multi-objective, multi-agent framework of a climate-economy model, providing a tool that can inform IPCC's synthesis reports. Such a framework can enable the design of equitable policies that account for regional disparities while balancing multiple aspects of climate change, including economic and environmental outcomes. (2) For the MOMARL community, JUSTICE offers a well-designed open-source implementation using the MOMALand API [Felten *et al.*, 2024b], allowing seamless integration with any RL algorithm. This provides a real-world testbed for evaluating and benchmarking future algorithms, supporting the development of robust and societal applications of reinforcement learning methods.

2 Background

Our interdisciplinary work is based on two foundational areas: climate modelling and reinforcement learning.

2.1 Integrated Assessment Models

IAMs offer a holistic approach to inform policy decisions on climate change [Gambhir *et al.*, 2019]. These models integrate socioeconomic, technological, and biogeochemical variables to represent interactions between simplified social and climate components [Rivadeneira and Carton, 2022]. IAMs primarily inform climate mitigation policies and are frequently used for policy evaluation and optimization [Mastrandrea, 2009]. Originating from William Nordhaus's DICE model [1992], which combined economic and climate models for global cost-benefit analyses of climate policies, IAMs analyze climate policy's impacts on both the economy and environment. DICE is a foundational framework, earning Nordhaus a Nobel Prize and being utilized by the US government to calculate the social cost of carbon, which measures the societal benefit of reducing CO_2 emissions [Grubb *et al.*, 2021]. DICE has inspired numerous IAMs and remains a valuable tool for assessing climate policies. A popular regional variation of DICE, known as RICE [Nordhaus and Yang, 1996], encompasses 12 regions, and its recent and enhanced version called RICE50+ [Gazzotti *et al.*, 2021] expands to 57 regions for greater regional resolution.

IAMs can be broadly categorized into two types: (1) highly aggregated Cost-Benefit models and (2) detailed process-based models [Van Beek *et al.*, 2020]. Cost-Benefit IAMs (CB-IAMs) are optimization models that identify near-term emission reduction pathways to maximize long-term benefits by considering high-level economic and climate interactions. DICE/RICE IAMs fall under this CB-optimization category. In contrast, Process-Based IAMs (PB-IAMs) are simulation models that offer detailed economic representations across multiple sectors to analyze the impacts of specific policies on economic, social, and environmental factors, with an emphasis on sector-specific environmental impacts, such as those in energy systems [Nikas *et al.*, 2019]. CB-IAMs are employed for global mitigation pathways, macroeconomic assessments of mitigation strategies, and strategic interactions reported by the IPCC's Working Group III (WGIII). Notably, the macroeconomic components of CB-IAMs form the foundation of many process-based models. The global mitigation pathways found by CB-IAMs can also be used to inform the simulation of PB-IAMs. The DICE/RICE family, despite its simplicity, has significantly influenced IPCC WGIII mitigation assessments in various reports, including the recent sixth assessment report, particularly in Chapters 3 and 14, which focus on Mitigation Pathways and International Cooperation, respectively [Shukla *et al.*, 2022].

2.2 Multi-Objective Multi-Agent RL

We formalize a MOMARL problem as a multi-objective multi-agent Markov decision process (MOMAMDP) with team reward [Rădulescu *et al.*, 2020], defined as a tuple $\big(\mathcal{S}, \mathcal{A}, F, \mathbf{R}\big)$, comprising $N \geq 2$ agents and $d \geq 2$ objectives:

- S is the state space.
- $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_N$ is the set of joint actions, where $\mathcal{A}_n$ denotes the action set of agent n.
- $F : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ is the probabilistic transition function, mapping each state–joint action pair to a distribution over next states.
- $\mathbf{R} = \mathbf{R}_1 \times \cdots \times \mathbf{R}_n$ represents the reward functions, where $\mathbf{R}_n : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}^d$ is the vector-valued reward function for agent n across d objectives.

Agents optimize policies π_n to maximize expected discounted returns:

$$\mathbf{v}_n^{\pi} = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \mathbf{R}_n(s_t, \mathbf{a}_t, s_{t+1}) \mid \boldsymbol{\pi}\right]$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ is the joint vector policy of the agents acting in the environment, γ is the discount factor, and $\mathbf{R}_n(s_t, \mathbf{a}_t, s_{t+1})$ is the vectorial reward obtained by agent n at timestep t for the joint action $\mathbf{a}_t \in \mathcal{A}$ at state $s_t \in \mathcal{S}$.

Solution Set Concept The value function $\mathbf{v}_n^{\boldsymbol{\pi}} \in \mathbb{R}^d$ offers a partial ordering over policies, as it is a vector. Identifying the optimal policy requires a utility function $u_n : \mathbb{R}^d \to \mathbb{R}$ to capture agents' preferences by mapping vectors to scalars.

The Pareto set, assuming the utility function is monotonically increasing, uses Pareto dominance ($\succ_P$), where a vector dominates another if it is at least equal in all objectives and strictly better in one. When agents share a team reward

($\mathbf{v}_1^{\boldsymbol{\pi}} = \mathbf{v}_2^{\boldsymbol{\pi}} = \cdots = \mathbf{v}_n^{\boldsymbol{\pi}} = \mathbf{v}^{\boldsymbol{\pi}}$), Pareto dominance applies directly. For a set of policies $\boldsymbol{\Pi}$, the Pareto set $P(\boldsymbol{\Pi})$ includes all undominated policies:

$$P(\boldsymbol{\Pi}) = \{\boldsymbol{\pi} \in \boldsymbol{\Pi} \mid \nexists \boldsymbol{\pi}' \in \boldsymbol{\Pi} : \mathbf{v}^{\boldsymbol{\pi}'} \succ_P \mathbf{v}^{\boldsymbol{\pi}}\}.$$

The Pareto front $F(P)$ contains value vectors for Pareto-optimal policies in $P(\boldsymbol{\Pi})$. When the utility function is a positively weighted linear sum, the solution set forms the convex hull of $\mathbf{v}^{\boldsymbol{\pi}}$ [Felten *et al.*, 2024a].

3 The JUSTICE Model

The JUSTICE IAM is a modular and efficient framework created to assess various modeling assumptions related to economic growth, damage functions, abatement costs, and social welfare. It allows experimentation with different uncertainties without incurring high computational costs. We provide an outline of the model and its MOMARL formulation.

3.1 High-Level Outline

JUSTICE incorporates the economy, damage, and abatement modules from the RICE50+ model [Gazzotti *et al.*, 2021] and integrates them with the FAIR climate model [Smith *et al.*, 2023], facilitating probabilistic assessments of climate policies across various climate sensitivity scenarios. FAIR is a simplified emulator of complex climate models grounded in climate physics that accurately reproduces historical climate data and is utilized by the IPCC [Leach *et al.*, 2021; Shukla *et al.*, 2022]. Unlike other IAMs, JUSTICE identifies Pareto-optimal policies in a multi-objective framework, assessing their robustness across various socioeconomic and climate uncertainties and considering the distributional impacts of the Pareto-optimal policies.

JUSTICE operates on a yearly resolution (instead of a 5-year interval used in RICE50+) and covers 57 independent regions (Table 1 in Appendix A[2]). JUSTICE utilizes the Shared Socioeconomic Pathway (SSP) dataset [Riahi *et al.*, 2017] for exogenous regional economic growth, carbon intensity, and population projections, and the Representative Concentration Pathway (RCP) dataset [Meinshausen *et al.*, 2011] informs emission trajectories for other greenhouse gases and aerosols. The SSPs and RCPs provide a consistent framework for integrating socioeconomic and climate research, representing various future scenarios through different socioeconomic narratives and climate forcings. The exogenous SSP data is denoted with a superscript $(^*)$ in the equations.

Figure 1 provides an overview of the JUSTICE IAM. The economy sub-model employs the neoclassical Cobb-Douglas production function [Gazzotti *et al.*, 2021], as in the DICE IAM, to compute economic growth using exogenous labour and total factor productivity data along with capital stock. The resulting economic output feeds the Emissions sub-model, which calculates CO_2 emissions at each timestep based on output and carbon intensity, defined as the fossil fuel share in energy production. The CO_2 emissions from various regions are input into the FAIR climate model, which aggregates them to calculate global mean surface temperature, expressed in °C above pre-industrial levels. This global temperature is then downscaled to regional temperature rise using a data-driven statistical downscaler. The downscaled regional temperature is used by the Damage Function to compute the fraction of economic output damaged by temperature increases in every region. Additionally, the Emissions sub-model allows for CO_2 mitigation, with associated costs calculated by the Abatement module, based on the emission control rate chosen by agents (or regions) at each timestep. Detailed descriptions and equations for each model component are available in Appendix B [2].

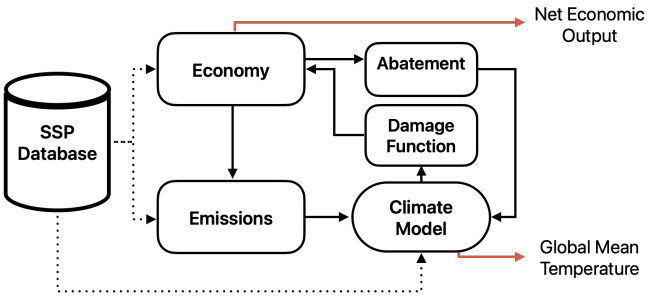


Figure 1: Overview of JUSTICE. The main outputs of the model have been highlighted with red arrows.

3.2 Multi-Agent Multi-Objective Formulation

We model JUSTICE as a MOMAMDP with team reward as specified below.

Agents We model 12 agents, each representing a macro-region that groups countries following RICE region specification [Nordhaus and Yang, 1996]. A map of these macro regions with regional mapping table is provided in Appendix A [2]. While JUSTICE can generate data for 57 entities, modelling a MOMARL problem with 57 regions is computationally expensive and complicates analysis.

Actions Each agent's action is a two-dimensional discrete vector, where the actions of agent n are:

- Emission Control Rate (ECR) $\in \{0, 0.1, 0.2, \ldots, 1.0\}$,
- Savings Rate (SR) $\in \{0, 0.1, 0.2, \ldots, 1.0\}$.

ECR represents the percentage of emissions removed from the atmosphere relative to the baseline SSP emissions. It is used in the Emissions module to assess emissions reduction, with the cost of this reduction being calculated in the Abatement module. SR determines the level of investment, indicating the amount of capital saved and reinvested into the economy, which in turn influences the economic output of each agent (a higher savings rate leads to higher economic output). Each agent undertakes these actions annually.

Observations The agents can observe their local outcomes, the global temperature, and the mitigation rates set by all the agents at the previous timestep. For n-th agent at timestep t, the observation and their ranges are as follows:

$$\Omega_{t,n} = Y_{t,n}^{GROSS} \cdot \omega_{t,n}, \quad \Omega_{t,n} \in [0, \infty), \tag{1}$$

[2]Online Appendix: https://doi.org/10.5281/zenodo.15424271

where $\Omega_{t,n}$ is economic damage by the climate change in trillion USD adjusted to 2005 Purchasing Power Parity (PPP), $Y_{t,n}^{GROSS}$ is the gross economic output and $\omega_{t,n}$ is the damage fraction for agent n at timestep t and is calculated in the Damage module of JUSTICE.

$$\Lambda_{t,n} = \zeta_{t,n} \cdot \varepsilon^*_{t,n} \cdot \left(\frac{AC1_{t,n}}{2} ECR^2_{t,n} + \frac{AC2_{t,n}}{5} ECR^5_{t,n} \right), \quad \Lambda_{t,n} \in [0, \infty), \tag{2}$$

where $\Lambda_{t,n}$ is the abatement cost (cost for the regional economy to mitigate CO_2), in trillion USD adjusted to 2005 PPP, $\zeta_{t,n}$ is the statistical correction factor, $AC1_{t,n}$ and $AC2_{t,n}$ are the region-specific abatement coefficients of agent n at timestep t calculated in the Abatement module of JUSTICE.

$$Y^{NET}_{t,n} = Y^{GROSS}_{t,n} - \Omega_{t,n} - \Lambda_{t,n}, \quad Y^{NET}_{t,n} \in [0, \infty), \tag{3}$$

where $Y^{NET}_{t,n}$ is the net economic output of agent n at timestep t in trillion USD adjusted to 2005 PPP and is calculated in the Economy module of JUSTICE.

$$\varepsilon_{t,n} = {CI^*}_{t,n} \cdot Y^{GROSS}_{t,n} \cdot (1 - ECR_{t,n}) + {AFOLU^*}_{t,n}, \quad \varepsilon_{t,n} \in [0, \infty), \tag{4}$$

where $\varepsilon_{t,n}$ represents total emissions (Annual CO_2 emissions in Gigatonne, $GtCO_2$ per year), ${CI^*}_{t,n}$ is the carbon intensity (exogenous SSP data), ${AFOLU^*}_{t,n}$ is the Agriculture, Forestry, and Other Land Uses emissions (exogenous SSP data) of agent n at timestep t and is calculated in the Emissions module of JUSTICE.

$$GMT_t = \textit{FAIR}\left(\sum_{n \in N} \varepsilon_{t-1,n} \right), \quad GMT_t \in [0, \infty), \tag{5}$$

where GMT_t is the rise in the global average surface temperature since the pre-industrial era at timestep t (in degrees Celsius) outputted by the FAIR model in the climate module.

$$RMT_{t,n} = TC1_{t,n} + TC2_{t,n} \cdot GMT_t, \quad RMT_{t,n} \in [0, \infty), \tag{6}$$

where $RMT_{t,n}$ is the rise in the regional average surface temperature since the pre-industrial era (in degrees Celsius), $TC1_{t,n}$ and $TC2_{t,n}$ are the region-specific downscaler coefficients for agent n at timestep t and is calculated in the Downscaler of the Climate module.

$$\mathbf{VECR_{t,n}} = \{\text{ECR}_{t-1,1}, \text{ECR}_{t-1,2}, \ldots, \text{ECR}_{t-1,N}\}, \quad \mathbf{VECR_{t,n}} \in \{0, 0.1, 0.2, \ldots, 1.0\}^N, \tag{7}$$

where $\mathbf{VECR_{t,n}}$ is a vector of emission control rates adopted by the agents at the previous timestep that agent n observes at timestep t.

Rewards The rewards are modelled as team rewards, which assume collaboration among agents, where all agents receive the same return vector for executing the policy, reflecting their shared goal of improving global outcomes. This setup can be easily adjusted to individual rewards if needed. In our approach, each agent receives a two-dimensional vector of continuous rewards based on the following metrics:

- **Inverse Global Temperature (IGT):**
$$IGT_{t,n} = 1/GMT_t, \quad IGT_{t,n} \in [0, 1]. \tag{8}$$
- **Global Economic Output (GEO):**
$$GEO_{t,n} = \sum_{n=1}^{N} Y^{NET}_{t,n}, \quad GEO_{t,n} \in [0, \infty). \tag{9}$$

In the RL setup, agents maximize their reward. Thus, we use the inverse of the Global Mean Temperature to reflect the goal of minimizing temperature rise. Similarly, we use the absolute value of the net economic output to reflect our aim of maximizing economic performance.

Starting State The JUSTICE simulation is initialized in 2015, using the SSP-2 data. SSP-2 is commonly used in IAM literature and it is the continuation of current emission trends into the future.

Episode Termination Each episode consists of 285 timesteps, with each timestep representing one year, starting from 2015 and continuing until 2300. This extended timeframe accounts for the lag effects of the climate response, allowing the evaluation of damages not only for the current period but also for the centuries that follow.

4 Experimental Setup

We train our agents using Multi-Objective Multi-gent Proximal Policy Optimization (MOMAPPO)[Felten *et al.*, 2024b], an extension of Multi-Agent PPO (MAPPO) [Yu *et al.*, 2022] designed for multi-objective settings. MOMAPPO decomposes a multi-objective problem into multiple single-objective problems using a weighted-sum scalarization function for simplicity. Rewards are normalized to ensure consistency across objectives, and 100 weight vectors are uniformly sampled to generate diverse solutions. For each weight, MOMAPPO trains a multi-agent policy using MAPPO, evaluating its performance and adding non-dominated policies to the solution set. MOMAPPO is trained for 1 million global steps, with evaluations every (approximately) 20,000 steps for 10 random seeds. Appendix C[2] includes additional details.

4.1 Performance Indicators

Evaluating and comparing solution sets in MOMARL is more complex than in single-agent, single-objective RL due to the lack of inherent ordering of solution sets and intertwined performance across objectives. This added complexity leads to varied evaluation methods in MOMARL. To study convergence, we use two most commonly used approaches from both MORL and MARL domains for their suitability in MOMARL settings [Hayes *et al.*, 2022] and to study the equality of distribution between agents we use GINI index:

Hypervolume (↑) [Zitzler *et al.*, 2003] represents the region or (hyper-)volume between the points in the solution set and a reference point. The reference point indicates the lower bound for each objective. A solution set can be assessed by comparing its hypervolume with that of competing algorithms or the true Pareto front, if known.

Expected utility (↑) When the decision-maker's utility function u is linear, the expected utility (EU) metric [Rădulescu *et al.*, 2020] can be used to represent the expected utility over a distribution of reward weights W.

GINI Index (↓) The GINI index is a measure of inequality that quantifies disparities among agents based on a specific metric, with 0 indicating perfect equality and 1 indicating maximal inequality; we employ Concept-1 GINI by [Milanovic, 2011] to assess international inequality, where each agent represents a region and the index reflects whether their metrics are converging.

4.2 RICE50+ Comparison

We compare our results with the default RICE50+ model, as outlined in Section 2.1. This model consists of 57 regions and integrates optimization directly within the global climate simulation. The optimization follows a single-objective approach, employing a social welfare function as the global objective and using a single representative agent to optimize welfare, rather than a multi-agent multi-objective framework. The results presented in the next section are obtained using standard IAM methods, specifically non-linear programming within the General Algebraic Modeling Language (GAMS) for policy optimization. Unlike our approach, RICE50+ does not perform inter-temporal optimization at each timestep. Instead, it optimizes over the entire time horizon by using a social welfare function as a proxy for consumption per capita. Since consumption per capita is directly derived from net economic output (Equation 16a in Appendix B[2]), we extract the net economic output from RICE50+ to compare with our objectives.

5 Results

The objectives of our experiments are to (1) verify the convergence of JUSTICE MOMARL, (2) compare the solutions between JUSTICE and the RICE50+ model, and (3) analyze key solutions from the Pareto set. We present results in line with these objectives.

5.1 Convergence

Figure 2 shows that our agents converge and demonstrate consistent training over time, with both performance metrics growing and eventually stabilizing. Note that both the hypervolume and Expected Utility metrics appear large due to their exponential scaling with the number of objectives and the achievable value ranges, particularly influenced by the high values of the GEO objective.

5.2 Solutions

Figure 3 shows the Pareto set of solutions produced by JUSTICE and the single RICE50+ solution (red star). We transform the objective values for simplicity: Total Global Economic Output represents the GEO objective, and Global Average Annual Temperature corresponds to the mean of the inverse of IGT (inverted to retrieve the temperature).

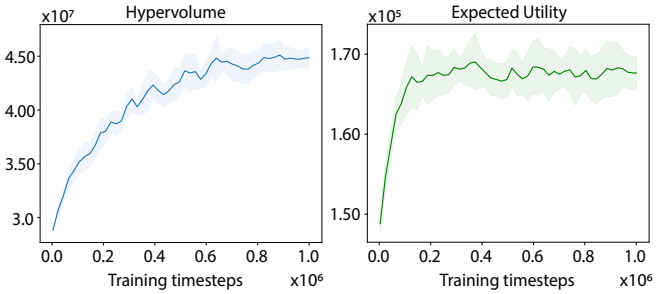


Figure 2: Mean hypervolume and expected utility over training steps (shaded area represents standard deviation).

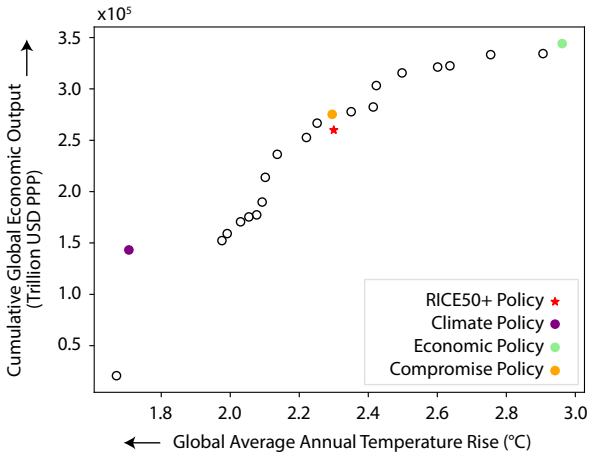


Figure 3: Pareto set of policies obtained by JUSTICE (across 10 random seeds) with the RICE50+ polict for comparison. Arrows indicate the direction of preference for the objectives.

JUSTICE produces 22 solutions (policies), shown as circles in Figure 3. Among these, three are highlighted: Climate Policy (purple), Economic Policy (green), and Compromise Policy (yellow). Although the purple policy is not the most extreme in terms of climate performance, it achieves substantial economic gains with only a slight increase in temperature compared to the extreme climate solution. Therefore, it is chosen as the Climate Policy. This example illustrates how the multi-objective approach supports decision-making by presenting a range of trade-offs.

The RICE50+ solution lies close to the JUSTICE Pareto front, roughly in the middle, indicating a balance between economy and temperature rise. However, the JUSTICE Compromise Policy dominates the RICE50+ solution, albeit by a small margin. Thus, JUSTICE not only offers a range of solutions but also a slightly better solution than RICE50+.

5.3 Comparison of Key Solutions

We perform a comprehensive analysis of the highlighted policies in Figure 3—the three JUSTICE policies and the

RICE50+ policy. To do so, we plot the key IAM outcomes from 2015 to 2300 for these policies in Figure 4.

The economic output trajectories are shown in left panel of Figure 4. As expected, the best economic policy achieves the highest output. The best climate policy starts with the lowest output but eventually aligns with RICE50+ levels, as its rapid and deep mitigation efforts stabilize the climate and reduce damages, stimulating economic growth after 2150. The compromise policy presents an interesting trade-off; although it sacrifices slightly more output in the near term compared to RICE50+, it leads to significantly higher long-term growth, surpassing RICE50+ by hundreds of trillions of dollars, by reducing climate damages through immediate mitigation.

The temperature trajectories are shown in the center panel of Figure 4. As expected, the best economic policy results in the highest temperature rise, while the best climate policy yields the lowest. Under the compromise policy, temperatures exceed 2°C by the end of the 2100s, reaching about 3.5°C by 2300. The RICE50+ temperature projections are significantly lower than those of the JUSTICE compromise policy over the long term. This discrepancy highlights a common criticism of cost-benefit IAMs: their use of simple climate models inadequately captures the complexities and uncertainties of climate sensitivity, leading to smooth projections that miss abrupt, high-impact events [Mastrandrea and Schneider, 2001; Stanton *et al.*, 2009] temperature rises, leading to policy inertia and delayed action [Füssel, 2006; Stern, 2022]. We address these issues by using the FAIR model, which captures nonlinear dynamics and feedback mechanisms, effectively handling uncertainty in climate sensitivity (see Appendix B[2]). FAIR produces temperature scenarios that align with more complex Earth System Models featured in the IPCC [O'Neill *et al.*, 2016].

The abated emissions trajectories are shown in the right panel of Figure 4. All three JUSTICE policies favour rapid near-term mitigation, unlike RICE50+, which delays mitigation and peaks only around 2150. This finding aligns with the criticisms of traditional CB-IAMs for undervaluing early mitigation efforts. Among JUSTICE policies, the best climate policy (purple trajectory) achieves the highest cumulative abated emissions. Both the best economic and compromise policies show similar mitigation levels, though the compromise policy emphasizes earlier action. This finding also aligns with the IPCC's recommendations for rapid near-term mitigation to limit global warming and promote sustainable growth with minimal climate damage [Shukla *et al.*, 2022].

5.4 Equity Analysis

Figure 5 shows the GINI index for emissions among 12 macro-regional agents, illustrating the distribution of mitigation and the future emissions budget. The emissions metric is used because it captures the agent's primary action of emissions control. A basic setup in JUSTICE is employed without explicit equity objectives (such as Utilitarianism in RICE50+) to examine how disaggregating objectives and including multiple agents affect equity outcomes. For consistency, RICE50+ results from 57 regions have been aggregated to match our 12-region model.

As seen in Figure 5, the best climate and economic policies yield higher emissions inequality than the compromise policy. In the climate policy, inequality tends to be higher on average compared to other policies, as developing countries are potentially required to limit emissions despite their growth needs. In contrast, the GINI index for the economic policy is higher than the compromise because high-emitting developed countries are likely to increase emissions to drive economic growth. The compromise policy has the lowest average GINI index and a more equitable emissions distribution compared to RICE50+, which, despite a similar starting point, sees inequality rise sharply to around 0.5 by 2100. This setup demonstrates that the objectives that agents optimize and the chosen Pareto-optimal policy shape the allocation of mitigation burdens and emissions budgets across agents, setting the MOMARL approach apart from single-objective models. Fair policy design requires distinct equity objectives that convert outputs from JUSTICE, such as regional damage and abatement costs, into an equity score while further disaggregating regions to better represent smaller nations. Nonetheless, the MOMARL approach is a promising step towards equitable policy design.

6 Discussion

Our results demonstrate how the optimization setup in IAMs can significantly affect the resulting climate policy recommendations. We compare JUSTICE with the traditional single-objective optimization of RICE50+, a successor of widely used RICE IAM [Nordhaus and Yang, 1996] for studying optimal mitigation pathways. Our findings underscore the advantages of multi-objective optimization, which offers a range of solutions compared to single-objective optimization. We do not make any policy recommendations in this study; rather, we illustrate how policy outcomes are sensitive to the framing of the optimization problem and highlight how equity considerations vary with different solutions and objectives. This flexibility is crucial when agents have unequal economic capacities and face disproportionate climate impacts in an uncertain future.

JUSTICE can have real-world impact in several ways.

Decision Support The multi-objective multi-agent nature makes JUSTICE a powerful decision-support tool for climate (e.g., COP) negotiations. In this use case, each JUSTICE agent can represent a country or a coalition with associated preferences, simulating various policy options and implications. This enables stakeholders to propose, reassess, and refine actions, promoting transparent and robust negotiations.

Scientific Discourse Our model can enrich scientific (e.g., IPCC) reports by providing complementary insights that capture dimensions often overlooked by conventional models. By exploring additional strategies through multi-objective multi-agent reinforcement learning, our approach broadens the scope of potential solutions and offers innovative pathways for regions to understand the consequences of their actions and implement more equitable climate policies.

Broader Engagement While IAMs are crucial for policy recommendations, their technical nature and costly software licenses (e.g., models written in GAMS) often limit accessibility. By offering our framework as open source, we em-

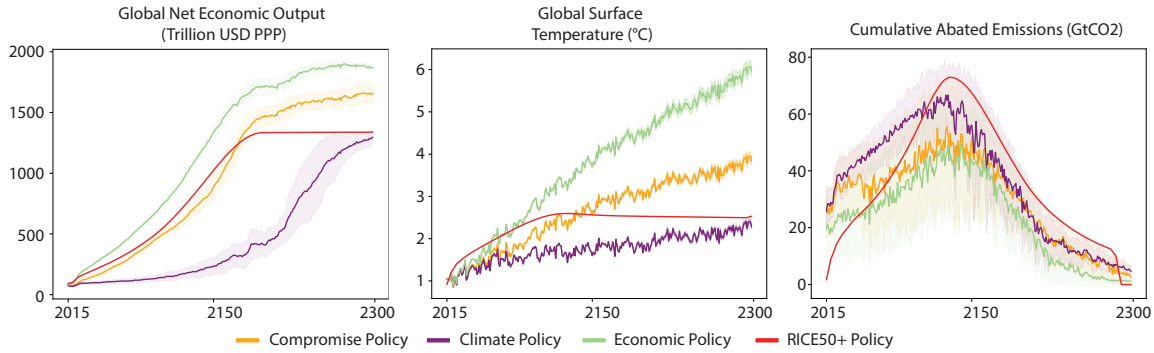


Figure 4: Performance of selected JUSTICE policies over time (years on x-axis) for three important indicators: Global Net Economic Output, Temperature, and Cumulative Abated Emissions. The indicators show mean values and standard deviation (over 10 seeds) for selected JUSTICE and RICE50+ policies over time.

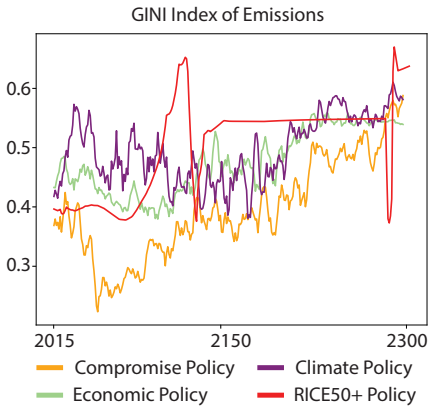


Figure 5: GINI Index of Emissions Over Time for 12 Regions (years on x-axis).

power stakeholders to explore different scenarios and policy impacts. A more user-friendly interface could enhance this further. Further, our approach bridges research (e.g., IAM and AI) communities, demonstrating the benefits of building on models and algorithms developed by each other.

6.1 Limitations and Directions

The following are key limitations of our work. Addressing these limitations opens interesting avenues for future work.

Reward and Utility The MOMARL algorithm we employ makes two key simplifying assumptions. First, a team-based reward structure, where all agents receive the same global reward. This is a simplification since, in practice, countries are likely to prioritize their individual interests. Second, it requires a linear utility function for the multi-objective aspect, which simplifies real-world scenarios. To our knowledge, this is the only MOMARL algorithm suitable for our setup of continuous state and reward spaces, and a discrete action space (see Table 2 in [Felten *et al.*, 2024b]). Relaxing these assumptions, i.e., exploring individual reward structures and other utility functions, is an important direction for future work.

Scalability We perform analyses considering 12 macro-regions of the world. Although this is a substantial improvement over the single-agent assumption of traditional IAMs, a more fine-grained resolution may be necessary to derive insights on equity. This requires a substantial speedup of current MOMARL algorithms. One direction in this regard is to experiment with alternative weight-sampling techniques, such as adaptive weight sampling [Felten *et al.*, 2024a].

Communication Our current model assumes cooperation among regions, whereas real climate negotiations involve complex bargaining and conflicting agendas. Exploring different communication strategies among agents, and modeling trade and capital transfers, could improve the model's ability to reflect real-world dynamics. Further, training the agents on different climate ensembles of FAIR, which represent varying plausible climate sensitivities, would allow the agents to take actions that are robust under climate uncertainty.

Policy Implementation We acknowledge that the policies generated by JUSTICE may be unrealistic, as both RICE50+ and JUSTICE assume mitigation begins at the start of the model run, and the drastic mitigation suggested by JUSTICE is infeasible due to geopolitical dynamics, technological inertia, and policy implementation challenges. However, the utility of JUSTICE is in providing a range of alternatives, which can serve as a scientific basis for stakeholder deliberations.

6.2 Conclusions

We introduce the JUSTICE framework, the first to integrate IAM with MOMARL. Our approach stands apart from the models typically featured in climate reports—models upon which policymakers rely to inform decisions and negotiate actions. Unlike traditional models, which often oversimplify climate change into problems with a single agent and objective, our model embraces the complexity of the real world by acknowledging its multi-agent, multi-objective nature. Our experiments show that JUSTICE produces flexible policies and allows detailed exploration of equity compared to RICE50+. Additionally, our model is open-source, implemented in Python, and designed to be accessible for exploration and experimentation of RL algorithms for an important real-world problem.

Contribution Statement

Palok Biswas, Zuzanna Osika, Isidoro Tamassia, and Adit Whorra made equal contributions to this study and are designated as co-first authors. Jazmin Zatarain Salazar, Jan Kwakkel, Frans A. Oliehoek and Pradeep K. Murukannaiah, reviewed the final manuscript.

References

[Asayama, 2024] Shinichiro Asayama. The history and future of ipcc special reports: A dual role of politicisation and normalisation. *Climatic Change*, 177(9):137, 2024.

[Bromley and Beattie, 1973] Daniel W Bromley and Bruce R Beattie. On the incongruity of program objectives and project evaluation: An example from the reclamation program. *American Journal of Agricultural Economics*, 55(3):472–476, 1973.

[Burke *et al.*, 2018] Marshall Burke, W Matthew Davis, and Noah S Diffenbaugh. Large potential reduction in economic damages under un mitigation targets. *Nature*, 557(7706):549–553, 2018.

[Cointe *et al.*, 2019] Béatrice Cointe, Christophe Cassen, and Alain Nadaï. Organising policy-relevant knowledge for climate action: integrated assessment modelling, the ipcc, and the emergence of a collective expertise on socioeconomic emission scenarios. *Science & Technology Studies*, 2019.

[Faus Onbargi, 2022] Alexia Faus Onbargi. The climate change–inequality nexus: towards environmental and socio-ecological inequalities with a focus on human capabilities. *Journal of Integrative Environmental Sciences*, 19(1):163–170, 2022.

[Felten *et al.*, 2024a] Florian Felten, El-Ghazali Talbi, and Grégoire Danoy. Multi-objective reinforcement learning based on decomposition: A taxonomy and framework. *Journal of Artificial Intelligence Research*, 79:679–723, 2024.

[Felten *et al.*, 2024b] Florian Felten, Umut Ucak, Hicham Azmani, Gao Peng, Willem Röpke, Hendrik Baier, Patrick Mannion, Diederik M Roijers, Jordan K Terry, El-Ghazali Talbi, et al. Momaland: A set of benchmarks for multi-objective multi-agent reinforcement learning. *arXiv preprint arXiv:2407.16312*, 2024.

[Ferrari *et al.*, 2022] Luca Ferrari, Angelo Carlino, Paolo Gazzotti, Massimo Tavoni, and Andrea Castelletti. From optimal to robust climate strategies: expanding integrated assessment model ensembles to manage economic, social, and environmental objectives. *Environmental Research Letters*, 17(8):084029, 2022.

[Füssel, 2006] HM Füssel. Logical and empirical flaws in applications of simple climate-economy models. *Kluwer Academic Publishers*, 2006.

[Gambhir *et al.*, 2019] Ajay Gambhir, Isabela Butnar, Pei-Hao Li, Pete Smith, and Neil Strachan. A review of criticisms of integrated assessment models and proposed approaches to address these, through the lens of beccs. *Energies*, 12(9):1747, 2019.

[Gambhir *et al.*, 2022] Ajay Gambhir, Gaurav Ganguly, and Shivika Mittal. Climate change mitigation scenario databases should incorporate more non-iam pathways. *Joule*, 6(12):2663–2667, 2022.

[Gazzotti *et al.*, 2021] Paolo Gazzotti, Johannes Emmerling, Giacomo Marangoni, Andrea Castelletti, Kaj-Ivar van der Wijst, Andries Hof, and Massimo Tavoni. Persistent inequality in economically optimal climate policies. *Nature Communications*, 12(1):3421, 2021.

[Grubb *et al.*, 2021] Michael Grubb, Claudia Wieners, and Pu Yang. Modeling myths: On dice and dynamic realism in integrated assessment models of climate change mitigation. *Wiley Interdisciplinary Reviews: Climate Change*, 12(3):e698, 2021.

[Hayes *et al.*, 2022] Conor F Hayes, Roxana Rădulescu, Eugenio Bargiacchi, Johan Källström, Matthew Macfarlane, Mathieu Reymond, Timothy Verstraeten, Luisa M Zintgraf, Richard Dazeley, Fredrik Heintz, et al. A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems*, 36(1):26, 2022.

[Kasprzyk *et al.*, 2016] Joseph R Kasprzyk, Patrick M Reed, and David M Hadka. Battling arrow's paradox to discover robust water management alternatives. *Journal of Water Resources Planning and Management*, 142(2):04015053, 2016.

[Leach *et al.*, 2021] Nicholas J Leach, Stuart Jenkins, Zebedee Nicholls, Christopher J Smith, John Lynch, Michelle Cain, Tristram Walsh, Bill Wu, Junichi Tsutsui, and Myles R Allen. Fairv2. 0.0: a generalized impulse response model for climate uncertainty and future scenario exploration. *Geoscientific Model Development*, 14(5):3007–3036, 2021.

[Lönngren and Svanström, 2016] Johanna Lönngren and Magdalena Svanström. Systems thinking for dealing with wicked sustainability problems: Beyond functionalist approaches. *New developments in engineering education for sustainable development*, pages 151–160, 2016.

[Marangoni *et al.*, 2021] Giacomo Marangoni, Jonathan R Lamontagne, Julianne D Quinn, Patrick M Reed, and Klaus Keller. Adaptive mitigation strategies hedge against extreme climate futures. *Climatic Change*, 166(3):37, 2021.

[Mastrandrea and Schneider, 2001] Michael D Mastrandrea and Stephen H Schneider. Integrated assessment of abrupt climatic changes. *Climate Policy*, 1(4):433–449, 2001.

[Mastrandrea, 2009] Michael D Mastrandrea. Calculating the benefits of climate policy: examining the assumptions of integrated assessment models. *Pew Center on Global Climate Change Working Paper*, 2009.

[Meinshausen *et al.*, 2011] Malte Meinshausen, Steven J Smith, Katherine Calvin, John S Daniel, Mikiko LT Kainuma, Jean-Francois Lamarque, Kazuhiko Matsumoto, Stephen A Montzka, Sarah CB Raper, Keywan Riahi, et al. The rcp greenhouse gas concentrations and

their extensions from 1765 to 2300. *Climatic change*, 109:213–241, 2011.

[Milanovic, 2011] Branko Milanovic. *Worlds apart: Measuring international and global inequality*. Princeton University Press, 2011.

[Nikas *et al.*, 2019] Alexandros Nikas, Haris Doukas, and Andreas Papandreou. A detailed overview and consistent classification of climate-economy models. *Understanding risks and uncertainties in energy and climate policy: Multidisciplinary methods and tools for a low carbon society*, pages 1–54, 2019.

[Nordhaus and Yang, 1996] William D Nordhaus and Zili Yang. A regional dynamic general-equilibrium model of alternative climate-change strategies. *The American Economic Review*, pages 741–765, 1996.

[Nordhaus, 1992] William D Nordhaus. An optimal transition path for controlling greenhouse gases. *Science*, 258(5086):1315–1319, 1992.

[O'Neill *et al.*, 2016] Brian C O'Neill, Claudia Tebaldi, Detlef P Van Vuuren, Veronika Eyring, Pierre Friedlingstein, George Hurtt, Reto Knutti, Elmar Kriegler, Jean-Francois Lamarque, Jason Lowe, et al. The scenario model intercomparison project (scenariomip) for cmip6. *Geoscientific Model Development*, 9(9):3461–3482, 2016.

[Pozo *et al.*, 2020] Carlos Pozo, A Galán-Martín, D Cortés-Borda, Marta Sales-Pardo, Adisa Azapagic, R Guimerà, and Gonzalo Guillén-Gosálbez. Reducing global environmental inequality: Determining regional quotas for environmental burdens through systems optimisation. *Journal of cleaner production*, 270:121828, 2020.

[Rădulescu *et al.*, 2020] Roxana Rădulescu, Patrick Mannion, Diederik M Roijers, and Ann Nowé. Multi-objective multi-agent decision making: a utility-based analysis and survey. *Autonomous Agents and Multi-Agent Systems*, 34(1):10, 2020.

[Radulescu, 2024] Roxana Radulescu. The world is a multi-objective multi-agent system: Now what? In *27th European Conference on Artificial Intelligence*, pages 32–38. IOS Press, 2024.

[Riahi *et al.*, 2017] Keywan Riahi, Detlef P Van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C O'neill, Shinichiro Fujimori, Nico Bauer, Katherine Calvin, Rob Dellink, Oliver Fricko, et al. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global environmental change*, 42:153–168, 2017.

[Rising *et al.*, 2022] James Rising, Marco Tedesco, Franziska Piontek, and David A Stainforth. The missing risks of climate change. *Nature*, 610(7933):643–651, 2022.

[Rivadeneira and Carton, 2022] Natalia Rubiano Rivadeneira and Wim Carton. (in) justice in modelled climate futures: A review of integrated assessment modelling critiques through a justice lens. *Energy Research & Social Science*, 92:102781, 2022.

[Shukla *et al.*, 2022] P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, and J. Malley. Climate change 2022: Mitigation of climate change. contribution of working group iii to the sixth assessment report of the intergovernmental panel on climate change, 2022.

[Smith *et al.*, 2023] Christopher J Smith, Alaa Al Khourdajie, Pu Yang, and Doris Folini. Climate uncertainty impacts on optimal mitigation pathways and social cost of carbon. *Environmental Research Letters*, 18(9):094024, 2023.

[Stanton *et al.*, 2009] Elizabeth A Stanton, Frank Ackerman, and Sivan Kartha. Inside the integrated assessment models: Four issues in climate economics. *Climate and Development*, 1(2):166–184, 2009.

[Stern, 2022] Nicholas Stern. A time for action on climate change and a time for change in economics. *The Economic Journal*, 132(644):1259–1289, 2022.

[Van Beek *et al.*, 2020] Lisette Van Beek, Maarten Hajer, Peter Pelzer, Detlef van Vuuren, and Christophe Cassen. Anticipating futures through models: the rise of integrated assessment modelling in the climate science-policy interface since 1970. *Global Environmental Change*, 65:102191, 2020.

[van der Wijst *et al.*, 2021] Kaj-Ivar van der Wijst, Andries F Hof, and Detlef P van Vuuren. On the optimality of 2° c targets and a decomposition of uncertainty. *Nature communications*, 12(1):2575, 2021.

[Wei *et al.*, 2013] Yi-Ming Wei, Le-Le Zou, Kai Wang, Wen-Jin Yi, and Lu Wang. Review of proposals for an agreement on future climate policy: Perspectives from the responsibilities for ghg reduction. *Energy Strategy Reviews*, 2(2):161–168, 2013.

[Yu *et al.*, 2022] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 35:24611–24624, 2022.

[Zhang *et al.*, 2022] Tianyu Zhang, Andrew Williams, Soham Phade, Sunil Srinivasa, Yang Zhang, Prateek Gupta, Yoshua Bengio, and Stephan Zheng. Ai for global climate cooperation: modeling global climate negotiations, agreements, and long-term cooperation in rice-n. *arXiv preprint arXiv:2208.07004*, 2022.

[Zitzler *et al.*, 2003] Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M Fonseca, and Viviane Grunert Da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on evolutionary computation*, 7(2):117–132, 2003.