AI for Global Climate Cooperation: Modeling Global Climate Negotiations, Agreements, and Long-Term Cooperation in RICE-N

Tianyu Zhang[1,2]**, Andrew Williams**[1,2]**, Soham Phade**[6]**, Sunil Srinivasa**[6*]**, Yang Zhang**[2]**,**
Prateek Gupta[2,3,4]**, Yoshua Bengio**[1,2,5]**, Stephan Zheng**[6]

[1] Université de Montréal, [2] Mila, [3] University of Oxford,
[4] The Alan Turing Institute, [5] CIFAR, [6] Salesforce Research

Abstract

Comprehensive global cooperation is essential to limit global temperature increases while continuing economic development, e.g., reducing severe inequality or achieving long-term economic growth. Achieving long-term cooperation on climate change mitigation with n strategic agents poses a complex game-theoretic problem. For example, agents may negotiate and reach climate agreements, but there is no central authority to enforce adherence to those agreements. Hence, it is critical to design negotiation and agreement frameworks that foster cooperation, allow all agents to meet their individual policy objectives, and incentivize long-term adherence. This is an interdisciplinary challenge that calls for collaboration between researchers in machine learning, economics, climate science, law, policy, ethics, and other fields. In particular, we argue that machine learning is a critical tool to address the complexity of this domain. To facilitate this research, here we introduce RICE-N, a multi-region *integrated assessment model* that simulates the global climate and economy, and which can be used to design and evaluate the strategic outcomes for different negotiation and agreement frameworks. We also describe how to use multi-agent reinforcement learning to train rational agents using RICE-N. This framework underpins *AI for Global Climate Cooperation*, a working group collaboration and competition on climate negotiation and agreement design. Here, we invite the scientific community to design and evaluate their solutions using RICE-N, machine learning, economic intuition, and other domain knowledge. More information can be found on www.ai4climatecoop.org.

1 Introduction

The latest IPCC report (Pörtner et al. 2022) warns that it is "now or never" to stave off a climate disaster. Ecosystems are drastically changing: the Amazon rainforest is receding (Lovejoy and Nobre 2018) and polar ice sheets are melting (Boers and Rypdal 2021; DeConto et al. 2021). Extreme weather events are clear warning signs too, such as the recent uptick in coastal flooding and forest fires. These developments are increasingly being attributed to climate change and driving towards a system-wide tipping point.

Climate change is a global phenomenon, and it affects all. In response, private and public financing have driven technological innovation (e.g., in renewable energy) and community campaigns to drive systemic change. However, climate change mitigation investments vary in size and type across nations and global regions, due to various social and economic factors. For example, developing nations may need to focus first on the basic needs of their citizens, while developed nations have more funding and opportunities to prepare for the adverse impacts of climate change. Hence, climate change presents a classic case of "tragedy of the commons", where individual agents who pursue their own self-interest may lead to a destructive outcome for everyone.

As such, achieving and maintaining global cooperation is essential to achieve the Paris Agreement's long-term goal of limiting global temperature rise to well below $2°C$ above pre-industrial levels (DeConto et al. 2021). At the same time, it is important to maintain economic development, e.g., achieving growth and reducing inequality. For example, international trade treaties, foreign investment, and technology transfer may enable developing countries to meet net-zero commitments, while contributing to global economic growth. Such cooperation may be implemented through climate clubs (Nordhaus 2015), which may overcome barriers to taking action on climate change.

From a modeling point of view, achieving and maintaining global cooperation poses a complex game-theoretic problem involving cooperation, communication, and competition. This game can be modeled with n strategic agents, each agent representing a region or nation. Each agent is (boundedly) rational and implements policies to achieve its own socio-economic and climate objectives, which may be at odds with the objectives of the other agents. Agents interact through trade, diplomacy, or foreign aid and investments. Cooperation may happen through mutual negotiation and agreements.

However, a crucial issue is that there is no central entity

---

*Work done while at Salesforce.

to enforce cooperation or adherence to signed agreements in the real world. For these reasons, it is critical to design multilateral negotiations and agreements that best foster *continuing* cooperation towards mitigating climate change while allowing all parties to meet their individual policy objectives.

Such game-theoretic problems pose unsolved technical challenges. For instance, a key analysis in the 2022 IPCC report (Pörtner et al. 2022) predicts climate change under five different so-called Shared Socio-Economic Pathways (SSP). Each SSP uses a manually defined set of climate policies for each global region. However, a key limitation is that it is unclear whether these policies would be executed by rational actors, and thus how likely these scenarios are to materialize or how robust these scenarios are to agents who may change their behavior over time.

Contributions. In this paper, we propose a conceptual and practical framework to overcome these limitations and provide a strategic analysis of rational agent behaviors and their impact on climate change. We also introduce *AI for Global Climate Cooperation*, a community initiative to foster interdisciplinary research and real-world impact in this area. Specifically, we present the following:

- we argue that machine learning (ML) offers an attractive framework to analyze the strategic aspects of real world climate negotiations and agreements;

- we argue this research requires combining both existing and new tools: calibrated agent-based climate-economic simulations, multi-agent machine learning, deep reinforcement learning (RL), game theory, and integrating data;

- we urge the community to pursue interdisciplinary research, e.g., between economics, climate science, machine learning and other fields to understand the impact of strategic behavior on climate change;

- we introduce RICE-N, a calibrated climate-economic simulation that can model negotiations, agreements, and strategic behaviors between multiple regions;

- we describe how to use multi-agent reinforcement learning (MARL) to train rational agents in RICE-N;

- we introduce a working group collaboration to build momentum and to have actual policy impact through peer-reviewed publications authored by groups and policy briefs and communication with policymakers; and

- as a first step, we are organizing a competition to evaluate solutions, both through quantitative results and scientific assessment by a diverse, interdisciplinary expert jury. The objective of this competition is to design contracts and protocols that foster global climate cooperation, using a mix of machine learning tools, economic intuition, and domain knowledge. Participants will evaluate their solutions using RICE-N and a web platform run by the group.

For up-to-date information, see www.ai4climatecoop.org.

2 Related Work

Integrated assessment models. Climate-economic simulations, such as RICE-N, are known as Integrated Assessment Models (IAM). IAMs are commonly used in climate change policy analysis to simulate future scenarios, based on a joint climate-economic dynamics model that quantifies the effect of economic activity and CO_2 emissions on global temperatures and long-term economic development. A pioneering example is the Dynamic Integrated model of Climate and Economy (DICE) (Nordhaus 2007), which models the links among climate and economic factors, e.g., population growth, technological change, CO_2 emissions and concentrations, global temperatures, and economic damages. Notably, DICE models a single global region. The Regional Integrated Climate-Economy (RICE) model generalizes this to multiple heterogeneous regions, modeling heterogeneous population growth, climate change mitigation investments, and other factors. The RICE model was further extended to include trade and tariffs (Lessmann, Marschinski, and Edenhofer 2009).

Key limitations of IAMs include a lack of modeling decision-making under uncertainty, distributional analysis, technological change, and realistic economic damage functions (Farmer et al. 2015). Moreover, cooperation, communication, and competition between regions can significantly influence regional policies and their (long-term) effects, but are under-explored in IAMs. Our proposed framework, based on RICE-N, machine learning, and other domains, addresses the latter shortcoming and is inspired by agent-based modeling (Bonabeau 2002) as a bottom-up modeling framework and advances in MARL to find well-performing policies (Zheng et al. 2022) and to train human-level strategic agents (Silver et al. 2016; Vinyals et al. 2019).

Multi-agent aspects of climate change. Previous work has studied the connection between political economy, negotiations, and climate change. Empirical work has found that previous climate summits have had inconsistent or too little impact (Chan et al. 2022; Bakaki 2022). The impact of social dynamics has been studied in a stylized climate-social model, finding that public perception and institutional responsiveness are important to explain variations in emissions (Moore et al. 2022). Coalition-forming under climate negotiations and agreements have also been studied from a game-theory perspective (Zenker 2019). IAMs have also been used to study the impact of political bargaining on the economic burden required to meet climate targets (Rochedo et al. 2018). However, to the best of our knowledge, no work has analyzed the game-theoretic aspects of climate cooperation using machine learning and calibrated IAMs.

Strategic behavior and climate change. Game theory has long studied the collective behavior of self-interested agents, e.g., the tragedy of the commons (Hardin 1968), negotiation and agreements of agents with conflicting and common goals (Schelling 1980). Certain works have analyzed international negotiations on climate collaboration and agreements regarding economic activity and climate efforts, e.g., imposing tariffs on countries that do not mitigate sufficiently. Climate negotiations have been studied using mathematical games, e.g., coordination games or prisoner's dilemmas (De-Canio and Fremstad 2013). However, the reliability of such simplified models for real-world policy has been called into question. In particular, these games lack (i) a multilateral, rather than bilateral, setting, (ii) strategic behavior from

agents with multiple, possibly conflicting, goals, (iii) evolving climate dynamics and changing agent behavior that lead to non-equilibrum outcomes, and (iv) heterogeneity among agents (Madani 2013).

Subsequent work has gone beyond equilibrium analysis by modeling climate negotiations as a bargaining game in which agents learn, albeit in a highly simplified manner (Smead et al. 2014). Climate scenarios could be studied through the emergence of climate mitigation from these games with learning, in which regions can cooperate or compete (Greeven et al. 2016). Furthermore, other work has studied the difficulty of long term climate collaboration (Carney 2015), as well as potential mechanisms for overcoming associated issues (Nordhaus 2015).

Multi-agent reinforcement learning. MARL has emerged in recent years as an attractive framework that studies how to train rational agents that may communicate, cooperate, or compete. This is a rich area of research that intersects machine learning with game theory, economics, and other domains (Shoham and Leyton-Brown 2008). Games can be classified as cooperative, competitive, or a mixture of both. For instance, in fully cooperative games, agents learn to work together, e.g., to lower the carbon power consumption of heating, ventilation and air conditioning (HVAC) systems (Mai, Zhang, and Lesage-Landry 2021; Hanumaiah and Genc 2021; Yu et al. 2020), or in the game of Hanabi (Yu et al. 2021). On the other hand, in a competitive game, agents may need to find strategies to defeat opponents, e.g., in Diplomacy (Paquette et al. 2019) and Go (Schrittwieser et al. 2020). However, many games are neither purely competitive nor purely cooperative. Hence, multi-agent games can be simple to define, yet it may be hard to find optimal agent strategies using learning-based methods, e.g., in the iterative prisoner's dilemma and coin games (Lerer and Peysakhovich 2017; Foerster et al. 2017).

Recent work has explored the link between MARL and negotiation (Cao et al. 2018), as well as cooperation in social dilemmas and collaboration on climate change (Jaques et al. 2019; Chelarescu 2021; Le Gléau et al. 2022). As such, MARL is an attractive framework to analyze climate outcomes taking strategic behavior into account. However, previous work has largely considered highly stylized environments and has not yet been applied to rich calibrated climate-economic simulations; our work fills this gap.

3 The RICE-N Integrated Assessment Model

We introduce RICE-N, an IAM that further augments RICE with a framework for negotiation protocols, and also includes international trade and tariffs, following Lessmann, Marschinski, and Edenhofer (2009)). As such, RICE-N shares climate, economic and social characteristics with the real world.

In RICE-N, there are n regions, each modeled as an independent decision-making agent. Regions interact with each other and the environment through their actions: setting a savings rate, mitigation rate, trades and tariffs, and negotiation actions, for each time step (e.g., every 5 years).

RICE-N has two main components: *negotiation* and *climate-economic activity*, see Figure 1. The activity com-

ponent simulates the physical actions of the agents and the resulting evolution of the environment. The negotiation component simulates communication between regions, allowing them to influence each other's behaviors and form agreements. Agreements may in turn adjust the available actions for each region during the activity stage.

Each simulation episode consists of H steps, each step representing Δ years (e.g, $\Delta = 5$). Thus, the simulation spans a period of $H \times \Delta$ simulation-years. At every step, the simulation goes through the negotiation stages, and agreements are formed within the regions. The simulation then enters the activity stage where each region takes actions that are affected by the agreements formed during the negotiation stages.

For the public version of the simulation, we have fitted the structural parameters of the simulation to real data. However, the regions and their characteristics in the public version are not an identical one-to-one representation of the real world. Moreover, the regions are fictitious and do not represent real-world nations.

3.1 Base Dynamics

We now describe the climate-economic and trade dynamics of the simulation that are based on existing work.

Climate-economic dynamics overview. The *state of the world* is characterized by global variables such as the concentration of CO_2 levels in the Earth's atmosphere, and the average global temperature, as well as region-specific variables such as population, capital, technology level, carbon intensity of economic activity, and balance of trade (see Table 2 in Appendix J).

The climate dynamics model how CO_2 levels in the atmosphere impacts global temperatures. The economic dynamics model how technology levels, capital, population, and gross domestic production evolve. Notably, the climate dynamics impact the economic dynamics through a *damage function* which describes how higher temperatures lead to losses in capital.

These dynamics depend on savings and mitigation rates set by each agent, e.g., agents may choose to invest more in climate change mitigation, but this may lower economic productivity in the short-term.[1] As global CO_2 levels and temperatures affect all agents, these dynamics mean the decisions of each agent affect the climate-economic outcomes for other agents too.

The activity component encapsulates these dynamics. At every step, it does the following: The gross output production for each region is computed based on the state of the region, in particular, its capital investment, labor (or population), and technology factor. The net economic output is the gross output production reduced by climate damages from rising

---

[1]In economic terms, variables such as capital, balance of trade, carbon mass, and global temperature depend on the agents' actions and are called *endogenous* variables. On the other hand, variables such as population, technology level, carbon intensity of economic activity are called *exogenous*, i.e., their values do not depend on the agent actions. Note that the values of endogenous variables can vary across steps in a predetermined manner.
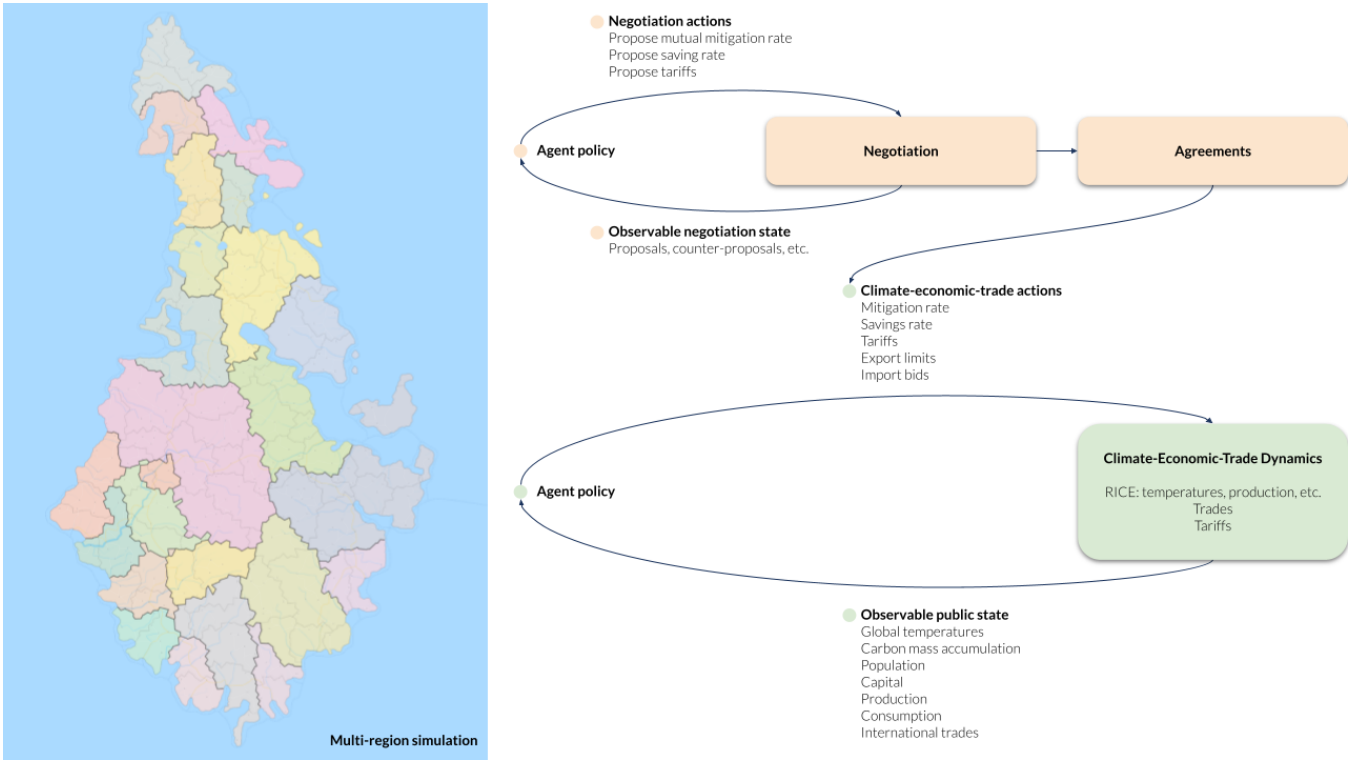
Figure 1: **Schematic overview of observations, actions, and components in RICE-N.** Each region (agent) uses a policy model to negotiate and make climate, economic, and trade decisions. For clarity, we show the flow of information and actions for a single agent only. The details of the negotiation protocol are up to competitors to implement. At a high-level, at each timestep, each policy first receives an observation of the negotiation and world state and decides how to negotiate. Negotiation may proceed for several iterations (at the same timestep). The outcome of all negotiations are agreements (or lack thereof), which may be between two or more agents. In particular, an agreement may influence the remaining actions that an agent can take in the climate and economic domains. For the same timestep, each agent then makes decisions with respect to the climate, economy, and trade.

global temperature, and the cost of efforts towards mitigation by this region. The region consumes domestic goods equal to the quantity of the net economic output that is left after capital investment and export. It also consumes foreign goods from imports. The consumption utility for each region from consuming domestic and foreign goods is computed using the Armington elasticity assumption that has become standard in international computable general equilibrium models (Armington 1969). This gives the *reward* corresponding to each region in every step. Please refer to Appendix A.1 and Table 3 in Appendix J for more details.

International trade and tariffs. RICE-N features international trade to exchange and transfer goods between agents, following (Lessmann, Marschinski, and Edenhofer 2009). Here, agents are modelled to seek diversity in their consumption according to the Armington assumption (Armington 1969), so they want to consume goods produced by other agents and are willing to export some of their own goods in exchange. Each agent specifies how many goods they would like to import and sees its orders (partially) filled, depending on the other agents' willingness to export.

In addition, agents can choose to impose import tariffs to restrict trading. Import tariffs restrict the consumption of

the imported goods that they are applied to, which implicitly increases their prices. These price increases make imported goods without import tariffs more attractive. Therefore, trade and tariffs force agents to engage with other regions and be strategic, and hence may incentivize negotiations and signing agreements. Please refer to Appendix A.2 for more details.

4 The Negotiation Component

A key objective of our working group collaboration and competition is to test different negotiation protocols that affect the actions during the activity stage and lead to socially better outcomes. As such, the negotiation component can include different stages of communication and the rules to form agreements between the regions. Note that RICE-N can work with an empty negotiation component, and indeed these provide important baselines for performance evaluation.

For a detailed exposition and formal details on negotiation protocols and agreements, see Section K in the Appendix.

Binding agreements via action masks. In the base implementation of RICE-N, we allow the negotiated agreements to control the range of actions allowed for the regions. The effect of an agreement can then be encoded in the form of

action masks that controls the allowed action space during the activity stage. For example, an agreement could state that a given region should implement a minimum of 20% mitigation rate. Then the action mask only allows setting mitigation rates above that level for this region.

Hence, it is important to ensure that the negotiation protocol provides each region with a way to reject any agreement that restricts its actions. For example, it shouldn't happen that a certain region is forced to implement a mitigation rate that it did not agree to (Mitchell 2003). Even if the agents are provided a choice to accept or reject an agreement, action masks could be implemented by the simulation designer to artificially restrict actions to make an agreed-upon agreement fully-binding. However, this is less realistic.

Non-binding agreements. Alternatively, an implementation can include agent observations that allow regions to check if the agreements were followed or not and learn to punish the regions that did not comply with them. Alternatively, a reputation system could be built based on the observations corresponding to agents complying with their agreements, with agents using these reputation scores to engage with each other for future agreements. Such implementations would be more realistic as compared to an implementation based on action masks because they do not assume any agreement enforcing entity.

Properties of negotiation protocols. A wide range of protocols can be implemented as part of the negotiation component. A key goal is to design a negotiation protocol so that agents may not only maximize their own utility but also care about the collective goal: mitigating global temperature rise. From a theoretical point of view, attractive properties of negotiation protocols could be:

- *incentive-compatible*: every agent can achieve the best outcome for themselves by acting according to their true preferences (Pavan, Segal, and Toikka 2014),
- *self-enforcing*: no external body or agent is required to enforce other agents to participate in negotiations and adhere to agreements (Telser 1980),
- *strategy-proof*: without information about other agents' actions, agents do at least as well by being truthful (Li 2017).

We now discuss several example negotiation protocols.

4.1 No Negotiations

Let us begin with the most simple negotiation protocol, the one with no negotiation. In this case, the simulation runs through the activity component at each step. Each region chooses its actions from the entire feasible range of actions without any restriction, since no agreements are formed. Each region aims to optimize for its individual rewards and thus this is the classic case of tragedy of the commons.

4.2 Unilateral Contracts

The negotiation component could include various constraints that do not require negotiation. For example, if two regions a and b enact mitigation rates μ_a and μ_b such that $\mu_a > \mu_b$,

then an agreement could entail that region a imposes a tariff $\tau_a = \alpha(\mu_a - \mu_b)$ on region b where α is a "mitigation correction" coefficient. This is a simple mechanism through which a region a could incentivize regions b to mitigate more.

4.3 Bilateral Negotiations on Mitigation Rates

The negotiation component can be composed of multiple stages of observations and actions that ultimately lead to agreements within the regions. Consider the following protocol: For each ordered pair of regions (i, j), region i makes a proposal $(\hat{\mu}_i, \hat{\mu}_j)$ and region j decides whether to accept the proposal or not. The above proposal means that, if accepted by region j, then an agreement is formed between region i and region j that says, region i will choose a mitigation rate μ_i at least as large as $\hat{\mu}_i$ and region j will choose a mitigation rate μ_j at least as large as $\hat{\mu}_j$ as their inputs to the activity component.

Thus, during negotiation, there will be $n(n-1)$ total proposals and decisions regarding their acceptance. After negotiation, each region i will select its mitigation rate μ_i that is greater than or equal to all of the mitigation rates $\hat{\mu}_i$ in the accepted proposals that region i was a part of.

This negotiation protocol can be implemented in two stages:

- *Proposal stage*: At this stage, each region i makes a proposal to every other region j.
  Observations: state observations for agent i.
  Actions: Proposals $(\hat{\mu}_i, \hat{\mu}_j)$ for every other region j.
- *Evaluation stage*: At this stage, each region observes the proposals made to it in the preceding stage and takes an action of accepting or rejecting each of the proposals.
  Observations: state observations for agent i, incoming and outgoing proposals for agent i.
  Actions: Accept or reject each received proposal.

At the end of the evaluation stage, the minimum required mitigation rate is computed for each region and the corresponding action mask is set. This ensures that the regions choose actions for the activity component that comply with the agreements formed in the negotiation component. Instead of using action masks, the implementation can allow agents to take any action, but include these actions as observations for the other agents so that they can learn to punish any agreement violations. For example, regions can increase the tariff rates on regions that violate agreements.

Visualizing outcomes. Figure 2 visualizes the effect of a baseline implementation of the bilateral negotiation protocol, with the agents trained independently using PPO (Schulman et al. 2017), an RL algorithm (see Section 5 for details). You can tell that the negotiations and agreements have a beneficial effect: global temperatures rise less than without negotiation. However, note that this baseline protocol may not achieve the best trade-off in climate-economic outcomes possible. Also note that this protocol is not necessarily robust against sim-to-real gaps, e.g., if the agent reward function is mis-specified or the dynamics are not calibrated well. A key objective of this initiative is to investigate such protocols in more detail and elucidate their properties.
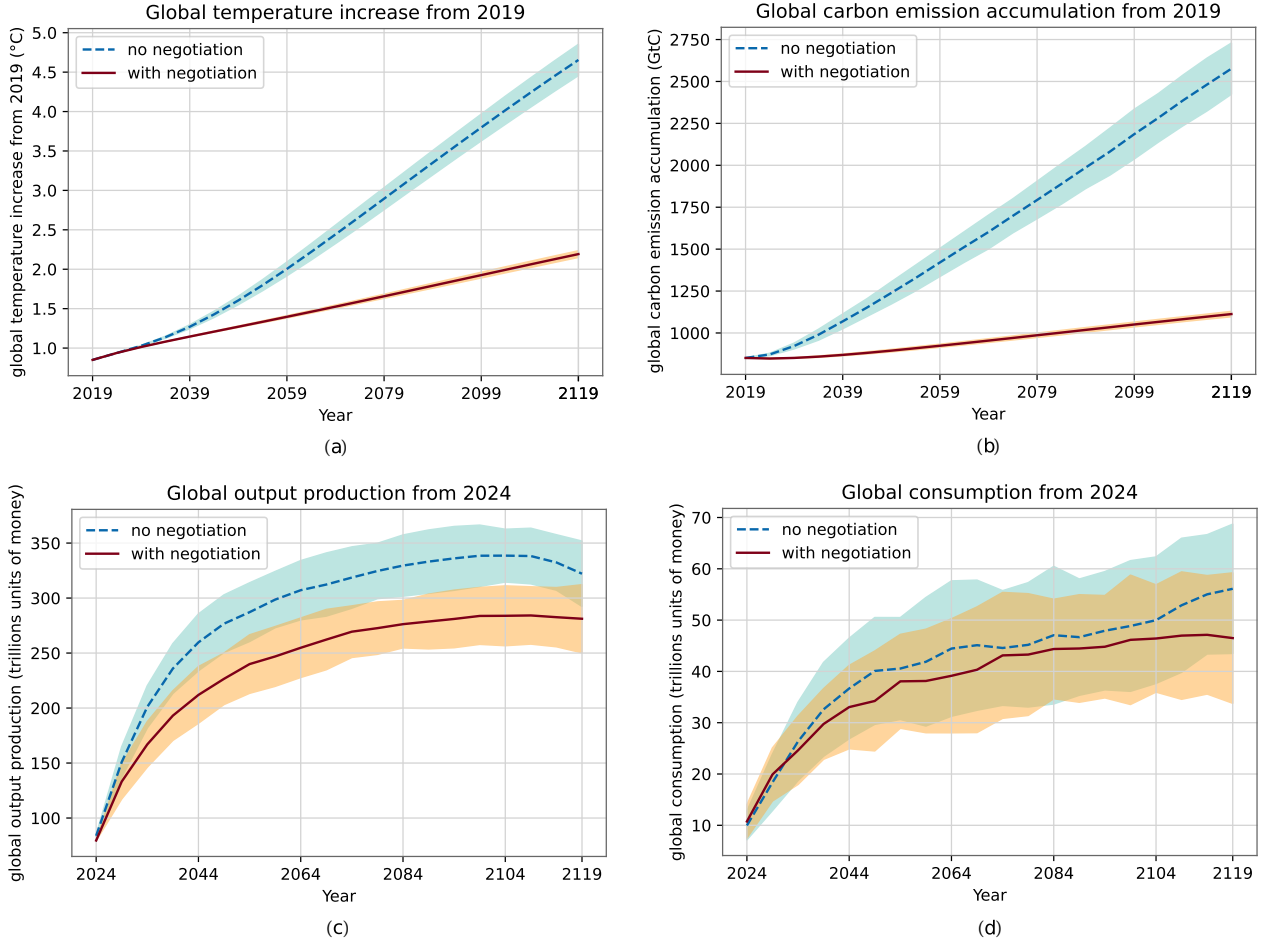
Figure 2: **Comparing outcomes with RL agents trained with and without a simple bilateral negotiation protocol.** Here we show the simulated outcomes under the simple bilateral protocol from Section 4.3, and agents that were independently trained using PPO. (a) The predicted increase in global temperature, with predictions starting in 2024 and initialization in 2019. (b) The predicted global carbon emission accumulation, with predictions starting in 2024 and initialization in 2019. (c) The predicted global output production (aggregated over regions) starting from 2024. (d) The predicted global consumption (aggregated over regions) starting from 2024. The solid blue and red lines represent the mean across 100 repetitions with random seeds without negotiation and with negotiation respectively; the shading represents 1.96 times the standard error. In each simulation episode, agents action are decided on and executed every 5 years.

4.4 Multilateral Negotiations

Multilateral negotiations imply communication amongst several regions simultaneously. Although more technically challenging to implement, such negotiation protocols offer more possibilities and added realism. A key goal of this competition is to explore and implement effective negotiation protocols. We now discuss a recent proposal as an example.

Climate clubs. Climate clubs (Nordhaus 2015) are one example of a multi-region agreement. Each region that is a part of a club agrees to 1) enact a minimum mitigation rate and 2) enforce a minimum import tariff on regions that are not part of the club. A high minimum import tariff acts as an incentive to join the climate club by applying a mitigation rate at least as high as the club's minimum.

The negotiation protocol would specify how clubs are

formed, e.g., by randomly selecting one region to make a proposal (μ_{min}, τ_{min}) to which all other regions can agree (thereby joining the climate club) or refuse. This could be augmented in many ways, e.g., by repeating it with the standalone regions until all regions form a climate club, by implementing an iterative proposal-evaluation framework, or by specifying a more sophisticated mechanism for climate club proposals than random selection.

Asymmetric negotiation protocols. Certain negotiation protocols, e.g., climate clubs, rely on the uniform treatment of trading partners to simplify their theoretical analysis (Nordhaus 2015). However, such limits do not apply to agent-based simulations (Farmer et al. 2015), and therefore the negotiation actions of regions in the simulator are not required to be applied uniformly across trading partners.

In fact, effective international cooperation is based on *climate justice* (Klinsky et al. 2017), which often requires that countries be treated differently based on their characteristics, e.g., historical cumulative emissions or exposure to potential climate damages (Assembly et al. 1992). For example, article 4.4 of the Paris Agreement states that developed regions, which have historically contributed emissions in a disproportionate manner relative to developing countries, should continue taking the lead when it comes to mitigation efforts (Schleussner et al. 2016). However, there is no agreed-on way to label regions as "developed" and "developing" in simulations such as RICE-N, nor is there unanimous agreement on the validity of that particular denomination. Rather, we invite the community to consider how the characteristics of the fictitious regions in RICE-N relate to the climate debate in the real world.

Moreover, the "fairness" of climate change legislation can be established through many different methodologies which treat regions differently (Dooley et al. 2021). We encourage the community to explore important opportunities for the innovation and study of negotiation protocols that implement forms of climate justice.

Although building any simulation requires making design choices, we emphasize that we do not seek to make normative statements about "fairness" and "justice" in this work.

4.5 Intelligent negotiation protocols and agreements

It is also possible to consider learned negotiation protocols rather than specifying a set of handcrafted rules. For example, the AI Economist's two-level RL system has a coadaptive social planner that learns tax policies that outperform the Saez taxation framework (Zheng et al. 2022). Negotiation protocols could similarly be designed to adapt to the behavior of the agents in its economy. A fixed negotiation protocol may have difficulty handling agents that exhibit strategic and adaptive behavior and might therefore fall short of climate-economic outcomes that could be reached by an AI-driven mediator.

In particular, identifying a useful sequence of negotiation protocols and agreements can be seen as an iterative discovery process, where we would like to sample and improve solutions according to the climate-economic reward they provide. Additionally, by sufficiently exploring the space of negotiations and agreements, one might identify many different solutions along the Pareto frontier of climate-economic outcomes, which enables studying the trade-offs between different (types of) solutions.

Recent work has explored the potential of an intelligent mediator in MARL (Liu et al. 2022). However, to the best of our knowledge, there is no work exploring machine-learning based solutions that combine communication, negotiation, and agreements in the presence of strategic behavior in the MARL setting.

5 Modeling Agent Behavior using ML

The negotiation protocols and the climate-economic dynamics of RICE-N define a game-theoretic setup between the different regions. We model the behavior of an agent i using its policy $\pi_i(a_t|o_t)$ that maps the agent's observations to a probability distribution over its actions.

Why use ML? Existing IAMs often use a predetermined policy for each agent that is fixed exogenously. Such an approach would require handcrafted agent policies for each different negotiation protocol and the reliability of the outcomes from the simulation will depend on the modeled policies.

In contrast, we assume each region to be a strategic agent that is interacting with the environment and other strategic agents in it. Here, we do not manually set the behavioral policy π_i, rather we use machine learning techniques to find the policy that optimizes the agent's objective, and hence derive the agent policies endogenously.

Specifically, the agents are assumed to be rational such that each agent i optimizes its policy $\pi_i(a_t|o_t)$ to maximize its long-term aggregate γ-discounted utility:

$$\max_{\pi_i} \mathbb{E}_{\pi_1,\ldots,\pi_n} \left[\sum_{t=0}^{H} \gamma^t r_{i,t} \right]. \qquad (1)$$

Here, $r_{i,t}$ is the utility of the region i at step t determined by its aggregate consumption $C_{i,t}$ as follows:

$$r_{i,t} = U_{i,t} = \frac{1}{1-\alpha} L_{i,t} \left(\frac{C_{i,t}}{L_{i,t}} \right)^{1-\alpha}, \qquad (2)$$

where $L_{i,t}$ is the population of the region i at step t. The discount factor γ models the long term value of rewards for the agents and as such they could differ across different agents. The discount factor for each region is often updated with the changing administration and in the absence of any consensus we fix γ to be homogeneous across agents as assumed in (Nordhaus 2015). The aggregate consumption $C_{i,t}$ in the above equation is obtained by combining the domestic and foreign goods consumption using the Armington model (Armington 1969):

$$C_{i,t} = \left(\psi^{\text{dom}}(C_{i,i,t})^\lambda + \sum_{j \neq i} \psi^{\text{for}}(C_{i,j,t})^\lambda \right)^{\frac{1}{\lambda}}, \qquad (3)$$

$$C_{i,j,t} = x_{i,j,t}(1 - \tau_{i,j,t}) \quad \forall j \neq i, \qquad (4)$$

where $C_{i,j,t}$ are the foreign goods consumed after imposing tariffs $\tau_{i,j,t}$ on the imported goods $x_{i,j,t}$ by region i from region j at step t. Furthermore, $\psi^{\text{dom}}, \psi^{\text{for}}$ are shared parameters and λ is the Armington elasticity parameter.

Multi-agent reinforcement learning. Simulating a rational agent requires computing the optimal policy for each agent. [2] Finding the optimal rational policy for each agent in response to complex environment dynamics and other agent policies naturally leads to MARL; Busoniu, Babuska, and De Schutter (2008) give a comprehensive overview of the topic. In short, MARL extends single-agent RL to find an optimal policy for each agent interacting in a dynamic environment to solve Equation 1.

---

[2]Although we assume the agents to be rational, other behavioral models could also be implemented in RICE-N.

The RL framework models how an agent's actions affect the state of the environment and its rewards (utilities). Thus, an RL agent has to learn to anticipate the long-term effects of its actions. This is especially true and challenging in multi-agent environments, e.g., in RICE-N, where agent actions affect key climate-economic metrics, e.g., global temperatures, capital investments, etc. In addition, MARL solution algorithms have to deal with additional challenges since each agent has to respond to the policies of other agents. This makes the task of finding the optimal policy a moving target (until a form of equilibrium is reached in the agent policies). Developing effective MARL is an ongoing research topic, and we encourage the community to explore MARL as a solution framework to analyze the short and long-term effects of (sub-)optimal behavior on the climate and economy.

Implementing RL agents. Our public code (Zhang et al. 2022) includes both CPU and GPU implementations of the full RL pipeline using two popular RL algorithms: A2C and PPO (Schulman et al. 2017). RICE-N can also be used with other RL implementations.

Our base implementation models each RL agent using a neural network policy that shares weights across agents, but uses region-specific inputs. The architecture of the network can be adjusted, e.g., the number of layers and the dimension of each layer. Agent policies use separate heads for each action. To distinguish between agents, the policy model's input contains agent-specific features, e.g., their population, capital, technology factor, damage function, and a one-hot representation of the region's index, as well as the public state of the world (e.g., climate conditions).

In addition, each agent gets information about negotiations, e.g., the latest proposals made to and by this region, or the minimum mitigation rate agreed upon by this region. How negotiations evolve depends on the specifics of the protocol and the different actions executed by the agent, e.g., proposals for other regions, decisions on proposals made by other regions, and setting mitigation and savings rates that may or may not be in line with what was agreed on. Depending on the negotiation protocol, not all observations and actions are relevant to each agent. Hence, it is recommended to use efficient implementations that only feed the information to agents that are relevant at a given time.

6 Evaluating Negotiations and Agreements

Evaluating solutions. A complete solution consists of a variation of RICE-N with a negotiation protocol, agreements, and agents (trained using RL or whose behavior is learned or defined through other means). We can then measure the performance of a solution via global economic and climate metrics. In particular, because there may be intrinsic trade-offs between these objectives, we are interested in the set of best trade-offs, i.e., the *Pareto frontier* that are achievable under a solution framework. Hence, we focus on estimating the area-under-curve of the Pareto frontier that is achievable for a given solution.

We define performance in this way, because the framework itself should not encode a preference for a certain set of climate and economic outcomes. Which specific set of outcomes is desired represents a social choice; in this work, we do not intend to make any normative statements about which specific outcome is preferred.

Objectives. The climate index $\mathcal{C}$ measures the temperature change over the course of 100 years, *compared to two extremal policies that do not mitigate at all or use a 100% mitigation rate*. The economic index $\mathcal{E}$ measures the increase in global productivity[3] (i.e., total GDP), similarly compared to two extremal policies. Each index is computed as a relative gain compared to extremal agent policies to get normalized, dimensionless metrics along both climate and economic dimensions. This allows for fair comparison and computation of an area-under-curve metric in the $(\mathcal{C}, \mathcal{E})$-space. A basic intuition is that these two extremal policies represent two extremal points (very high $\mathcal{C}$, very low $\mathcal{E}$, and vice versa). However, these extremal policies may not be Pareto-optimal, may not be optimal under every negotiation protocol, and may not be individual best-response strategies.

Empirical Pareto frontier. Each solution framework may generate a set of different outcomes $(\mathcal{C}, \mathcal{E})$, e.g., if the negotiation protocol has a free parameter that may change how agents behave (optimally). In addition, each specific outcome needs to account for randomness by averaging over Monte-Carlo roll-outs performed under different random seeds.[4]

Each solution thus yields a set of points in the $(\mathcal{C}, \mathcal{E})$-space, from which we can infer an empirical Pareto frontier: the set of points for which neither index can be increased without decreasing the other index. Specifically, a point a *Pareto-dominates* another point b if at least one index is strictly higher and the other index is at least as high, i.e., 1) $\mathcal{C}_a > \mathcal{C}_b$ and $\mathcal{E}_a \geq \mathcal{E}_b$, or 2) $\mathcal{C}_a \geq \mathcal{C}_b$ and $\mathcal{E}_a > \mathcal{E}_b$.

The empirical area-under-curve is defined using a *hypervolume indicator*: the area of a set of non-dominated policies in the objective space with respect to a predetermined reference point (Van Moffaert and Nowé 2014). In our case, the reference point is set by the extremal policies.

Besides giving a numerical value to compare different solutions in a multi-objective setting, this score definition has several attractive properties in a competition setting: 1) The hypervolume indicator encourages competitors to lie as close as possible to the true Pareto frontier, 2) it is strictly monotonic with respect to Pareto dominance such that a set of points that Pareto-dominates dominates another set must have a bigger hypervolume, and 3) it incentivizes diversification (along the Pareto frontier), i.e., competitors are incentivized to submit solutions that make different climate-economic trade-offs to increase the hypervolume of their set.

See Appendix G and Figure 5 for more details.

---

[3]Exploring additional, more granular measures of economic well-being within the simulator is an interesting area of exploration.

[4]For example, the base implementation of RICE-N has deterministic climate-economic dynamics, but the agents' behavior may be stochastic. In our competition, we use 10 simulation roll-outs under 3 random seeds, and determine the mean and standard deviation of the outcomes for evaluation purposes.

7 Engineering aspects and customization

The RICE-N simulation builds on the RICE model (Nordhaus 2015), a multi-region climate-economic simulation, by adding negotiation protocols, international trade, and support for strategic agents. RICE-N has n agents, each representing a region in the world or a group of (fictitious) countries that are assumed to make decisions as a single entity. To maintain realism, we calibrate the structural parameters and establish a reasonable range of agent-specific parameters using data from the World Bank API (WorldBank 2022b). Algorithm 1 in Appendix B shows the full flow of the activity component.

Code modularity. RICE-N is designed to be easily extendable and modular. In particular, RICE-N provides a framework to implement negotiation protocols as well as learning algorithms (e.g., RL agents) to enable designing negotiation protocols and study their effects on global cooperation.

The simulation code is structured so that custom negotiation protocols can be implemented as part of the negotiation component. An implementation may add agent observations and actions corresponding to the proposed negotiation protocol, and rules to generate the corresponding agreements. Notice that the agents can have access to these additional observations even during the activity stage. These observations are useful in letting agents adopt policies in response to the negotiations and agreements. For example, they can learn to punish agents that do not comply with the agreements. Besides, by using action masking, the simulation can ensure that the agreements are followed (we can also have mix where some agreements are enforced using action masking and some others are non-enforceable). Thus, the simulation code provides complete freedom in implementing any negotiation protocol of choice provided the corresponding agreements are appropriately connected to the activity stage action through additional observations or action masks.

Consistency checks. On the other hand, for the purpose of the competition, we fix the actions in the activity stage and the corresponding climate and economic dynamics. Thus, certain parts of the simulation should not be modified by competitors, e.g., the core climate and economic parameters and equation, such as the ones which affect carbon emissions and productivity. Changing these components may change the problem and prevent fair comparisons across submitted solutions. As such, our evaluation protocol and implementation of RICE-N include consistency checks that test submitted solutions; each submission requiring submitting the full code of the (modified) simulation and agent models. The activity stage is designed to follow the well established climate-economic dynamics in the RICE model augmented with trades and tariffs to include an essential influence in international negotiations. This way the simulation follows a simple model but is rich enough to implement interesting negotiation protocols.

Further, our implementation is set up to model each region as an RL agent that optimizes its discounted long-term rewards. Although other agent policy models are possible, we require that the competitors clarify how their proposed agent policy models also aim to optimize for the same goal, namely, the personal discounted long term reward for each agent. This ensures that the agents behave greedily and any improvements in global metrics are strictly a result of the implemented negotiation protocols. See section 5 for details.

8 Working Group and Competition

To build momentum for this research direction, we invite the community to join our working group collaboration and competition. The goal of this initiative is to help discover innovative, AI-based approaches to international climate negotiations that (i) go beyond traditional climate-economic trade-offs, (ii) are sufficiently sophisticated to have real world relevance, and (iii) are useful and relevant to policymakers. Specifically, to motivate rapid development and comparison of different solutions, we are organizing a competition in which participants use RICE-N as a test bed to develop and evaluate negotiation protocols.

The working group consists of: 1) organizers of the competition, 2) competition participants, 3) an expert jury, and 4) advisors. A concrete outcome and goal of this working group is to report the findings of the competition (vetted by an expert jury and internal review) in a peer-reviewed publication, co-authored by the competition participants and other working group members. The findings would then be used to make policy and protocol recommendations to policymakers. This represents a first step toward the development of effective international climate agreements in the real world.

The competition has three tracks. The first track quantitatively evaluates the performance of a negotiation protocol on climate and economic performance. The second track, in addition to the quantitative evaluation, involves a scientific assessment of feasibility by a diverse, interdisciplinary jury with expertise on machine learning, economics, climate science, computational social science, policy writing, international relations, law, and ethics. Negotiation protocols championed by such a diverse jury are likely to be more viable for real-world application and suitable for policymaking. The third track invites critical analysis and suggestions for improvements.

9 Discussion

We presented the our working group collaboration, competition, the RICE-N simulation, and a conceptual framework to include strategic negotiation into climate-economic modeling. To our knowledge, this is the first simulation and learning-based framework to study negotiation in climate-economic systems of this complexity. Given the modular structure of our implementation, RICE-N can be easily extended to include more climate, economic, or strategic features. As such, we believe this is a compelling step to develop AI for climate change and invite the community to build on this work.

The RICE climate and economic dynamics are relatively simple, but are a minimal implementation of an IAM that supports studying the strategic aspects of decentralized decision-making, negotiation, and their impact on the climate. Both the climate and economic components could be made more sophisticated, as implemented by a large number of existing IAMs. For example, one might add rich spatial climate dynamics, or model more parts of the real economy, e.g., international companies or markets with price discovery. The game-theoretic features could also be enhanced, e.g., by al-

lowing for forms of communication, more policy levers, more possible interactions (e.g., technology transfer), and others. The behavior models for the agents could also be modified, for example, by considering heterogeneous discounting coefficients, alternative utility functions with different shapes, etc. Besides, instead of considering a time span of 100 years for each agent's optimization, we can consider an election cycle, thus making the agent incentives more relevant from a practical point of view. However, simulating more complex dynamics also incurs significant computational cost. As such, we hope that future work will extend the RICE-N model towards using AI for climate change mitigation and social good.

10 Ethical Considerations

While the intention of this paper and the corresponding challenge is to stimulate innovative solutions to climate change, there are some unintended consequences that we would like to acknowledge and address here. These include the carbon footprint of running the simulation itself, the economic disparities that can exist with climate negotiations, and the potential extensibility of this simulation to the real world.

Carbon emissions from simulations.   First, it is important to acknowledge that running climate change simulations in RICE-N will inevitably release carbon emissions into our atmosphere. While the computational requirements of these simulations are much smaller than training large language models, they still exist. To mitigate this harm, we encourage participating teams to consider their energy use during experimentation, offsetting their carbon emissions if possible. For reference emission estimates of our code and suggestions for measurement tools, please see Appendix F.

Economic and climate disparities.   Second, as the World Bank states, "Climate change is deeply intertwined with global patterns of inequality" and yet "the most vulnerable are often also disproportionately impacted by measures to address climate change" (WorldBank 2022a). While it is important to determine ways that to mitigate climate change, it is equally important to ensure that vulnerable populations are not negatively impacted by climate change measures.

Technical aspects and limitations.   Last but not least, it is important to note that the predicted climate and economic predictions made in RICE-N may differ in a real-world setting due to externalities beyond the boundaries of the simulation. A fictional world is utilized in this competition to further illustrate the potential gap between simulation and reality, but the uncertainty of the results should be fully understood, especially before implementing any policies recommended by RICE-N.

References

Armington, P. S. 1969. A theory of demand for products distinguished by place of production. *Staff Papers*, 16(1): 159–178.

Assembly, U. N. G.; et al. 1992. Report of the United Nations Conference on Environment and Development. In *A/CONF*, volume 151, 26. Rio de Janeiro.

Bakaki, Z. 2022. The impact of climate summits. *Nature Climate Change*, 12(7): 611–612.

Boers, N.; and Rypdal, M. 2021. Critical slowing down suggests that the western Greenland Ice Sheet is close to a tipping point. *Proceedings of the National Academy of Sciences*, 118(21).

Bonabeau, E. 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, 99(suppl_3): 7280–7287.

Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; Layton, R.; VanderPlas, J.; Joly, A.; Holt, B.; and Varoquaux, G. 2013. API design for machine learning software: experiences from the scikit-learn project.

Busoniu, L.; Babuska, R.; and De Schutter, B. 2008. A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2): 156–172.

Cao, K.; Lazaridou, A.; Lanctot, M.; Leibo, J. Z.; Tuyls, K.; and Clark, S. 2018. Emergent communication through negotiation. *arXiv preprint arXiv:1804.03980*.

Carney, M. 2015. Breaking the tragedy of the horizon–climate change and financial stability. *Speech given at Lloyd's of London*, 29: 220–230.

Chan, S.; Hale, T.; Deneault, A.; Shrivastava, M.; Mbeva, K.; Chengo, V.; and Atela, J. 2022. Assessing the effectiveness of orchestrated climate action from five years of summits. *Nature Climate Change*, 1–6.

Chelarescu, P. 2021. Deception in Social Learning: A Multi-Agent Reinforcement Learning Perspective. *arXiv preprint arXiv:2106.05402*.

Comin, D. 2010. *total factor productivity*, 260–263. London: Palgrave Macmillan UK. ISBN 978-0-230-28082-3.

De Filippi, P.; and Hassan, S. 2018. Blockchain technology as a regulatory technology: From code is law to law is code. *arXiv preprint arXiv:1801.02507*.

DeCanio, S. J.; and Fremstad, A. 2013. Game theory and climate diplomacy. *Ecological Economics*, 85: 177–187.

DeConto, R. M.; Pollard, D.; Alley, R. B.; Velicogna, I.; Gasson, E.; Gomez, N.; Sadai, S.; Condron, A.; Gilford, D. M.; Ashe, E. L.; et al. 2021. The Paris Climate Agreement and future sea-level rise from Antarctica. *Nature*, 593(7857): 83–89.

Dooley, K.; Holz, C.; Kartha, S.; Klinsky, S.; Roberts, J. T.; Shue, H.; Winkler, H.; Athanasiou, T.; Caney, S.; Cripps, E.; et al. 2021. Ethical choices behind quantifications of fair contributions under the Paris Agreement. *Nature Climate Change*, 11(4): 300–305.

Farmer, J. D.; Hepburn, C.; Mealy, P.; and Teytelboym, A. 2015. A third wave in the economics of climate change. *Environmental and Resource Economics*, 62(2): 329–357.

Foerster, J. N. 2018. *Deep multi-agent reinforcement learning*. Ph.D. thesis, University of Oxford.

Foerster, J. N.; Chen, R. Y.; Al-Shedivat, M.; Whiteson, S.; Abbeel, P.; and Mordatch, I. 2017. Learning with Opponent-Learning Awareness.

Greeven, S.; Kraan, O.; Chappin, É. J.; et al. 2016. The emergence of climate change and mitigation action by society: an agent-based scenario discovery study. *Journal of Artificial Societies and Social Simulation*, 19(3) 9).

Hanumaiah, V.; and Genc, S. 2021. Distributed Multi-Agent Deep Reinforcement Learning Framework for Whole-building HVAC Control.

Hardin, G. 1968. The tragedy of the commons: the population problem has no technical solution; it requires a fundamental extension in morality. *science*, 162(3859): 1243–1248.

Jaques, N.; Lazaridou, A.; Hughes, E.; Gulcehre, C.; Ortega, P.; Strouse, D.; Leibo, J. Z.; and De Freitas, N. 2019. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In *International conference on machine learning*, 3040–3049. PMLR.

Kellett, C. M.; Weller, S. R.; Faulwasser, T.; Grüne, L.; and Semmler, W. 2019. Feedback, dynamics, and optimal control in climate economics. *Annual Reviews in Control*, 47: 7–20.

Klinsky, S.; Roberts, T.; Huq, S.; Okereke, C.; Newell, P.; Dauvergne, P.; O'Brien, K.; Schroeder, H.; Tschakert, P.; Clapp, J.; et al. 2017. Why equity is fundamental in climate change policy research. *Global Environmental Change*, 44: 170–173.

Lacoste, A.; Luccioni, A.; Schmidt, V.; and Dandres, T. 2019. Quantifying the Carbon Emissions of Machine Learning. *arXiv preprint arXiv:1910.09700*.

Le Gléau, T.; Marjou, X.; Lemlouma, T.; and Radier, B. 2022. Towards circular and asymmetric cooperation in a multi-player Graph-based Iterated Prisoner's Dilemma. In *14th International Conference on Agents and Artificial Intelligence*.

Lerer, A.; and Peysakhovich, A. 2017. Maintaining cooperation in complex social dilemmas using deep reinforcement learning.

Lessmann, K.; Marschinski, R.; and Edenhofer, O. 2009. The effects of tariffs on coalition formation in a dynamic global warming game. *Economic Modelling*, 26(3): 641–649.

Li, S. 2017. Obviously strategy-proof mechanisms. *American Economic Review*, 107(11): 3257–87.

Liu, D.; Shah, V.; Boussif, O.; Meo, C.; Goyal, A.; Shu, T.; Mozer, M.; Heess, N.; and Bengio, Y. 2022. Coordinating Policies Among Multiple Agents via an Intelligent Communication Channel. *arXiv preprint arXiv:2205.10607*.

Lovejoy, T. E.; and Nobre, C. 2018. Amazon tipping point.

Madani, K. 2013. Modeling international climate change negotiations more responsibly: Can highly simplified game theory models provide reliable policy insights? *Ecological Economics*, 90: 68–76.

Mai, V.; Zhang, T.; and Lesage-Landry, A. 2021. Multi-agent reinforcement learning for renewable integration in the electric power grid. In *NeurIPS 2021 Workshop on Tackling Climate Change with Machine Learning*.

Martin-Bariteau, F.; and Scassa, T. 2020. Artificial intelligence and the law in Canada. *Artificial Intelligence and the Law in Canada (Toronto: LexisNexis Canada, 2021)*.

Mitchell, R. B. 2003. International environmental agreements: a survey of their features, formation, and effects. *Annual review of environment and resources*, 28(1): 429–461.

Moore, F. C.; Lacasse, K.; Mach, K. J.; Shin, Y. A.; Gross, L. J.; and Beckage, B. 2022. Determinants of emissions pathways in the coupled climate–social system. *Nature*, 603(7899): 103–111.

Nordhaus, W. 2015. Climate clubs: Overcoming free-riding in international climate policy. *American Economic Review*, 105(4): 1339–70.

Nordhaus, W. 2018. Evolution of modeling of the economics of global warming: changes in the DICE model, 1992–2017. *Climatic Change*, 148(4): 623–640.

Nordhaus, W. D. 2007. A review of the Stern review on the economics of climate change. *Journal of economic literature*, 45(3): 686–702.

Paquette, P.; Lu, Y.; BOCCO, S. S.; Smith, M.; O.-G., S.; Kummerfeld, J. K.; Pineau, J.; Singh, S.; and Courville, A. C. 2019. No-Press Diplomacy: Modeling Multi-Agent Gameplay. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Pavan, A.; Segal, I.; and Toikka, J. 2014. Dynamic mechanism design: A myersonian approach. *Econometrica*, 82(2): 601–653.

Pörtner, H. O.; Roberts, D. C.; Adams, H.; Adler, C.; Aldunce, P.; Ali, E.; Begum, R. A.; Betts, R.; Kerr, R. B.; Biesbroek, R.; et al. 2022. Climate change 2022: impacts, adaptation and vulnerability.

Rochedo, P. R.; Soares-Filho, B.; Schaeffer, R.; Viola, E.; Szklo, A.; Lucena, A. F.; Koberle, A.; Davis, J. L.; Rajão, R.; and Rathmann, R. 2018. The threat of political bargaining to climate mitigation in Brazil. *Nature Climate Change*, 8(8): 695–698.

Schelling, T. C. 1980. *The Strategy of Conflict: with a new Preface by the Author*. Harvard university press.

Schleussner, C.-F.; Rogelj, J.; Schaeffer, M.; Lissner, T.; Licker, R.; Fischer, E. M.; Knutti, R.; Levermann, A.; Frieler, K.; and Hare, W. 2016. Science and policy characteristics of the Paris Agreement temperature goal. *Nature Climate Change*, 6(9): 827–835.

Schmidt, V.; Goyal, K.; Joshi, A.; Feld, B.; Conell, L.; Laskaris, N.; Blank, D.; Wilson, J.; Friedler, S.; and Luccioni, S. 2021. CodeCarbon: Estimate and Track Carbon Emissions from Machine Learning Computing.

Schrittwieser, J.; Antonoglou, I.; Hubert, T.; Simonyan, K.; Sifre, L.; Schmitt, S.; Guez, A.; Lockhart, E.; Hassabis, D.; Graepel, T.; Lillicrap, T.; and Silver, D. 2020. Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature*, 588(7839): 604–609.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Shoham, Y.; and Leyton-Brown, K. 2008. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; and Hassabis, D. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587): 484–489. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7587 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Computational science;Computer science;Reward Subject_term_id: computational-science;computer-science;reward.

Smead, R.; Sandler, R. L.; Forber, P.; and Basl, J. 2014. A bargaining game analysis of international climate negotiations. *Nature Climate Change*, 4(6): 442–445.

Telser, L. G. 1980. A theory of self-enforcing agreements. *Journal of business*, 27–44.

Van Moffaert, K.; and Nowé, A. 2014. Multi-objective reinforcement learning using sets of pareto dominating policies. *The Journal of Machine Learning Research*, 15(1): 3483–3512.

Vinyals, O.; Babuschkin, I.; Czarnecki, W. M.; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, D. H.; Powell, R.; Ewalds, T.; Georgiev, P.; Oh, J.; Horgan, D.; Kroiss, M.; Danihelka, I.; Huang, A.; Sifre, L.; Cai, T.; Agapiou, J. P.; Jaderberg, M.; Vezhnevets, A. S.; Leblond, R.; Pohlen, T.; Dalibard, V.; Budden, D.; Sulsky, Y.; Molloy, J.; Paine, T. L.; Gulcehre, C.; Wang, Z.; Pfaff, T.; Wu, Y.; Ring, R.; Yogatama, D.; Wünsch, D.; McKinney, K.; Smith, O.; Schaul, T.; Lillicrap, T.; Kavukcuoglu, K.; Hassabis, D.; Apps, C.; and Silver, D. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782): 350–354. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7782 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Computer science;Statistics Subject_term_id: computer-science;statistics.

Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17: 261–272.

WorldBank. 2022a. Social Dimensions of Climate Change.

WorldBank. 2022b. The World Bank Documents and Report API.

Yu, C.; Velu, A.; Vinitsky, E.; Wang, Y.; Bayen, A.; and Wu, Y. 2021. The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games.

Yu, L.; Sun, Y.; Xu, Z.; Shen, C.; Yue, D.; Jiang, T.; and Guan, X. 2020. Multi-Agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings.

Zenker, A. 2019. *International Climate Agreements Under Review: The Potential of Negotiation Linkage Between Climate Change and Preferential Free Trade*. Springer Nature.

Zhang, T.; Srinivasa, S.; Williams, A.; Phade, S.; Zhang, Y.; Gupta, P.; Bengio, Y.; and Zheng, S. 2022. RICE-N.

Zheng, S.; Trott, A.; Srinivasa, S.; Parkes, D. C.; and Socher, R. 2022. The AI Economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Science Advances*, 8(18): eabk2607.

Zhu, C.; Dastani, M.; and Wang, S. 2022. A Survey of Multi-Agent Reinforcement Learning with Communication. *arXiv preprint arXiv:2203.08975*.

Zitzler, E.; Thiele, L.; Laumanns, M.; Fonseca, C. M.; and Da Fonseca, V. G. 2003. Performance assessment of multi-objective optimizers: An analysis and review. *IEEE Transactions on evolutionary computation*, 7(2): 117–132.

Acknowledgements

Contributions

- SZ conceived the project;
- SZ, PG, and YB directed the project;
- All authors developed the conceptual and theoretical framework;
- TZ, AW, SP, SS developed the economic simulator and implemented the reinforcement learning platform;
- SS developed the evaluation pipeline.
- SS developed the GPU version of the simulation.
- TZ scraped data;
- TZ, SP performed calibration;
- TZ, AW, SP, SS performed experiments;
- TZ, SP, AW developed the tutorial notebook;
- YZ advised the work;
- All authors drafted, discussed, and commented on the manuscript.
- All authors reviewed the code.

A The Activity Component: Climate, Economics, Trade, and Tariffs

A.1 Climate and Economic Dynamics

We now describe the RICE-N dynamics developed from DICE and RICE models by (Nordhaus 2018; Kellett et al. 2019) that govern the evolution of the world state from time t to $t + 1$ for the different regions. Note that variables without an agent index are global quantities.

Carbon mass. The total carbon mass in the climate system is given by:

$$M_{t+1} = \Phi_M M_t + B_M \sum_i E_{i,t}, \tag{5}$$

$$E_{i,t} = E_t^{\text{Land}} + \sigma_{i,t}(1 - \mu_{i,t})Y_{i,t}, \tag{6}$$

$$M_t \doteq \begin{bmatrix} M_t^{\text{AT}} & M_t^{\text{UP}} & M_t^{\text{LO}} \end{bmatrix}^\top \in \mathbb{R}^3, \tag{7}$$

$$\Phi_M \doteq \begin{bmatrix} \zeta_{11} & \zeta_{12} & 0 \\ \zeta_{21} & \zeta_{22} & \zeta_{23} \\ 0 & \zeta_{32} & \zeta_{33} \end{bmatrix}, \tag{8}$$

$$B_M \doteq \begin{bmatrix} \xi_2 \\ 0 \\ 0 \end{bmatrix}. \tag{9}$$

This describes a three-reservoir model of the global carbon cycle, in which M_{AT} describes the average mass of carbo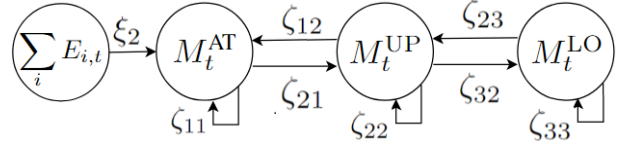n in the atmosphere, M_{UP} is the average mass of carbon in the upper ocean, and M_{LO} the average mass of carbon in the deep or lower ocean, see Figure 3. Φ_M is the Markov transition matrix describing how carbon transfer between different reservoirs. B_M describes how the weight of carbon emission affects the carbon accumulation in the reservoirs.

Figure 3: **The three-reservoir carbon mass model.**

Global temperature. Ultimately, increasing carbon mass leads to rising temperatures:

$$T_{t+1} = \Phi_T T_t + B_T F_t, \tag{10}$$

$$T_t \doteq \begin{bmatrix} T_t^{\text{AT}} & T_t^{\text{LO}} \end{bmatrix}^\top \in \mathbb{R}^2, \tag{11}$$

$$F_t = F_{2\times} \log_2 \left(\frac{M_t^{\text{AT}}}{M^{\text{AT},1750}} \right), \tag{12}$$

$$\Phi_T \doteq \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}, \tag{13}$$

$$B_T \doteq \begin{bmatrix} \xi_1 \\ 0 \end{bmatrix}. \tag{14}$$

Similar to the carbon mass dynamic, there are two layers in the energy balance model, see Figure 4. T_{AT} is the combined average temperature in atmosphere, land surface, and upper ocean (simply referred to as the "atmospheric layer" hereafter). T_{LO} is the temperature in the lower ocean. Φ_T is the Markov transition matrix describing how heat transfers between different layers. B_T describes how carbon mass contributes to the temperature increases.

Output production. The production in a region is given by the *total factor productivity (TFP)* (Comin 2010) formula:

$$Y_{i,t} = A_{i,t} K_{i,t}^\gamma L_{i,t}^{1-\gamma}. \tag{40}$$

Production depends on three factors: total factor productivity ("technology") A_t, capital K_t, and labor L_t. This production function is common in the economic literature and used in the DICE/RICE models. The capital elasticity $\gamma \in [0, 1]$ explains the different levels of contribution of capital and labor.

Population. The number of people in a region, denoted L_t grows as:

$$L_{i,t+1} = L_{i,t} \left(\frac{1 + L_{a;i}}{1 + L_{i,t}} \right)^{l_{g;i}}. \tag{38}$$

There are two parameters $L_{A;i}$ and $l_{g;i}$. $L_{a;i}$ represents the convergence population of region i and $l_{g;i}$ shows how fast the population $L_{i,t}$ converge to $L_{a;i}$. Please refer to the Appendix D for a more detailed analysis and the calibration procedure.

Level of technology. The technology factor A_t describes how efficient production is, i.e., how many units of output a region achieves given fixed capital and labor:

$$A_{i,t+1} = (e^\eta + g_{A;i} e^{-\delta_{A;i}\Delta(t-1)}) A_{i,t}. \qquad (39)$$

Here, η represents the long-term growth of economics which is usually larger than 0, g_A represents the short-term part of economics growth, and δ_A represents the speed of decay of short-term growth factor. Δ is the time difference between steps. We use $\eta = 0.33\%$ as in (Nordhaus 2018).

Capital. The amount of capital evolves as:

$$\Phi_K \doteq (1-\delta_K)^\Delta, \qquad (15)$$

$$K_{i,t+1} = \Phi_{K,i} K_{i,t} + \Delta \left(1 - a_1 T_t^{\text{AT}} - a_2 \left(T_t^{\text{AT}}\right)^2\right) \qquad (16)$$

$$\times \left(1 - \theta_{1;i,t} \mu_{i,t}^{\theta_2}\right) Y_{i,t} s_{i,t}. \qquad (17)$$

The evolution of the capital comes from two parts. The first part is capital inherited from the previous period with depreciation. In the second part, s_t is a control variable which represents the investment/savings rate (as a fraction of production). That is, as a base amount, the economy invests/saves a total of $Y_{i,t} s_{i,t}$ which yields new capital. This base amount is further modified by 2 multipliers: the damage function and mitigation/abatement costs, which are discussed below.

Damage function. The climate damage function represents the economic damage due to climate change, e.g., increases in the atmosphere temperature T_t^{AT}. That is, in Equation 36, the fraction of new capital is modified by the damage function

$$1 - a_1 T_t^{\text{AT}} - a_2 \left(T_t^{\text{AT}}\right)^2, \qquad (18)$$

following (Nordhaus 2015). That is, higher temperatures lead to less new capital. Similarly, $1 - \theta_{1;i,t} \mu_t^{\theta_2}$ is the fraction of new capital after taking into account carbon emission mitigation. Mitigating carbon emissions more (higher μ_t) means (dirty) production needs to be lowered, hence yields less new capital.

Mitigation (abatement) cost. Following (Kellett et al. 2019), for a mitigation rate $\mu_{i,t}$, the mitigation cost is

$$\theta_{1;i,t} \mu_{i,t}^{\theta_2} Y_{i,t} s_{i,t}, \qquad (19)$$

where $\theta_{1;i,t}$ is given by Equation 37. This represents the loss in capital growth due to a fraction of production being used for mitigation.

Carbon intensity of economic activity. A critical part of the model is the interaction between the climate and economic parts. Specifically, the RICE model describes how production leads to carbon emissions:

$$E_t^{\text{Land}} = E_{L0} \cdot (1-\delta_{EL})^{t-1}, \qquad (20)$$

$$E_{i,t} = E_t^{\text{Land}} + \sigma_{i,t} A_{i,t}(1-\mu_{i,t}) Y_{i,t} \qquad (21)$$

$$\sigma_{i,t+1} = \sigma_{i,t} e^{-g_{\sigma;i}(1-\delta_{\sigma;i})^{\Delta(t-1)}\Delta}. \qquad (22)$$

Here E_t^{land} is the carbon emission due to (changes in) land use, E_{L0} is the carbon emission in the base year, and δ_{EL}
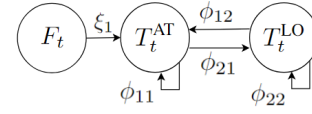

Figure 4: **The two-reservoir temperature model.**

is the speed of decrease of changes in land use. The rates $0 < \delta_{EL} < 1$, $0 < \delta_{L0} < 1$ are free parameters. Due to a lack of data, E_t^{land} is set to be the same for each region.

$E_{i,t}$ is the total carbon emission, E_t^{land} is emission from natural sources, while $E_{i,t} - E_t^{\text{land}}$ is emission caused by economic activity. $\sigma_{i,t} A_{i,t}$ is the effective carbon intensity of economic activity: a higher technology factor lead to higher emissions, but can be modulated by lower σ (which can be thought of as the degree of "clean" production). $\mu_{idxi,t} \in [0,1]$ is a control variable called the abatement (ratio), which represents the proportion of the economics contributing to reducing carbon emission. Furthermore, we have 2 parameters g_σ and δ_σ that are fitted to data. g_σ is the rate of decrease in carbon emissions.

A.2 Trade

We now describe the international trade dynamics and the resulting regional consumption and utilities. Regions trade by exporting their own consumption goods and importing other regions' consumption goods at a fixed unit price[5].

Agent actions. Each region i at time t must first specify a desired basket of consumption goods $\boldsymbol{b_{i,t}} = [b_{i,1,t}, ..., b_{i,k,t}]$ that they are willing to import from the other regions. These desired imports form a matrix of *bids* B_t such that the import bid by region i for goods from j at time t is $b_{i,j,t} \geq 0$, i.e., the amount of goods region i is willing to import from region j at time t is $b_{i,j,t}$.

Regions also set an upper bound $p_{i,t}^x \in [0,1]$ on the proportion of their own consumption goods that they are willing to export.

Tariffs. Regions can also choose to impose import tariffs on other regions. We denote an import tariff imposed by region i on a region j by $\tau_{i,j,t} \in [0,1]$. If region i imposes an import tariff $\tau_{i,j,t} \in [0,1]$ on region j, region i consumes

$$C_{i,j,t} = x_{i,j,t}(1-\tau_{i,j,t}), \qquad (23)$$

and $\tau_{i,j,t} x_{i,j,t}$ is added to a reserve fund specific to that region.

Consumption. Consumption of domestic goods $C_{i,i,t}$ is determined according to gross output, the savings rate and exports:

$$C_{i,i,t} = (1-s_{i,t}) Q_{i,t} - \sum_{j \neq i} x_{j,i,t}. \qquad (24)$$

The aggregated consumption $C_{i,t}$ at time t for region i is given by the Armington elasticity model (Lessmann,

---

[5]More generally, prices should be dynamic, but the current implementation does not support this.

Marschinski, and Edenhofer 2009)) as follows:

$$C_{i,t} = \left(\psi^{dom}(C_{i,i,t})^{\lambda} + \sum_{j \neq i} \psi^{for}(C_{i,j,t})^{\lambda} \right)^{\frac{1}{\lambda}}. \quad (25)$$

B RICE-N dynamics

At a high-level, Equations 34 and 35 capture climate dynamics (temperature and carbon mass), while Equations 36, 38, 39, and 40 capture economic dynamics. Finally, Equation 41 captures the carbon-intensity of production, providing a key link between the climate and economic sectors.

C Creating a 27-Region Simulation

We feature $n = 27$ fictitious regions in our public simulation. These are inspired by merging and splitting real-world countries, but are not exactly the same as real-world regions.

We used real data from the World Bank API (World-Bank 2022b), e.g., GDP, capital stock, population, and CO_2-quivalent (CO_2eq) emissions. Furthermore, the World Bank groups countries into regions, including Sub-Saharan Africa, South Asia, North America, the Middle East and North Africa, Latin America and the Caribbean, Europe and Central Asia, East Asia and Pacific. In each region, the different countries (or sub-regions) are classified into 4 income groups: high income, upper middle income, lower middle income, and low income.

Merging regions. We assume the GDP, capital stock, and population for the regions are additive. We also assume the gross CO_2eq emissions across the regions are additive. Thus, we have

$$K_m = \sum_i K_i, \quad (26)$$

$$L_m = \sum_i L_i, \quad (27)$$

$$Y_m = \sum_i Y_i, \quad \text{where} \quad Y_i := A_i K_i^{\gamma} L_i^{1-\gamma}, \quad (28)$$

$$A_m = \frac{Y_m}{K_m^{\gamma} L_m^{1-\gamma}}, \quad (29)$$

$$\sigma_m = \frac{\sum_i \sigma_i Y_i}{Y_m}. \quad (30)$$

Note that the production function is not scale-invariant:

$$Y_t = (A_t K_t)^{\gamma} (A_t L_t)^{1-\gamma} \quad (31)$$

$$c \cdot Y_t = (c \cdot A_t K_t)^{\gamma} (c \cdot A_t L_t)^{1-\gamma} \quad (32)$$

$$\neq (c \cdot A_t)(c \cdot K_t)^{\gamma}(c \cdot L_t)^{1-\gamma}, \quad \forall c > 0. \quad (33)$$

Hence, one cannot get the technology after merging multiple regions by simply adding the individual technology levels. Rather, the combined technology factor is imputed from the combined productions, labor, and capital.

Splitting large regions. To avoid huge economies that dominate the fictitious world, we split large economies into pieces

based on predetermined fractions c_i and random sampled A_i:

$$\sum_i c_i = 1, \quad (42)$$

$$L_i = c_i L_m, \quad (43)$$

$$Y_i = c_i Y_m, \quad (44)$$

$$K_i = \frac{Y_i}{A_i L_i^{1-\gamma}}, \quad (45)$$

$$\sigma_i = \sigma_m. \quad (46)$$

D Model Calibration

The structural parameters of the RICE-N simulation were calibrated to meet the following objectives:

1. Temperatures match the real data in different versions of RICE-N with 1 region, 27 regions, and 189 regions (all raw regions), under 0% and 100% mitigation. To make the computational cost more affordable, we use the 27-region version for the competition.

2. The optimistic-pessimistic temperature outcomes fit the projects of Shared Socioeconomic Pathways in the IPCC Sixth Assessment Report (Pörtner et al. 2022) (2 - 5 deg Celsius increase in the year of 2100). Each region optimizes the target without negotiation and direct cooperation in the pessimistic case. In the optimistic case, regions negotiate with each other using the baseline bilateral negotiation protocol. Please also notice that, in the extreme pessimistic case that regions ignore the climate change at all and always choose 0% mitigation and 100% savings, the temperature leads to approximately 7 deg Celsius increase in the year of 2100.

The parameters that we estimated and the corresponding estimation methods are listed below:

- The dynamic parameters for total factor productivity A: g_A and δ_A.

- The capital K: for the regions whose capital data is not available, we use a KNN regressor (Buitinck et al. 2013) to estimate it.

- The dynamic parameters for population L: l_g; similarly, for the regions whose convergence population data is not available, we use a KNN regressor to estimate it.

- The initial carbon intensity σ_0: for the regions whose capital data is not available, we use a KNN regressor to estimate it.

- KNN regressor: Because all regions have GDP and population data, we use them as features. For each region that lacks emission data and capital data, we find the nearest 5 neighbors according to its GDP and population. We use the average of the 5 neighbors' emission data and capital data as the estimated values.

Algorithm 1: Activity Component (implemented by `Climate_and_economy_simulation_step()`) Note that we only list input state variables and omit model parameters.

Require: exogenous emissions, land emissions, intensity, production factor, labor, capital, previous global temperature, previous government balance
Require: actions : mitigation rates, saving rates, tariffs, export rate limit, desired imports
  for each region **do**
    mitigation cost ← f(intensity)   ▷ Equation 37
    damages ← f(previous global temperature)   ▷ Equation 18
    abatement cost ← f(mitigation rate, mitigation cost)   ▷ Equation 19

    production ← f(production factor, capital, labor)   ▷ Equation 40
    gross output ← f(damages, abatement cost, production)   ▷ Equation 40
    government balance ← f(interest rate, previous government balance)
    investment ← f(saving rate, gross output)   ▷ Using Equation 36

    scaled imports ← f(gross output, desired imports)   ▷ Equation 52
    debt ratio ← f(previous government balance)   ▷ Equation 53
    scaled imports ← f(scaled imports, debt ratio)   ▷ Equation 54
  end for
  for each region **do**
    max potential exports ← f(gross output, investment, export rate limit) ▷ Equation 55
    Scaled imports ← f(scaled imports, max potential exports)   ▷ Equation 56
  end for
  for each region **do**
    tariff-ed imports, tariff revenue ← f(scaled imports, tariffs)   ▷ Equation 23
    domestic consumption ← f(savings, gross output, scaled imports) ▷ Equation 24
    aggregate consumption ← f(domestic consumption, tariff-ed imports) ▷ Equation 25
    utility ← f(labor, aggregate consumption)   ▷ Equation 2
    government balance ← f(imports, exports)   ▷ Equation 57
  end for
  temperature ← f(previous temperature, previous carbon mass, exogenous emissions)
  carbon mass ← f(previous carbon mass, intensity, mitigation rate, production, land emissions)
  for each region **do**
    capital ← f(capital, investment)   ▷ Equation 36
    labor ← f(labor)   ▷ Equation 38
    production factor ← f(capital)   ▷ Equation 39
    carbon intensity ← f(carbon intensity)   ▷ Equation 41
  end for

$$T_{t+1} = \Phi_T T_t + B_T \left(F_{2\times} \log_2 \left(\frac{M_t^{\text{AT}}}{M^{\text{AT},1750}} \right) + F_t^{\text{EX}} \right), \tag{34}$$

$$M_{t+1} = \Phi_M M_t + B_M \left(\sum_i \sigma_{i,t}(1 - \mu_{i,t}) Y_{i,t} + E_t^{\text{Land}} \right), \tag{35}$$

$$K_{i,t+1} = \Phi_{K,i} K_{i,t} + \Delta \left(1 - a_1 T_t^{\text{AT}} - a_2 \left(T_t^{\text{AT}} \right)^2 \right) \left(1 - \theta_{1;i,t} \mu_{i,t}^{\theta_2} \right) Y_{i,t} s_{i,t}, \tag{36}$$

$$\theta_{1;i,t} = \frac{p_b}{1000 \cdot \theta_2} \left(1 - \delta_{pb} \right)^{t-1} \cdot \sigma_{i,t}, \tag{37}$$

$$L_{i,t+1} = L_{i,t} \left(\frac{1 + L_{a;i}}{1 + L_{i,t}} \right)^{l_{g;i}}, \tag{38}$$

$$A_{i,t+1} = (e^\eta + g_{A;i} e^{-\delta_{A;i} \Delta(t-1)}) A_{i,t}, \tag{39}$$

$$Y_{i,t} = A_{i,t} K_{i,t}^\gamma L_{i,t}^{1-\gamma}, \tag{40}$$

$$\sigma_{i,t+1} = \sigma_{i,t} e^{-g_{\sigma;i}(1-\delta_{\sigma;i})^{\Delta(t-1)}\Delta}. \tag{41}$$

D.1 Population dynamic calibration

Denoting $L_{\infty;i} := \lim_{t \to \infty} L_{i,t}$, in the limit $t \to \infty$ we have:

$$L_{\infty;i} = L_{\infty;i} \left(\frac{1 + L_{a,i}}{1 + L_{\infty;i}} \right)^{l_{g;i}}, \qquad (47)$$

$$1 = \left(\frac{1 + L_{a;i}}{1 + L_{\infty}} \right)^{l_{g;i}}. \qquad (48)$$

As long as $l_{g;i}$ is not zero, $L_{\infty;i} = L_{a;i}$. Thus, $L_{a;i}$ is the long-term population size and a free parameter that is fitted to data. Assuming $\{L_{i,t}\}_{t=1,2,\dots}$ is monotonically increasing or decreasing, the absolute value of $l_{g;i}$ represents how fast it converges to $L_{a;i}$. The closer $L_{i,t}$ is to monotonically increasing or monotonically decreasing in the real data, the easier it is to fit $l_{g;i}$ and $L_{a;i}$.

To fit the population parameters, we take logs on both sides of Equation 38:

$$\log L_{i,t+1} = $$
$$\log L_{i,t} + l_{g;i}(\log(1 + L_{a;i}) - \log(1 + L_{i,t+1})), \quad (49)$$

where $\log L_{i,t+1} - \log L_{i,t}$ and $\log(1 + L_{i,t})$ are given by the data. $\log(1 + L_{i,t})$ and $l_{g;i}$ can then be estimated by linear regression.

D.2 Technology dynamic calibration

We estimate both g_A and δ_A from the existing data $\{A_t\}_{i=1\cdots n}$ by solving a regression problem:

$$g_{a;i}^{*}, \delta_{a;i}^{*} = \arg\max_{g_{a;i}, \delta_{A,i}} \mathcal{L}_{i,t} \qquad (50)$$

$$\mathcal{L}_{i,t} = ||A_{i,t} - (\exp\eta + g_{A,i}\exp(-\delta_{A,i}\Delta(t-1)))A_{i,t}||^2. \qquad (51)$$

This can be solved by numerical optimization algorithms, e.g., as provided in SciPy (Virtanen et al. 2020).

Because the emissions data from the World Bank API do not fit the form of the σ dynamic as assumed by DICE2016, use the DICE2016 parameter values for g_σ and δ_σ.

E Implementation details

Consistency checks. On the other hand, for the purpose of the competition, we fix the actions in the activity stage and the corresponding climate and economic dynamics. Thus, certain parts of the simulation should not be modified by competitors, e.g., the core climate and economic parameters and equation, such as the ones which affect carbon emissions and productivity. Changing these components may change the problem and prevent fair comparisons across submitted solutions. As such, our evaluation protocol and implementation of RICE-N include consistency checks that test submitted solutions; each submission requiring submitting the full code of the (modified) simulation and agent models. The activity stage is designed to follow the well established climate-economic dynamics in the RICE model augmented with trades and tariffs to include an essential influence in international negotiations. This way the simulation follows a simple model but is rich enough to implement interesting negotiation protocols.

Further, our implementation is set up to model each region as an RL agent that optimizes its discounted long-term rewards. Although other agent policy models are possible, we require that the competitors clarify how their proposed agent policy models also aim to optimize for the same goal, namely, the personal discounted long term reward for each agent. This ensures that the agents behave greedily and any improvements in global metrics are strictly a result of the implemented negotiation protocols. See Section 5 for details on the agent model.

CPU vs GPU simulations. We provide both CPU and GPU versions of the base RICE-N, but evaluate all submissions using the CPU version to reduce the cost of evaluation. The CPU and GPU implementations of the base version of RICE-N yield consistent results. The GPU version of the simulation is meant to accelerate training.

F CO$_2$ Emission

Participants who use Google Colab have a carbon efficiency of 0.37 kg CO$_2$eq/kWh. This was measured on 0.35 hours of computation on a Tesla T4 (300W TDP) node.

The emissions produced by going through the tutorial and training a baseline model are estimated to be 0.01 kg CO$_2$eq of which 100% are offset by the cloud provider.

Estimates were obtained using the Machine Learning Impact calculator (Lacoste et al. 2019) available at https://mlco2.github.io/impact#compute. We encourage participants to use tools such as CodeCarbon (Schmidt et al. 2021) (available at https://codecarbon.io) as illustrated in the tutorial notebook to track their carbon emissions and report it as part of their publications.

G Submission and evaluation details

Structure of a submission. Submissions consist of the full simulation code, agent behavioral models, and other components that may be needed to generate full simulation roll-outs.

If participants use the GPU version of the simulation, they should also provide a CPU version and ensure that any models trained on the GPU version and any customization yield consistent results on a CPU.

Evaluation metrics. Solutions are evaluated by computing the area under the empirical Pareto frontier produced by that solution. The area computation is explained in Figure 5.

H WarpDrive Support

Performing multi-agent RL is significantly faster using GPU acceleration. We provide a GPU version of the simulation and an end-to-end GPU-based RL training loop using WarpDrive. For details, see the instruction on the Github repo https://github.com/mila-iqia/climate-cooperation-competition.

I Trade constraints

To ensure that total imports and total exports match, three constraints are enforced on regions' trade flows.

Table 1: Calibrated parameters for 27 regions

Region ID	A_0	K_0	L_0	L_a	$delta_A$	g_A	l_g	$sigma_0$
1	1.872	0.239	476.878	669.594	0.139	0.122	0.034	0.456
2	8.405	3.304	68.395	93.497	0.188	0.103	0.058	0.529
3	3.558	0.109	64.122	135.074	0.161	0.127	0.026	0.816
4	1.927	1.424	284.699	465.308	0.244	0.134	0.024	1.221
5	8.111	0.268	28.141	23.574	0.163	0.106	-0.057	0.290
6	4.217	3.184	548.754	560.054	0.170	0.095	0.080	0.302
7	2.491	0.044	46.489	59.988	0.058	0.049	0.037	0.420
8	2.525	1.080	69.194	100.016	0.346	0.079	0.029	1.010
9	2.460	0.184	513.737	1867.771	1.839	0.462	0.017	0.310
10	12.158	2.642	38.101	56.990	0.131	0.063	0.020	0.350
11	0.993	0.160	522.482	1830.325	0.086	0.065	0.019	0.235
12	5.000	2.289	165.293	230.191	0.183	0.071	0.027	0.419
13	29.854	2.020	165.751	216.927	0.088	0.075	-0.002	0.254
14	23.315	3.039	109.395	143.172	0.088	0.075	-0.002	0.254
15	29.854	0.687	56.355	73.755	0.088	0.075	-0.002	0.254
16	10.922	0.606	705.465	532.497	0.096	0.168	-0.016	0.781
17	9.634	0.608	465.607	351.448	0.096	0.168	-0.016	0.781
18	8.621	0.453	239.858	181.049	0.096	0.168	-0.016	0.781
19	3.190	0.129	690.002	723.513	0.054	0.068	-0.013	0.949
20	2.034	0.381	455.401	477.518	0.054	0.068	-0.013	0.949
21	13.220	16.295	502.410	445.861	0.252	0.074	-0.033	0.170
22	3.190	0.044	234.601	245.994	0.054	0.068	-0.013	0.949
23	6.387	1.094	317.880	287.533	0.194	0.237	-0.053	0.840
24	2.481	0.090	94.484	102.997	0.203	0.201	0.037	1.665
25	10.853	17.554	222.891	168.351	0.005	0.000	-0.012	0.285
26	4.135	1.002	103.294	87.418	0.158	0.123	-0.063	0.601
27	2.716	1.034	573.818	681.210	0.097	0.101	0.043	0.638

1. For each region i, if the region's total desired imports from other regions exceed its own gross output, then the imports are scaled to sum up to the region's gross output. We enforce the constraint that $\sum_{i \neq j} b_{i,j,t} \leq Q_{i,t}$, which is to say that a region may not import more goods than its current gross output capacity. This constraint helps the agents avoid insurmountable debt, thereby stabilizing trade balances over the entire time period while also easing learning. If a region's desired imports exceed its production capacity, then its import bids are scaled down to size :

$$b_{i,j,t} \leftarrow b_{i,j,t} \min \left\{ 1, \frac{Q_{i,t}}{\sum_{i \neq j} b_{i,j,t}} \right\}. \qquad (52)$$

2. Regions are allowed to carry a (positive or negative) trade balance $D_{i,t}$. At the start of each new time step, each region's trade balance, positive or negative, accumulates interest at a fixed rate of 10%. Based on this balance, a region's debt-to-initial-capital ratio is determined and the imports are scaled according to this ratio:

$$d_{i,t} = 10 \frac{D_{i,t}}{K_0}, \qquad (53)$$

$$b_{i,j,t} \leftarrow b_{i,j,t}(1 + d_{i,t}). \qquad (54)$$

3. If other regions' total desired imports from region i exceed region i's upper bound on exports $x_{i,t}^{\max}$, then the bids for goods from region i are scaled proportionally to $x_{i,t}^{\max}$. Otherwise, each region receives its full import bid from region i. In other words, region i cannot export more goods at time t than it could consume at time t, so other regions will import less from region i.

$$x_{i,t}^{\max} = \min(p_{i,t}^x Q_{i,t}, Q_{i,t} - I_{i,t}), \qquad (55)$$

$$x_{i,j,t} = b_{i,j,t} \min \left\{ 1, \frac{\min(x_{i,t}^{\max})}{\sum_{j \neq i} b_{i,j,t}} \right\}. \qquad (56)$$

After all constraints have been applied, the trade balance for the next period is calculated:

$$D_{i,t+1} = D_{i,t} + \Delta \left(\sum_{j \neq i} x_{j,i,t} - \sum_{j \neq i} x_{i,j,t} \right). \qquad (57)$$

J Parameters and variables

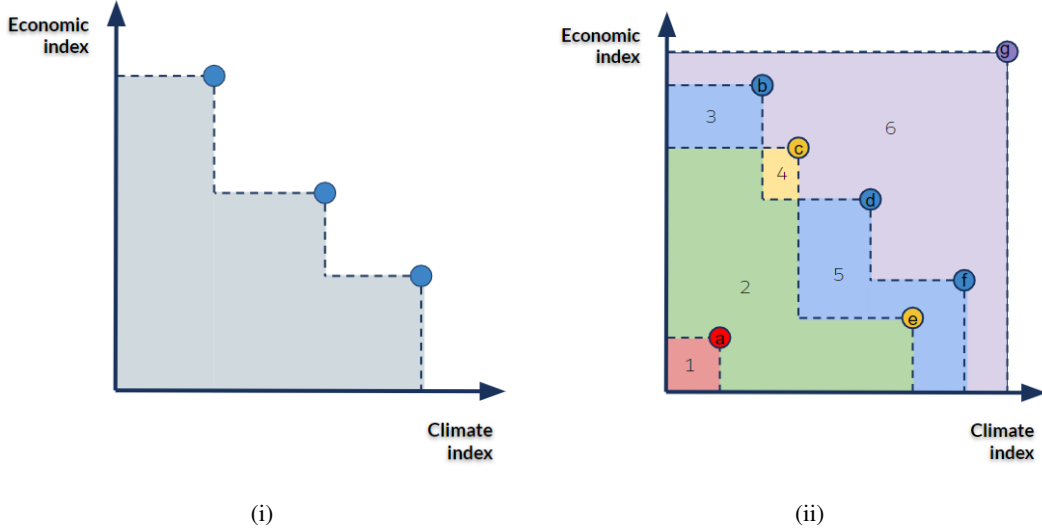Tables 1, 2, 3, 4, and 5 list all (calibrated) parameters and variables.

Figure 5: **A visualization of how teams are ranked using the hypervolume indicator (Zitzler et al. 2003).** (i) A single team's set of submissions is represented by a set of points of the same color. The hypervolume indicator is the area of a set of non-dominated policies (Van Moffaert and Nowé 2014) in the 2-dimensional $(\mathcal{C}, \mathcal{E})$-space, as represented by the shaded area. (ii) Multiple sets of solutions ({a}, {b,d,f}, {c,e}, {g})proposed by various teams. Teams are ranked by the value of their hypervolume indicators: a higher value is a better score. A first team submits point (a), so their set of solutions contains one point. The hypervolume indicator of this solution corresponds to area 1. A second team submits points (c) and (e). Their score corresponds to the sum of areas 2 and 4. A third team submits points (b), (d) and (f), and so their score is the sum of areas 2, 3 and 5. A final team submits a single point (g). The hypervolume indicator of this point corresponds to the sum of all areas 1 to 6. As such, the team that submitted point (g) is ranked highest, the team that submitted points (b), (d) and (f) is second, the team that submitted points (c) and (e) is third, and the team that submitted point (a) is last.

K Formal Description for Negotiations and Agreements

The aim of this section is to specify a working mathematical definition of a *negotiation protocol* in the context of the simulator and competition. Given the interdisciplinary nature of this work, we fully expect these definitions to grow in the future.

Consider a set of agents representing regions in RICE-N. These agents are seeking to individually optimize their economic outcomes, but the sum of their economic outcomes depends on them collaborating toward a common climate goal. Hence, such a setting can be understood as a variable sum, or mixed-motive, game (Schelling 1980).

K.1 Agent actions

Consider n agents with k-dimensional action spaces acting from $t = 0$ to $t = T$.

Their **collective action space** at time t is a kn-dimensional space, where each agent's action space is a k-dimensional subspace of the collective action space. The collective action space at time t is denoted $\mathcal{A}^t$, and agent i's action space is denoted $\mathcal{A}_i^t$. At each timestep t, an nk-dimensional vector is sampled from $\mathcal{A}^t$, with each of the n agents specifying how to sample k values.

Analogously, the **collective action space over time** is a knT-dimensional space, where each agent's action space

over time is a kT-dimensional subspace of the collective action space over time. The collective action space over time is denoted $\mathcal{A}$ and agent i's action space over time is denoted $\mathcal{A}_i$. At the end of a run, an nkT-dimensional vector will have been sampled from the simulator detailing each action of each agent at each timestep.

K.2 Agreements

The definition of an *agreement* in the context of international environmental agreements usually draws on the definition of a treaty in the 1969 Vienna convention on the law of treaties: "an international agreement concluded between States in written form and governed by international law" (Mitchell 2003). Notably, states participating in the agreement express a "consent to be bound" (Mitchell 2003).

Algorithmically defining agreements is challenging for numerous reasons. In particular, legal agreements are drafted in natural language, which can not be directly interpreted by computers. This would seemingly encourage the design of agreements through a mathematical lens to eliminate ambiguity. However, in agreements, ambiguity is not a bug, it's a feature. Many agreements are intentionally drafted to be generic enough to generalize to a broad set of situations (Martin-Bariteau and Scassa 2020), "some of which could not have been foreseen at the time of drafting" (De Filippi and Hassan 2018).

Variable	Type	Symbol	Description
Carbon Mass	Global, endogenous	$M_t, [M_t^{AT}, M_t^{UP}, M_t^{LO}]$	A three-dimensional vector that indicates the average carbon accumulation in the atmosphere, upper oceans, and lower oceans.
Temperature	Global, endogenous	$T_t, [T_t^{AT}, T_t^{LO}]$	A two-dimensional vector that indicates the average temperature of the atmosphere and the lower ocean.
Population	Regional, exogenous	$L_{i,t}$	Population and the labor in a region.
Technology	Regional, exogenous	$A_{i,t}$	Technology factor in the production function of a region.
Capital	Regional, endogenous	$K_{i,t}$	Total capital accumulated by a region.
Carbon intensity of economic activity	Regional, exogenous	$\sigma_{i,t}$	A scalar coefficient that gives the emissions resulting from economic production.
Balance of trade	Regional, endogenous	$D_{i,t}$	Surplus or deficit from international trade activities.
Cost of mitigation efforts	Global, endogenous	$\theta_{1;i,t}$	An estimate of the cost of mitigation efforts.
Emission due to land use	Regional, exogenous	E_t^{Land}	Carbon emission for land use in a specific region.

Table 2: **World-state variables.** Global type variables correspond to the entire world, whereas regional type variables correspond to each region. Endogenous variables are those which are affected by the agent actions, whereas exogenous variables are those that are predetermined and not affected by agent actions. Note that the values of endogenous variables can vary across steps in a predetermined manner. **Notation: indices are separated from subscripts referring to a name by semicolons (;). For instance, the parameter θ_1 varies in time t and by region i, which is denoted as $\theta_{1;i,t}$.**

Variable	Symbol	Description
Savings rate	$s_{i,t}$	The fraction of output production to be invested in capital.
Mitigation rate	$\mu_{i,t}$	The fraction of mitigation efforts by a region.
Import tariffs	$\tau_{i,j,t}$	The fraction of imports that are converted to tariff revenue.
Export limits	$p_{i,t}^x$	The fraction of domestic production that regions are willing to export.
Import bids	$b_{i,j,t}$	The amount of production each region is willing to import from other regions.

Table 3: **Agent-action variables.**

K.3 Commitments, enforcement and incentives

Although regions can commit to a course of action, there is no supranational body that can enforce regions to respect commitments. Regions therefore always have the option to breach the conditions in their agreements. Hence, regions may need to engage in strategic behavior, such as mutual enforcement or providing incentives to other regions, to maintain cooperation and adherence to agreements. For instance, they might use trade and tariffs to reward or punish other regions. However, understanding to what extent such behaviors are rational and strategically advantageous is an open challenge.

K.4 Formalizing agreements

In light of this, we provide working definitions of proposed and enacted agreements in RICE-N.

A **proposed agreement** between agents $I \subseteq \{1, ..., n\}$ is a set of constraints on the union of subspaces $\{\mathcal{A}_i\}_{i \in \mathcal{I}}$.

An **enacted agreement** between agents $I \subseteq \{1, ..., n\}$ is a set of constraints on the sum of subspaces $\{\mathcal{A}_i\}_{i \in \mathcal{I}}$ in which each agent $i \in \mathcal{I}$ agrees to the constraints that apply to their subspace $\mathcal{A}_i$.

Naturally, an agreement enacted at time t can apply constraints to the entire space $\mathcal{A}$, but actions that have already been drawn from $\sum_0^{t-1} \sum_{i \in \mathcal{I}} \mathcal{A}_i$ must be taken as given:

agreements can not change the past. For example, an agreement at time t can constrain the sum total of emissions over a full run of the simulator, but it can not alter emissions from $t = 0$ to $t = t - 1^6$.

A key distinction between *proposed* agreements and *enacted* agreements is that, while proposed agreements can be generated in any manner, enacted agreements must be voluntarily accepted by the agents constrained by the agreement, expressing the "consent to be bound" aspect of the Vienna convention on the law of treaties (Mitchell 2003). In other words, a proposed agreement cannot be enacted (become an enacted agreement) without the consent of all implicated agents. If a proposed agreement is not enacted, then its constraints are not applied in the simulation.

Currently, an enacted agreement only applies changes to $\mathcal{A}^t$. In the next time period, new proposed agreements are submitted and evaluated, which enables regions to not follow through on their previous agreements.

---

[6] In practice, reported carbon emissions are influenced by carbon accounting practices. Distinguishing between ground truth emissions, regions' private climate-economic information and what they make publicly available within the simulator is an interesting area of exploration.

Variable	Symbol	Description
Initial population	$L_{0;i}$	The initial population for a specific region.
Population convergence target	$L_{a;i}$	The estimated convergence population for a specific region.
Population convergence rate	$l_{g;i}$	How fast the current population converges.
Initial capital	$K_{0;i}$	The initial capital for a specific region.
Initial carbon intensity	$\sigma_{0;i}$	The initial carbon intensity for a specific region.
Carbon intensity parameters	$g_{\sigma;i}$ and $\delta_{\sigma;i}$	The decay speed of the carbon intensity.
Initial technology factor	$A_{0;i}$	The initial carbon technology factor for a specific region.
Technology factor parameter	$g_{i,A}$ and $\delta_{i,A}$	The update pattern of the technology factor.
Initial land use emission	$E_{L0;i}$	The initial land use emission for a specific region.
Land use emission parameter	$\delta_{EL;i}$	The depreciation rate for the land use emission in a specific region.

Table 4: **Agent-specific constants.**

Variable	Symbol	Description
Capital elasticity of production	γ	The contribution from capital and population to the economy.
Armington substitution parameter	λ	How substitutable consumption goods from different regions are.
Long term welfare discount rate	ρ	How much short-term welfare is weighted versus long-term welfare.
capital depreciation rate	Φ_K	The capital depreciation rate.
Backstop technology	p_b	Price of a backstop technology that can remove carbon dioxide from the atmosphere.
Backstop technology parameter	δ_{pb}	The decay speed of the cost of backstop technology.
Mitigation efficiency parameter	θ_2	The efficiency loss component of mitigation
Domestic share parameter	ψ^{dom}	The relative preference for domestic goods
Foreign share parameter	ψ^{for}	The relative preference for foreign goods

Table 5: **Global constants.**

K.5   Agreements through negotiation protocols

A **negotiation protocol** is an algorithmic procedure to propose agreements. The goal of such a protocol is to propose agreements that agents will enact and that lead to optimal climate-economic outcomes by balancing collective climate and individual economic objectives.

Negotiation protocols currently consist of two (potentially iterative) phases: 1) generating proposals, and 2) evaluating proposals.

Generating proposals.   Agreements can be generated in any way that participants see fit. Such agreements can be proposed 1) *internally*, which is to say that agents participating in the negotiation protocol can send messages to other agents in which they propose agreements, or 2) *externally*, where an external body (represented within the negotiation protocol itself) proposes an agreement to a subset of participating agents.

Evaluating proposals.   Agents participating in the simulation must evaluate proposals that concern them to determine whether they wish to accept them or refuse them. Importantly, a proposed agreement that is refused by an agent i must not apply constraints to the action space $\mathcal{A}_i$ of the agent that refused it. However, an agreement proposed internally by a participating agent is de facto considered accepted by that agent. For example, an agent might partition their own action space and propose an agreement to another agent while guaranteeing that they will only choose actions within a certain partition based on the other agent's acceptance or refusal.

Repeating the process.   Generation and evaluation can iterate in a negotiation protocol. For example, an agent i could initially propose an agreement to an agent j. The agent j can refuse this agreement, and propose an alternative agreement back to i, or to another agent k.

Generating a proposed agreement that agents will enact can be seen as a double constraint satisfaction problem: the goal of a negotiation protocol is to identify a set of constraints in the action space that are accepted by the participating agents. Agents map proposed agreements implicating them to either 0 or 1 depending on acceptance or refusal. The agreement is itself a set of constraints on the action space. Therefore, generating a proposed agreement is equivalent to generating a set of constraints on the action space that is mapped to the intersection of the acceptance space for all implicated agents.

This implies a curse of dimensionality when it comes to generating agreements: since an agreement fails unless all concerned agents accept, the more agents an agreement is proposed to, the less likely it is that the agreement is enacted.

K.6   Communication

Within RICE-N, agents may communicate by exchanging messages, which consist of numerical values that do not directly affect the climate-economic dynamics of the simulator. However, such messages might influence subsequent actions that directly impact the dynamics.

The goal of the competition can therefore be broadly understood as attempting to structure communication within the simulation in a way that, e.g., agents might share information about goals or priorities, their willingness to cooperate, with the end goal of aligning incentives and maximizing their individual objectives, and ultimately global climate and economic goals.

Modeling and learning how to communicate in multi-agent

RL is its own subfield. We refer readers to (Zhu, Dastani, and Wang 2022) for a recent review of the topic, and (Foerster 2018) for an in-depth technical introduction.