Multi-Agent Reinforcement Learning Simulation for Environmental Policy Synthesis

Blue Sky Ideas Track

James Rudd-Jones
Centre for Artificial Intelligence,
Department of Computer Science,
University College London
London, UK
james.rudd-jones.22@ucl.ac.uk

Mirco Musolesi
Centre for Artificial Intelligence,
Department of Computer Science,
University College London
London, UK
Department of Computer Science and
Engineering, University of Bologna
Bologna, Italy
m.musolesi@ucl.ac.uk

María Pérez-Ortiz
Centre for Artificial Intelligence,
Department of Computer Science,
University College London
London, UK
maria.perez@ucl.ac.uk

ABSTRACT

Climate policy development faces significant challenges due to deep uncertainty, complex system dynamics, and competing stakeholder interests. Climate simulation methods, such as Earth System Models, have become valuable tools for policy exploration. However, their typical use is for evaluating potential polices, rather than directly synthesizing them. The problem can be inverted to optimize for policy pathways, but the traditional optimization approaches often struggle with non-linear dynamics, heterogeneous agents, and comprehensive uncertainty quantification. We propose a framework for augmenting climate simulations with Multi-Agent Reinforcement Learning (MARL) to address these limitations. We identify key challenges at the interface between climate simulations and the application of MARL in the context of policy synthesis, including reward definition, scalability with increasing agents and state spaces, uncertainty propagation across linked systems, and solution validation. Additionally, we discuss challenges in making MARL-derived solutions interpretable and useful for policy-makers. Our framework provides a foundation for more sophisticated climate policy exploration while acknowledging important limitations and areas for future research.

KEYWORDS

Multi-Agent Reinforcement Learning; Earth System Models; Integrated Assessment Models; Policy Making

1 INTRODUCTION

Climate policy derivation represents one of society's most difficult governance challenges, characterized by deep uncertainty, competing stakeholder interests, and complex interdependencies across social, economic, and environmental systems [44]. At the national level, climate policy formulation typically involves analyzing emissions reduction potential, assessment of technological pathways, and evaluation of policy actions [13]. Multi-stakeholders are accounted for to ensure a cohesive approach, but extending to the continental or global scale makes this process more challenging. This conventional approach faces further fundamental difficulties due to the socio-environmental domain that can be summarized threefold. Firstly, the global climate system has significant temporal latency between action and downstream effects. Many of these effects are not monitored or hard to capture, further increasing this latency and making causal links between policy implementation and observable outcomes hard to solidify [21]. This temporal disconnect can span decades, making it very challenging to evaluate policy effectiveness and adjust strategies in response to emerging data [37]. Secondly, climate policies often generate unevenly distributed impacts across different socioeconomic groups and regions, potentially exacerbating existing inequalities and creating political resistance [4]. Thirdly, the presence of tipping points and feedback loops in the Earth system introduces non-linear dynamics that traditional policy analysis struggles to address [16]. Current approaches to climate policy development rely heavily on international negotiations and evidence synthesis from various scientific disciplines. Organizations such as the IPCC coordinate massive efforts to provide policy-relevant insights [32]. However, this process can be slow, politically constrained, and sometimes fails to capture the full range of possible policy interventions [52], including the deep uncertainty inherent in climate projections [25].

Simulation-based approaches have emerged as crucial tools for addressing these limitations. They enable policy-makers to explore potential outcomes in a risk-free environment, facilitating the evaluation of various policy combinations and their long-term implications [51]. Being able to quantify the forecasted effects of policies
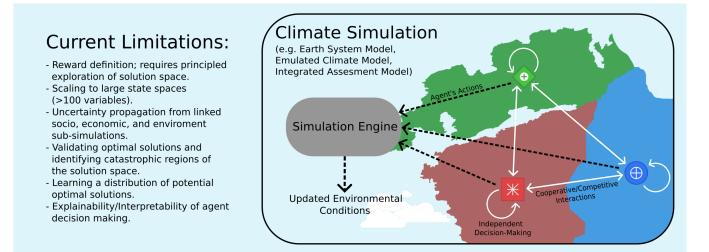
Figure 1: On the right: A conceptual framework of a multi-agent climate simulation system. The outer climate simulation is imbued with socio-economic agents, in this case three geographical regions. Agents make independent decisions while engaging in cooperative/competitive interactions (white arrows). A central simulation engine (grey) processes agents' actions (dashed black lines) for the linked socio, economic, and environmental simulations, updating the environmental conditions accordingly. On the left: An enumeration of six key technical open challenges currently facing such systems.

not only guides the policy derivation process but also provides evidence to address concerns raised by critics [2]. Earth System Models (ESMs) provide a crucial simulation framework, offering high-resolution representations of atmospheric, oceanic, and terrestrial processes [8]. For example, they have directly informed national carbon budgets and adaptation strategies by simulating changes in temperature, precipitation, and extreme weather events [41]. However, their utility in policy-making is generally constrained as the influence of society is often not an internal process in these models. Exogenous variables, such as atmospheric carbon, are changed over time to "simulate" anthropogenic effects, with limited feedback on how the changing climate would affect society [43]. Instead, Integrated Assessment Models (IAMs) combine socio-economic and environmental components to model climate socio-economic interactions, representing the current state-of-the-art in climate policy simulation. Generally, they use simplified environmental processes derived from ESMs but the incorporation of linked socio-economic simulations allows for a more defined feedback loop between the environment and society. Notable examples include the DICE model [27], REMIND [18], and GEMINI-E3 [1]. On the UN website they publicly list twenty-nine IAMs used for their decision making [48], as an example GEMINI-E3 was used to analyze the future CO_2 emission trajectories of the national policies made at the COP26 Glasgow conference [49]. IAMS can be preferable to ESMs as it allows a more direct way to evaluate policy interventions as cascading effects between environment and society are linked. Importantly the problem can be "flipped"; *we can optimize for an idealistic policy.* We focus in on IAMs due to this feature, as in ESMs this is not as connected. In ESMs we could define optimizing criteria, but the resultant solution has no explicit connection to a policy and there are no clear "actions". Traditional optimization approaches for IAMs (for example Model Predictive Control [9])

solve for a set utility function. However, may rely on simplifying assumptions, discrete state spaces, and local linear approximations to make the problem computationally tractable [3, 46]. These solvers typically only explore a limited portion of the possible solution space, potentially missing innovative policy combinations [31]. On top of this, most models use highly aggregated representations of socio-economic agents, failing to capture the diversity of stakeholder behaviors and interactions [22]. Finally, these approaches in IAMs often focus on finding a single "optimal" solution, which may not adequately capture the full range of possible outcomes [57]. This in itself can fail to capture some notion of uncertainty that exists in both climate system responses and socioeconomic developments.

These limitations point to the need for more sophisticated approaches that can better handle uncertainty, complex system dynamics, interactions between heterogeneous agents, and the vast solution space of possible policy combinations. Modern computational techniques, particularly Reinforcement Learning (RL) and Multi-Agent Reinforcement Learning (MARL), offer promising capabilities to address these challenges. RL approaches can overcome these limitations by providing more robust exploration of policy spaces, not requiring linear assumptions, and improved handling of uncertainty. MARL approaches can better represent agent heterogeneity for resilient policies. For a single agent scenario, the recent work of [46] and [58] applied an RL agent into an IAM, that generated policy guidance pathways towards a defined "economic and environmental positive future". Both [46] and [58] use a singular agent, hence assuming a "unified" earth, in which there is a collectively shared goal, restricting the potential for agent heterogeneity. For agent heterogeneity, [60] created the RICE-N model, itself an extension of the Regional Integrated model of Climate and the Economy (RICE) model developed in [24] that models twelve

global regions. The RICE-N model imbues an IAM with trade and negotiation dynamics for further interactions between agents [60]. Further, [38] extended the environment in [58] towards multiple agents, using the intepretability gained from the low-dimensional environment to understand policy pathways.

2 TOWARDS A FRAMEWORK FOR MARL DERIVED CLIMATE POLICY

ESMs and IAMs can be used to validate potential policy interactions in a safe simulated space. More importantly we can work the other way and derive optimal policies. Whilst ESMs can be anthropogenically influenced, we focus on IAMs as a case study since they have a more direct connection between socio and environmental systems that better suite optimization. RL is able to model non-linear stochastic systems, greatly aligning its use with IAM dynamical systems, and accounting for long term reward well posits its use in policy trajectory derivation. Importantly as a first step, how can we reformulate IAMs as an RL problem, or more concretely can we setup an IAM as a Markov Decision Process (MDP) [36] for RL or Stochastic Game (SG) [40] for MARL? IAMs already share significant structural similarities with RL environments; they inherently require objective functions if optimized, that can be used as a reward function. Further, actions can be defined as varying model parameters over time to reflect some policy trajectory. For example, investment decisions, policy implementations, and technological adoption rates. To summarize, all that is needed to utilize RL/MARL with an IAM is defining a set of actions, and accessing outputs of a running simulator or porting the simulator into a programming language that is suited for RL/MARL research.

Why Use MARL? Anthropogenic climate change inherently involves heterogeneous actors with mixed motives, strategic interactions, and complex cooperation-competition dynamics - elements that single-agent RL cannot adequately capture. Mirroring major criticisms in the IAM literature where economic and behavioral dynamics are poorly represented by a singular entity [19, 22]. Further, MARL's ability to model long-term strategic behavior while accounting for multiple interacting entities makes it particularly suitable for climate policy modeling [12, 15].

What are we Optimizing with MARL? A clarification upon moving to MARL is how we view the optimization. We believe there are two main views that necessitate certain restrictions on the MARL application. *Firstly*, one could focus on purely modeling all heterogeneous agents interacting together - we want to understand the dynamics between agents as some form of sequential social dilemma (SSD). MARL has emerged as a powerful framework for studying SSDs and "tragedy of the commons" type scenarios, offering insights into emergent social dynamics and equilibrium behavior. Unlike traditional game-theoretic approaches that often rely on one-shot interactions, MARL enables repeated sequential decisions with complex state dynamics [15]. This approach has proven particularly valuable in commons problems, where agents must balance immediate individual rewards against long-term collective welfare [12]. These studies collectively suggest that MARL not only serves as a tool for finding Game-Theoretic equilibria without heavy assumptions but also provides insights into the mechanisms that drive the evolution of social behavior and cooperation in complex,

dynamic environments [33, 47]. Results from this view-point can help domain specialists understand the effects of certain simulation or agent parameters on an IAM equilibrium.

Secondly, a more applicative approach entails planning policy pathways for one or a subset of the agents, and using representative models of other entities. The resultant pathways can be used to guide decisions by policy makers, as the framework provides evidence of their validity. As an example use case, the other agents in the simulation (that we have no agency over in reality) could be trained using imitation learning on historical data to represent in-silico versions of real-world entities. The MARL agent(s) can learn a best response to these pre-trained agents dependent on the validity not only of the IAM, but also the representative agents. This approach shares similarities with the domain of Ad-Hoc Teamwork [45, 53, 54], which explores how agents can learn to interact with novel and unknown partners. This second framework necessitates the use of decentralized algorithm approaches as agents must not share model parameters, rewards, or observations.

3 OPEN CHALLENGES

Reward Definition. A common issue with any optimization approach is the objective function, which requires some bias towards a chosen target variable or parameter [35]. Motivating these is tricky and is up to the modelers discretion [35]. Similarly for RL and MARL we have the definition of a reward function, which can be an arbitrary choice. There are two sides of the coin here, rewards in RL/MARL can be more abstract than objective functions. For example a binary reward for reaching a certain end state, that encodes complex objectives like sustainability and equity, but may mask important nuances of the underlying simulation [42]. We could set rich reward signals such as directly relating to a normalized economic or environmental variable. In this domain even dense rewards may receive many sequential negative "poor" rewards (e.g., strict legislation) until a "positive" outcome is reached. This is a typical deep exploration task in RL (e.g. deepsea [30]), which generally require uncertainty based exploration methods to fully search the solution space [6]. Uncertainty based exploration in RL is an extensive field [7, 28, 29], but is not so explored in MARL [11, 20, 39, 61]. Due to the non-stationarity in the transition function there seems even more of a need for principled exploration. On top of this, the climate can change rapidly as tipping points are reached requiring quick behavioral responses from the other agents, both exacerbating non-stationarity [5].

Scalability. RL and MARL approaches can be more adept in the complex IAM environment over traditional optimal control approaches, but can similarly struggle with scalability. Particularly in MARL where increasing the number of agents can lead to an exponential growth of state and action spaces in centralized training approaches [59]. Decentralized approaches which are required for the *second MARL viewpoint*, are generally able to scale more easily as model parameters are not shared. However, coordination becomes more challenging and thus requires more complex methodologies that impact scaling, especially as numbers of agents grow beyond hundreds. Further compounding the issue, large scale IAMs such as the Intertemporal General Equilibrium Model [26] that has 4000 endogenous variables, lead to a very large state space for any

RL application. Reaching these IAMs not only raises challenges in the algorithmic side, but these large-scale simulators require much larger computational budgets, further increasing training times. The difficulty lies in coupling these systems - economic shocks influence social responses, which affect climate policy implementation, creating complex cascade effects [23].

Uncertainty Representation. Understanding model uncertainty is inherent to the downstream use, it is imperative to understand when our temporal predictions become no better than random choices. IAMs contend with significant sources of uncertainty from many sources. Each separate socio-economic and environmental model has its own epistemic uncertainty stemming from modeling assumptions and aleatoric uncertainty from inherent noise in the underlying process. As these models are all linked there is further uncertainty in their interactions, particularly climate sensitivity, damage functions, and economic growth projections [10, 34]. Further, IAMs bridge multiple temporal and spatial scales, from short-term economic decisions to long-term climate dynamics. Uncertainty compounds across these scales, making it difficult to accurately represent interactions between local impacts and global processes [50]. Decision making becomes intractable as many climate risks exhibit heavy-tailed probability distributions [56]. As discussed earlier, uncertainty-driven methods in RL can bring deeper exploration of the solution space and have a chance to adapt to these regions of greater uncertainty. However, typically exploration is guided by epistemic uncertainty of the RL agent and does not account for the aleatoric uncertainty stemming from the simulator [7]. Importantly, how can we factor in the multiple sources of uncertainty from the linked models, to get better calibrated notions of uncertainty in the suggested solutions?

Solution Validation. Validating adaptions of IAMs to MDPs or SGs can be fairly straightforward, simulation dynamics can be compared, usually with minimal discrepancy as shown in [38]. Comparing traditional optimization approaches with the RL and MARL solution is fairly similar. Distance metrics between potential pathways, or final visitation states quantify the performance of the framework. The real challenge lies in validating simulation solutions with real world applicability, connecting to broader questions in simulation intelligence and model validation [14]. Perfect validation against real-world outcomes, particularly for long-term climate predictions, is fraught with risk due to the inherent stochasticity of such systems. Importantly, it can be easier to identify and validate the unfeasible or undesirable trajectories, which may enter areas of the solution space that we can have more assurance on being incorrect. Mapping out dangerous trajectories provides valuable insights for decision-makers to avoid catastrophic scenarios. Quantifying and understanding negative outcomes can be as crucial as identifying optimal solutions.

Distribution of Solutions. Traditional optimization approaches usually provide only one optimal solution without describing the robustness of said approach [57]. Due to the inherent uncertainties of IAMs, and the challenges from exactly following a projected pathway, solutions that have more tolerance for error, in that they are more resilient to shocks or tipping points, are preferable. In RL/MARL this is possible by evaluating optimal agent policies with various initial states to get a distribution of solutions. It can be a cumbersome approach, as if the optimal solution is found to be non-robust, the agents must be retrained and the process iterated. Instead recent work by [17] finds a diverse range of optimal solutions in complex highly dynamic environments.

Explainability & Interpretability - Explainability is an extensive challenge in any use of Machine Learning, but especially so in RL/MARL when using complex non-linear models. While traditional control methods like MPC offer clearer insights into their decision-making through explicit optimization objectives and constraints, deep RL methods often operate as black boxes. Most approaches provide post-hoc explanations that attempt to provide understanding between agent actions and certain states. Our framework provides policy trajectories in the solution space of the IAM, but these larger state spaces easily extend beyond human interpretable three-dimensional spaces. How are downstream users able to interpret solutions, are we visualize policy trajectories in high-dimensional spaces?

4 LIMITATIONS
Primarily, the MARL solution can only be as valid as the underlying simulator; improved optimization solutions cannot transcend the constraints of an imperfect model. This becomes particularly problematic in these environmental domains characterized by unprecedented scenarios or complex emergent phenomena, such as extreme weather patterns as we pass climate tipping points [5]. Current climate simulators may struggle to accurately model the feedback loops between anthropogenic caused rising temperatures and the downstream effects, leading to potentially dangerous blind spots in MARL-based policy recommendations. While there has been progress in open-ended learning within RL frameworks [55], this capability must be matched by equally adaptable simulation environments - a requirement that may exceed current modeling capabilities. However, these limitations do not entirely diminish the value of our approach; since our framework is agnostic to the underlying world system model (ESM or IAM or more), it can readily incorporate improved simulators as they become available, ensuring adaptability across various domains and modeling paradigms.

5 CONCLUSION
This paper presents a framework for enhancing climate policy exploration through the integration of MARL with IAMs. While MARL offers promising capabilities for handling non-linear dynamics, agent heterogeneity, and uncertainty quantification, significant challenges remain. The framework's success hinges on addressing key challenges including principled exploration in sparse reward settings, scalability of both algorithms and simulations, and the propagation of uncertainty across coupled systems. Moreover, translating MARL-derived solutions into actionable policies requires advances in explainability and visualization techniques. We highlight three promising research directions: **1.** uncertainty-driven MARL algorithms to handle the multiple sources of uncertainty inherent in IAMs; **2.** MARL for extensive state spaces that may rely upon linked emulations; **3.** explainable RL techniques specifically tailored for large state spaces. We hope that this position paper drives efforts towards the improvement of simulation derived climate policy, supporting political guidance to return the Earth's trajectory onto a habitable and stable future.

REFERENCES

[1] Alain Bernard and Marc Vielle. 2008. GEMINI-E3, a General Equilibrium Model of International–National Interactions Between Economy, Energy and the Environment. *Computational Management Science* 5, 3 (2008), 173–206.

[2] Paul Cairney. 2016. *The Politics of Evidence-based Policy Making.* Springer.

[3] Guillaume Deffuant and Nigel Gilbert. 2011. *Viability and Resilience of Complex Systems: Concepts, Methods and Case Studies from Ccology and Society.* Springer.

[4] Noah S Diffenbaugh and Marshall Burke. 2019. Global Warming has Increased Global Economic Inequality. *Proceedings of the National Academy of Sciences* 116, 20 (2019), 9808–9813.

[5] Isaiah Farahbakhsh, Chris T Bauch, and Madhur Anand. 2024. Tipping Points in Coupled Human–Environment System Models: A Review. *Earth System Dynamics* 15, 4 (2024), 947–967.

[6] Matthew Fellows. 2021. *Bayesian and Variational Inference for Reinforcement Learning.* Ph.D. Dissertation. University of Oxford.

[7] Mattie Fellows, Brandon Kaplowitz, Christian Schroeder de Witt, and Shimon Whiteson. 2023. Bayesian Exploration Networks. *arXiv preprint arXiv:2308.13049* (2023).

[8] Gregory Flato, Jochem Marotzke, Babatunde Abiodun, Pascale Braconnot, Sin Chan Chou, William Collins, Peter Cox, Fatima Driouech, Seita Emori, Veronika Eyring, et al. 2014. Evaluation of Climate Models. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (2014), 741–866.

[9] Carlos E Garcia, David M Prett, and Manfred Morari. 1989. Model Predictive Control: Theory and Practice—A Survey. *Automatica* 25, 3 (1989), 335–348.

[10] Kenneth Gillingham, William D Nordhaus, David Anthoff, Geoffrey Blanford, Valentina Bosetti, Peter Christensen, Haewon McJeon, John Reilly, and Paul Sztorc. 2015. Modeling Uncertainty in Climate Change: A Multi-Model Comparison. *National Bureau of Economic Research Working Paper Series* w21637 (2015).

[11] Jianye Hao, Tianpei Yang, Hongyao Tang, Chenjia Bai, Jinyi Liu, Zhaopeng Meng, Peng Liu, and Zhen Wang. 2023. Exploration in Deep Reinforcement Learning: From Single-Agent to Multiagent Domain. *IEEE Transactions on Neural Networks and Learning Systems* (2023).

[12] Edward Hughes, Joel Z Leibo, Matthew Phillips, Karl Tuyls, Edgar Dueñez-Guzman, Antonio García Castañeda, Iain Dunning, Tina Zhu, Kevin McKee, Raphael Koster, et al. 2018. Inequity Aversion Improves Cooperation in Intertemporal Social Dilemmas. In *Proceedings of the 32nd Conference on Advances in Neural Information Processing Systems (NeurIPS'18).*

[13] Mike Hulme and John Turnpenny. 2004. Understanding and Managing Climate Change: the UK Experience. *Geographical Journal* 170, 2 (2004), 105–115.

[14] Alexander Lavin, David Krakauer, Hector Zenil, Justin Gottschlich, Tim Mattson, Johann Brehmer, Anima Anandkumar, Sanjay Choudry, Kamil Rocki, Atılım Güneş Baydin, et al. 2021. Simulation Intelligence: Towards a New Generation of Scientific Methods. *arXiv preprint arXiv:2112.03235* (2021).

[15] Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. 2017. Multi-Agent Reinforcement Learning in Sequential Social Dilemmas. In *Proceedings of the 16th Conference on Autonomous Agents and Multagent Systems (AAMAS'17).*

[16] Timothy M Lenton, Johan Rockström, Owen Gaffney, Stefan Rahmstorf, Katherine Richardson, Will Steffen, and Hans Joachim Schellnhuber. 2019. Climate Tipping Points—Too Risky to Bet Against. *Nature* 575, 7784 (2019), 592–595.

[17] Borja G Leon, Francesco Riccio, Kaushik Subramanian, Peter R Wurman, and Peter Stone. 2024. Discovering Creative Behaviors through DUPLEX: Diverse Universal Features for Policy Exploration. In *arXiv preprint.*

[18] Gunnar Luderer, Robert C Pietzcker, Christoph Bertram, Elmar Kriegler, Malte Meinshausen, and Ottmar Edenhofer. 2013. Economic Mitigation Challenges: How Further Delay Closes the Door for Achieving Climate Targets. *Environmental Research Letters* 8, 3 (2013), 034033.

[19] Kaveh Madani. 2013. Modeling International Climate Change Negotiations More Responsibly: Can Highly Simplified Game Theory Models Provide Reliable Policy Insights? *Ecological Economics* 90 (2013), 68–76.

[20] Anuj Mahajan, Tabish Rashid, Mikayel Samvelyan, and Shimon Whiteson. 2019. Maven: Multi-Agent Variational Exploration. In *Proceedings of the 33rd Conference on Advances in Neural Information Processing Systems (NeurIPS'19).*

[21] H Damon Matthews, Katarzyna B Tokarska, Zebedee RJ Nicholls, Joeri Rogelj, Josep G Canadell, Pierre Friedlingstein, Thomas L Frölicher, Piers M Forster, Nathan P Gillett, Tatiana Ilyina, et al. 2020. Opportunities and Challenges in Using Remaining Carbon Budgets to Guide Climate Policy. *Nature Geoscience* 13, 12 (2020), 769–779.

[22] Jean-Francois Mercure, Hector Pollitt, Andrea M Bassi, Jorge E Viñuales, and Neil R Edwards. 2016. Modelling Complex Systems of Heterogeneous Agents to Better Design Sustainability Transitions Policy. *Global environmental change* 37 (2016), 102–115.

[23] Deyu Ming, Daniel Williamson, and Serge Guillas. 2023. Deep Gaussian Process Emulation using Stochastic Imputation. *Technometrics* 65, 2 (2023), 150–161.

[24] William D Nordhaus. 2010. Economic Aspects of Global Warming in a Post-Copenhagen Environment. *Proceedings of the National Academy of Sciences* 107,

[25] William D Nordhaus. 2013. *The Climate Casino: Risk, Uncertainty, and Economics for a Warming World.* Yale University Press.

[26] William D Nordhaus. 2013. Integrated Economic and Climate Modeling. In *Handbook of Computable General Equilibrium Modeling.* Vol. 1. Elsevier, 1069–1131.

[27] William D Nordhaus. 2017. Revisiting the Social Cost of Carbon. *Proceedings of the National Academy of Sciences* 114, 7 (2017), 1518–1523.

[28] Brendan O'Donoghue. 2023. Efficient Exploration via Epistemic-Risk-Seeking Policy Optimization. In *Proceedings of the 40th International Conference on Machine Learning (ICML'23).*

[29] Ian Osband, John Aslanides, and Albin Cassirer. 2018. Randomized Prior Functions for Deep Reinforcement Learning. In *Proceedings of the 32nd Conference on Advances in Neural Information Processing Systems (NeurIPS'18).*

[30] Ian Osband, Yotam Doron, Matteo Hessel, John Aslanides, Eren Sezener, Andre Saraiva, Katrina McKinney, Tor Lattimore, Csaba Szepesvari, Satinder Singh, et al. 2020. Behaviour Suite for Reinforcement Learning. In *Proceedings of the 8th International Conference on Learning Representations (ICLR'20).*

[31] Ilona M Otto, Jonathan F Donges, Roger Cremades, Avit Bhowmik, Richard J Hewitt, Wolfgang Lucht, Johan Rockström, Franziska Allerberger, Mark McCaffrey, Sylvanus SP Doe, et al. 2020. Social Tipping Dynamics for Stabilizing Earth's Climate by 2050. *Proceedings of the National Academy of Sciences* 117, 5 (2020), 2354–2365.

[32] Camille Parmesan, Mike D Morecroft, and Yongyut Trisurat. 2022. *Climate Change 2022: Impacts, Adaptation and Vulnerability.* Ph.D. Dissertation. GIEC.

[33] Alexander Peysakhovich and Adam Lerer. 2018. Consequentialist Conditional Cooperation in Social Dilemmas with Imperfect Information. In *Proceedings of the 6th International Conference on Learning Representations (ICLR'18).*

[34] Robert S Pindyck. 2013. Climate Change Policy: What do the Models tell us? *Journal of Economic Literature* 51, 3 (2013), 860–872.

[35] Robert S Pindyck. 2017. The Use and Misuse of Models for Climate Policy. *Review of Environmental Economics and Policy* (2017).

[36] Martin L Puterman. 2014. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* John Wiley & Sons.

[37] Katharine L Ricke and Ken Caldeira. 2014. Maximum Warming Occurs About One Decade After a Carbon Dioxide Emission. *Environmental Research Letters* 9, 12 (2014), 124002.

[38] James Rudd-Jones, Fiona Thendean, and María Pérez-Ortiz. 2024. Crafting Desirable Climate Trajectories with RL Explored Socio-Environmental Simulations. *arXiv preprint arXiv:2410.07287* (2024).

[39] Lukas Schäfer, Oliver Slumbers, Stephen McAleer, Yali Du, Stefano V Albrecht, and David Mguni. 2023. Ensemble Value Functions for Efficient Exploration in Multi-Agent Reinforcement Learning. In *Proceedings of the Adaptive and Learning Agents Workshop (ALA 2023).*

[40] Lloyd S Shapley. 1953. Stochastic Games. *Proceedings of the National Academy of Sciences* 39, 10 (1953), 1095–1100.

[41] Theodore G Shepherd, Emily Boyd, Raphael A Calel, Sandra C Chapman, Suraje Dessai, Ioana M Dima-West, Hayley J Fowler, Rachel James, Douglas Maraun, Olivia Martius, et al. 2018. Storylines: an Alternative Approach to Representing Uncertainty in Physical Aspects of Climate Change. *Climatic Change* 151 (2018), 555–571.

[42] David Silver, Satinder Singh, Doina Precup, and Richard S Sutton. 2021. Reward is Enough. *Artificial Intelligence* 299 (2021), 103535.

[43] Will Steffen, Johan Rockström, Katherine Richardson, Timothy M Lenton, Carl Folke, Diana Liverman, Colin P Summerhayes, Anthony D Barnosky, Sarah E Cornell, Michel Crucifix, et al. 2018. Trajectories of the Earth System in the Anthropocene. *Proceedings of the National Academy of Sciences* 115, 33 (2018), 8252–8259.

[44] Nicholas Stern, Joseph Stiglitz, and Charlotte Taylor. 2022. The Economics of Immense Risk, Urgent Action and Radical Change: Towards New Approaches to the Economics of Climate Change. *Journal of Economic Methodology* 29, 3 (2022), 181–216.

[45] Peter Stone, Gal A Kaminka, Sarit Kraus, and Jeffrey S Rosenschein. 2010. Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination. In *Proceedings of the 24th Conference on Artificial Intelligence (AAAI'10).*

[46] Felix M Strnad, Wolfram Barfuss, Jonathan F Donges, and Jobst Heitzig. 2019. Deep Reinforcement Learning in World-Earth System Models to Discover Sustainable Management Strategies. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29, 12 (2019), 123122.

[47] Andrea Tacchetti, H Francis Song, Pedro AM Mediano, Vinicius Zambaldi, János Kramár, Neil C Rabinowitz, Thore Graepel, Matthew Botvinick, and Peter W Battaglia. 2019. Relational Forward Models for Multi-Agent Learning. In *Proceedings of the 7th International Conference on Learning Representations (ICLR'19).*

[48] UN. 2023. Integrated Assessment Models (IAMs) and Energy-Environment-Economy (E3) models | UNFCCC. https://unfccc.int/topics/mitigation/workstreams/response-measures/modelling-tools-to-assess-the-impact-of-the-implementation-of-response-measures/integrated-assessment-models-iams-and-energy-environment-economy-e3-models. (Accessed on 13/11/2024).

26 (2010), 11721–11726.

[49] Dirk-Jan van de Ven, Shivika Mittal, Ajay Gambhir, Robin D Lamboll, Haris Doukas, Sara Giarola, Adam Hawkes, Konstantinos Koasidis, Alexandre C Köberle, Haewon McJeon, et al. 2023. A Multimodel Analysis of Post-Glasgow Climate Targets and Feasibility Challenges. *Nature Climate Change* 13, 6 (2023), 570–578.

[50] Frederick van der Ploeg and Armon Rezai. 2019. Simple Rules for Climate Policy and Integrated Assessment. *Environmental and Resource Economics* 72, 1 (2019), 77–108.

[51] Bob CC van der Zwaan, Hilke Rösler, Tom Kober, Tino Aboumahboub, Katherine V Calvin, David EHJ Gernaat, Giacomo Marangoni, and David McCollum. 2013. A Cross-Model Comparison of Global Long-Term Technology Diffusion Under a 2 C Climate Change Control Target. *Climate Change Economics* 4, 04 (2013), 1340013.

[52] David G Victor, Frank W Geels, and Simon Sharpe. 2019. Accelerating the Low Carbon Transition. *The Case for Stronger, More Targeted and Coordinated International Action. Brookings Institution* (2019).

[53] Caroline Wang, Arrasy Rahman, Ishan Durugkar, Elad Liebman, and Peter Stone. 2024. N-Agent Ad Hoc Teamwork. *arXiv preprint arXiv:2404.10740* (2024).

[54] Jianhong Wang, Yang Li, Yuan Zhang, Wei Pan, and Samuel Kaski. 2024. Open Ad Hoc Teamwork with Cooperative Game Theory. *arXiv preprint arXiv:2402.15259* (2024).

[55] Rui Wang, Joel Lehman, Aditya Rawal, Jiale Zhi, Yulun Li, Jeffrey Clune, and Kenneth Stanley. 2020. Enhanced Poet: Open-Ended Reinforcement Learning through Unbounded Invention of Learning Challenges and their Solutions. In *Proceedings of the 37th International Conference on Machine Learning (ICML'20).*

[56] Martin L Weitzman. 2011. Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change. *Review of Environmental Economics and Policy* 5, 2 (2011), 275–292.

[57] John Weyant. 2017. Some Contributions of Integrated Assessment Models of Global Climate Change. *Review of Environmental Economics and Policy* (2017).

[58] Theodore Wolf, Nantas Nardelli, John Shawe-Taylor, and Maria Perez-Ortiz. 2023. Can Reinforcement Learning support policy makers? A preliminary study with Integrated Assessment Models. *arXiv preprint arXiv:2312.06527* (2023).

[59] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The Surprising Effectiveness of PPO in Cooperative Multi-Agent Games. In *Proceedings of the 36th Conference on Advances in Neural Information Processing Systems (NeurIPS'22).*

[60] Tianyu Zhang, Andrew Williams, Soham Phade, Sunil Srinivasa, Yang Zhang, Prateek Gupta, Yoshua Bengio, and Stephan Zheng. 2022. AI for Global Climate Cooperation: Modeling Global Climate Negotiations, Agreements, and Long-Term Cooperation in RICE-N. *arXiv preprint arXiv:2208.07004* (2022).

[61] Luisa Zintgraf, Sam Devlin, Kamil Ciosek, Shimon Whiteson, and Katja Hofmann. [n.d.]. Deep Interactive Bayesian Reinforcement Learning via Meta-Learning. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'21).*