InvestESG: A Multi-agent Reinforcement Learning Benchmark for Studying Climate Investment as a Social Dilemma

Xiaoxuan Hou[*1], **Jiayi Yuan**[*2], **Natasha Jaques**[2,3]
[1]Foster School of Business, University of Washington
[2]Paul G. Allen School of Computer Science and Engineering, University of Washington
[3]Google Deepmind
xxhou@uw.edu, jiayiy9@cs.washington.edu, jzl@google.com
nj@cs.washington.edu

Abstract

InvestESG is a novel multi-agent reinforcement learning (MARL) benchmark designed to study the impact of Environmental, Social, and Governance (ESG) disclosure mandates on corporate climate investments. Supported by both PyTorch and JAX implementation, the benchmark models an intertemporal social dilemma where companies balance short-term profit losses from climate mitigation efforts and long-term benefits from reducing climate risk, while ESG-conscious investors attempt to influence corporate behavior through their investment decisions, in a scalable and hardware-accelerated manner. Companies allocate capital across mitigation, greenwashing, and resilience, with varying strategies influencing climate outcomes and investor preferences. Our experiments show that without ESG-conscious investors with sufficient capital, corporate mitigation efforts remain limited under the disclosure mandate. However, when a critical mass of investors prioritizes ESG, corporate cooperation increases, which in turn reduces climate risks and enhances long-term financial stability. Additionally, providing more information about global climate risks encourages companies to invest more in mitigation, even without investor involvement. Our findings align with empirical research using real-world data, highlighting MARL's potential to inform policy by providing insights into large-scale socio-economic challenges through efficient testing of alternative policy and market designs.

1 Introduction

Climate change poses a persistent threat to the natural ecosystem and the global economy. Addressing climate change requires coordinated efforts, particularly from large corporations, which are reportedly responsible for over 70% of global industrial greenhouse gas emissions [32]. However, emissions mitigation presents a social dilemma [40], where the benefits of reduced emissions are shared globally, but the cost of reducing emissions reduces profits for individual companies [48, 15, 8]. As corporations are inherently self-interested, they are unlikely to reduce emissions voluntarily without external incentives or regulations.

Numerous policies have been proposed to address this challenge, among which mandatory Environmental, Social, and Governance (ESG) disclosure has recently ignited a vigorous global debate. The Securities and Exchange Commission's (SEC) proposal, which would require publicly traded companies to disclose climate-related risks and greenhouse gas emissions, attracted over 15,000 comments, making it one of the most contentious proposals in the SEC's history [56, 13]. This intense debate has led to an indefinite delay in the proposal's implementation [55]. Similar delays

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022.

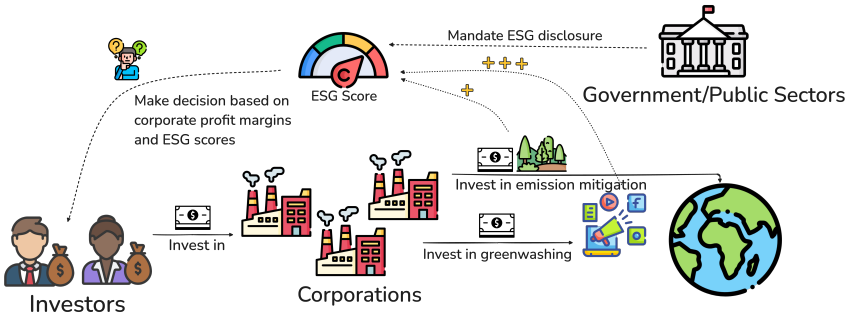


Figure 1: The **InvestESG** Environment. Corporations choose how much to invest in mitigating emissions, which affects their ESG Score. Climate-conscious investors can see ESG Scores when deciding how much to invest in each company. However, companies can engage in greenwashing to inexpensively and falsely improve ESG scores without actually mitigating climate change. InvestESG is a social dilemma, where selfish, profit-motivated corporations will not invest in mitigation without further incentives, leading to increased climate risks and decreased global wealth.

are also unfolding in the European Union and Korea [6, 38]. This highlights the need for thorough research to effectively inform the design and implementation of these policies.

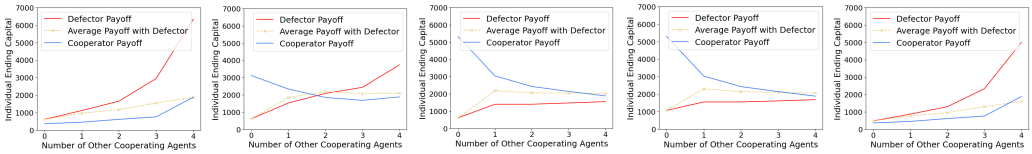
Traditional economics and policy research relies on either empirical analysis—which does not enable testing possible new policies [21, 42, 39]—or theoretical economics models, which are often limited to scenarios with only two agents (e.g., Friedman et al. 29), or single-period games (e.g., Pástor et al. 50). In contrast, Multi-Agent Reinforcement Learning (MARL) enables simulating complex interactions between multiple agents over extended time periods, under diverse hypothesized policy settings. Leveraging MARL to address large-scale socio-economic questions is a growing field [33]. Prior work has demonstrated the potential of MARL to design effective taxation schemes that enhance both equality and productivity [66], highlighting its relevance for tackling social challenges.

We propose using multi-agent reinforcement learning (MARL) to explore the impact of the ESG disclosure policy. We introduce **InvestESG**, an open-source MARL benchmark, to examine how profit-driven corporations balance short-term profits with long-term climate investments and whether ESG-informed investor choices influence corporate behavior. The simulation involves two agent types: companies and investors. Companies allocate funds to mitigation, greenwashing, and resilience, while investors choose investment portfolios based on their preferences for financial returns and ESG benefits. Our experiments with a state-of-the-art MARL implementation yield findings that align with real-world empirical evidence and provide novel insights.

More broadly, we demonstrate the potential of using a MARL framework to inform policy debates in the field of climate change. Our aim is not to claim that our model fully represents real-world complexity. It remains a "first-principles" model; however, it captures aspects that are challenging to integrate into a single framework using traditional economic approaches. To further improvement our environment setup to accommodate different use cases, we plan to seek advice from climate change research experts within the CCAI community. We present InvestESG as a challenging benchmark for the machine learning community; for researchers that are interested in developing MARL algorithms that can solve social dilemmas, InvestESG represents a social dilemma environment that could inform policymaking. We will provide the code for the benchmark and agent baselines in open-source.

2 The InvestESG Environment and Experiments

In this study, we introduce InvestESG, a MARL benchmark environment to analyze the impact of ESG disclosure mandates on corporate investment in climate efforts. The simulation involves two types of agents: companies and investors. At each time step, companies decide how much investment to allocate toward mitigation, which incurs higher short-term costs, thus reducing profits, and affects their attractiveness to investors. However, these investments reduce emissions and lower the overall climate risk for society, leading to improved long-term financial performance for all entities. Investors derive utility from two main metrics: (1) the profits generated by the companies they invest in, and (2) the increased utility from investing in ESG-friendly firms depending on their preference for ESG investments. Investors allocate their capital based on their knowledge of each company's total capital,



(a) 0/3 Conscious Investors (b) 2/3 Conscious Investors (c) 3/3 Conscious Investors (d) 3/3 Conscious & Resilience Spending (e) 3/3 Conscious & Greenwashing

Figure 2: Schelling diagrams demonstrating that the environment constitutes a social dilemma. The graphs compare payoffs between cooperation (mitigation, blue lines) and defection (no mitigation, red lines) for a focal company, given varying number of other cooperating companies. Yellow lines represent the average payoff across all companies when the focal company defects. Subfigure (a) illustrates the selfish scenario, where all three investors consistently prioritize financial returns ($\alpha^{\mathcal{I}_j} = 0,$ for $j = 1, 2, 3$). Here, defection always yields higher payoffs for the focal company than cooperation, leading all companies to defect. However, widespread defection results in lower overall profits, as the average payoff (yellow) increases with greater cooperation. Subfigure (b) and (c) correspond to two and three infinitely ESG-conscious investors ($\alpha^{\mathcal{I}_j} \approx \infty$), respectively. In (b), cooperation yields higher payoffs than defection for the focal company when few others cooperate. In (c), cooperation outperforms defection in all cases. Therefore, (b-c) demonstrate how investor behavior can transform the environment, eliminating the social dilemma by aligning corporate incentives with mitigation. Subfigures (d) and (e) build on (c) with three ESG-conscious investors. Subfigure (d) introduces resilience spending, while (e) adds greenwashing. The latter reintroduces a social dilemma, where corporations again avoid mitigation.

past returns, climate risk exposure, and ESG score [1] from the previous step. The environment is implemented in both PyTorch and JAX.

Climate change as a social dilemma. Companies' trade-off between short-term private costs and long-term collective benefits create a social dilemma within the environment. To illustrate this, we use empirical Schelling diagrams [34], which plot the focal-company payoffs from following either a cooperative or defecting policy, depending on the number of other cooperating company agents in a 5-company, 3-investor environment. In our case, cooperation represents a policy which consistently invests 0.5% of capital in mitigation and 0 in greenwashing or resilience building, while a defector policy takes zero action in all of mitigation, greenwashing, and resilience. Figure 2a to 2c reveal that the environment constitutes a social dilemma when investors are profit-motivated, which can be mitigated as more investors become ESG-conscious. However, if companies can defect by greenwashing, where companies can portray themselves as climate friendly at a low cost without genuine mitigation, the environment reverts back to a social dilemma.

Experiments with state-of-the-art MARL baselines suggest highly ESG-conscious investors are required to effectively incentivize mitigation efforts. We train the state-of-the-art MARL algorithm, IPPO [19], to learn to optimize the company and investor behavior. By default, 5 companies are modeled alongside 3 investors, with a total beginning market wealth of $98 trillion, which is comparable to the global stock market cap [61]. All the results are averaged result based on 3 runs of different random seeds, with the standard error plotted in the shaded region. Our results are consistent when scaled up to 25 companies and 25 investors as shown in the Appendix.

To evaluate the effectiveness of ESG disclosure mandates in incentivizing emissions mitigation by companies, we conduct three experiments: (1) *Status Quo*: All agents are profit-motivated, and no ESG scores are released; (2) *Status Quo with Mandate*: ESG scores are disclosed, but investors remain profit-driven ($\alpha = 0$); (3) *Mandate with ESG-Conscious Investors*: ESG scores are disclosed, and investors are ESG-conscious ($\alpha > 0$). In all scenarios, companies only choose mitigation effort levels, with greenwashing and resilience spending disabled. Figure 3 tracks evaluation metrics over training. The final system climate risks for the *Status Quo with Mandate* and low-ESG-conscious scenarios ($\alpha = 0.5$ and 1) are similar to those in the *Status Quo*, indicating that mandatory ESG disclosure alone does not significantly incentivize mitigation when investors prioritize profits. However, when investors have a high level of ESG-consciousness ($\alpha = 10$), companies invest significantly in mitigation, reducing climate risk and increasing market wealth. In this setting, a few leading companies attract the majority of ESG-conscious investments, forming a positive feedback loop

[1] The ESG score represents the fraction of capital that companies appear to allocate toward mitigation efforts, including both genuine mitigation and greenwashing investments. The concept of greenwashing will be introduced later in the discussion.

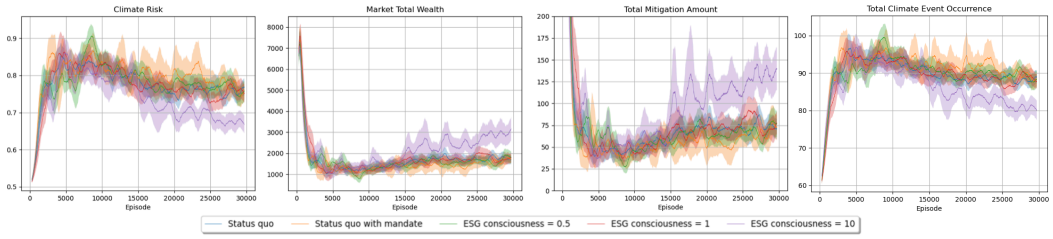


Figure 3: Climate risk, total market wealth, total mitigation amount and climate event occurrence over the course of training for the 5-company-3-investor case. We compare the status quo scenario with solely profit-driven investors (investors with ESG consciousness level of 0), both with and without the ESG disclosure mandate, to scenarios involving three ESG-conscious investors with ESG consciousness level of $\alpha = 0.5$, $\alpha = 1$, and $\alpha = 10$. These results indicate that merely disclosing ESG scores is insufficient to resolve the social dilemma if investors are not interested in investing in climate-friendly companies. Only when investors have a high level of ESG consciousness—$\alpha = 10$, in this case—does the ESG mandate make a difference in increasing the level of mitigation.

where better market performance draws further investments. These findings align with research showing that ESG disclosure mandates encourage climate-friendly efforts driven by societal and stakeholder pressures [14, 30, 25, 28, 11, 35], and that fund managers are willing to sacrifice financial returns for ESG benefits [39].

In addition to the main results presented above, our results further show that (1) heterogeneous investor preferences lead to divergence in agent strategies for both company and investor agents; (2) the possibility of greenwashing does not significantly undermine mitigation efforts; and (3) providing additional information about climate risk increases mitigation efforts. We present these results in more details in the Appendix. Beyond the findings under these default model settings, we also explore settings where (1) company agents are allowed to spend on building their own resilience, (2) company agents are initialized with real-world company actions, (3) agents' decisions are locked in for five years to simulate capital flexibility challenges, (4) companies' climate resilience parameter $L_t^{\mathcal{C}_i}$ are random and vary across events and companies, and (5) a stricter bankruptcy mechanism. These experiments yield directional conclusions consistent with those presented in the main text. Detailed results can be found in the Appendix.

3 Impact and Relevance

We demonstrate the potential of using a MARL framework to inform policy debates in the field of climate change. Importantly, our results are compatible broadly with the existing theoretical and empirical literature that examines this question, which shows its validity in predicting directional behaviors of rational decision makers. Assessing the effectiveness of a policy is inherently challenging, due to the fact that policy experiments are often prohibitively expensive and impractical to conduct, and even when they are feasible, it can be extremely time-consuming. Given the urgency of addressing climate change, our work provides a new vector for studying this problem, creating a simulated environment where a broad range of regulations can be explored and tested efficiently to provide novel insights into the problem. **For policymakers**, InvestESG shows various policy insights, e.g., mandatory ESG disclosure, paired with highly ESG-conscious investors, can drive corporate mitigation efforts. **For economics and policy researchers**, InvestESG introduces MARL as a promising tool to complement traditional empirical and theoretical methods, allowing scalable policy testing in a simulated environment. Our model predicts agent behaviors consistent with empirical evidence and uncovers novel insights. **For the machine learning community**, InvestESG presents a novel multi-agent benchmark, fostering the development of RL algorithms that tackle complex social dilemmas, competition, and long-term strategy—pushing forward AI applications in real-world, high-impact domains. We encourage the machine learning community to develop algorithmic innovations for InvestESG that can inspire practical actions to address climate change.

To further improvement our environment setup to accommodate different use cases, we plan to seek advice from climate change research experts within the CCAI community. Additionally, we plan to organize a competition where participants can submit solutions for designing regulations or incentives aimed at achieving an overall reduction in global emissions. For the competition, we will involve climate experts to evaluate the solutions for their real-world feasibility.

References

[1] John P. Agapiou, Alexander Sasha Vezhnevets, Edgar A. Duéñez-Guzmán, Jayd Matyas, Yiran Mao, Peter Sunehag, Raphael Köster, Udari Madhushani, Kavya Kopparapu, Ramona Comanescu, DJ Strouse, Michael B. Johanson, Sukhdeep Singh, Julia Haas, Igor Mordatch, Dean Mobbs, and Joel Z. Leibo. Melting pot 2.0, 2023. URL https://arxiv.org/abs/2211.13746.

[2] Amir Amel-Zadeh and George Serafeim. Why and how investors use esg information: Evidence from a global survey. *Financial analysts journal*, 74(3):87–103, 2018.

[3] Matteo Bettini, Ryan Kortvelesy, and Amanda Prorok. Controlling behavioral diversity in multi-agent reinforcement learning. *arXiv preprint arXiv:2405.15054*, 2024.

[4] Matteo Bettini, Amanda Prorok, and Vincent Moens. Benchmarl: Benchmarking multi-agent reinforcement learning. *Journal of Machine Learning Research*, 25(217):1–10, 2024.

[5] Alexander Bisaro and Jochen Hinkel. Governance of social dilemmas in climate change adaptation. *Nature Climate Change*, 6(4):354–359, 2016.

[6] Bloomberg News. EU to Delay ESG Reporting Rule for Some Sectors by Two Years, 2024. URL https://www.bloomberg.com/news/articles/2024-02-07/eu-to-delay-esg-reporting-rule-for-some-sectors-by-two-years?embedded-checkout=true.

[7] Frances E Bowen. Environmental visibility: a trigger of green organizational response? *Business strategy and the environment*, 9(2):92–107, 2000.

[8] James M Buchanan and Wm Craig Stubblebine. Externality. In *Inframarginal Contributions to Development Economics*, pages 55–73. World Scientific, 2006.

[9] Barbara Buchner. Annual finance for climate action surpasses usd 1 trillion, but far from levels needed to avoid devastating future losses, 2023. URL https://www.climatepolicyinitiative.org/press-release/annual-finance-for-climate-action-surpasses-usd-1-trillion-but-far-from-levels-needed-to-av Accessed 27-11-2024.

[10] Micah Carroll, Rohin Shah, Mark K. Ho, Thomas L. Griffiths, Sanjit A. Seshia, Pieter Abbeel, and Anca Dragan. On the utility of learning about humans for human-ai coordination, 2020. URL https://arxiv.org/abs/1910.05789.

[11] Yi-Chun Chen, Mingyi Hung, and Yongxiang Wang. The effect of mandatory csr disclosure on firm profitability and social externalities: Evidence from china. *Journal of accounting and economics*, 65(1):169–190, 2018.

[12] Soo-Haeng Cho, Xin Fang, Sridhar Tayur, and Ying Xu. Combating child labor: Incentives and information disclosure in global supply chains. *Manufacturing & Service Operations Management*, 21(3):692–711, 2019.

[13] CNBC. The sec votes this week on controversial climate change rule: Here's what's at stake, 2024. URL https://www.cnbc.com/2024/03/04/the-sec-votes-this-week-on-controversial-climate-change-rule-heres-whats-at-stake.html.

[14] Denis Cormier and Michel Magnan. Corporate environmental disclosure strategies: determinants, costs and benefits. *Journal of Accounting, Auditing & Finance*, 14(4):429–451, 1999.

[15] Carl J Dahlman. The problem of externality. *The journal of law and economics*, 22(1):141–162, 1979.

[16] Aswath Damodaran. Historical returns on stocks, bonds, and bills, 2024. URL https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/histretSP.html. New York University Stern School of Business.

[17] Aswath Damodaran. Capital expenditures by sector, 2024. URL https://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/capex.html. Accessed 27-11-2024.

[18] Sebastião Vieira de Freitas Netto, Marcos Felipe Falcão Sobral, Ana Regina Bezerra Ribeiro, and Gleibson Robert da Luz Soares. Concepts and forms of greenwashing: A systematic review. *Environmental Sciences Europe*, 32:1–12, 2020.

[19] Christian Schroeder De Witt, Tarun Gupta, Denys Makoviichuk, Viktor Makoviychuk, Philip HS Torr, Mingfei Sun, and Shimon Whiteson. Is independent learning all you need in the starcraft multi-agent challenge? *arXiv preprint arXiv:2011.09533*, 2020.

[20] Magali A Delmas and Michael W Toffel. Organizational responses to environmental demands: Opening the black box. *Strategic management journal*, 29(10):1027–1055, 2008.

[21] Anil R Doshi, Glen WS Dowell, and Michael W Toffel. How firms respond to mandatory information disclosure. *Strategic Management Journal*, 34(10):1209–1231, 2013.

[22] Ishan P. Durugkar, Clemens Rosenbaum, Stefan Dernbach, and Sridhar Mahadevan. Deep reinforcement learning with macro-actions, 2016. URL https://arxiv.org/abs/1606.04615.

[23] Javier El-Hage. Fixing esg: Are mandatory esg disclosures the solution to misleading esg ratings? *Fordham J. Corp. & Fin. L.*, 26:359, 2021.

[24] European Investment Bank. What drives firms' investment in climate action? evidence from the 2022-2023 eib investment survey. Technical report, European Investment Bank, 2023. URL https://www.eib.org/en/publications/20230114-what-drives-firms-investment-in-climate-change. Accessed: 2024-11-26.

[25] Eugene F Fama and Kenneth R French. Disagreement, tastes, and asset prices. *Journal of financial economics*, 83(3):667–689, 2007.

[26] J Doyne Farmer, Cameron Hepburn, Penny Mealy, and Alexander Teytelboym. A third wave in the economics of climate change. *Environmental and Resource Economics*, 62:329–357, 2015.

[27] Peter Fiechter, Jörg-Markus Hitz, and Nico Lehmann. Real effects of a widespread csr reporting mandate: Evidence from the european union's csr directive. *Journal of Accounting Research*, 60(4):1499–1549, 2022.

[28] Henry L Friedman and Mirko S Heinle. Taste, information, and asset prices: Implications for the valuation of csr. *Review of Accounting Studies*, 21:740–767, 2016.

[29] Henry L Friedman, Mirko Stanislav Heinle, and Irina Luneva. A theoretical framework for esg reporting to investors. *Available at SSRN 3932689*, 2021.

[30] Ramin Gamerschlag, Klaus Möller, and Frank Verbeeten. Determinants of voluntary csr disclosure: empirical evidence from germany. *Review of managerial science*, 5:233–262, 2011.

[31] David Gardiner and Associates. Nearly half of fortune 500 companies engaged in major climate initiatives. https://www.dgardiner.com/fortune-500-climate-initiatives-2023/#:~:text=There%20are%20now%20239%20Fortune,of%20the%20U.S.%20Fortune%20500, 2023. Accessed: 2024-11-26.

[32] Paul Griffin and CR Heede. The carbon majors database. *CDP carbon majors report 2017*, 14, 2017.

[33] Uri Hertz, Raphael Koster, Marco Janssen, and Joel Z Leibo. Beyond the matrix: Experimental approaches to studying social-ecological systems. 2023.

[34] 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. Inequity aversion improves cooperation in intertemporal social dilemmas. *Advances in neural information processing systems*, 31, 2018.

[35] Ioannis Ioannou and George Serafeim. Corporate sustainability: A strategy? *Harvard Business School Accounting & Management Unit Working Paper*, (19-065), 2019.

[36] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Çaglar Gülçehre, Pedro A. Ortega, DJ Strouse, Joel Z. Leibo, and Nando de Freitas. Intrinsic social motivation via causal influence in multi-agent RL. *CoRR*, abs/1810.08647, 2018. URL http://arxiv.org/abs/1810.08647.

[37] Basak Kalkanci and Erica L Plambeck. Managing supplier social and environmental impacts with voluntary versus mandatory disclosure to investors. *Management Science*, 66(8):3311–3328, 2020.

[38] Korea Economic Daily. Korea to enhance esg disclosures as part of corporate sustainability efforts, 2023. URL https://www.kedglobal.com/esg/newsView/ked202310160022. Accessed: 2024-08-22.

[39] Philipp Krueger, Zacharias Sautner, Dragon Yongjun Tang, and Rui Zhong. The effects of mandatory esg disclosure around the world. *Journal of Accounting Research*, 2021.

[40] Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas. *arXiv preprint arXiv:1702.03037*, 2017.

[41] Joel Z Leibo, Edgar A Dueñez-Guzman, Alexander Vezhnevets, John P Agapiou, Peter Sunehag, Raphael Koster, Jayd Matyas, Charlie Beattie, Igor Mordatch, and Thore Graepel. Scalable evaluation of multi-agent reinforcement learning with melting pot. In *International conference on machine learning*, pages 6187–6199. PMLR, 2021.

[42] Jun Li and Di Wu. Do corporate social responsibility engagements lead to real environmental, social, and governance impact? *Management Science*, 66(6):2564–2588, 2020.

[43] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30, 2017.

[44] Antoine Marot, Benjamin Donnot, Gabriel Dulac-Arnold, Adrian Kelly, Aïdan O'Sullivan, Jan Viebahn, Mariette Awad, Isabelle Guyon, Patrick Panciatici, and Camilo Romero. Learning to run a power network challenge: a retrospective analysis. *CoRR*, abs/2103.03104, 2021. URL https://arxiv.org/abs/2103.03104.

[45] Christopher Marquis, Michael W Toffel, and Yanhua Zhou. Scrutiny, norms, and selective disclosure: A global study of greenwashing. *Organization Science*, 27(2):483–504, 2016.

[46] V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, editors. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Intergovernmental Panel on Climate Change, 2021. In Press.

[47] MSCI Inc. Msci esg ratings, 2024. URL https://www.msci.com/sustainable-investing/esg-ratings. Accessed: 2024-09-19.

[48] Mancur Olson Jr. *The Logic of Collective Action: Public Goods and the Theory of Groups, with a new preface and appendix*, volume 124. harvard university press, 1971.

[49] Climate Impact Partners. Fortune global 500 climate commitments, 2024. URL https://www.climateimpact.com/news-insights/fortune-global-500-climate-commitments/. Accessed: 2024-11-26.

[50] L'uboš Pástor, Robert F Stambaugh, and Lucian A Taylor. Sustainable investing in equilibrium. *Journal of financial economics*, 142(2):550–571, 2021.

[51] A Mitchell Polinsky and Steven Shavell. Mandatory versus voluntary disclosure of product risks. *The Journal of Law, Economics, & Organization*, 28(2):360–379, 2012.

[52] H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, and B. Rama, editors. *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022. doi: 10.1017/9781009325844.

[53] Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021. URL http://jmlr.org/papers/v22/20-1364.html.

[54] Mikayel Samvelyan, Tabish Rashid, Christian Schroeder De Witt, Gregory Farquhar, Nantas Nardelli, Tim GJ Rudner, Chia-Man Hung, Philip HS Torr, Jakob Foerster, and Shimon Whiteson. The starcraft multi-agent challenge. *arXiv preprint arXiv:1902.04043*, 2019.

[55] SEC. The enhancement and standardization of climate-related disclosures for investors; delay of effective dates, 2024. URL https://www.federalregister.gov/documents/2024/04/12/2024-07648/the-enhancement\protect\penalty\z@-and-standardization-of-climate-related-disclosures-for-investors\protect\penalty\z@-delay-of-effective. Accessed: 2024-08-22.

[56] SEC. Sec adopts rules to enhance and standardize climate-related disclosures for investors, Mar 2024. URL https://www.sec.gov/news/press-release/2024-31.

[57] Yin Shi and Xiaoni Li. An overview of bankruptcy prediction models for corporate firms: A systematic literature review. *Intangible Capital*, 15(2):114–127, 2019. ISSN 1697-9818. doi: 10.3926/ic.1354. URL https://www.intangiblecapital.org/index.php/ic/article/view/1354.

[58] P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, and J. Malley, editors. *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022. doi: 10.1017/9781009157926.

[59] Alfonso Siano, Agostino Vollero, Francesca Conte, and Sara Amabile. "more than words": Expanding the taxonomy of greenwashing after the volkswagen scandal. *Journal of business research*, 71:27–37, 2017.

[60] S&P Global. S&p global esg scores, 2024. URL https://www.marketplace.spglobal.com/en/datasets/s-p-global-esg-scores-(171). Accessed: 2024-09-19.

[61] WFE Statistics. Market Capitalisation Q3 2023. *World Federation of Exchanges*, 2023. URL https://focus.world-exchanges.org/articles/market-capitalisation-q3-2023. Accessed: 2024-09-27.

[62] Yue Wu, Kaifu Zhang, and Jinhong Xie. Bad greenwashing, good greenwashing: Corporate social responsibility and information transparency. *Management Science*, 66(7):3095–3112, 2020.

[63] Jiachen Yang, Ang Li, Mehrdad Farajtabar, Peter Sunehag, Edward Hughes, and Hongyuan Zha. Learning to incentivize other learning agents. *CoRR*, abs/2006.06051, 2020. URL https://arxiv.org/abs/2006.06051.

[64] Zhi Yang, Thi Thu Huong Nguyen, Hoang Nam Nguyen, Thi Thuy Nga Nguyen, and Thi Thanh Cao. Greenwashing behaviours: Causes, taxonomy and consequences based on a systematic literature review. *Journal of business economics and management*, 21(5):1486–1507, 2020.

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

[66] Stephan Zheng, Alexander Trott, Sunil Srinivasa, David C. Parkes, and Richard Socher. The AI economist: Optimal economic policy design via two-level deep reinforcement learning, 2021. URL `https://arxiv.org/abs/2108.02755`.

4 Appendix

4.1 Related Work

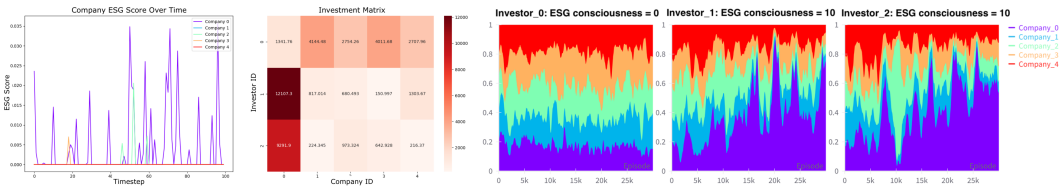
Creating a benchmark environment that analyzes the interplay between corporations and investors and their impacts on climate change mitigation requires various domain knowledge and connects multiple streams of studies. We closely examined three streams of literature.

Conventional economic methods are limited by either generalizability or tractability. Existing ESG disclosure related research in economics, business, and public policy rely on either empirical data [21, 42, 39] or simplified theoretical models [51, 37, 12, 50, 29]. Empirical analyses, while grounded in real-world data, struggle with generalizability and testing counterfactual policies. Theoretical models provide formalized equilibria and explore counterfactuals but are limited by tractability, modeling either multiple agents in single-period games (e.g., Pástor et al. 50) or only two agents over limited time periods (e.g., Friedman et al. 29). By proposing a MARL framework, our method overcomes these limitations by enabling the simulation of complex, multi-agent systems over extended time horizons under diverse policy settings, which allows for emergent behaviors and better captures complex socio-economic dynamics among diverse agents [33].

Current MARL benchmarks and social dilemma environments were not designed to model specific policy problems. Various MARL benchmarks have been created to study multi-agent coordination and cooperation; however, they are often limited to simplistic particle simulations [43] or videogames [1, 10, 54, 4, 3], which have little direct real-world implication. Sequential social dilemmas (SSD) [40] are spatially and temporally extended multi-agent environments in which the payoff to an individual agent for defecting is higher, but if all agents defect the payoff is lower. SSDs go beyond traditional game-theoretic environments like Prisoner's Dilemma because the complexity of the solving the SSD depends on not only addressing the misalignment between individual and collective rationality, but doing so when the negative consequences of short-sighted actions may take a long time to manifest. Prior research has examined methods to promote cooperation in SSDs by incorporating inequity aversion in agents [34], rewarding agents for influencing others' actions [36], enabling agents to provide incentives to others [63], and controlling for behavioral diversity [3]. Many of these studies are inspired by factors that drive human cooperation in social dilemmas. However, much of this research has been conducted in environments with no direct real-world implications (e.g. Leibo et al. 40, 41). In contrast, our environment directly addresses the problem of climate change, and is designed with critical trade-offs around a certain policy-making question. We hope to encourage researchers interested in addressing social dilemmas to focus on an environment with potential impact on the problem of climate change [5].

RL benchmarks can be effective in focusing AI research on climate change issues. Learning to Run a Power Network (L2RPN) is a single-agent RL benchmark focused on improving power grid efficiency [44]. This work has spawned multiple competitions, and it is currently hosted by the Electric Power Research Institute (EPRI) in conjunction with several other energy companies, government agencies, and universities[2]. L2RPN continues to foster innovative and meaningful collaboration across institutions, showing the potential impact of this type of simplified, simulated RL benchmark. However, L2RPN is focused on single-agent RL, whereas we explore a multi-agent, multi-party social dilemma. The only other MARL climate benchmark we are aware of was proposed by [65], and is focused on studying the dynamics of international climate negotiations, by incorporating an integrated assessment model simulating global climate and economic systems. In contrast, our environment, InvestESG, was designed with a focus on a more targeted policy question

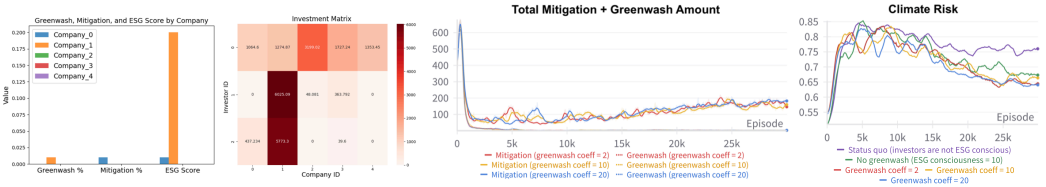
[2] `https://www.epri.com/l2rpn`



(a) Investment matrix for investors with different level of ESG consciousness: 0, 10, 10

(b) Investors with different level of ESG consciousness make different investment decisions.

Figure 4: Investigating the effects of the level of ESG consciousnessness in the case of 5 companies and 3 investors. ESG consciouness levels are $\alpha^{\mathcal{I}_0} = 0$, $\alpha^{\mathcal{I}_1} = \alpha^{\mathcal{I}_2} = 10$. In (a) and (b), Company 0 is the leading mitigator out of all the 5 companies. The figure plots the investment distribution for each investor, showing that more climate-conscious investors focus on investing in the more climate-conscious companies, mirroring the market bifurcation results of the Schelling diagrams.



(a) ESG-conscious investors can be distracted by greenwashing.

(b) Mitigation, greenwashing, and climate risk for greenwashing coefficient = 2, 10, 20.

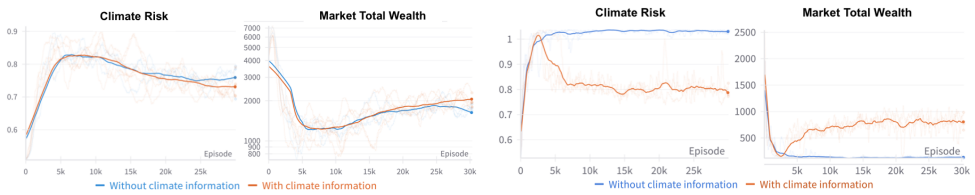
Figure 5: (a) shows ESG-conscious investors can be distracted by greenwashing, and heavily invest in a company that greenwashes. Here we examine a scenario where Company 0 is hard-coded to invest in real mitigation while Company 1 only invests in greenwashing. Investors have ESG consciousness level of 0, 1, 10. Part (b) shows that when both companies and investors are IPPO agents, regardless of greenwashing cost, companies initially explore greenwashing and mitigation equally but quickly abandon greenwashing and invest in mitigation to attract ESG-conscious investors. In these experiments, the final climate risk is similar to the value when greenwashing is not enabled, suggesting greenwashing does not hinder mitigation investment.

regarding ESG disclosure mandates, making our approach more directly applicable to an ongoing and highly debated policy issue.

4.2 Additional Results

Heterogeneous investor preference. Research shows that investors have varying preferences for ESG efforts and respond in different ways [2]. In this scenario, we initialized three investors with different levels of ESG-consciousness, with Investor 0 representing the solely profit-seeking investor with ESG-preference set to $\alpha^{\mathcal{I}_0} = 0$, and Investor 1 and 2 representing highly ESG-conscious investors with $\alpha^{\mathcal{I}_1} = \alpha^{\mathcal{I}_2} = 10$, which matches the scenario presented in the Schelling diagram in Figure 2b. Figure 4a and 4b highlight a divergence in both investor and company behavior. Profit-driven investor 0 distributes investments evenly across companies, while ESG-focused investors (1 and 2) favor climate-conscious firms, such as Company 0, which prioritizes mitigation. Such divergence extends to company strategies: Company 0 learns to attract more ESG investment by focusing on mitigation, while others prioritize financial returns at the expense of some investor interest.

Option to greenwash. Building on research that examines the existence and extent of greenwashing [62, 45, 59], and the concern that the ESG disclosure policy can backfire with greenwashing [23], we explore whether companies adjust their strategies when greenwashing is permitted. Based on the Schelling diagram in Figure 2d and Figure 4b, we predict that investors would be misled by greenwashing, prompting companies to prioritize greenwashing over mitigation due to its lower cost, leading to waste of resources. However, this prediction is not observed when simulating agents with IPPO. Figure 5b plots results under varying greenwashing costs, represented by coefficient β in Equation 3, where a larger β indicates cheaper greenwashing. All investors have an ESG-consciousness level of $\alpha = 1$. The results show that regardless of greenwashing cost, companies initially explore greenwashing and mitigation equally but quickly drop greenwashing as the main strategy.



(a) More information for investors and companies (b) More information for companies (no investors)

Figure 6: Effect of providing more information about climate risk to both investors and companies in the default 5-company-3-investor case (a), or companies only in a 5-company-0-investor case (b). These results show that simply providing more information about climate risk to companies can help them coordinate to increase mitigation efforts.

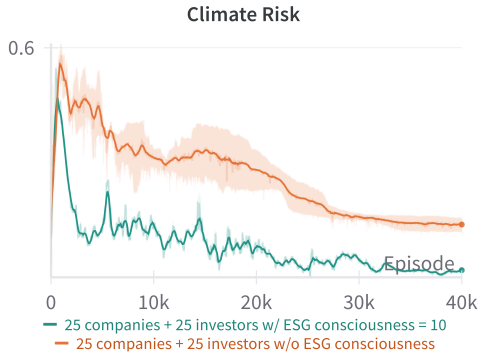
This likely occurs because, in the early episodes, investors have not yet linked ESG scores to their investment strategies, so do not respond to greenwashing efforts with increased investment. Instead, greenwashing presents an immediate cost for companies, which does not pay off with reduced climate risk, so they quickly learn to avoid it. In contrast, even without immediate investor rewards, companies may recognize the long-term benefits of mitigation (which leads to higher collective returns), and be incentivized to continue those efforts at first, even if they later learn to defect. This mirrors reality, where if investors are slow to adjust their strategies, companies may abandon greenwashing early but persist in low levels of mitigation for either its long-term climate benefits or the foresight of regulations and societal pressure that will eventually nudge them towards sustainable operation anyway. This is consistent with some empirical literature stating that despite the emergence of some greenwashing, ESG disclosure mandates overall encourage mitigation [27].

Additional information. In this scenario, we want to observe the effect of having additional climate-related information available on mitigation behavior. Therefore, we provide the climate event probability and climate event occurrences as additional information in the observation space for both companies and investors. In the default 5-company-3-investor scenario, having additional climate-related information reduces the ending system climate risks, as shown in Figures 6a. Figure 6b show similar effects of having additional climate-related information in the scenario where investors are not present. This result suggests that climate-related information helps both companies and investors make better environmental decisions. Even in the absence of ESG-focused investors, educating companies about the overall system improves outcomes. This aligns with literature showing that raising awareness encourages positive corporate responses [20, 7].

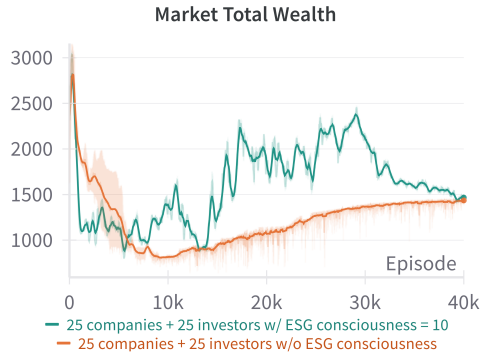
Scale up the number of agents. Figure 7a - 7b show the experiment results when the number of company and investor agents are scaled up to 25-by-25. The increased number of agents reveal the same directional story as the main results shown in Figure 3, where highly ESG-conscious investors motivate mitigation efforts from companies, resulting in lower ending climate risk and higher total market wealth. Figure 7c-7d zoom into a single episode of the 10-company, 10-investor case, with investor ESG-consciousness set to 0 and 10, respectively. In the ideal case where all investors are highly ESG-conscious, one leading mitigating company attracts the majority of the investment, which is consistent with the pattern shown in Figure 4a.

Effects of resilience spending. In this case, we allow companies to invest in resilience spending to improve their own robustness to climate events, without affecting the global climate risk. Although investors continue to respond to companies' ESG scores, resilience spending is excluded from the calculation of ESG scores. As shown in Figure 8a, when resilience is allowed, companies invest significantly more in resilience than in mitigation. By investing in resilience, companies are more resilient to the climate event and therefore are able to maintain more capital to invest in actual mitigation, compared to the case when resilience is not allowed, reflected in 8c. Therefore mitigation spending when resilience is allowed is actually slightly higher compared to when it is not an option, resulting in comparable final climate risk in Figure 8b. This suggests that companies are financially incentivized to prioritize resilience investments, which can actually enable a greater commitment to climate mitigation efforts.

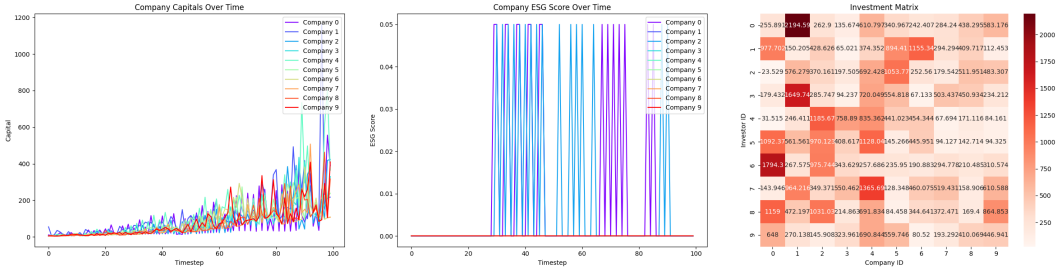
Seeding companies with real-world data. To ground the agents in real-world data, we seeded company agents' actions with real-world corporate behaviors. According to [31, 24, 49], approximately 50% of large companies are currently investing in climate mitigation. Globally, about 1% of GDP is allocated to climate finance annually [9]. Since GDP can be roughly viewed as the counterpart



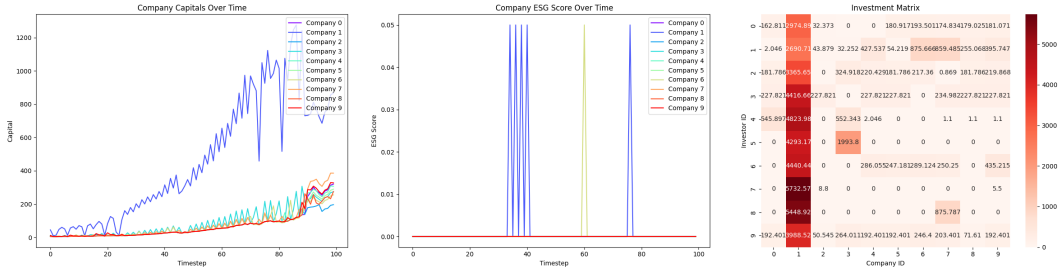
(a) Ending climate risk of 25 companies and 25 investors.



(b) Ending total market wealth of 25 companies and 25 investors.



(c) Company capital over time, company ESG score over time, and investment matrix of 10 companies and 10 investors with ESG consciousness = 0 in one episode.

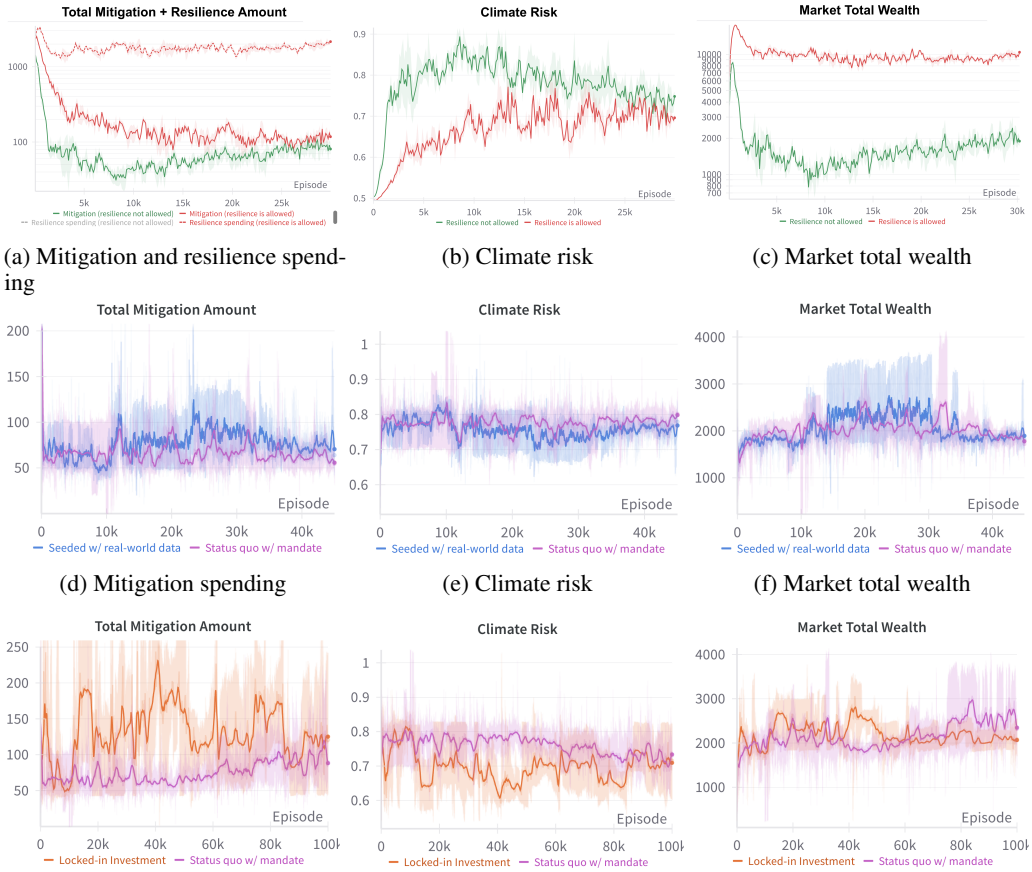


(d) Company capital over time, company ESG score over time, and investment matrix of 10 companies and 10 investors with ESG consciousness = 10 in one episode.

Figure 7: (a)(b) shows the final climate risk and market total wealth for the case of 25 companies and 25 investors. Similar to the default 5-company-and-3-investor case, when investors are highly conscious, the final climate risk would be decreased. (c)(d) shows the ending episode company capitals over time, company ESG score over time, and investment matrix of environments with 10 companies. When investors are ESG conscious, the investments are more concentrated on the mitigating company compared to the case when the investors are not ESG conscious. Consequently, the capitals of mitigating companies are much larger compared to non-mitigating companies in the case when the investors are ESG conscious, while the capitals of mitigating companies are comparable to non-mitigating companies when the investors are not ESG conscious.

of a company's sales, and based on data from an NYU database [17], which estimates the average sales-to-capital ratio across sectors to be between 0.8 and 1.28, we approximate sales and capital to be of similar magnitudes. Consequently, we seeded 50% of companies to invest between 0.5% and 1% of their total capital into mitigation, reflecting real-world investment levels. As shown in Figures 8d, 8e, and 8f, when seeded with real data, the total corporate mitigation efforts and resulting climate risk levels eventually align closely with the baseline scenario.

Lock-in investments. In reality, the decision-making processes of both companies and investors can be less flexible than modeled, where companies and investors update their strategies annually. To reflect the capital inflexibility, we implemented a 5-year lock-in period for agent decisions. This



(a) Mitigation and resilience spending (b) Climate risk (c) Market total wealth

(d) Mitigation spending (e) Climate risk (f) Market total wealth

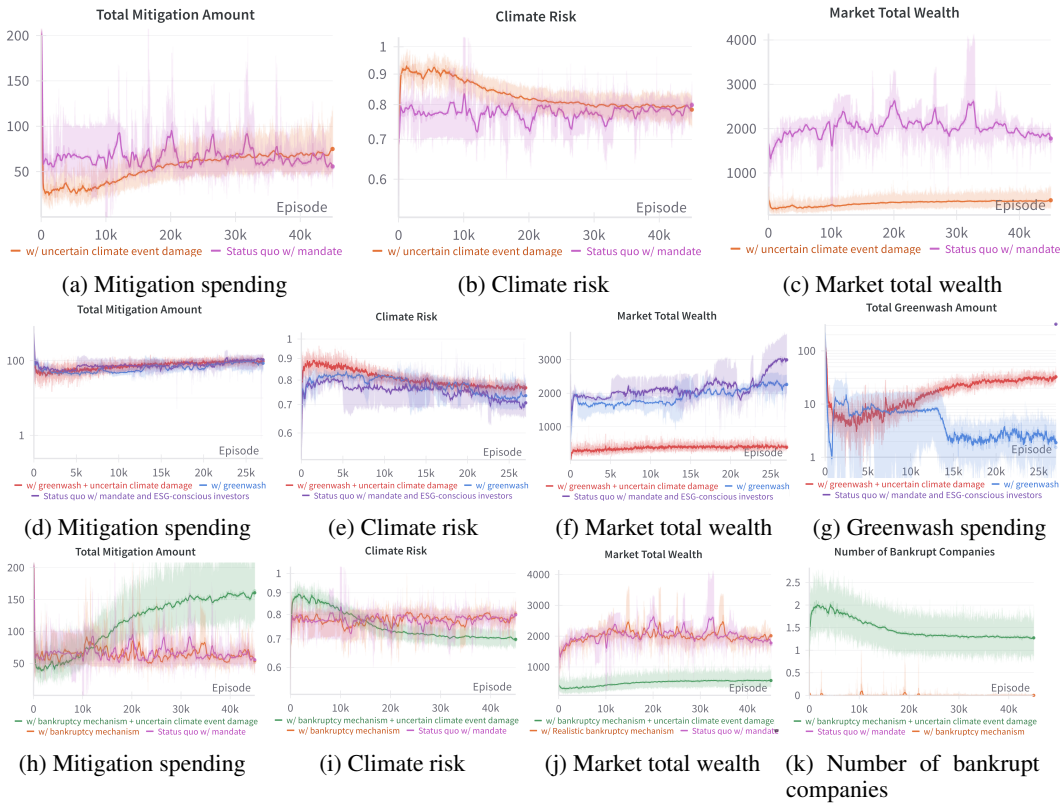
(g) Mitigation spending (h) Climate risk (i) Market total wealth

Figure 8: (a)-(c) Effect of allowing resilience spending. (d)-(f) Effect of seeding with real-world data. (g)-(i) Effect of 5-year lock-in period for agent decisions. All comparisons are made against the default case with an ESG disclosure mandate, in which investors have zero ESG-consciousness, and company acitions are restricted to mitigation efforts only.

approach allows the climate to evolve more rapidly than agents can respond. Despite this constraint, the results remain consistent with the directional findings presented in the main text, as illustrated in Figures 8g, 8h and 8i, suggesting that agents would achieve similar results using a macro-action reinforcement learning approach [22]. However, the learning curves for the locked-in investment cases exhibit greater volatility and slower convergence compared to the default case, indicating that capital inflexibility indeed increases the challenges of addressing climate change.

Uncertain climate event damage. Given the significant uncertainty surrounding the economic damages of climate change [26], we conducted additional experiments where companies' resilience parameters, $L_t^{\mathcal{C}_i}$, representing economic losses from extreme climate events, were modeled as Gaussian random variables $L_t^{\mathcal{C}_i} \sim \mathcal{N}(\mu, \sigma)$ clipped within the range [0,1], varying across both events and companies. This randomness allows for extreme climate damages, including the potential bankruptcy of some companies, which could incentivize risk-averse strategies where companies engage in greater mitigation efforts. Conversely, the randomness complicates the ability of company agents to learn the long-term benefits of mitigation, potentially encouraging short-sighted behaviors or greenwashing.

Figures 9a to 9c illustrate the effects of uncertain climate event damage compared to the default mandate case, where no investors are ESG-conscious, and company actions are restricted to mitigation only. Both scenarios result in similar levels of total mitigation and climate risk. However, agents in the uncertain damage scenario require more time to learn the optimal mitigation level due to the increased uncertainty in the environment. Despite the similarities, the uncertain damage scenario



(a) Mitigation spending (b) Climate risk (c) Market total wealth

(d) Mitigation spending (e) Climate risk (f) Market total wealth (g) Greenwash spending

(h) Mitigation spending (i) Climate risk (j) Market total wealth (k) Number of bankrupt companies

Figure 9: (a)-(c) Effect of uncertain climate event damage. (d)-(g) Effect of uncertain climate event damage on greenwash spending. (h)-(k) Effect of a more strick bankruptcy mechanism and its combination with uncertain climate event damage.

results in significantly lower total market wealth due to potentially some high-tail losses in early years. When considering the wealth discrepancy, companies in the uncertain damage scenario ultimately adopt a more aggressive mitigation strategy relative to their capital. This behavior suggests that the high uncertainty may have prompted risk-averse strategies.

Uncertain climate event damage also incentivizes the companies to focus more on short-term benefits by attracting investments, compared to the case with fixed damage when the companies first explore greenwashing strategy and then settle down on doing little greenwashing. As shown in 9d to 9g, when threatened with uncertain climate event damage, companies increase their greenwashing amount to attract immediate investments from ESG-conscious investors.

Stricter bankruptcy mechanism. Additionally, to make our model more enriched, we implemented a more strict bankruptcy mechanism for company agents: if a company agent has a margin worse than negative 10% for 3 consecutive years, a red flag signaling potential financial distress, is deemed as bankrupt [57]. As shown in the comparison between orange and pink lines in 9h, 9i and 9j, the new bankruptcy mechanism does not cause significant difference in the level of mitigation from the status quo with mandate case. Given that the market growth rate is high, companies are not severely threatened by bankruptcy despite that the new bankruptcy mechanism is in place, and therefore refrain from investing in climate mitigation.

However, when the stricter bankruptcy mechanism is combined with uncertain climate event damage, as depicted by the green lines in Figures 9h and 9k, a significant increase in mitigation and more bankruptcies are observed. This indicates that the combination of uncertain climate event damage and the bankruptcy mechanism creates an immediate risk of bankruptcy for company agents, thereby strongly incentivizing their mitigation efforts.

4.3 Technical details of the InvestESG environment

The simulation starts in 2021 and runs through 2120, with each period t corresponding to one year. The environment includes two main components: (1) an evolving climate and economic system, and (2) two types of agents: M company agents $\mathcal{C}_i$ for $i \in \{1, \ldots, M\}$ and N investor agents $\mathcal{I}_j$ for $j \in \{1, \ldots, N\}$. **Climate and Economic Dynamics.** The environment is characterized by three climate risk parameters: extreme heat probability (P_t^h), heavy precipitation probability (P_t^p), and drought probability (P_t^d) in year t. Initial climate risks are set to $P_0^h = 0.28$, $P_0^p = 0.13$, and $P_0^d = 0.17$ following the IPCC estimate [46], resulting in an overall climate risk of $P_0 = 0.48$, the probability of at least one adverse climate event in a year. Without mitigation efforts, these risks increase linearly over time, reaching the IPCC's 4 scenario scenario by 2100, which corresponds to the 80th period in our 100-period simulation. By that point, we observe $\overline{P}_{80}^h = 0.94$, $\overline{P}_{80}^p = 0.27$, and $\overline{P}_{80}^d = 0.41$ (or an overall climate risk of $\overline{P}_{80} = 0.97$). We then extrapolate this trend through to period 100. Figure 10a depicts how increased climate risks and adverse climate events increase over time in a scenario where companies are solely profit-motivated. Company agents can mitigate the growth of climate risk by investing in emissions reduction. The change in climate risk P_t^e for event $e \in \{h, p, d\}$ in year t is governed by the function:

$$P_t^e = \frac{\mu_e t}{1 + \lambda_e U_{t,m}} + P_0^e, \quad \text{for } e \in \{h, p, d\}, \tag{1}$$

where $U_{t,m}$ is the cumulative mitigation spending from all agents by period t. If $U_{t,m} = 0$, risks increase linearly to reach $\overline{P}_{80}^e$ by 2100 as explained earlier. When $U_{t,m} > 0$, the growth rate of climate risk decreases. The parameters λ_e are calibrated based on [58], which estimates that an annual mitigation investment of \$2.3 trillion is required to achieve IPCC's 1.5 scenario. The model fits λ_e so that such investment levels would yield 1.5 scenario climate risks by 2100. Climate events are modeled as independent Bernoulli processes, allowing for multiple events within a year (red dashed lines in Figure 10a). Let $X_{t,h}, X_{t,p}, X_{t,e} \in \{0, 1\}$ represent the occurrence of each climate event, and X_t denote the total number of events in period t, determined by Equation 2.

$$X_t = X_{t,h} + X_{t,p} + X_{t,e}, \quad \text{where } X_{t,e} \sim \text{Bernoulli}(P_t^e), \quad \text{for } e \in \{h, p, d\}. \tag{2}$$

In addition to the evolving climate risks, the environment incorporates a baseline *economic growth rate* γ, set to 10% by default, aligned with the historical average annual return of the S&P 500 over the past century [16]. Company agents' capital levels $K_t^{\mathcal{C}_i}$ grow at rate γ each year, barring climate events. If an adverse event occurs, company agents lose a portion of their total capital according to their respective *climate resilience* parameter $L_t^{\mathcal{C}_i}$. If economic losses drive a company's remaining capital into negative territory, the company is declared bankrupt.

Company Action Space. Each company agent $\mathcal{C}_i$ in period t selects actions from a continuous vector $\mathbf{u}_t^{\mathcal{C}_i} = (u_{t,m}^{\mathcal{C}_i}, u_{t,g}^{\mathcal{C}_i}, u_{t,r}^{\mathcal{C}_i})$, where $u_{t,m}^{\mathcal{C}_i}$ represents the share of capital allocated to mitigation, $u_{t,g}^{\mathcal{C}_i}$ to greenwashing, and $u_{t,r}^{\mathcal{C}_i}$ to building climate resilience. The action space for each company is defined as a continuous 3-dimensional unit cube, $\mathcal{U}_t^{\mathcal{C}_i} = [0, 1]^3$. If the sum of the three ratios exceeds 1 in any period, the company agent is deemed to be overspending, resulting in bankruptcy. Mitigation spending directly reduces the system-wide climate risk as explained above. Greenwashing involves deceptive marketing or accounting tactics that allow companies to appear climate-friendly at a low cost without providing any real benefits to society [18, 64]. Resilience spending enhances the company's climate resilience by lowering its vulnerability $L_t^{\mathcal{C}_i}$, but it does not reduce emissions and therefore does not mitigate system-wide climate risk. In the default setting, we disable greenwashing and resilience spending to focus on testing companies' mitigation efforts. .

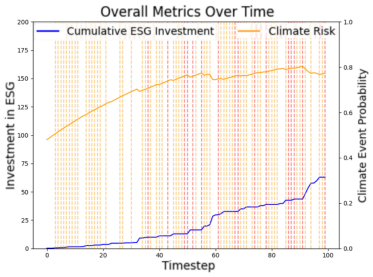
Investor Action Space. Investor agents first select the companies they want to invest in. Specifically, agent $\mathcal{I}_j$ in period t selects an action from a binary vector of length M, $\mathbf{a}_t^{\mathcal{I}_j} = (a_{t,1}^{\mathcal{I}_j}, \ldots, a_{t,M}^{\mathcal{I}_j})$, corresponding to the M company agents. Each entry $a_{t,i}^{\mathcal{I}_j} = 1$ indicates that investor $\mathcal{I}_j$ invests in company $\mathcal{C}_i$ in period t, and $a_{t,i}^{\mathcal{I}_j} = 0$ otherwise. The investor's action space at time t is thus $\mathcal{A}_t^{\mathcal{I}_j} = \{0, 1\}^M$. Once the choices are made, investors capitals are distributed equally among these chosen companies, as will be detailed later in Equation 4.

Modeling ESG Disclosure. With the ESG disclosure mandate in place, each company agent receives an updated ESG score $Q_{t+1}^{\mathcal{C}_i}$ in period t, calculated as

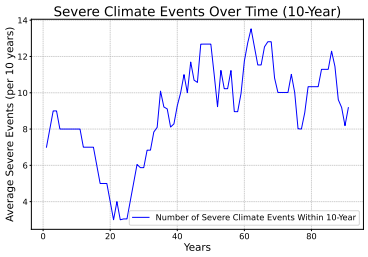
$$Q_{t+1}^{\mathcal{C}_i} = u_{t,m}^{\mathcal{C}_i} + \beta u_{t,g}^{\mathcal{C}_i}, \tag{3}$$

where $\beta > 1$ indicates that greenwashing is cheaper than genuine mitigation in terms of building an ESG-friendly image. This mirrors simple ESG ratings provided by some agencies, which typically range from 1 to 100 [60] or use letter-based ratings [47].

State and Observation Space. The environment simulates a partially observable Markov game $\mathcal{M}$ defined over a continuous, multi-dimensional state space. The system state at period t is characterized by the three climate risk parameters $\mathbf{P}_t = (P_t^h, P_t^p, P_t^d)$, each company agent's state vector $\mathbf{S}_t^{\mathcal{C}_i} = (K_t^{\mathcal{C}_i}, Q_t^{\mathcal{C}_i}, L_t^{\mathcal{C}_i})$, where $K_t^{\mathcal{C}_i}$ is the capital level, $Q_t^{\mathcal{C}_i}$ is the ESG score, and $L_t^{\mathcal{C}_i}$ is the climate resilience. Each investor agent's state is represented by their investment portfolio and cash levels $\mathbf{S}_t^{\mathcal{I}_j} = (H_{t,1}^{\mathcal{I}_j}, \ldots, H_{t,M}^{\mathcal{I}_j}, C_t^{\mathcal{I}_j})$, where $H_{t,i}^{\mathcal{I}_j}$ represents investor $\mathcal{I}_j$'s holdings in company $\mathcal{C}_i$, and $C_t^{\mathcal{I}_j}$ is the investor's cash level. The full system state at period t is thus $\mathcal{S}_t = \left(\mathbf{P}_t, \{\mathbf{S}_t^{\mathcal{C}_i}\}_{i=1}^M, \{\mathbf{S}_t^{\mathcal{I}_j}\}_{j=1}^N\right)$. All company and investor agents share a common observation space, denoted as $\mathcal{O}_t$. In the default setting, $\mathcal{O}_t = \left(\{\mathbf{S}_t^{\mathcal{C}_i}\}_{i=1}^M, \{\mathbf{S}_t^{\mathcal{I}_j}\}_{j=1}^N\right)$. Extensions that incorporate additional observable information, like climate risk, are explored in experiment results.



(a) Environment dynamics over the course of a single 100 year episode.



(b) Average number of severe climate event occurrences over 10-year period.

Figure 10: Status quo scenario where all agents are only profit-motivated. In (a), mitigation spending (blue curve) is minimal, leading climate risk (yellow curve) to increase over time. Adverse weather event occurrences are shown as dotted lines; red lines indicate multiple adverse events in a single year. In (b), increasing climate risk leads to more frequent occurrence of severe climate events over time.

State Transition. The environment's state transition $\mathcal{T}$ proceeds as follows. At the beginning of period t, investors collect their investment holdings from period $t-1$ and redistribute their capital according to $\mathbf{a}_t^{\mathcal{I}_j}$. Denote $||a_t||_1^{\mathcal{I}_j} = \sum_{i=1}^M a_{t,i}^{\mathcal{I}_j}$ as the number of companies investor $\mathcal{I}_j$ invests in during period t and let $K_t^{\mathcal{I}_j} = \sum_{i=1}^M H_{t,i}^{\mathcal{I}_j} + C_t^{\mathcal{I}_j}$ represent the total capital of investor $\mathcal{I}_j$ at the start of period t. Companies reach an interim capital level after returning old investments, $\sum_{j=1}^N H_{t,i}^{\mathcal{I}_j}$, and receiving new ones, with the investment from investor $\mathcal{I}_j$ calculated as $a_{t,i}^{\mathcal{I}_j} \frac{K_t^{\mathcal{I}_j}}{||a_t||_1^{\mathcal{I}_j}}$ or 0 if the investor opts out of investing, as shown in Equation 4.

$$K_{t+1,interim}^{\mathcal{C}_i} = K_t^{\mathcal{C}_i} - \sum_{j=1}^N H_{t,i}^{\mathcal{I}_j} + \sum_{j=1}^N a_{t,i}^{\mathcal{I}_j} \frac{K_t^{\mathcal{I}_j}}{||a_t||_1^{\mathcal{I}_j}}, \quad \text{for } i = 1, \ldots, M \tag{4}$$

Companies then make climate-related spending using the interim capital, as described in Equations 5 to 6. Here, $U_{t,m}$ represents the cumulative mitigation spending by all company agents up to period t, while $U_{t,r}^{\mathcal{C}_i}$ denotes the cumulative resilience spending by company $\mathcal{C}_i$.

$$U_{t,m} = U_{t-1,m} + \sum_{i=1}^M u_{t,m}^{\mathcal{C}_i} \times K_{t+1,interim}^{\mathcal{C}_i} \tag{5}$$

$$U_{t,r}^{\mathcal{C}_i} = U_{t-1,r}^{\mathcal{C}_i} + u_{t,r}^{\mathcal{C}_i} \times K_{t+1,interim}^{\mathcal{C}_i} \quad \text{for } i = 1, \ldots, M \tag{6}$$

While $U_{t,m}$ is then plugged into Equation 1 where system climate risks are updated. Equation 7 states that a company's climate resilience, $L_t^{\mathcal{C}_i}$, scales with the proportion of cumulative resilience investment relative to its capital, and that increasing resilience becomes progressively more challenging

due to diminishing returns [52].

$$L_t^{\mathcal{C}_i} = L_0^{\mathcal{C}_i} \exp\left(-\eta^{\mathcal{C}_i} \frac{U_{t-1,r}^{\mathcal{C}_i} + u_{t,r}^{\mathcal{C}_i} K_{t+1,interim}^{\mathcal{C}_i}}{K_{t+1,interim}^{\mathcal{C}_i}}\right), \quad \text{for } i = 1, \ldots, M. \tag{7}$$

At the same time, the occurrence of climate events are simulated according to Equation 2, and companies receive updated ESG scores based on Equation 3. Afterwards, companies' profit margins, $\rho_t^{\mathcal{C}_i}$, are computed using Equation 8, factoring in climate-related spendings $u_{t,m}^{\mathcal{C}_i}$, $u_{t,g}^{\mathcal{C}_i}$, $u_{t,r}^{\mathcal{C}_i}$, default economic growth γ, and losses due to climate events, which are influenced by companies' climate vulnerability $L_t^{\mathcal{C}_i}$ and the number of climate events X_t as defined in Equation 2:

$$\rho_t^{\mathcal{C}_i} = (1 - u_{t,m}^{\mathcal{C}_i} - u_{t,g}^{\mathcal{C}_i} - u_{t,r}^{\mathcal{C}_i})(1 + \gamma)(1 - X_t L_t^{\mathcal{C}_i}) - 1, \quad \text{for } i = 1, \ldots, M. \tag{8}$$

Company capitals $K_{t+1}^{\mathcal{C}_i}$, investor holdings $H_{t+1,i}^{\mathcal{I}_j}$, and investor cash positions $C_{t+1}^{\mathcal{I}_j}$ are updated according to Equations 9 to 11. Company capitals $K_{t+1}^{\mathcal{C}_i}$ are updated according to Equation 9 by scaling the interim capital levels by profit margin.

$$K_{t+1}^{\mathcal{C}_i} = (1 + \rho_t^{\mathcal{C}_i}) K_{t+1,interim}^{\mathcal{C}_i}, \quad \text{for } i = 1, \ldots, M. \tag{9}$$

Equation 10 adjusts investor holdings based on company profit margins in their portfolios.

$$H_{t+1,i}^{\mathcal{I}_j} = a_{t,i}^{\mathcal{I}_j} (1 + \rho_t^{\mathcal{C}_i}) \frac{K_t^{\mathcal{I}_j}}{||a_t||_1^{\mathcal{I}_j}}, \quad \text{for } i = 1, \ldots, M, j = 1, \ldots, N. \tag{10}$$

If an investor chooses not to invest, all capital remains as cash, as shown in Equation 11

$$C_{t+1}^{\mathcal{I}_j} = \begin{cases} 0, & \text{if } ||a_t||_1^{\mathcal{I}_j} \neq 0 \\ K_t^{\mathcal{I}_j}, & \text{if } ||a_t||_1^{\mathcal{I}_j} = 0 \end{cases}, \quad \text{for } j = 1, \ldots, N. \tag{11}$$

Rewards. The single-period reward for company $\mathcal{C}_i$ is solely based on its profit margin, given by $r_t^{\mathcal{C}_i} = K_{t+1}^{\mathcal{C}_i} - K_{t+1,interim}^{\mathcal{C}_i}$ for $i = 1, \ldots, M$, reflecting the assumption that companies are profit-driven. The reward for investor $\mathcal{I}_j$ is $r_t^{\mathcal{I}_j} = \frac{K_{t+1}^{\mathcal{I}_j} - K_t^{\mathcal{I}_j}}{K_t^{\mathcal{I}_j}} + \alpha^{\mathcal{I}_j} \frac{\sum_{i=1}^M H_{t+1,i}^{\mathcal{I}_j} Q_{t+1}^{\mathcal{C}_i}}{\sum_{i=1}^M K_{t+1}^{\mathcal{I}_j}}$ for $j = 1, \ldots, N$. The first component is the portfolio return ratio, and the second component represents the weighted average ESG score of the investor's portfolio adjusted by the investor's ESG preference, $\alpha^{\mathcal{I}_j}$.

Social Outcome Metrics. We evaluate agent performance based on two key social outcome metrics: the final climate risk level, P_{100}, defined as $P_{100} = 1 - (1 - P_{100}^h)(1 - P_{100}^p)(1 - P_{100}^d)$, and the total market wealth at the end of the period, W_{100}, defined as $W_{100} = \sum_{i=1}^M K_{100}^{\mathcal{C}_i} + \sum_{j=1}^N K_{100}^{\mathcal{I}_j}$.

4.4 Implementation details

4.5 Independent-PPO

To test how self-interest agents learn to respond to incentives in the environment, we employ a state-of-the-art MARL algorithm based on Independent PPO. Each agent has its own policy parameters, and agents do not share parameters among themselves. This is because we are interested in simulating companies and investors as independent, selfishly motivated agents that specialize in maximizing their own expected reward. The policy model is a simple Multi-Layer Perceptron (MLP) network, with input as the capital, resilience and margin of each company, along with the investments and capital of each investor. Additional information can be added to the observation. All company and investor agents share a common observation space. To implement Independent-PPO with different roles, we built upon the Stable Baseline 3 repository [53]. Each policy model is an MLP network with two layers of size 256 and 128 and with tanh activation layers, and update each policy after 5 episodes during the training. See Table 1 for more details.

In the default setting, we assign equal amount of initial capitals to companies and investors, which is a rough representation of the current market. For each experimental scenario, we run the learning algorithms for 30k episodes, over 3 trials with different random seeds.

MLP layers	Activation layers	PPO n_steps	PPO learning rate	PPO entropy coefficient	Gradient Clipping
256, 128	tanh	500	3e-5	0.01	0.2

Table 1: IPPO policy training parameters. The rest of the parameters are the default as described in Stable Baselines 3 [53].