INVESTESG: A MULTI-AGENT REINFORCEMENT LEARNING BENCHMARK FOR STUDYING CLIMATE INVESTMENT AS A SOCIAL DILEMMA

Anonymous authors
Paper under double-blind review

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. 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. 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 global stability, with droughts, storms, fires, and flooding becoming more intense and frequent (Christopher B Field & Dahe, 2012), leading to disruption of the natural ecosystem and significant impacts on the global economy. Addressing climate change requires coordinated efforts across multiple sectors, particularly from large corporations, which are reportedly responsible for over 70% of global industrial greenhouse gas emissions (Griffin & Heede, 2017). While adaptation—preparing for the inevitable consequences of climate change—tends to be party-specific and often driven by financial incentives, mitigation—reducing emissions—presents a de facto social dilemma (Leibo et al., 2017), where the benefits of reduced emissions are shared globally yet the costs are borne locally (Olson Jr, 1971; Dahlman, 1979; Buchanan & Stubblebine, 2006). 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 these, mandatory Environmental, Social, and Governance (ESG) disclosures have recently been hotly debated in the United States. The Securities and Exchange Commission's (SEC) proposal, which would require publicly traded companies to disclose climate-related risks, mitigation strategies, and greenhouse gas emissions from their operations, has attracted over 15,000 comments, making it one of the most contentious proposals in the SEC's history (SEC, 2024b; CNBC, 2024). This has resulted in an indefinite delay in enactment of the policy to allow for further discussion (SEC, 2024a). The U.S. is not alone in facing such pushback; similar delays are unfolding in the European Union and Korea (Bloomberg News, 2024; Korea Economic Daily, 2023). 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 (Doshi et al., 2013; Li & Wu, 2020; Krueger et al., 2021)—or theo-

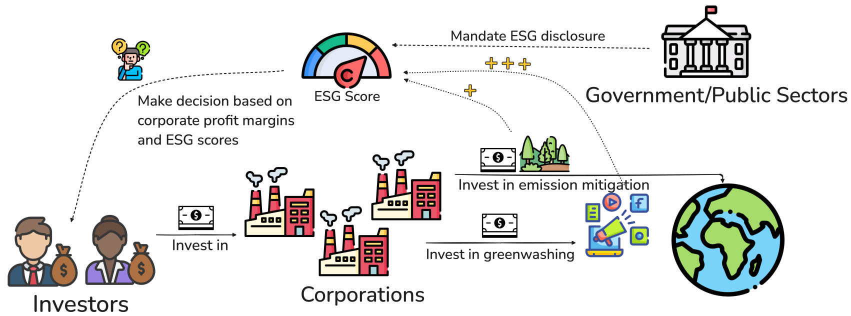


Figure 1: The **InvestESG** Environment. Corporations choose how much to invest in mitigating emissions, which affects their ESG Score. Climate-conscious investors can see ESG Scores when deciding how much to invest in each company. However, companies can engage in greenwashing to inexpensively and falsely 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.

retical economics models, which are often limited to scenarios with only two agents (e.g., Friedman et al. 2021), or single-period games (e.g., Pástor et al. 2021). 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 (Hertz et al., 2023). Prior work has demonstrated the potential of MARL to design effective taxation schemes that enhance both equality and productivity (Zheng et al., 2021), highlighting its relevance for tackling real-world 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 decide whom to invest in based on their preferences for financial returns versus ESG benefits. This creates an intertemporal social dilemma (Hughes et al., 2018), where both agent classes must weigh immediate and local costs against long-term and global gains. Using Schelling diagrams, we demonstrate that in a fully profit-driven environment, a social dilemma arises, but with sufficiently ESG-conscious investors with enough capital, climate mitigation becomes optimal for some or all corporations. However, if companies are able to *greenwash* to cheaply increase ESG scores without genuine mitigation, the environment once again becomes a social dilemma.

Our experiments with a state-of-the-art MARL implementation yield findings that align with real-world empirical evidence and provide novel insights. First, a sufficient number of highly ESG-conscious investors are needed to incentivize corporate mitigation efforts. In such cases, climate-focused companies emerge and attract most climate-conscious investments, while others prioritize profit maximization. When only some investors are climate-focused, the market bifurcates, with mitigating companies aligning with conscious investors. Additionally, sharing climate risk information helps companies to increase mitigation efforts, even when investors are not present.

More broadly, we demonstrate how a MARL framework can inform policy debates in the field of climate change. 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. However, the findings revealed by our environment match existing empirical results, suggesting the simulation has high ecological validity. 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. 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 have real-world impact. We will provide the code for the benchmark and agent baselines in open-source. We aim to show that MARL algorithmic design can inspire real-world actions that can be leveraged to tackle climate change.

2 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 (Doshi et al., 2013; Li & Wu, 2020; Krueger et al., 2021) or simplified theoretical models (Polinsky & Shavell, 2012; Kalkanci & Plambeck, 2020; Cho et al., 2019; Pástor et al., 2021; Friedman et al., 2021). 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. 2021) or only two agents over limited time periods (e.g., Friedman et al. 2021). 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 (Hertz et al., 2023).

Current MARL benchmarks and social dilemma environments were not designed to model real-world problems. Various MARL benchmarks have been created to study multi-agent coordination and cooperation; however, they are often limited to simplistic particle simulations (Lowe et al., 2017) or videogames (Agapiou et al., 2023; Carroll et al., 2020; Samvelyan et al., 2019), which have little direct real-world implication. Sequential social dilemmas (SSD) (Leibo et al., 2017) 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 (Hughes et al., 2018), rewarding agents for influencing others' actions (Jaques et al., 2018), and enabling agents to provide incentives to others (Yang et al., 2020a). 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. 2017; 2021). In contrast, our environment directly addresses the problem of climate change, and is benchmarked to real-world climate and economic parameters. We hope to encourage researchers interested in addressing social dilemmas to focus on an environment with potential impact on the problem of climate change.

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 (Marot et al., 2021). 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[1]. 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 multi-agent RL climate benchmark we are aware of was proposed by Zhang et al. (2022), 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 regarding ESG disclosure mandates, making our approach more directly applicable to an ongoing and highly debated policy issue.

3 The InvestESG Environment

InvestESG is a MARL benchmark environment designed to evaluate the impact of ESG disclosure mandates on corporate climate investments. We model the environment as an intertemporal social dilemma where company agents face short-term profit losses from climate mitigation efforts, but

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

these actions reduce long-term climate risks, benefiting all. 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 (Masson-Delmotte et al., 2021), 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 $4\,°\mathrm{C}$ 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 2 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:

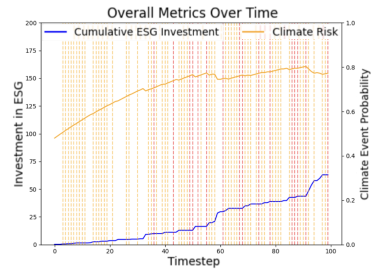


Figure 2: Environment dynamics over the course of a single 100 year episode, for the status quo scenario where all agents are only profit-motivated. Mitigation spending (blue curve) is minimal, leading climate risk (yellow curve) to increase over time. Adverse weather event occurrences are shown as dotted lines; red lines indicate multiple adverse events in a single year.

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

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 Shukla et al. (2022), which estimates that an annual mitigation investment of \$2.3 trillion is required to achieve IPCC's $1.5\,°\mathrm{C}$ scenario. The model fits λ_e so that such investment levels would yield $1.5\,°\mathrm{C}$ 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 2). 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 (Damodaran, 2024). 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 $L_t^{\mathcal{C}_i}$.

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 a negative capital balance and 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 (de Freitas Netto et al., 2020; Yang et al., 2020b). 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. These options are later enabled to examine their effects, as will become clear in Section 4.

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 (S&P Global, 2024) or use letter-based ratings (MSCI Inc., 2024).

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 $\mathcal{S}_t = \Big(\mathbf{P}_t, \{\mathbf{S}_t^{\mathcal{C}_i}\}_{i=1}^M, \{\mathbf{S}_t^{\mathcal{I}_j}\}_{j=1}^N\Big)$. All company and investor agents share a common observation space, denoted as $\mathcal{O}_t$. In the default setting, $\mathcal{O}_t = \Big(\{\mathbf{S}_t^{\mathcal{C}_i}\}_{i=1}^M, \{\mathbf{S}_t^{\mathcal{I}_j}\}_{j=1}^N\Big)$. Extensions that incorporate additional observable information, like climate risk, are explored in Section 4.

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 (Pörtner et al., 2022).

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

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:

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

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.

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

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

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

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

$$C_{t+1}^{\mathcal{I}_j} = K_t^{\mathcal{I}_j} - \sum_{i=1}^{M} H_{t+1,i}^{\mathcal{I}_j}, \quad \text{for } j = 1, \dots, N. \tag{11}$$

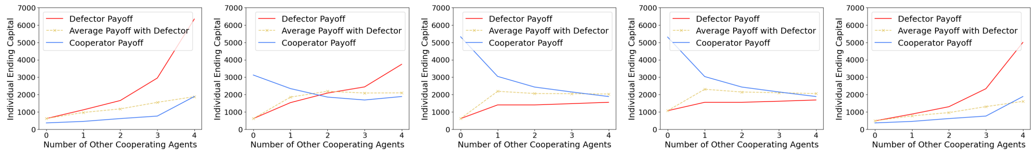
Rewards. The single-period reward for company $\mathcal{C}_i$ is solely based on its profit margin, given by $r_t^{\mathcal{C}_i} = K_{t+1}^{\mathcal{C}_i} - K_{t+1,interim}^{\mathcal{C}_i}$ for $i = 1, \dots, M$, reflecting the assumption that companies are profit-driven. The reward for investor $\mathcal{I}_j$ 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 $j = 1, \dots, N$. 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 ESGpreference, $\alpha^{\mathcal{I}_j}$.

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 $W_{100} = \sum_{i=1}^{M} K_{100}^{\mathcal{C}_i} + \sum_{j=1}^{N} K_{100}^{\mathcal{I}_j}$.

3.1 The trade-off for the agents frames climate change as a social dilemma

With companies balancing between short-term costs and long-term benefits, as well as self-interest and social welfare, and investors weighing investment returns against ESG preferences, these trade-offs create a social dilemma within the environment. To illustrate this, we use empirical Schelling diagrams (Hughes et al., 2018), which plot the focal-company payoffs from following either a co-operative or defecting policy, depending on the number of other cooperating company agents in a 5-company, 3-investor environment.

In our case, cooperation represents altruistic investment in mitigating emissions. We simulate a co-operative policy as one which consistently invests 0.5% of capital in mitigation and 0 in greenwashing or resilience building. In Figure 3a to 3c, a defector policy takes zero action in all of mitigation, greenwashing, and resilience. As shown in Figure 3a, when investors are solely profit-driven, the payoff for following a defector policy is always higher than the cooperator payoff. However, if all agents fail to cooperate, overall payoffs will be extremely low, showing that the environment represents a social dilemma, which by definition, involves a conflict between immediate self-interest and longer-term collective interests. Figures 3b and 3c illustrate how the payoffs for cooperation and defection evolve as the proportion of *ESG-conscious investors* with positive ESG preferences increases. Notably, Figure 3b suggests the possibility of a bifurcated equilibrium when investor preferences are mixed. Specifically, climate-friendly companies may attract ESG-conscious investors, while other companies free ride and attract profit-motivated investors. Figure 3c shows that when all investors prioritize ESG-friendly firms, cooperation dominates defection in all cases. Therefore, with enough ESG-conscious investors the environment is no longer a social dilemma, and companies are actually greedily motivated to invest in mitigation to attract investors.



(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 3: Schelling diagrams demonstrating that the environment constitutes a social dilemma. 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, the payoff for an individual company for defecting (not investing in reducing emissions) is always higher than cooperating (mitigating), so companies will selfishly choose to defect. However, if all companies defect, overall profits will be lower. Subfigure (b) and (c) correspond to two and three infinitely ESG-conscious investors ($\alpha^{\mathcal{I}_j} \approx \infty$), respectively. In (b), when few other companies cooperate (corresponding to points 0, and 1 on the x-axis), cooperation yields a higher individual payoff than defection for the focal company. In (c), cooperation outperforms defection in all cases. Therefore, (b-c) demonstrate that investor behavior can change the properties of the environment such that it no longer represents a social dilemma, in that companies may be greedily motivated to mitigate to attract investors. Subfigure (d) and (e) have the same setting as (c) with three ESG-conscious investors. In addition, (d) adds the option for resilience spending to (c). Subfigure (e) adds the option of greenwashing to (c), which once again converts the environment to a social dilemma in which corporations will no longer invest in mitigation.

Greenwashing and Climate Resilience Investment. Although Figure 3c appears promising, the option to greenwash and spend on resilience complicates the trade-offs companies face. In Figure 3e, we define the defection policy as investing 0.5% of capital annually in resilience. This makes defection slightly more attractive compared to Figure 3c, as resilience spending results in higher capital gains than taking no action. Figure 3d presents an alternative scenario where defecting companies invest 0.3% annually in greenwashing[2]. In this case, cooperation becomes significantly more challenging, because companies can attract ESG-conscious investors without actually mitigating their emissions. The addition of greenwashing turns the environmental back into a social dilemma.

3.2 Multi-agent Reinforcement Learning Baselines

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 (IPPO) (De Witt et al., 2020; Yu et al., 2022). 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.

By default, 5 companies are modeled with an initial capital of \$10 trillion each, alongside 3 investors, each starting with \$16 trillion. This creates a total beginning market wealth of \$98 trillion, which is comparable to the global stock market cap (WFE Statistics, 2023). We conduct the experiments in such a 5-company, 3-investor environment by default unless specified otherwise. All the following results are averaged result based on 3 runs of different random seeds, with the standard error plotted in the shaded region.

4 Results

The status quo environment presents a social dilemma. For all experiments, our evaluation metric is (1) ending system climate risk, P_{100}, and (2) ending market wealth, W_{100}. We first model the agents' behavior under the status quo where there are no ESG disclosure mandates. All the agents are solely profit-motivated and no ESG score is released. We monitor how the evaluation metrics evolve over the course of training in Figure 4. The company agents begin training by acting randomly, leading to more frequent mitigation. Total market wealth and total reward for individual agents is highest in the first 200 episodes, which is the optimal collective outcome. As the training progresses, the company agents learn to not mitigate in order to selfishly maximize their return. Conflict breaks out when agents discover that, in a situation of great climate risk, investing in mitigation

[2]For the Schelling diagram, we set the greenwashing coefficient $\beta = 2$. As a result, 0.3% spending on greenwashing leads to a higher ESG score than 0.5% spent on mitigation.

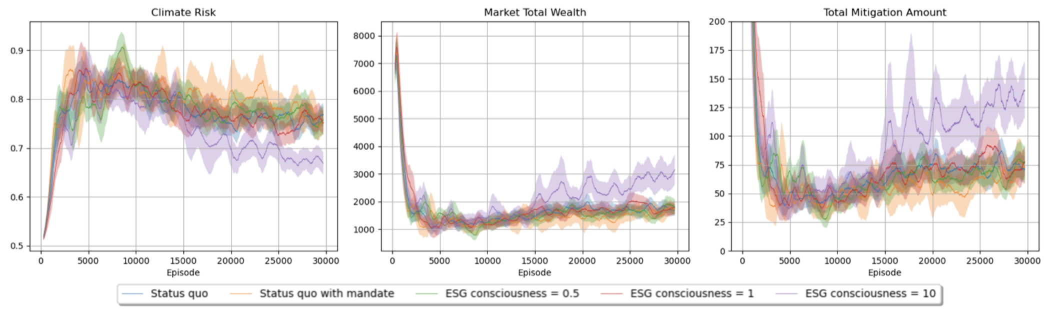
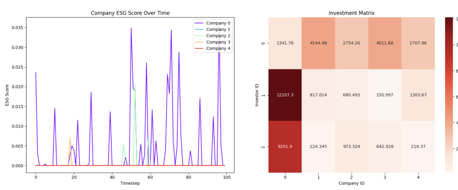
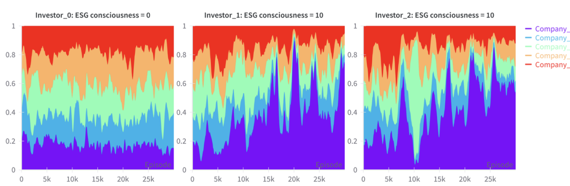


Figure 4: Climate risk, total market wealth, and total mitigation amount over the course of training for the 5-company-3-investor case. We compare the status quo scenario with solely profit-driven investors (investors with ESG consciousness level of 0), both with and without the ESG disclosure mandate, 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.



(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.

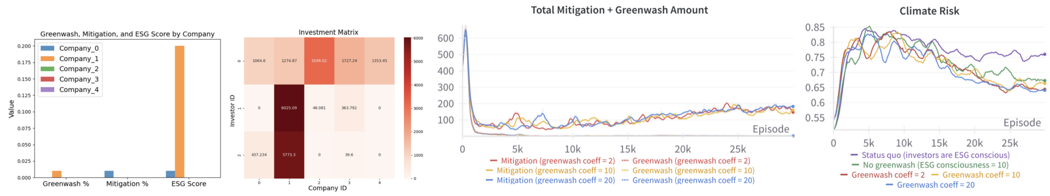
Figure 5: Investigating the effects of the level of ESG consciousnessness in the case of 5 companies and 3 investors. ESG conscioueness levels are $\alpha^{\mathcal{I}_0} = 0$, $\alpha^{\mathcal{I}_1} = \alpha^{\mathcal{I}_2} = 10$. In (a) and (b), Company 0 is the leading mitigator out of all the 5 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.

is marginally more beneficial than not, which is shown by a steady and slow increase in total market wealth, converging around \$2,000 trillion. However, the profit-motivated strategy for the individual agents achieves a sub-optimal collective outcome, as is characteristic of a social dilemma.

Does ESG score alone incentivize mitigation? As shown in Figure 4, the ending system climate risks for the status quo scenario is around 0.75. To model the effects of ESG disclosure mandate, we add each company's annual ESG score to agents' observation space, calculated according to Equation 3. The learning curve shown in Figure 4 reveals a similar level of climate risk, indicating that mandatory ESG disclosure alone does not significantly incentivize emission mitigation when investors care only about profits.

How conscious do investors need to be? Studies have found that ESG disclosure mandates promote climate-friendly efforts driven by societal and stakeholder pressures (see e.g., Cormier & Magnan 1999; Gamerschlag et al. 2011; Fama & French 2007; Friedman & Heinle 2016; Chen et al. 2018; Ioannou & Serafeim 2019), and that investment fund managers are willing to sacrifice financial returns for ESG benefits (Krueger et al., 2021). Therefore, we study the effects of investors with different level of ESG consciousness on the emission mitigation level. As shown in Figure 4, when all investors have a high level of ESG consciousness ($\alpha = 10$), we can see a significant difference where investors incentivize companies to invest in mitigation, lowering final climate risk and increasing market wealth. In this case, a few leading mitigating companies attract the majority of ESG-conscious investment while others remain inactive, as shown in Appendix Figure 9c. However, ESG consciousness levels of 0.5 and 1 (valuing climate investment and profits equally) are insufficient to resolve the dilemma.

What if investors have different levels of ESG-consciousness? Research shows that investors have varying preferences for ESG efforts and respond in different ways (Amel-Zadeh & Serafeim,



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

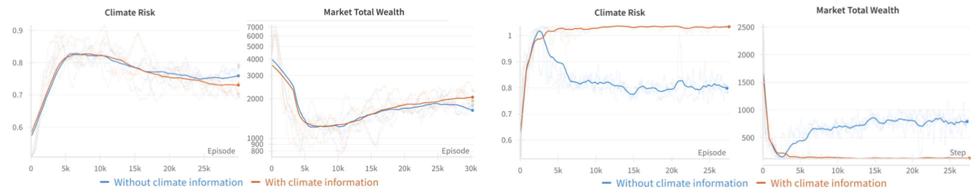
(b) Mitigation, greenwashing, and climate risk for greenwashing coefficient = 2, 10, 20.

Figure 6: (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.

2018). 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 3b. Figure 5a and 5b highlight a bifurcation 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. This bifurcation 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.

Will companies cheat with greenwashing? Building on research that examines the existence and extent of greenwashing (Wu et al., 2020; Marquis et al., 2016; Siano et al., 2017), and the concern that the ESG disclosure policy can backfire with greenwashing (El-Hage, 2021), we explore whether companies adjust their strategies when greenwashing is permitted. Based on the Schelling diagram in Figure 3d and Figure 5b, 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 6b 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 abandon greenwashing. 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 (Fiechter et al., 2022).

Can providing additional information about climate risk affect behavior? 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 7a. Figure 7b 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



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

Figure 7: 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.

improves outcomes. This aligns with literature showing that raising awareness encourages positive corporate responses (Delmas & Toffel, 2008; Bowen, 2000).

5 Discussion and Conclusion

In this paper, we introduce InvestESG, a MARL environment simulating long-term interactions between companies and investors under varying ESG disclosure policies. **For policymakers**, InvestESG shows that mandatory ESG disclosure, paired with informed, ESG-conscious investors, can drive corporate mitigation efforts. Additionally, providing high-quality system-wide information effectively motivates action from both corporations and investors. **For economics and policy researchers**, InvestESG introduces MARL as a powerful 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 example, greenwashing poses a lesser challenge than anticipated, and providing system-level climate risk information helps resolve the dilemma even without investor involvement. These findings position InvestESG as a promising, ecologically valid platform for testing more complex and nuanced policies. **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 real-world actions to address climate change.

Limitations. We acknowledge that InvestESG simplifies various aspects of real-world climate evolution, financial markets, and regulatory frameworks; for example, the cost of mitigation efforts varies by market sector and geographic location, and investor ESG preferences can fluctuate with market conditions and social trends. These simplifications were intentional to enable flexible experimentation on the core dynamics between companies and investors at this stage. However, we anticipate ongoing improvements, as outlined below.

Future Work. Our goal is to spark discussions and evolve InvestESG into a robust tool for policy design, with input from experts in machine learning, climate change, and economics. We envision three key areas for future work: **(1) Richer environment.** For instance, we plan to incorporate employees and customers, who may also be attracted to more climate-friendly companies. We also envision capturing complex market and environmental dynamics between industries, and enhancing investor decision-making to move beyond binary choices to better reflect real-world complexities. **(2) Enhanced policy design.** We propose testing policies that strengthen ESG disclosure mandates, such as guidelines on Scope 1 and 2 emissions, which, while costly, may ease the evaluation of companies' efforts. We also suggest exploring ways to redesign the simple ESG score to better inform investors and motivate corporate actions. **(3) Private-sector initiatives.** Given the challenges of passing new regulations, we aim to explore if companies and investors can self-regulate through coordination. For instance, we suggest testing if collective funds or peer monitoring can enhance mitigation or lead to collusion. We invite researchers interested in multi-agent learning to collaborate with us on advancing these ideas and turning InvestESG into a comprehensive policy tool.

6 Ethics Statement

This research is directly aimed at improving human well-being by addressing the existential threat posed by climate change, one of the most pressing issues of our time. Our work seeks to develop a fully reproducible, open-source multi-agent reinforcement learning (MARL) benchmark, enabling the highest standard of transparency and scientific rigor. The benchmark is designed to be freely accessible to all researchers, and performant enough to run efficiently on commonly available GPU resources, fostering inclusivity. By simulating real-world social dilemmas between corporations and investors, we aim to spark novel solutions for reducing emissions and mitigating climate change risks. No human subjects are involved in this research.

7 Reproducibility Statement

To promote reproducibility, we provide our source code and corresponding configuration files used for running the experiments as supplementary material. The details about installation and sample code for running the experiment is also included in the supplementary material. To encourage engagement and exploration from the machine learning community, we will open source the code for the InvestESG environment and baselines via Github.

References

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.

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.

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.

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

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

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.

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.

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.

Thomas F Stocker Christopher B Field, Vicente Barros and Qin Dahe. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2012.

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.

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

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

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.

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.

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.

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

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.

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

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

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.

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.

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

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

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

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

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.

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

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`.

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.

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.

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

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.

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*, pp. 6187–6199. PMLR, 2021.

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.

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.

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`.

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.

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 (eds.). *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.

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

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.

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

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

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 (eds.). *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.

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`.

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.

SEC. The enhancement and standardization of climate-related disclosures for investors; delay of effective dates, 2024a. 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.

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

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 (eds.). *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.

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.

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.

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.

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

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

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, 2020b.

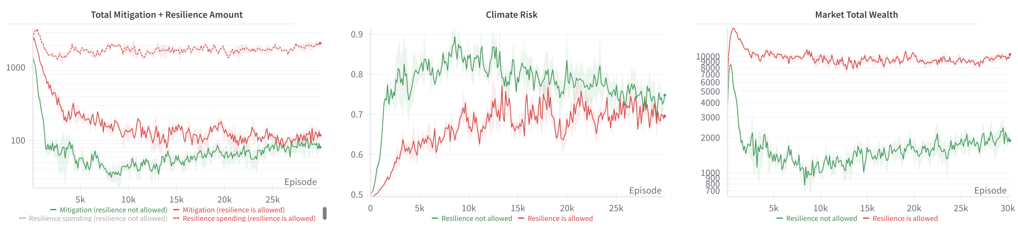
Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 35:24611–24624, 2022.

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.

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.

MLP layers	Activation layers	PPO n_steps	PPO learning rate	PPO entropy coefficient	Gradient Clipping
256, 128	tanh	500	3×10^{-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 (Raffin et al., 2021).



(a) Mitigation and resilience spending (b) Climate risk (c) Market Total Wealth

Figure 8: Effect of allowing resilience spending. Compared to the default case except that resilience spending is allowed.

In the Appendix, we will cover the following details of our work.

- **Implementation details** Appendix 8: Implementation details about multi-agent reinforcement learning experiments.
- **Additional experiments** Appendix 9: Details about additional experiment, including the impacts of resilience spending on the stakeholders' behavior and mitigation level in 9.1 and the effects of the number of agents in 9.2.

8 Implementation details

8.1 Independent-PPO

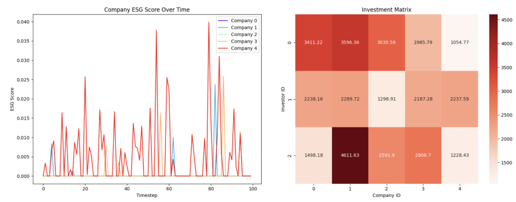
To test how self-interest agents learn to respond to incentives in the environment, we employ a state-of-the-art MARL algorithm based on Independent PPO. Each agent has its own policy parameters, and agents do not share parameters among themselves. This is because we are interested in simulating companies 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 as in 3.2. 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 (Raffin et al., 2021). 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.

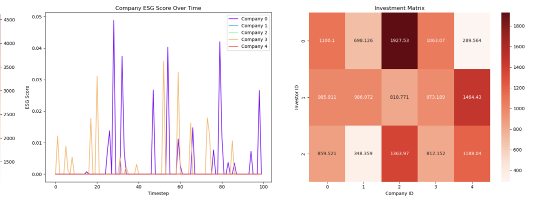
9 Additional Experiments

9.1 What are the effects of resilience spending?

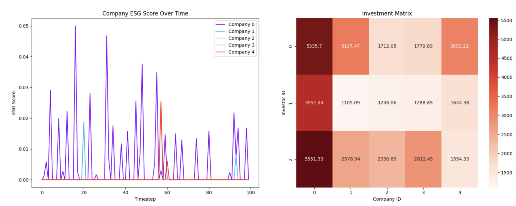
In this case, we allow the company to invest in resilience spending to improve their own robustness to climate events, without affecting the global climate risk. Although investors continue to respond to companies' ESG scores, resilience spending is excluded from the calculation of ESG scores. As



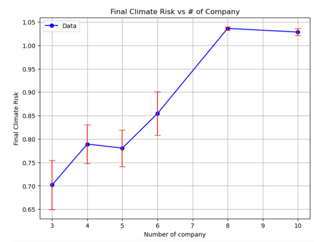
(a) ESG consciousness = 0.5



(b) ESG consciousness = 1



(c) ESG consciousness = 10



(d) Ending climate risk increases as the number of shareholders increases.

Figure 9: (a)-(c) shows the ending episode investment matrix for investors with different level of ESG consciousness. (d) shows ending climate risk vs. number of companies. When the number of companies increases, each company has less capital, and therefore companies have to spend a greater portion of their capital to see the same mitigation effect.

shown in Figure 9d, when resilience is allowed, companies invest significantly more in resilience than in mitigation. By investing in resilience, companies are more resistant 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.

9.2 How does the number of companies affect the overall mitigation efforts?

We change the number of companies while keeping the initial wealth distribution between companies and investors as 50-50. As shown in, the ending system climate risk increases with the number of companies. This is because when the number of companies increases, each company has less capital, and therefore companies have to spend a greater portion of their capital to see the same mitigation effect. This case reflects the difficulty of coordinated efforts for emission efforts in reality, as the number of shareholders increases, they have less motivation to invest in mitigation.