A Unifying Theory of Foreign Intervention in Domestic Climate Policy.

[Latest version here]

Juan Moreno-Cruz, University of Waterloo, jmorenoc@uwaterloo.ca

Anthony Harding, Harvard University, tonyharding@hks.harvard.edu

> "*Free-riding* occurs when a party receives the benefits of a public good without contributing to the costs. In the case of the international climate-change policy, countries have an incentive to rely on the emissions reductions of others without taking proportionate domestic abatement."

Nordhaus (2015)

> "[Free-driving] A second less-familiar externality shows up in the scary form of geo- engineering the stratosphere(…)The challenge with this second global externality also appears to be enormous, because here too so much is at stake, and because it also seems difficult to reach an international governing agreement."

Weitzman (2015)

Abstract

We propose a theory of climate-policy motivated foreign intervention to study different forms of international climate governance in the presence of power imbalance. Foreign countries have at least three options to intervene in another country's domestic climate policy: i.) Agreements with Extraction; ii.) Agreements with Transfers; and iii.) Agreements with Sanctions. We distill the fundamental properties of different climate policy options into a simple parameterization and examine the incentivizes and preferences for each type of foreign intervention. We find that the preference for the type of foreign intervention depends critically on the policy externality of different domestic climate policies.

Acknowledgements: We acknowledge support from the LAD Climate Fund and useful comments and feedback from Jonathan Weiner and other participants in workshops at Resources for the Future, the Latin American Association of Environmental and Resource Economists, and the Canadian Economics Association. Moreno-Cruz acknowledges additional support from the Canada Research Chair program and the National Science Foundation grant NSF #1948154.

1. Introduction

In a warming world, the fates of all countries are intertwined. Given the globalized nature of the economic system and Earth's climate system, there is a need for global cooperation to reduce the impacts of climate change. Cooperation on reducing greenhouse gas emissions has been insufficient. The global public good nature of emissions reductions results in the sub-optimal allocation of effort to limit the harmful impacts of climate change. This lackluster progress on emissions reductions, including negative emissions, has opened the door to other forms of climate policy options that deal with impacts rather than the root cause of the problem; namely, adaptation and solar geoengineering. Against this background of multiple possible climate strategies, there is an international political economy literature that treats mitigation, adaptation, and solar geoengineering as intrinsically different.[1] Countries also have multiple options to influence the climate policy decisions of other countries, such as i.) voluntary international environmental agreements, ii.) agreements with transfers, and iii.) agreements with sanctions. While there could be good reasons to maintain these strategies as separate (Jinnah, Morrow and Nicholson, 2021), we hypothesize that by embedding mitigation, adaptation and geoengineering into a more extensive set of international governance mechanisms, we can learn new insights that could help break the deadlock in international climate negotiations.[2]

Our research question is under which political, economic, and technological circumstances do countries decide to intervene and what form of foreign intervention do they choose? To answer this question, we propose a unifying framework that considers diverse forms of domestic climate policy and multiple international governance mechanisms.

We consider an environment where a powerful country, the Hegemon, can threaten a weak country, the Target, to induce the behavior desired by the Hegemon. Intervention can be costly and the Hegemon can alter the decision

[1] Some recent entries have considered an optimal portfolio approach to climate policy (Aldy and Zeckhauser, 2020; Moreno-Cruz, Wagner and Keith, 2018; Ricke and Moreno-Cruz 2020; Belaia, Moreno-Cruz and Keith, 2021; Harding, Belaia and Keith, 2022), the international political economics literature continues to place the different approaches to managing climate change into siloes.

[2] The Climate Overshoot Commission was recently convened to look at the governance of accelerated adaptation, carbon dioxide removal and solar geoengineering. Our paper speaks directly to that effort.

of the Target to the extent that is possible short of direct military intervention.[3] We develop a game-theoretical framework to capture this strategic environment. The game has two stages. In the first stage, the Target chooses the domestic climate policy it wants to implement, considering the possibility of foreign intervention. In the second stage, the Hegemon decides whether to intervene and the type of intervention. The solution concept is that of subgame-perfect equilibrium.

We show that the form of foreign intervention chosen by the Hegemon depends mainly on the *magnitude* and *nature* of the policy externality that the Target imposes on the Hegemon. The magnitude of the policy externality depends on *exposure* and *preference asymmetry*. Exposure refers to the degree of influence of any domestic policy on foreign nations. Preference asymmetry is the difference in the preferred policy outcome between countries. The nature of the externality can take the form of under-provision due to *free-riding* or over-provision due to *free-driving*. When countries have equal power, the possibility of reaching an agreement is limited and occurs when either country want to curtail the other country's excessive use of a given climate management portfolio. When we introduce power, we find the Hegemon can increase participation by imposing an agreement that is in principle costly, but it can extract all the gains from the move to the optimal allocation. When there are no rents to extract, the Hegemon pursues either transfers or sanctions.

Our paper contributes to the literature on international environmental agreements that started with the work of Scott Barrett (1993, 2003). Since then, several other publications on the topic followed expanding on this seminal contribution. Some authors have introduced heterogeneity (e.g. McGinty, 2007), uncertainty (e.g. Finus 2013), and other complication into the analysis of climate change coalitions. Overall, the combined lesson learned from this literature is that stable self-enforced coalitions are not large enough (see Finus and McGinty (2019) for an exception.). For this reason, alternatives to self-enforced climate agreements have been introduced in academic and policy circles. Among these proposals is the idea

[3] We limit our foreign intervention options to diplomatic channels. While the possibility of direct conflict is not often discussed in the international political economy of climate change, it is discussed in the context of solar geoengineering (Schelling, 1996). We leave this tantalizing possibility for future research.

of transfers (Carraro, Eychmans and Finus, 2006; Bosetti et al. 2013) and trade sanctions (Barrett, 1997; Nordhaus, 2015). We consider all these forms of agreement simultaneously in our unifying framework.

Adaptation was mostly left to the fringes of climate change research for almost a decade, with some very notable exceptions (e.g. Mendelsohn, 2000). As a result, there are very limited entries in the literature that consider adaptation as part of an international environmental agreement (Lazkano, Marrouch , and Nkuiya, 2016; Li and Rus, 2019). Our paper expands this literature by presenting adaptation as another form of climate management that is subject to international governance or that affects the governance of other climate policy options by altering the exposure to foreign policy or preference asymmetries of countries.

Adaptation measures are often constrained to national or regional plans, and thus, they are assumed to be mostly private goods and less likely to suffer from under provision. Insufficient international finance and aid, however, risks the deployment of successful adaptation strategies. Adaptation measures are also likely to become transboundary issues as vulnerable populations look for ways to preserve their livelihoods (Black et al., 2011; also see Waldinger, (2022) for a historical perspective). Another clear example of this possibility is the damming of the Mekong River by China, altering the flow to Myanmar, Laos, Thailand, Cambodia and Vietnam (Eyler 2020). Thus, through financing, migration and large infrastructure projects, adaptation becomes subject to international governance (Khan and Munira, 2021) and should be considered with other, more global, forms of climate management.

More recently, solar geoengineering has entered the conversation to address the urgency of delayed action on climate change and to limit the impacts of unmitigated emissions (Aldy et al., 2021, Field et al., 2021). Solar geoengineering brings with it novel risks and governance challenges. The recent interest in the international political economy of geoengineering started with Barret (2008) and has since then explored other issues (Heyen, Horton, Moreno-Cruz 2019; Moreno-Cruz 2015; Ricke, Moreno-Cruz, Caldeira 2013; Rickels et al. 2020; Sayegh, Heutel, Moreno-Cruz 2021; Millar-Ball 2012; Urpelainen 2012; Heyen, Lehtomaa 2021) with a focus on voluntary international environmental agreements. Of particular interest for us is the free-driving externality that results from the low-cost, high-leverage nature of solar geoengineering techniques. Tackling the free-driver is, from an international governance perspective, the most complex issue associated with solar geoengineering (Parson and Reynolds 2021; Reynolds 2021). Here we show it is not uniquely related to solar geoengineering and also show there are options already existing in the policy repertoire to govern the free-rider.

This paper also contributes to the international political economy literature (Aidt, Albornoz and Hauk 2021). Much of the international political economy literature in climate change compartmentalizes the study of different interventions and focuses on each possibility in isolation. In the framework we propose here, we consider multiple channels of foreign intervention, thus placing the likelihood of voluntary cooperation and economic transfers or sanctions into the same context. We demonstrate that the type of foreign intervention is a strategic choice and a function of the technical and political characteristics of the source of the negative externalities (Eguia 2021).

The rest of the paper proceeds as follows. In section 2, we discussed how different climate management strategies, while diverse in terms of costs, benefits, and technological pathways, can be represented by in international governance as simple the degree of asymmetry between preferred outcomes between countries and the degree of exposure of one country to other country's climate strategy. In section 3 we introduce the model, define policy externality, and highlight the assumptions governing our modeling approach. In section 4 we analyze the equilibria that exist when power is balanced between countries. In section 5, we introduce the foreign intervention options that are available when a country holds substantial power over others. We characterize the space where different interventions are preferred depending on the characteristics of the climate policy externality.

2. Technological landscape

Policymakers and the academic community often consider mitigation, adaptation and solar geoengineering options as intrinsically different and subject to independent international governance mechanisms and regulations. For mitigation, the goal is to increase provision, the goal for solar geoengineering is to limit provision, and the goal for adaptation is to fill the gap in provision of the other two. Yet, these are not neatly disjoint sets as it comes to international governance.

For example, consider emissions reduction strategies. Think of these as low-emissions technologies. These technologies while becoming increasingly cheaper, require a system-level transition that makes them overall costly. There are, however, moderate private benefits that arise, such as the co-benefits from improved air quality (Gallagher and Holloway, 2020). At the same time, there are limited policy externalities because the global effects of domestic emissions reductions are small and only a concerted effort influences the climate. Direct Air Capture (DAC) is a form of negative emissions that satisfies the same role as emissions reductions, at least until zero emissions are achieved. DAC are high-cost techniques with the capacity to substantially reduce the burden of emissions reductions across countries. One single country, with enough effort, can reduce the negative effects of the most marginalized countries. This implies DAC has high policy externalities. Of course, there are domestic climate policies that can affect other policies. For example, consider the case of rare materials and the need for battery storage. If a country, say Canada, deems its materials an object of national security, it could implement a policy that bans the export of those materials thus increasing the costs of acquiring them for the rest of the world. There are no direct externalities for such a policy, but it will still affect the costs of foreign climate policy, including energy systems transitions to near-zero emissions technologies.

Climate policies traditionally understood as adaptation are also quite diverse in terms of their underlying characteristics, at least in dimensions relevant for foreign intervention in domestic climate policy. For example, if countries defend their coasts by building sea walls or dikes, the policy externalities are very low. The costs are high as are the private benefits. Another form of adaptation to changing climate is to switch crops to more heat resistant or less water intensive crops. These changes are highly decentralized, although they can be coordinated via national policies. In either case, this change in the crop composition could affect the balance of trade in global markets. This represents moderate policy impacts in other countries. A third option could be damming a river to increase irrigation or for hydroelectric power. Cost could be high, private benefits are high. In principle, this policy would not have foreign implications, unless the case of a transboundary river like the Mekong. In this case, decisions made upstream have direct externalities imposed on downstream countries.

This leads us to the final category – that of solar geoengineering. Two characteristics make solar geoengineering a different object of global governance relative to mitigation and adaptation. First, the effects and impacts of solar geoengineering are not uniformly distributed worldwide. Countries have different preferences regarding their climate and solar geoengineering. Second, implementing solar geoengineering can be done unilaterally and without the need for global consensus. Low costs of global deployment make it high leverage. There are many forms of solar geoengineering discussed in the literature. The two more prominent are marine cloud brightening and stratospheric aerosol injection. Marine cloud brightening consists of spraying microscopic droplets of sea salt in the lower atmosphere thus seeding clouds that are on balance more reflective of solar radiation than they are absorptive of heat radiating energy (Field et al., 2020). These techniques have predominantly local effects and could be designed to affect regional climates, but they also have the capacity to affect the climate in other regions, such as through teleconnections (Ricke et al., 2021), or even the whole planet. Stratospheric aerosol injection is a high-leverage, low-cost technique that has limited or no isolated local effects but can alter the climate at a global scale at a very low cost absorptive.

It is hopefully clear from our discussion above that while our assessment is somehow subjective, there is enough variation across possible interventions to merit a more general, unified approach to the question of international governance that is not restricted to narrow technical classifications and highlights the needs of a framework to think about these interventions in a more comprehensive setting.

3. The model

We consider a two-country world with a Hegemon (H) and a Target (R). Each country has a domestic climate policy lever, g_i for $i \in \{H, R\}$, that they use to minimize the damages from climate change. There are no positivity constraints on these policy levers. We think of domestic climate policy as policy portfolios implemented by a country. We are interested in these policies' combined effect on the climate. Countries can have policy portfolios with negative effects on the climate, let's say by subsidizing fossil fuel production more than near-zero emissions technology. In the context of solar geoengineering, we can think of negative policies as counter-geoengineering.[4]

Countries minimize total costs of climate change, $TC_i(g_i, g_j)$, that are taken as the sum of the damages from climate change, $D_i(g_i, \gamma_{ji} g_j; \Delta_i)$, and the private costs of implementing climate policy, $c_i(g_i)$. That is,

$$TC_i(g_i, g_j) = D_i(g_i, \gamma_{ji} g_j; \Delta_i) + c_i(g_i) \qquad \textit{Eq. 1}$$

Each country has a preferred amount of policy intervention that reduces their damages to zero, here captured by $\Delta_i > 0$.[5] Damages from climate change in each country increase in deviations from their preferred climate. Policy g_i can reduce climate damages in country i by either shifting the realized climate closer to their preferred state, i.e. mitigation and solar geoengineering, or by shifting the preferred climate closer to the realized climate, i.e. adaptation. Importantly, the policy of country j can also have a direct effect on country i's damages. We measure the strength of the *exposure* of country j to country i's policy by $\gamma_{ji} \in \mathbb{R}^+$.

We assume simple quadratic forms for each component of the total costs:

$$D_i(g_i, \gamma_{ji} g_j; T_i) = \frac{1}{2}\delta\left(\Delta_i - g_i - \gamma_{ji} g_j\right)^2 \qquad \textit{Eq. 2}$$

$$c_i(g_i) = \frac{1}{2}\tau {g_i}^2 \qquad \textit{Eq. 3}$$

[4] Imagine a set of policy options indexed by $k = 1, \dots, K$. The net effect of policies on the climate is given by $g_i = g(g_{i1}, \dots, g_{ik}, \dots, g_{iK})$. The cost of a given portfolio is $c_i(g_i) = c(g_{i1}, \dots, g_{ik}, \dots, g_{iK})$. We assume each country implements a portfolio the minimizes the cost of the policy portfolio for any given g_i chosen in the international stage.

[5] That is, when the combined policy implemented by two countries, $g_i + \gamma_{ji} g_j = \Delta_i$, damages for country i are zero.

Definition 1. We define $\Delta = \Delta_R - \Delta_H$ as the *preference asymmetry* between countries and $\gamma = \gamma_{HR} - \gamma_{RH}$ and the *exposure divergence* between countries.

Definition 2. We define ***Policy Externality*** as a function $F_i(\gamma, \Delta)$ for $i \in \{H, R\}$ that jointly captures differences in exposure and differences in preferred climate policies.

While there are many technology parameters that affect the outcomes, such as private implementation costs and benefits, we assume the only differences between countries' climate policies are due to strength of the **policy externality**.

Assumption 1. a.) Global average temperature is too high for the two countries, $\Delta > -\Delta_H$, and b.) The own effect of a policy is always larger than the indirect effect from the other country's policy, $\gamma_{HR}\gamma_{RH} < \left(\frac{\delta}{\delta+\tau}\right)^2$.

Assumption 1a ensures that countries have an incentive to implement some amount of combined policy that reduces the temperature. That is, in this setting there are no winners from climate change. Assumption 1b ensures the SPNE equilibrium is stable.

4. No-Intervention Benchmarks

To begin, we assume that countries are symmetric, $\Delta = 0$ and $\gamma = 0$. This does not to imply that the policy externality is zero, just that it is equal across countries: $F(\gamma = 0, \Delta = 0) \geq 0$. Later, we show how the Hegemon's policy changes as preferences and exposure diverge, that is when $\Delta \neq 0$ and $\gamma \neq 0$.

We begin by analyzing the solution for an uncoordinated equilibrium, U, with no foreign intervention. The Hegemon, in its position of power, moves in anticipation of the response it will elicit from the Target. The solution concept is that of Sub-game Perfect Nash Equilibrium (SPNE). Define the SPNE policy as g_i^U for $i \in \{H, R\}$ and the corresponding total costs associated with these policies as $TC_i(g_i^U, g_j^U) \equiv TC_i^U$ for $i \in \{H, R\}$.

Figure 1 displays the strategic decision space for the game. We represent the equilibrium policy decisions of the two countries in terms of their best responses.[6] The policy of the Target, g_R, is on the horizontal axis, and the policy of the Hegemon, g_H, is on the vertical axis. As mentioned above, these policy portfolios can have a net positive or negative effect on the climate depending on a country's energy and climate policies.

The best response function of the Hegemon is given by $g_H = BR_H(g_R)$ and the Target's best response is given by $g_R = BR_R(g_H)$. The SPNE lies at the intersection of these best response functions, (g_R^U, g_H^U). The resulting total costs of climate and policy faced by the Hegemon are given by $TC_H(g_H^U, g_R^U) \equiv TC_H^U$. The total costs of the Target are $TC_R(g_R^U, g_H^U) \equiv TC_R^U$. Iso-cost curves are depicted as ellipses, which follows directly from our choice of quadratic functional forms. These ellipses are centered where the $g_j = BR_j(g_i) = 0$, so that $TC_j^U = 0$. Higher total costs are captured by expanding concentric ellipses.

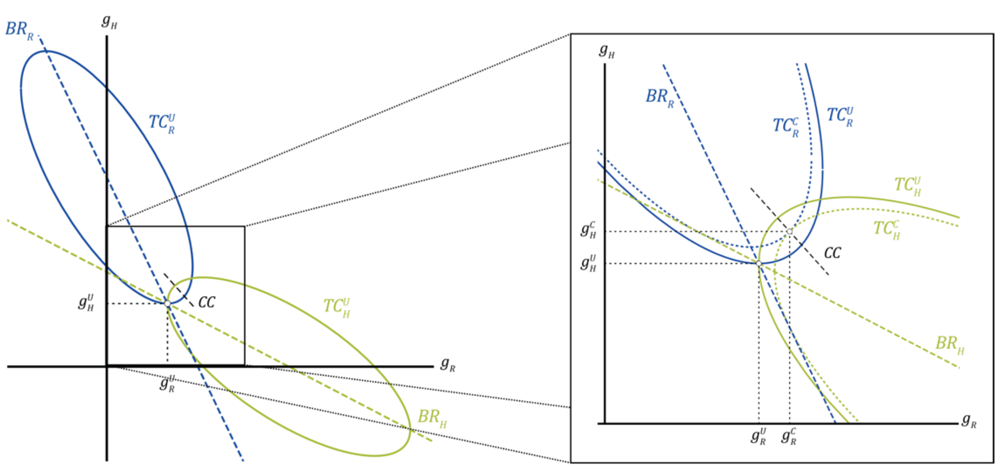


Figure 1. SPNE for symmetric countries.

Continuing with the symmetric case, we analyze the case of a coordinated equilibrium without intervention, C. In this equilibrium, countries minimize the joint total cost, leading to the optimal equilibrium allocation. The resulting policies are denoted $\{g_R^C, g_H^C\}$. The resulting total

[6] Our presentation of the results was inspired by two classic papers by Brander and Spencer that study the role of R&D in industrial competition (Brander and Spencer 1985,1987). We also borrowed from (Aidt, Albornoz, and Hawk, 2021) to depic our solution sin the policy space. While our context and analysis are substantially different, we are intellectually indebted to them as we worked on developing the intuition for our own work.

costs for this coordinated equilibrium are given by $TC_H(g_H^C, g_R^C) \equiv TC_H^C$ and $TC_R(g_R^C, g_H^C) \equiv TC_R^C$. These costs are represented by smaller concentric ellipses relative to the SPNE costs in the expanded panel on the right-side of **Figure 1**.

4.1. Free-driving vs Free-Riding

In the symmetric case, the coordinated equilibrium is interior to the uncoordinated equilibrium iso-cost curves for the Hegemon, $TC_H^U > TC_H^C$, and the Target, $TC_R^U > TC_R^C$. That is, the coordinated equilibrium lies inside the Pareto set and it is therefore a Pareto improvement. When both countries do more to address climate change, their total costs decrease. When decisions are uncoordinated, countries do not consider the externalities of their policy. Given the negative slope of the best response functions, the policies by the two countries are strategic substitutes. Thus, each country has an incentive to reduce their contribution to the public good, inducing a higher amount of policy on the other country. Of course, as it is well known in these situations, the other country has incentive to behave the same way. This captures the **Free-riding externality** that results in too little of the public good contributed in equilibrium. This is the outcome often discussed in the context of emissions reductions.

Free-riding equilibria are the only possibility when countries are symmetric. If we introduce asymmetry, it is possible to find situations where the coordinated equilibrium is outside the Pareto set. In fact, for large enough asymmetries, it is possible for one country to implement too much of a given policy, forcing the other country to implement countervailing actions. We show an example of this possibility in **Error! Reference source not found.**. The amount of policy implemented by the Target is too high and the Hegemon policy is then negative. While this idea of countervailing policies is not typically encountered in discussions of emissions cuts or adaptation, it has been raised as an issue in the context of solar geoengineering.[7] There it has been coined the **Free-driver externality** because, in equilibrium, a country over-provides policy to the detriment of others.

[7] These equilibria with countervailing policies are examples of the climate clash discussed in Heyen et al. (2020).

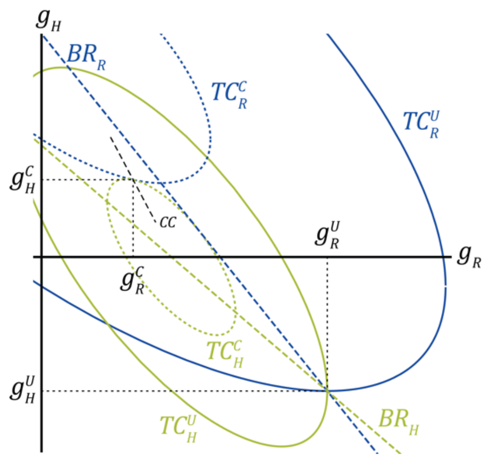
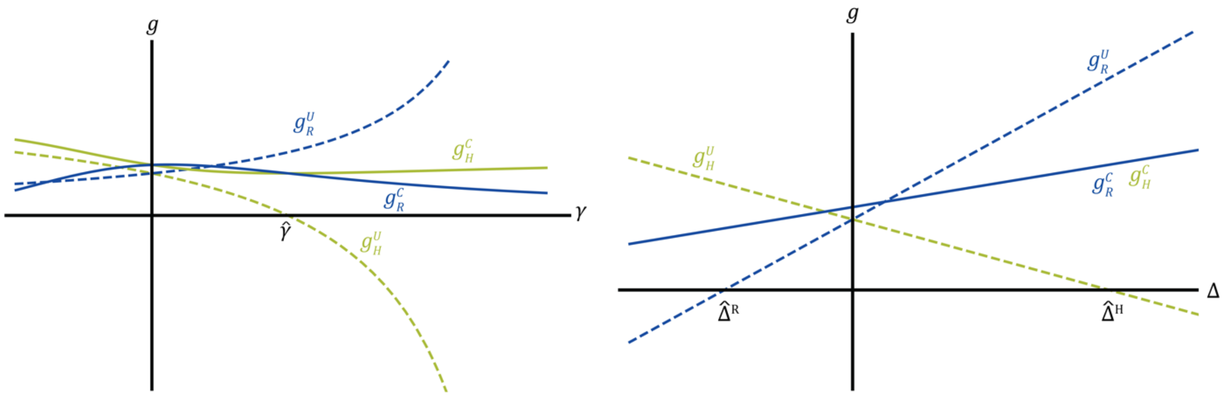


Figure 2. Example asymmetric countries with free-driving SPNE.

When an equilibrium is a free-riding equilibrium, countries are better off increasing total policy provision. When an equilibrium is a free-driving equilibrium, countries are better off reducing the net amount of their policy provision.

The only source of asymmetry in our context is the policy externality $F(\gamma,\Delta)$ and equilibria can be fully characterized as a function of preference divergence Δ and exposure divergence γ. In Figure 3, we illustrate how changes in the exposure γ or preference asymmetry Δ can affect policy outcomes in both the uncoordinated and coordinated equilibria. In Figure 3A, we fix preference asymmetry and show how the equilibrium policy choices change with the exposure of the Hegemon to the Target's policy. Starting from the symmetric case where $\gamma = 0$, as γ increases g_H decreases. The Hegemon is free-riding on the efforts of the Target as spillovers from the Target's policy grow. The Target's policy increases with γ to compensate for the free-riding. When $\gamma > \hat{\gamma}$ the free-riding equilibrium becomes a free-driving equilibrium. The Target country implements policy in excess of the amount preferred by the Hegemon, so the Hegemon begins to implement a net negative amount of policy to compensate for the excess.



A. Exposure Divergence B. Climate Preference Divergence

Figure 3. Asymmetric SPNE under increased policy externalities.

In Figure 3B we fix policy exposure asymmetry and show how equilibrium policy choices change with the Target's preferred policy. Starting from the symmetric case, as Δ increases the Target prefers a cooler world relative to the Hegemon. So, the Target implements more policy and the Hegemon free-rides on the Target's efforts. Eventually, for $\Delta > \hat{\Delta}^H$, the amount of policy implemented by the Target becomes excessive for the Hegemon, so the Hegemon implements countervailing policies to compensate. The free-riding becomes free-driving. This result is reversed when Δdecreases starting from the symmetric case. As the Target prefers a warmer climate relative to the Hegemon, they implement a lower amount of policy, free-riding on the efforts of the Hegemon. Eventually, for $\Delta < \hat{\Delta}^R$, the policy of the Hegemon becomes too much for the Target, so they implement countervailing policy to compensate. The free-riding becomes free-driving.

Drawing from these comparisons, whether an equilibrium is a free-driving or free-riding equilibrium depends critically on both preference asymmetry and policy exposure asymmetry. We can generalize our findings to the two-dimensional policy externality space $F(\Delta, \gamma)$.

Proposition 1. (Equilibrium policy outcomes). There exists some $\widehat{\Delta}^R(\gamma)$ and $\widehat{\Delta}^H(\gamma)$ such that

(i) *Target Free-driver:* if $\Delta = \widehat{\Delta}^H(\gamma)$ then $g_H = 0$ and if $\Delta < \widehat{\Delta}^H(\gamma)$ then $g_R > 0$ and $g_H < 0$.

(ii) *Hegemon Free-driver:* if $\Delta = \widehat{\Delta}^R(\gamma)$ then $g_R = 0$ and if $\Delta < \widehat{\Delta}^R(\gamma)$ then $g_R < 0$ and $g_H > 0$.

(iii) *Free-rider:* if $\widehat{\Delta^R}(\gamma) < \Delta < \widehat{\Delta^H}(\gamma)$ then $g_R > 0$ and $g_H > 0$.

Proof. See the Appendix.

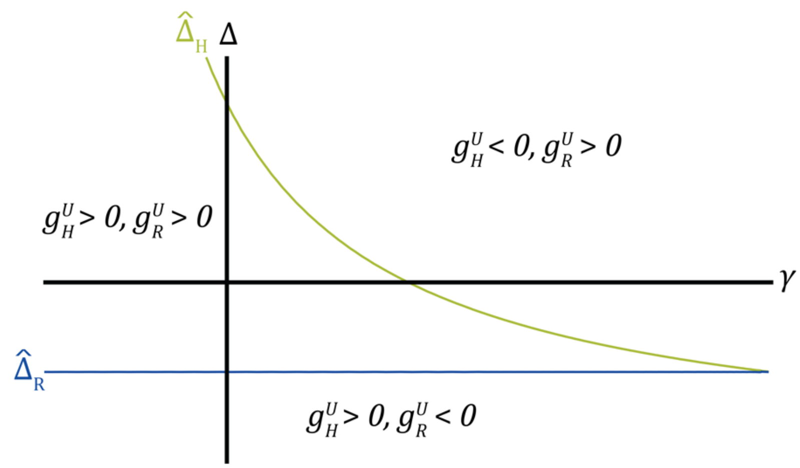


Figure 4. Policy outcomes in the policy externality space.

In **Error! Reference source not found.** we show the nature of the equilibrium i s a function of the policy externality, $F(\Delta, \gamma)$. The horizontal axis captures asymmetry in policy exposure, γ. The vertical axis captures asymmetry in preferences, Δ. In the analysis below, we are interested in how the Hegemon responds to changes in the Target's nature and magnitude of the policy externality. To that effect, we fixed the policy externality γ_{HR} so that changes in γ are capturing changes in γ_{RH}. We similarly fix Δ_H so that changes in Δ capture changes in Δ_R.

Using **Error! Reference source not found.**, we can explain the intuition b ehind Proposition 1. Let's start at the origin, with the symmetric case ($\gamma = 0, \Delta = 0$). Here, the Hegemon and the Target both implement net positive policies to reduce climate damages in both countries. However, as explained above, there is too little policy contribution in equilibrium. Free riding continues to occur with small deviations from the symmetric case. When $\Delta = \widehat{\Delta}^R(\gamma)$, the Hegemon provides cooling exactly equal to the preferred cooling by the Target, with no contribution or cost to the Target. $\widehat{\Delta}^R$ is independent of the exposure of the Hegemon to the Target and thus is a horizontal line. Similarly, when $\Delta = \widehat{\Delta}^H(\gamma)$, the Target provides cooling exactly equal to the preferred cooling by the Hegemon, with no contribution or cost to the Hegemon. $\widehat{\Delta}^H(\gamma)$ is decreasing in exposure of the Hegemon to the Target because as exposure increases, a given amount of cooling by the Target has a larger effect on the Hegemon. We call the region between these two curves, $\widehat{\Delta}^R(\gamma) < \Delta < \widehat{\Delta}^H(\gamma)$, the Free-rider space because at least one of the countries is underproviding effort to limit climate change in equilibrium.

If Δ decreases below $\widehat{\Delta}^R$, the Hegemon's policy becomes too much for the Target and the Target begins to engage in countervailing policy. Thus, we call the region such that $\Delta < \widehat{\Delta}^R(\gamma)$ the Hegemon Free-driver space. Alternatively, as Δ increases above $\widehat{\Delta}^H(\gamma)$, policy provision by the Target becomes too much for the Hegemon and the Hegemon begins to engage in countervailing policy. Thus, we call the region $\Delta > \widehat{\Delta}^H(\gamma)$ the Target Free-driver space.

4.2. Preferences Between Coordinated vs Uncoordinated Equilibria

Countries enter an agreement that implements the coordinated equilibrium with the expectation of improving on their uncoordinated equilibrium. These agreements are commonly known in the climate change literature as Self-enforcing International Environmental Agreements, introduced in the seminal work by Barrett (1994). An agreement is stable if it is incentive compatible for both players, for which self-enforcing agreements get their name. We define incentive compatibility as lower total cost in the Coordinated equilibrium relative to the Uncoordinated equilibrium,

$$TC_i^C \leq TC_i^U \text{ for } i \in \{R, H\}.$$

Here, as with most international environmental agreement models, we assume equal power between countries. In the next section, we relax this assumption to explore the role of power imbalances.

In the symmetric case, we see both countries prefer the coordinated equilibrium over the uncoordinated equilibrium. In this case, an agreement is incentive compatible and stable. But this is not always the case. Whether or not an agreement is stable depends on the policy externality, $F(\Delta,\gamma)$.

Proposition 2. (Coordinated (C) versus Uncoordinated (U) Equilibrium Outcomes). There exists some $\overline{\Delta}_H^{CU}(\gamma)$, $\underline{\Delta}_H^{CU}(\gamma)$, $\overline{\Delta}_R^{CU}(\gamma)$, $\underline{\Delta}_R^{CU}(\gamma)$, and $\overline{\overline{\Delta}}_H^{CU}(\gamma)$ such that

(i) If $\underline{\Delta}_H^{CU}(\gamma) < \Delta < \overline{\Delta}_H^{CU}(\gamma)$ then $U \succ C$

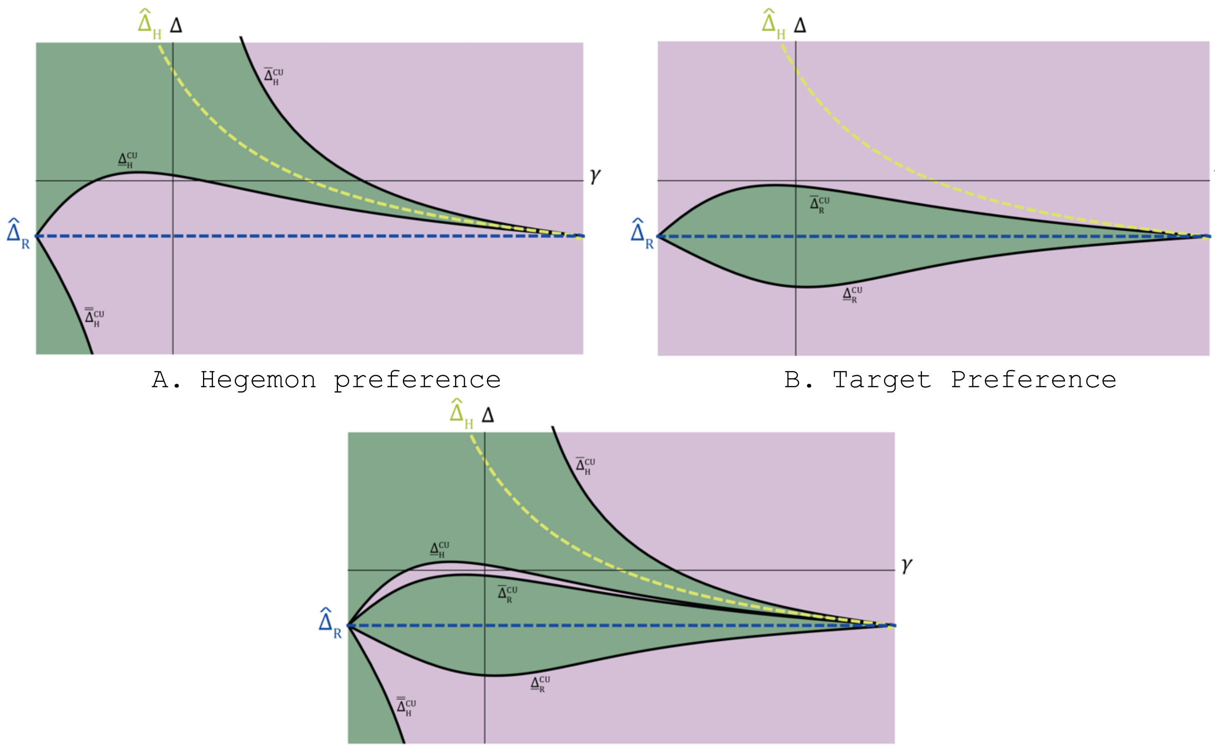
(ii) If $\Delta < \overline{\overline{\Delta}}_H^{CU}(\gamma)$ then $U \succ C$

(iii) If $\underline{\Delta}_R^{CU}(\gamma) < \Delta < \overline{\Delta}_R^{CU}(\gamma)$ then $C \succ U$ but C is not a stable equilibrium

(iv) Else, $C \succ U$

Proof. See the Appendix.

Figure 6 illuminates the intuition behind Proposition 2 through a decomposition of each countries' incentive to cooperate. Let's first consider the Figure 6incentives of the Figure 6Figure 6Figure 6Hegemon, Figure 6Figure 6A. Consider the space above the curve $\widehat{\Delta}_H$, which demarks $g_H^U = 0$. From Proposition 1, this is the Target Free-Driver space. Along the curve $\widehat{\Delta}_H$ the Hegemon's total costs are zero in the Uncoordinated equilibrium. The Target country provides the Hegemon with their preferred climate at no cost to the Hegemon. Thus, they prefer this equilibrium to the Coordinated equilibrium where they must provide policy to share costs with the Target. Increasing Δ from this curve into the Target Free-driver space, eventually, clashing with the Target becomes too costly and the Hegemon prefers to cooperate. For $\Delta > \overline{\Delta}_{\mathrm{H}}^{\mathrm{CU}}(\gamma)$ the Hegemon is willing to concede a small positive policy provision in exchange for reigning in the Target.



A. Hegemon preference B. Target Preference

C. Stable coordinated vs. uncoordinated equilibria

Figure 5. Coordinated vs. uncoordinated preference in the policy externality space.

Next, let's consider when Δ Figure 6decreases from the curve $\widehat{\Delta}_H$. As Δ decreases, policy provision by the Hegemon increases. Eventually, when $\Delta < \underline{\Delta}_H^{CU}(\gamma)$ the Hegemon prefers cooperation, now to encourage the Target to contribute more to the joint policy effort. The exception to this is when the Hegemon's exposure to the Target is low. When $\Delta < \overline{\overline{\Delta}}_H^{CU}$, the Hegemon has little exposure to the Target's policy, so they have no desire to cooperate and increase their policy provision when they receive little benefit for their cooperation efforts. While the Hegemon prefers cooperation for $\overline{\overline{\Delta}}_H^{CU} < \Delta < \underline{\Delta}_H^{CU}(\gamma)$, this is not always the case for the Target, and for an agreement to be stable, both countries must have incentive to cooperate.

Now consider the incentives of the Target,Figure 6Figure 6Figure 6Figure 6Figure 6 Figure 6Figure 6B. Along the curve $\widehat{\Delta}_F$, where $g_R^U = 0$, the total costs of the Target country are minimized in the Uncoordinated equilibrium. The Hegemon provides the Target with their preferred climate at no cost to the Target. For deviations of Δ from this curve, either positively or negatively, costs for the Target increase. As Δ increases, the Target prefers a cooler climate and increases policy provision. Once $\Delta > \overline{\Delta}_R^{CU}(\gamma)$ the preferred cooling is sufficient that the Target is willing to engage in the Coordinated equilibrium to encourage more policy provision by the Hegemon. Alternatively, as Δ decreases, the Target engages in countervailing policy to compensate for cooling from the Hegemon's policy. Once $\Delta < \underline{\Delta}_R^{CU}(\gamma)$ the Target costs of countervailing policy become sufficient that the Target prefers the Coordinated equilibrium to discourage policy provision by the Hegemon. Between these curves, $\underline{\Delta}_R^{CU}(\gamma) < \Delta < \overline{\Delta}_R^{CU}(\gamma)$, the Target does not have incentive to cooperate, even though the Hegemon does. In this range, the Target requires additional incentive from the Hegemon, opening the door for foreign intervention.

5. Strategic Intervention Strategies

In this section, we introduce power asymmetries by allowing the Hegemon to intervene in the domestic policy set by the Target. When there are power asymmetries, the options of the Hegemon go beyond the Uncoordinated and Coordinated equilibria discussed above.

Most of the work on international environmental agreements assumes countries have equal power, and an agreement is successful only if all countries prefer to participate in the resulting arrangement. The agreement negotiation takes the form, either implicitly or explicitly, of a Nash bargain where final allocation depends on countries' outside option. Here, we make power asymmetry explicit. The Hegemon can induce more activity if the Target is free-riding in the Uncoordinated equilibrium or restraint the Target's activity if it is free-driving.

We consider three options for possible foreign interventions available to the Hegemon to impose its will on the Target country. *First*, the Hegemon uses its power to propose a take-it-or-leave-it offer to the Target and extracts all the gains from entering the agreement. We refer to this sort of intervention as ***Agreement with Extraction***. *Second*, when there are no rents to extract from the Target, the Hegemon needs to shift its strategy to convincing the Target to participate in the agreement. Hence, the Hegemon promises a reward in exchange of an action taken by the Target. We refer to this type of intervention as ***Agreements with Transfers***, such as the Climate Investment Fund. *Third*, the Hegemon imposes a penalty on the Target if its actions do not align with the interest of the Hegemon. These penalties can take many forms, but trade tariffs or financial constraints are the most discussed so far in the literature, such as those envisioned in climate clubs (Nordhaus, 2015). We refer to this type of intervention as ***Agreements with Sanctions***. We use our framework to analyze these three types of policy intervention in turn and show how the selection of intervention by the Hegemon depends on the nature and magnitude of the policy externality.

5.1. Agreement Interventions with Extraction

The hegemon proposes a take-it-or-leave-it set of policy outcomes that minimizes its costs net of the gains from the agreement that it extracts from the Target. The Hegemon's objective function with extraction is defined as

$$TC_H\big(g_R, g_H(g_R)\big) - M\big(g_R, g_H(g_R)\big)$$

where $M\big(g_R, g_H(g_R)\big)$ are the rents extracted from the Target by the Hegemon. To maintain stability of the equilibrium, the Hegemon can only extract as much value as to leave the Target indifferent between the proposed agreement and the Uncoordinated equilibrium. That is,

$$TC_R\big(g_R, g_H(g_R)\big) + M\big(g_R, g_H(g_R)\big) \leq TC_R^U .$$

We assume the Target takes the offer when they are indifferent. If the gains of an Agreement are positive, the Hegemon extracts value $M > 0$. Otherwise, the Hegemon cannot extract value from the Target and maintain a stable outcome, so $M = 0$. Thus, the gains extracted by the Hegemon can be expressed as,

$$M\big(g_R, g_H(g_R)\big) = \min\{0, TC_R(g_R, g_H) - TC_R^U\}.$$

The Hegemon's objective when considering an Agreement with Extraction is given by

$$\{g_H^E, g_R^E\} = \arg\min\{TC_H(g_R, g_H) + TC_R(g_R, g_H) - TC_R(g_R^U, g_H^U)\} \quad \text{Eq. 4}$$

It follows from Eq. 5 that the Hegemon ends up proposing a solution that implements the same globally optimal allocation of the coordinated equilibrium, but in this case it appropriates all the gains from moving towards the coordinated outcome. By extracting the gains under coordinate policy outcomes, the Hegemon always weakly prefers an Agreement with Extraction to the Coordinated equilibrium. When extraction is positive, this is a strong preference.

As an initial benchmark, we can compare the Hegemon's preference for an Agreement with Extraction to the Uncoordinated equilibrium.

Proposition 3. (Agreement with Extractions (E) versus Uncoordinated (U) Equilibrium Outcomes). There exists some $\overline{\Delta}^{EU}(\gamma)$ and $\underline{\Delta}^{EU}(\gamma)$ such that

(i) If $\underline{\Delta}^{EU}(\gamma) < \Delta < \overline{\Delta}^{EU}(\gamma)$ then $U \succ E$

(ii) Else, $E \succ U$

Proof. See the Appendix.

Figure 6. Agreement with Extraction vs Uncoordinated Equilibria Outcomes illustrates Proposition 3 in the policy exposure space. First consider the region between $\underline{\Delta}^{EU}(\gamma)$ and $\overline{\Delta}^{EU}(\gamma)$. This range of the space is identical to the space between $\underline{\Delta}^{CU}(\gamma)$ and $\overline{\Delta}^{CU}(\gamma)$ in Figure 5. In this range, the Target does not have incentive to participate in the Coordinated equilibrium. They would rather free-ride on the Hegemon's efforts or engage in a small amount of countervailing policy effort. This is still true here. With extraction, the Hegemon can take more from the Target, but they cannot provide additional incentive to the Target to participate.

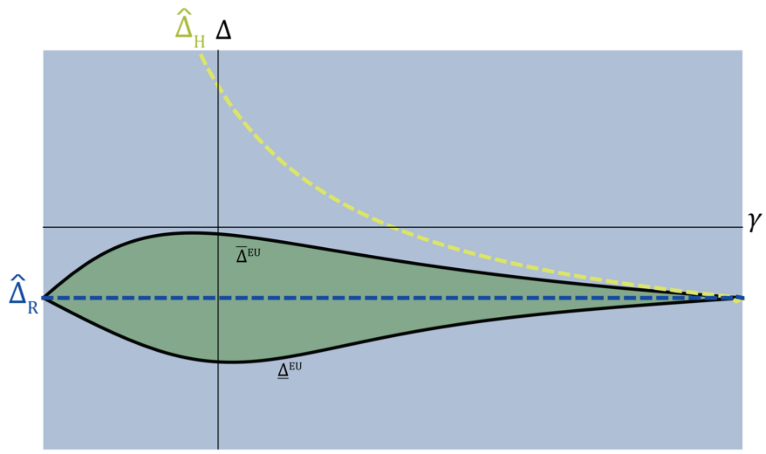


Figure 6. Agreement with Extraction vs Uncoordinated Equilibria Outcomes

Outside of this range, the Hegemon chooses the value extracted to leave the target indifferent between an Agreement with Extraction and the Uncoordinated equilibrium. When the Hegemon previously preferred the Coordinated equilibrium to the Uncoordinated equilibrium, they similarly prefer the Agreement with Extraction to the Uncoordinated equilibrium as extraction can only improve their outcome. In addition, in the range that the Hegemon previously did not prefer the Coordinated Equilibrium, the gains from extraction are sufficient to make the Agreement with Extraction preferrable to the Uncoordinated equilibrium. Thus, an Agreement with Extraction is the preferred intervention method of the Hegemon when it is possible.

5.2. Policy Interventions with Transfers

In an Agreement with Transfers, the Hegemon can offer a reward, W, to the Target country in exchange for a policy that aligns more with the interests of the Hegemon. The simplest way to think about this is the Hegemon transferring cash to the target in exchange for cooperation in an agreement. Of course, this takes many other forms: global investment funds, technological transfers, and investment deals such as the clean development mechanism. The Hegemon will only offer a reward when they prefer cooperation in an agreement and the Target is unwilling to participate without additional incentives. The objective function of the Target includes a transfer of the following form:

$$TC_R(g_R, g_H(g_R)) - W\big(g_R, g_H(g_R)\big) \qquad \text{Eq. 5}$$

where $W(g_R, g_H(g_R))$ is the Reward the Hegemon offers as a function of the policy choice of the Target and $g_H(g_R)$ is the Hegemon's best response to that choice. The transfer associated with the reward needs to be incentive compatible, so that the Target is not worse off with the transfer relative to the uncoordinated outcome. This implies

$$TC_R(g_R, g_H(g_R)) - W\big(g_R, g_H(g_R)\big) \leq TC_R^U$$

If it is a positive amount, the Hegemon transfers to the target the exact amount that makes this equation binding. This leaves the Target indifferent between accepting the transfer or not. We assume they take it in the case of indifference. The transfer is then given by

$$W\big(g_R, g_H(g_R)\big) = \max\{0, TC_R(g_R, g_H(g_R)) - TC_R^U\} \qquad \text{Eq. 6}$$

The problem of the Hegemon is now equivalent to minimizing the joint total cost of the two countries so that

$$\{g_H^T, g_R^T\} = \arg\min\{TC_H(g_H, g_R) + TC_R(g_R, g_H\ ) - TC_R(g_R^U, g_H^U)\} \qquad \text{Eq. 7}$$

This is the same objective function as in the Agreement with Extraction. The difference is who gets the rents from the coordinated allocation. In the extraction case, the Hegemon captures the rents, in the transfers the Target captures the rents. Since the Hegemon has power, they set rewards just to make the Target indifferent between joining the agreement and the Uncoordinated equilibrium. In the absence of power, Nash bargaining would set the value of rewards.

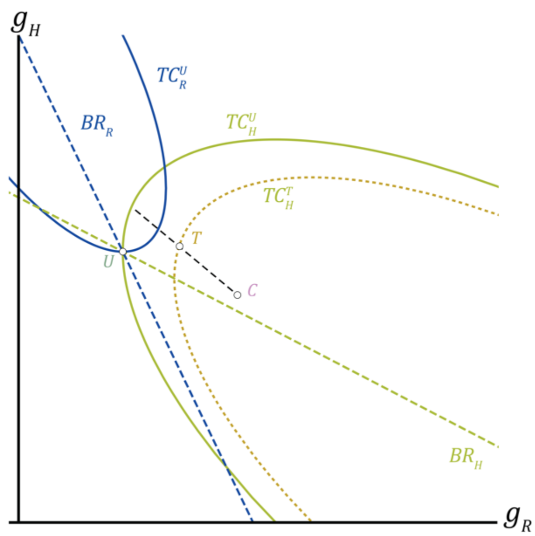


Figure 7. Policy Outcomes – Transfers

Figure 7. Policy Outcomes – Transfers illustrates an example of an agreement with positive transfers. The Hegemon prefers the Coordinated equilibrium, C, to the Uncoordinated equilibrium, U, but the Target does not. The Hegemon can move closer to their provision under the global optimum, along the contract curve, with the transfer of a reward, shown as the equilibrium T.

We now return to our question of the Hegemon's foreign intervention preference by comparing their preference for an agreement, now with transfers, to the Uncoordinated equilibrium and analyzing how this depends on preference asymmetry and policy exposure.

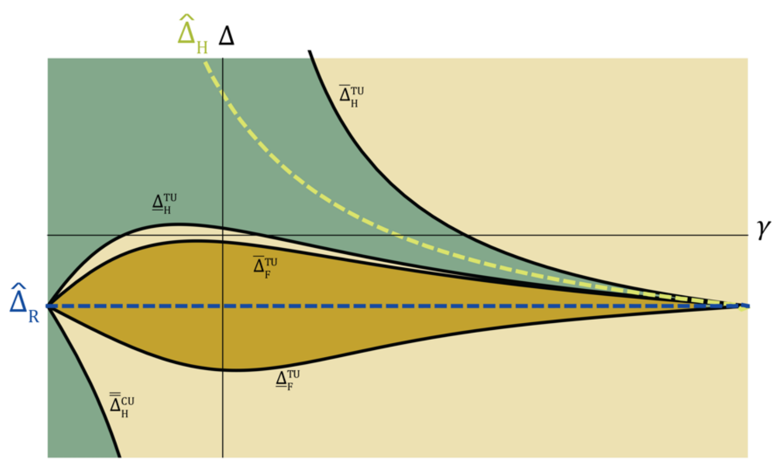


Figure 8. Agreement with Transfers vs Uncoordinated Equilibria Outcomes

Proposition 4. (Agreement with Transfers (T) versus Uncoordinated (U) Equilibrium Outcomes). There exists some $\overline{\Delta}_H^{TU}(\gamma)$, $\underline{\Delta}_H^{TU}(\gamma)$, $\overline{\Delta}_R^{TU}(\gamma)$, $\underline{\Delta}_R^{TU}(\gamma)$, and $\overline{\overline{\Delta}}_H^{TU}(\gamma)$ such that

(i) If $\underline{\Delta}_H^{TU}(\gamma) < \Delta < \overline{\Delta}_H^{TU}(\gamma)$ then $U \succ T$

(ii) If $\Delta < \overline{\overline{\Delta}}_H^{TU}(\gamma)$ then $U \succ T$

(iii) If $\underline{\Delta}_R^{TU}(\gamma) < \Delta < \overline{\Delta}_R^{TU}(\gamma)$ then $T \succ U$ and $W > 0$

(iv) Else, $T \succ U$ and $W = 0$

Proof. See the Appendix.

Figure 8. Agreement with Transfers vs Uncoordinated Equilibria Outcomes illustrates Proposition 4. To understand the intuition of the proposition, first not that when transfers are zero, the equilibrium for Agreement with Transfers is identical to the Coordinated equilibrium. Thus, for this area of the policy exposure space, the Hegemon's preference for an Agreement with Transfers is identical to their preference for the Coordinated equilibrium and how this compares to the Uncoordinated Equilibrium is as in described in Proposition 2. We focus here on what happens when transfers are non-zero. This is when the Target requires additional incentive to participate in the Coordinated equilibrium, Figure 5b. This is the region described by Proposition 2(iii) and Proposition 4(iii), which is shaded darker yellow in Figure 6. While the original Agreement is unstable, here the Hegemon offers rewards to make the Target indifferent with the coordinated option. Given the cost of these transfers, the Hegemon still prefers this intervention to the uncoordinated equilibrium.

5.3. Policy Interventions with Sanctions

The Hegemon could threaten the Target with imposing a sanction if the policy of the Target does not align with the preferences of the Hegemon. Sanctions take many forms, but as mentioned above, trade tariffs are the preferred method in the interconnected world we currently inhabit.[8] The objective function of the Target includes a sanction of the following form:

[8] Perhaps the closest to an energy-related sanction has been the sanctions imposed on Iran to deter them from developing a nuclear program. Russia's threats to curtail natural gas sells to Europe is another recent example, although of course it is not motivated by energy or climate issues.

$$C_R\big(g_R, g_H(g_R)\big) + \sigma L\big(g_R, g_H(g_R)\big) \qquad \text{Eq. 8}$$

where $L\big(g_R, g_H(g_R)\big)$ is the sanction the Hegemon imposes as a function of the policy choice of the Target and as before, $g_H(g_R)$ is the Hegemon's best response. Here, we introduce the parameter $\sigma \in [0,1]$ to capture the capacity of the Hegemon to inflict damages in the Target country via sanctions. For example, if the Target can divert trade flows via other trade partners, then the effects of the sanction are diluted but the costs of imposing the sanction remain the same.

The sanction needs to be incentive compatible, so that the Target is at least as well off as under Nash by behaving like the Hegemon demands. This implies

$$C_R\Big(g_R^S, g_H(g_R^S)\Big) + \sigma L(g_R^S, g_H^S) \le C_R\big(g_R,\ g_H(g_R)\big) + \sigma L\big(g_R,\ g_H(g_R)\big)$$

If the Hegemon's threat of a sanction is credible, the Target should respond by behaving the way the Hegemon demands and then $L(g_R^S, g_H^S) = 0$. The sanction is then given by

$$L(g_R, g_H) = \max\big\{0, (1/\sigma)\big[C_R\big(g_R,\ g_H(g_R)\big) - C_R(g_R^U, g_H^U)\big]\big\} \qquad \text{Eq. 9}$$

The problem of the Hegemon is now

$$\{g_H^S, g_R^S\} = \arg\min\ \{C_H(g_H, g_R) + (1/\sigma)[C_R(g_R,\ g_H) - C_R(g_R^U, g_H^U)]\} \qquad \text{Eq. 10}$$

The Hegemon will only threaten sanctions when they prefer cooperation in an agreement and the Target is unwilling to participate without additional incentive. This is the space described in Proposition 2 and shown in Figure 5. When $\sigma = 1$ the equilibrium outcome is the same allocation as under the coordinated equilibrium. As the effectiveness of the sanctions decline, $\sigma < 1$, equilibrium policies move along the contract curve away from the

Hegemon's preferred outcome. How far along the curve ultimately depends on the value of σ.

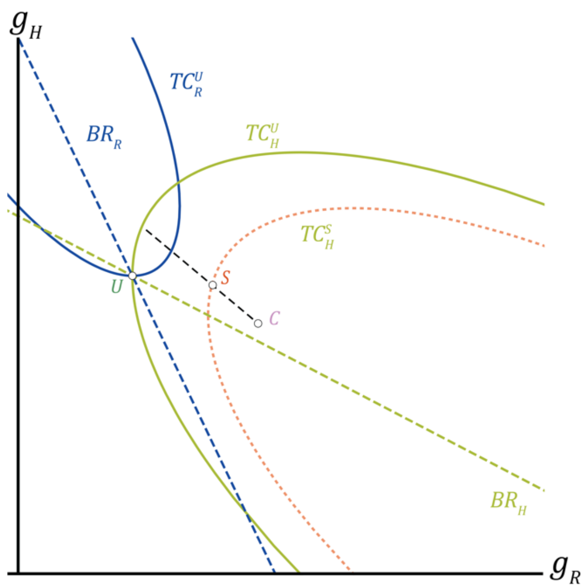


Figure 9. Policy Outcomes – Sanctions

Figure 9. Policy Outcomes – Sanctions provides an example of an equilibrium with sanctions when the Hegemon prefers the Coordinated equilibrium C but the Target needs additional incentives. With the threat of sanctions, the Hegemon can move the equilibrium S closer to the Coordinated equilibrium along the contract curve. We consider costly sanctions, $\sigma < 1$, so the Hegemon cannot move the equilibrium all the way to the coordinated equilibrium.

We again return to our question of when the Hegemon prefers and agreement, now with sanctions, to the non-coordinated equilibrium and how this depends on preference asymmetry and policy exposure.

Proposition 5. (Agreement with Sanctions (S) versus Uncoordinated(U) Equilibrium Outcomes). There exists some $\overline{\Delta}_H^{SU}(\gamma;\sigma)$, $\underline{\Delta}_H^{TU}(\gamma;\sigma)$, $\overline{\Delta}_R^{SU}(\gamma;\sigma)$, $\underline{\Delta}_R^{SU}(\gamma;\sigma)$, and $\overline{\overline{\Delta}}_H^{SU}(\gamma;\sigma)$ such that

(i) If $\underline{\Delta}_H^{SU}(\gamma;\sigma) < \Delta < \overline{\Delta}_H^{SU}(\gamma;\sigma)$ then $U \succ S$

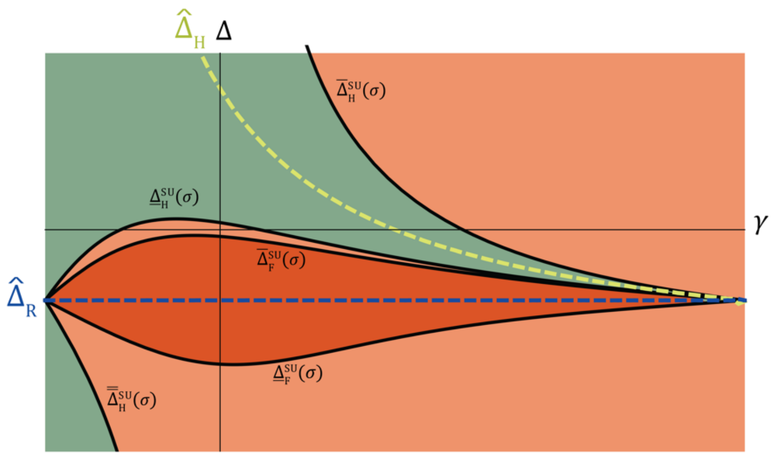
(ii) If $\Delta < \overline{\overline{\Delta}}_H^{SU}(\gamma;\sigma)$ then $U \succ S$

(iii) If $\underline{\Delta}_R^{SU}(\gamma;\sigma) < \Delta < \overline{\Delta}_R^{SU}(\gamma;\sigma)$ then $S \succ U$ and $S > 0$

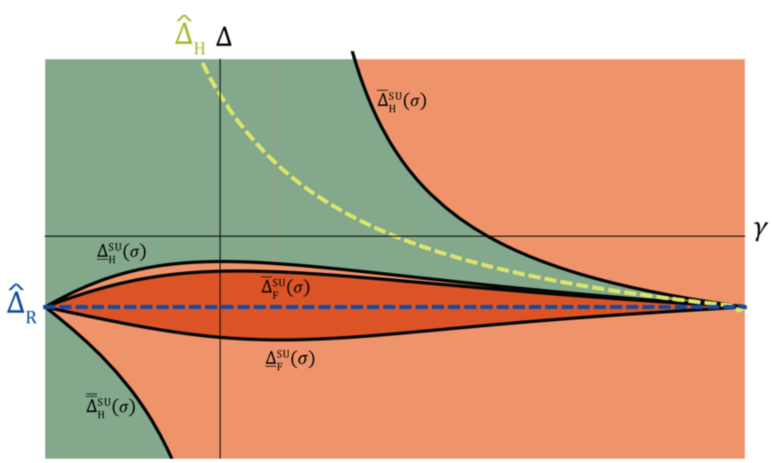
(iv) Else, $S \succ U$ and $S = 0$

Proof. See the Appendix.

Figure 10. illustrates the results given by Proposition 5. Let's first consider costless sanctions, Panel A, and then we will discuss how outcomes change when sanctions become costly, Panel B. For $\sigma = 1$, outcomes are the same as the Coordinated equilibrium when the Target prefers the Coordinated equilibrium to the Uncoordinated equilibrium. Thus, $\overline{\Delta}_H^{SU}(\gamma, 1) = \overline{\Delta}_H^{CU}(\gamma)$, $\underline{\Delta}_H^{SU}(\gamma, 1) = \underline{\Delta}_H^{CU}(\gamma)$, and $\overline{\overline{\Delta}}_H^{SU}(\gamma, 1) = \overline{\overline{\Delta}}_H^{CU}(\gamma)$. But, unlike in the Coordinated equilibrium, in the range $\underline{\Delta}_F^{SU}(\gamma, 1) < \Delta < \overline{\Delta}_F^{SU}(\gamma)$, the threat of sanction from the Hegemon is sufficient to incentivize the Target to provide optimal policy effort when it would not be incentive compatible in the absence of sanctions. This makes Sanctions stable in this region.



A. Costless sanctions $(\sigma = 1)$



B. Costly sanctions $(0 < \sigma < 1)$

Figure 10. Agreement with Sanctions vs Uncoordinated Equilibria Outcomes

As σ increases, sanctions become costly, so the threat of sanctions carries less power. As a result, the equilibrium policy outcome moves along the contract curve towards the Total Cost minimizing policy provision for the Target and, typically, further from the Total Cost minimizing policy provision of the Hegemon. It follows that the Hegemon's preference for an Agreement with Sanctions weakens as sanctions become more costly. For example, in the extreme case of infinitely costly sanctions, $\sigma = 0$, the Hegemon will always prefer the Uncoordinated equilibrium regardless of Policy Externality. Figure 10b shows how moderately costly sanctions change the Hegemon's preference for an Agreement with Sanctions relative to the Uncoordinated equilibrium.

5.4. Ranking of Interventions under different Policy Externalities

We can now compare how the Hegemon would rank each of the policy interventions depending on their policy externality. To grasp the intuition, let's first start with an example. Figure 11. Policy Outcomes combines the Agreement with Transfers and Agreement with Sanctions equilibria from Figure 7. Policy Outcomes – Transfers and Figure 9. Policy Outcomes – Sanctions on a single figure. From the perspective of the Hegemon, the ranking of preference is $C \sim E \succ S \succ T \succ U$, but C and E are not incentive compatible for the Target so the preferred intervention is sanctions. By offering transfers, the Hegemon incentivizes the Target to increase policy provision, however this comes at the cost of the transfers. With sanctions, there is no actual transfer of resources from the Hegemon to the Target. Thus, the Hegemon can do better by threatening the Target with sanctions, even if sanctions do not reach the optimal policy provision.

This raises the question of whether this ordering of policy preference is always the same. A quick glance at the symmetric case proves this is not the case, as the Hegemon prefers the agreement intervention without sanctions or rewards under symmetry. Proposition 6 illustrates how preference asymmetry and policy exposure jointly determine the ordering of policy preference.

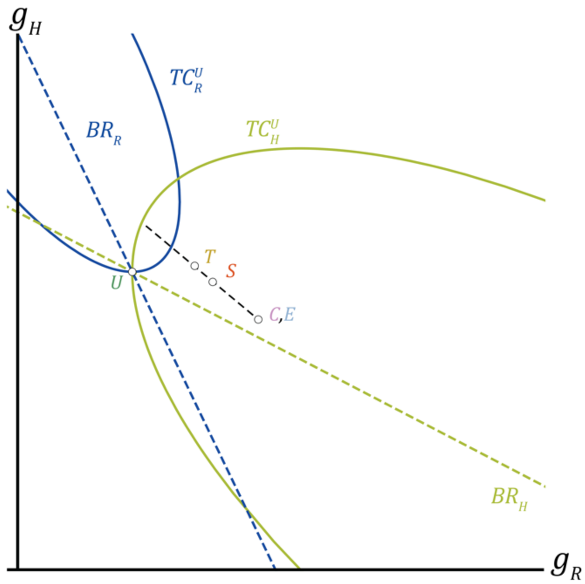


Figure 11. Policy Outcomes

Proposition 6. (Ordering of policy intervention preference)

(i) If $\Delta > \overline{\Delta}^{\mathrm{TE}}(\gamma)$ or $\Delta < \underline{\Delta}^{\mathrm{TE}}(\gamma)$ then $E \succ T \succcurlyeq S$

(ii) If $\overline{\Delta}^{ST}(\gamma;\sigma) < \Delta < \overline{\Delta}^{TE}(\gamma)$ or $\underline{\Delta}^{ST}(\gamma;\sigma) > \Delta > \underline{\Delta}^{TE}(\gamma)$ then $T \succ S$

(iii) If $\underline{\Delta}^{ST}(\gamma;\sigma) < \Delta < \overline{\Delta}^{ST}(\gamma;\sigma)$ then $S \succ T$

Proof. See the Appendix.

Figure 12. Comparing Sanctions and Rewards. illustrates the results given by Proposition 6. Let's first consider the regions where the Target prefers coordination. From Proposition 3, in these regions the Hegemon always prefers an Agreement with Extraction to the Uncoordinated equilibrium. In these regions where the Target also prefers Coordination over the Uncoordinated equilibrium, Agreements with transfers or costless sanctions are equivalent to the Coordinated equilibrium. In an Agreement with Extraction, the Hegemon can do even better than Coordination by extracting the gains of the Target. Thus, the Hegemon also prefers the Agreement with Extraction to either transfers or sanctions in these regions.

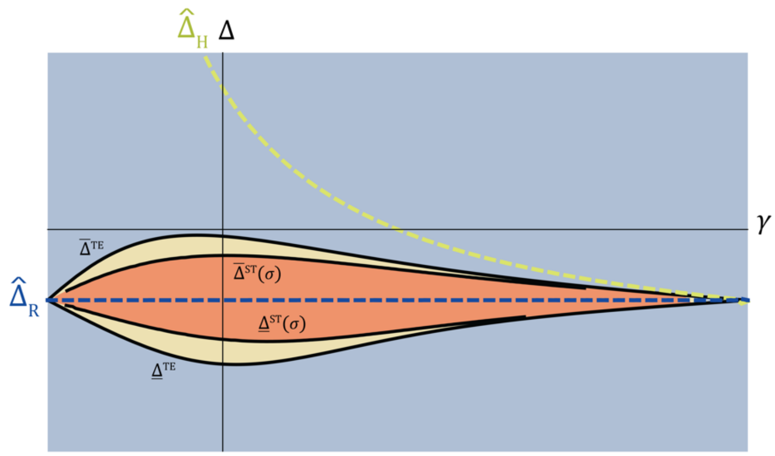


Figure 12. Comparing Sanctions and Rewards.

Now consider the region where the Target requires additional incentive to participate, $\underline{\Delta}^{TE}(\gamma,1) < \Delta < \overline{\Delta}^{TE}(\gamma)$. In this region, the Coordinated equilibrium and the Agreement with Extraction equilibrium are unstable. Thus, the Hegemon must choose between an Agreement with Transfers, with Sanctions, or the Uncoordinated equilibrium. An Agreement with Transfers is always preferred to the Uncoordinated equilibrium in this region per Proposition 4. The additional cost of transfers are less than the gains of coordination for the Hegemon. When sanctions are costless, the Hegemon always prefers sanctions to transfers because equilibrium policy outcomes are identical but the Hegemon does not have to transfer anything of value to the Target. They rely only on the threat of sanctions to provide incentive. However, as sanctions become costly, the Hegemon must trade off the cost of transfers and the cost of sanctions. With moderately costly sanctions, as illustrated in Figure 12, the Hegemon will prefer transfers when they are small and sanctions when transfers are large.

Taken together, we can see how power influences preferences for foreign intervention in domestic climate policy. When incentives for coordination are aligned, the Hegemon will use its power to extract as much as it can from the Target. And when the Target requires additional incentive to cooperate, the Hegemon will choose the intervention that provides that incentive at the least cost.

6. Conclusion

Global governance is needed to reduce the impacts of climate change. The type of governance depends on characteristics of the climate policy. For policies such as emissions mitigation, cooperation is needed to overcome the free-riding problem and increase policy provision. For policies such as solar geoengineering, coordination is needed to overcome the free-driving problem and reign in policy provision.

In this paper, we develop a unified theory of foreign intervention on domestic climate policy. Within this unified framework we analyze how preferences for different forms of intervention depend on the policy externality. By characterizing policy externality as a function of the exposure to climate interventions and preference asymmetry regarding desired objective, we remove the artificial silos around foreign climate intervention options and climate change policy options.

Within our unified framework, we specifically compare preference for alternative foreign intervention strategies for two classes of policy, those with a free-riding equilibrium and those with a free-driving equilibrium. In the absence of power asymmetries, we find countries join an agreement when one of the countries is free-driving too much in excess of the preferred optimal outcome. This result occurs when exposure to a domestic policy is high or when preferences diverge substantively. When we introduce power asymmetries, we find the powerful country can induce more cooperation from the weak country either by offering transfers or threatening sanctions. When the powerful country does not have power to directly influence domestic policy, let's say by intervening militarily, then their options need to provide enough incentives for the weak country to join. This implies that any equilibrium with intervention leads to allocations that are closer to the optimal allocation and thus reduces overall climate costs. Of course, this leaves out important questions regarding justice and distributional issues that are outside of the scope of our paper.

We are at a crossroad on climate policy: how do we move forward with seemingly risky technologies like solar geoengineering and accelerated adaptation? At the same time, we have hit a roadblock: climate negotiations have been staling and there is very little meaningful progress coming from the international community. In this paper we find a general ranking of preferred foreign intervention options that is a function of policy characteristics that are not linked to specific technological possibilities but are only characterized by the overall impacts on the climate and how they affect other countries. This framework then offers an alternative to the traditional siloed approach to international governance of climate policy. The results we present, while maybe contrary to the current wisdom, offer an alternative way to look at international governance and we hope are intriguing enough to engender further exploration.

7. References

[1]. Aidt TS, Albornoz F, Hauk E. Foreign influence and domestic policy. Journal of Economic Literature. 2021 Jun;59(2):426- 87.

[2]. Aldy JE, Zeckhauser R. Three prongs for prudent climate policy. Southern Economic Journal. 2020 Jul;87(1):3-29.

[3]. Aldy JE, Felgenhauer T, Pizer WA, Tavoni M, et al. Social science research to inform solar geoengineering. Science. 2021 Nov 12;374(6569):815-8.

[4]. Barrett S. Self-enforcing international environmental agreements. Oxford economic papers. 1994 Oct 1:878-94.

[5]. Barrett S. The strategy of trade sanctions in international environmental agreements. Resource and Energy Economics. 1997 Nov 1;19(4):345-61.

[6]. Barrett S. Environment and statecraft: The strategy of environmental treaty-making: The strategy of environmental treaty-making. OUP Oxford; 2003 Jan 9.

[7]. Barrett S. The incredible economics of geoengineering. Environmental and resource economics. 2008 Jan;39(1):45-54.

[8]. Belaia M, Moreno-Cruz JB, Keith DW. Optimal climate policy in 3D: Mitigation, carbon removal, and solar geoengineering. Climate Change Economics. 2021 Aug 2;12(03):2150008.

[9]. Black R, Bennett SR, Thomas SM, Beddington JR. Migration as adaptation. Nature. 2011 Oct;478(7370):447-9.

[10]. Burke M, Hsiang SM, Miguel E. Climate and conflict. Annu. Rev. Econ.. 2015 Aug 2;7(1):577-617.

[11]. Bollinger B, Gillingham K. Learning-by-doing in solar photovoltaic installations. Available at SSRN 2342406. 2019 Feb 19.

[12]. Bosetti V, Carraro C, De Cian E, Massetti E, Tavoni M. Incentives and stability of international climate coalitions: An integrated assessment. Energy Policy. 2013 Apr 1;55:44-56

[13]. Carraro C, Eyckmans J, Finus M. Optimal transfers and participation decisions in international environmental agreements. The Review of International Organizations. 2006 Dec;1(4):379-96.

[14]. Eyler B. Science Shows Chinese Dams Are Devastating the Mekong. Foreign Policy April 22, 2020 Accessed January 15, 2022: https://foreignpolicy.com/2020/04/22/science-shows-chinese-dams-devastating-mekong-river/

[15]. Field C, Cheung W, Dilling L, Frumhoff P, Greely H, Hordequin M, Hurrell J, Light A, Lin A, MacMartin D, McHenry R. Reflecting Sunlight: Recommendations for Solar Geoengineering Research and Research Governance. National Academies of Sciences, Engineering, and Medicine; 2021 Mar 25.

[16]. Finus M, McGinty M. The anti-paradox of cooperation: diversity may pay!. Journal of Economic Behavior & Organization. 2019 Jan 1;157:541-59.

[17]. Finus M, Pintassilgo P. The role of uncertainty and learning for the success of international climate agreements. Journal of Public Economics. 2013 Jul 1;103:29-43.

[18]. Gallagher CL, Holloway T. Integrating air quality and public health benefits in US decarbonization strategies. Frontiers in public health. 2020 Nov 19;8:563358.

[19]. Harding AR, Belaia M, Keith DW. The value of information about solar geoengineering and the two-sided cost of bias. Climate Policy. 2022 Jul 1:1-1.

[20]. Heyen D, Horton J, Moreno-Cruz J. Strategic implications of counter-geoengineering: Clash or cooperation?. Journal of Environmental Economics and Management. 2019 May 1;95:153-77.

[21]. Heyen D, Lehtomaa J. Solar geoengineering governance: a dynamic framework of farsighted coalition formation. Oxford Open Climate Change. 2021;1(1):kgab010.

[22]. Horton JB, Reynolds JL. The international politics of climate engineering: a review and prospectus for international relations. International Studies Review. 2016 Sep 1;18(3):438-61.

[23]. Hsiang SM, Burke M, Miguel E. Quantifying the influence of climate on human conflict. Science. 2013 Sep 13;341(6151).

[24]. Jinnah S, Morrow D, Nicholson S. Splitting Climate Engineering Governance: How Problem Structure Shapes Institutional Design. Global Policy. 2021 Apr;12:8-19.

[25]. Khan MR, Munira S. Climate change adaptation as a global public good: implications for financing. Climatic Change. 2021 Aug;167(3):1-8.

[26]. Lazkano I, Marrouch W, Nkuiya B. Adaptation to climate change: how does heterogeneity in adaptation costs affect climate coalitions?. Environment and Development Economics. 2016 Dec;21(6):812-38.

[27]. Li H, Rus HA. Climate change adaptation and international mitigation agreements with heterogeneous countries. Journal of the Association of Environmental and Resource Economists. 2019 May 1;6(3):503-30.

[28]. McGinty M. International environmental agreements among asymmetric nations. Oxford Economic Papers. 2007 Jan 1;59(1):45-62.

[29]. Mendelsohn R. Efficient adaptation to climate change. Climatic Change. 2000 Jun;45(3):583-600.

[30]. Millard-Ball A. The tuvalu syndrome. Climatic Change. 2012 Feb;110(3):1047-66.

[31]. Moreno-Cruz JB. Mitigation and the geoengineering threat. Resource and Energy Economics. 2015 Aug 1;41:248-63.

[32]. Moreno-Cruz J, Wagner G, Keith D. An economic anatomy of optimal climate policy. CESifo Working paper 7059. May 2018

[33]. Parson EA, Reynolds JL. Solar geoengineering governance: Insights from a scenario exercise. Futures. 2021 Sep 1;132:102805.

[34]. Reynolds JL. Linking solar geoengineering and emissions reductions: strategically resolving an international climate change policy dilemma. Climate Policy. 2021 Oct 26:1-6.

[35]. Ricke, Katharine; Moreno-Cruz, Juan (2020): Geo-Wedges: A Portfolio Approach to Geoengineering the Climate. In : Reference Module in Earth Systems and Environmental Sciences, vol. 531: Elsevier, p. 362.

[36]. Ricke KL, Moreno-Cruz JB, Caldeira K. Strategic incentives for climate geoengineering coalitions to exclude broad participation. Environmental Research Letters. 2013 Feb 12;8(1):014021.

[37]. Ricke K, Ivanova D, McKie T, Rugenstein M. Reversing Sahelian Droughts. Geophysical Research Letters. 2021 Jul 28;48(14):e2021GL093129.

[38]. Rickels W, Quaas MF, Ricke K, Quaas J, Moreno-Cruz J, Smulders S. Who turns the global thermostat and by how much? Energy Economics. 2020 Sep 1;91:104852.

[39]. Schelling TC. The Economic Diplomacy of Geoengineering. Climatic Change. 1996; 33:303-307

[40]. Sayegh S, Heutel G, Moreno-Cruz J. Regulating Geoengineering: International Competition and Cooperation. Climate Geoengineering: Science, Law and Governance 2021 (pp. 229-247). Springer, Cham.

[41]. Urpelainen J. Geoengineering and global warming: a strategic perspective. International Environmental Agreements: Politics, Law and Economics. 2012 Nov;12(4):375-89.

[42]. Waldinger M. The economic effects of long-term climate change: evidence from the Little Ice Age. Journal of Political Economy. 2022 Sep 1;130(9):2275-314.

[43]. Weitzman ML. A voting architecture for the governance of free-driver externalities, with application to geoengineering. The Scandinavian Journal of Economics. 2015 Oct;117(4):1049-68.

8. Appendix

8.1. Proofs

Proof of Proposition 1.

$g_H^U = 0$ when $\widehat{\Delta}^H(\gamma) = \Delta^H\left(\frac{2}{\gamma} - 1\right)$

$g_R^U = 0$ when $\widehat{\Delta}^R(\gamma) = -\frac{\Delta^H}{2}$

Proof of Proposition 2.

Proof of each of the delimitating curves come from the conditions of incentive compatibility for the Hegemon and the Target. These can be described as follow.

Part i. Hegemon

The Hegemon is indifferent between the Coordinated and Uncoordinated equilibria when

$\Delta = \Delta^H \frac{(a \pm b)}{c}$ where $a = 16 - 4\gamma + 31\gamma^2 + 26\gamma^3 - 33\gamma^4 + 21\gamma^5 - 7\gamma^6$, $b = \frac{\gamma^2(9+2\gamma^2)}{\sqrt{\frac{\gamma^2(9+2\gamma^2)}{(20-13\gamma+10\gamma^2-2\gamma^3)^2}}}$ and $c = -32 + 16\gamma - 16\gamma^2 + 24\gamma^3 + 12\gamma^4 - 6\gamma^5 + 7\gamma^6$

Part ii. Target

The Target is indifferent between the Coordinated and Uncoordinated equilibria when

$\Delta = \Delta^H \frac{(a \pm b)}{c}$ where $a = 36 - 32\gamma + 52\gamma^2 + \gamma^3 - 10\gamma^4 + 9\gamma^5 - \gamma^6$, $b = \frac{\gamma^2(10+\gamma^2)}{\sqrt{\frac{\gamma^2(9+2\gamma^2)}{(20-13\gamma+10\gamma^2-2\gamma^3)^2}}}$ and $c = -72 + 32\gamma - 72\gamma^2 + 20\gamma^3 - 11\gamma^4 - 8\gamma^5 + \gamma^6$

Proof of Proposition 3.

The difference in total costs for the Hegemon between the Agreement with Extraction and Uncoordinated Equilibria can be expressed as the difference in the sum of Total Costs for both countries in the coordinated and uncoordinated equilibria:

$$TC_H^U - TC_H^E = TC_H^U - (TC_H^C - M)$$

$$= TC_H^U - \left(TC_H^C - (TC_R^U - TC_R^C)\right)$$
$$= (TC_H^U + TC_R^U) - (TC_H^C + TC_R^C) \geq 0$$

The inequality in the final line comes from the following reasoning. By definition, both countries choose policy to minimize joint total costs in the coordinated equilibrium. Thus, the joint total costs in the Uncoordinated equilibrium are always at least weakly larger than the joint total costs in the Coordinated equilibrium. Thus, the Hegemon always weakly prefers the Agreement with Extraction. However, stability requires incentive compatibility for both the Hegemon and the Target. Conditions for incentive compatibility stemming from the Target, $\overline{\Delta}^{EU}$ and $\underline{\Delta}^{EU}$, follow from Part ii of the Proof of Proposition 2.

Proof of Proposition 4.

Transfers are set by the Hegemon to leave the Target at least as well of as in the Uncoordinated equilibrium. Thus, the only condition for stability is incentive compatibility for the Hegemon. Let's first consider the case of positive Transfers.

Transfers are positive when the Target requires additional incentive to provide policy consistent with the Coordinated equilibrium. Thus, the curves $\overline{\Delta}_R^{TU}$ and $\underline{\Delta}_R^{TU}$ which distinguish the location of positive transfers in the policy exposure space are given by the indifference of the Target between the Coordinated and Uncoordinated equilibria. Proof of this condition is as in Part ii of the Proof of Proposition 2.

Let's now consider stability from the perspective of the Hegemon. When Transfers are zero, i.e. $\Delta > \overline{\Delta}_R^{TU}$ and $\Delta < \underline{\Delta}_R^{TU}$, the equilibrium outcome is the Coordinated equilibrium. Thus, proof of delimiting curves $\overline{\Delta}_H^{TU}$, $\underline{\Delta}_H^{TU}$, and $\overline{\overline{\Delta}}_H^{TU}$ follow Part i of the Proof of Proposition 2. When transfers are positive, the Hegemon prefers the Agreement with Transfers to the Uncoordinated Equilibrium if the Total Cost with transfers is less than the Total Cost in the Uncoordinated Equilibrium. We can express this as

$$TC_H^U - TC_H^T = TC_H^U - (TC_H^C - W)$$
$$= TC_H^U - \left(TC_H^C - (TC_R^U - TC_R^C)\right)$$

$$= (TC_H^U + TC_R^U) - (TC_H^C + TC_R^C) \geq 0$$

Reasoning for the inequality in the final line is as in the Proof of Proposition 3.

Proof of Proposition 5.

Proof of each of the delimitating curves come from the conditions of incentive compatibility for the Hegemon and the Target. These can be described as follow.

Part i. Hegemon

The Hegemon is indifferent between the Agreement with Sanctions and Uncoordinated equilibria when $\Delta = \Delta^H \frac{(a \pm b)}{c}$ where,

$$a = \Big(4 + 6\sigma^3\gamma_{RH}^4 + 8\sigma(4 - 2\gamma_{RH} + \gamma_{RH}^2) - \sigma^4\gamma_{RH}^3(-16 + 20\gamma_{RH} - 9\gamma_{RH}^2 + \gamma_{RH}^3) + \sigma^2\gamma_{RH}(-16 + 44\gamma_{RH} - 15\gamma_{RH}^2 + 4\gamma_{RH}^3)\Big)$$

$$b = (-4 + \gamma)\gamma(1 + (4 - 2\gamma + \gamma^2)\sigma + \gamma^2\sigma^2)\sqrt{1 + 8\sigma + 2\gamma^2\sigma^2}$$

$$c = -32 + 16\gamma - 16\gamma^2\sigma + 2\gamma^5(1 - 4\sigma)\sigma^2 + 8\gamma^3\sigma(1 + 2\sigma) + \gamma^6\sigma^2(1 + 4\sigma + 2\sigma^2) + 2\gamma^4\sigma(1 - 3\sigma + 8\sigma^2)$$

Part ii. Target

The Target is indifferent between the Agreement with Sanctions and Uncoordinated equilibria when $\Delta = \Delta^H \frac{(a \pm b)}{c}$ where,

$$a = (4 + 8(4 - 2\gamma + \gamma^2)\sigma + \gamma(-16 + 44\gamma - 15\gamma^2 + 4\gamma^3)\sigma^2 + 6\gamma^4\sigma^3 - \gamma^3(-16 + 20\gamma - 9\gamma^2 + \gamma^3)\sigma^4)$$

$$b = (-4 + \gamma)\gamma\sigma\sqrt{2 + 8\sigma + \gamma^2\sigma^2}(1 + (4 - 2\gamma + \gamma^2)\sigma + \gamma^2\sigma^2)$$

$$c = -8 - 16(4 - 2\gamma + \gamma^2)\sigma + \gamma^2(-56 + 20\gamma - 7\gamma^2)\sigma^2 - 12\gamma^4\sigma^3 + \gamma^4(8 - 8\gamma + \gamma^2)\sigma^4$$

Proof of Proposition 6.

This proof is the culmination of the proofs for the previous propositions. First consider the regions in the policy exposure space $\Delta > \overline{\Delta}_R^{TE}$ and $\Delta < \underline{\Delta}_R^{TE}$. Comparing the Agreement with Extractions to the Agreement with Transfers, the Hegemon strictly prefers the Agreement with Extractions. The equilibrium for the Agreement with Transfers in this region of the space is equivalent to the Coordinated equilibrium. In the Agreement with Extractions, the

Hegemon's outcome is determined by the Coordinated equilibrium plus the value extracted from the gains, $M > 0$, of the Target. Thus, the Hegemon will always prefer the Agreement with Extraction in these regions because of the additional value from extracting gains to the Target. Additionally comparing an Agreement with Sanctions, when $\sigma = 1$, the Agreement with Transfers and the Agreement with Sanctions are equivalent in these regions because there is no transfer of material in either Agreement. However, when $\sigma < 1$, the Hegemon prefers an Agreement with Transfers because of the weakened power of costly sanctions.

Second, consider the region in the policy exposure space $\underline{\Delta}_R^{TE} < \Delta < \overline{\Delta}_R^{TE}$. Here, the Agreement with Exposure is unstable, as discussed in the Proof of Proposition 3. From the Proof of Proposition 4, Transfers are always stable and preferred to the Uncoordinated equilibrium. When $\sigma = 1$ sanctions are costless, so the Hegemon prefers the Agreement with Sanctions, where no material of value is transferred to the Target, to an Agreement with Transfers. As σ decreases, sanctions become costly and the Hegemon weighs the cost of sanctions against the cost of transfers. The Hegemon is indifferent between costly transfers and costly sanctions in this region when $\Delta(\sigma) = \Delta^{\mathrm{H}} \frac{(a \pm b)}{c}$ where,

$$a = 84. - 56.\gamma + 36.\gamma^2 - 9.\gamma^3 + 1.\gamma^4 + (32. - 48.\gamma + 112.\gamma^2 - 78.\gamma^3 + 64.\gamma^4 - 18.\gamma^5 + 2.\gamma^6)\sigma + (64. - 128.\gamma + 216.\gamma^2 - 188.\gamma^3 + 175.\gamma^4 - 103.\gamma^5 + 50.\gamma^6 - 11.\gamma^7 + 1.\gamma^8)\sigma^2 + \gamma^2(32. + 112.\gamma - 192.\gamma^2 + 172.\gamma^3 - 78.\gamma^4 + 22.\gamma^5 - 2.\gamma^6)\sigma^3 + \gamma^4(4. + 16.\gamma - 17.\gamma^2 + 9.\gamma^3 - 1.\gamma^4)\sigma^4$$

$$b = \gamma(-1760. + 2224.\gamma - 1822.\gamma^2 + 892.\gamma^3 - 241.\gamma^4 + 34.\gamma^5 - 2.\gamma^6 + (-11200. + 20960.\gamma - 23932.\gamma^2 + 17376.\gamma^3 - 8780.\gamma^4 + 2988.\gamma^5 - 634.\gamma^6 + 76.\gamma^7 - 4.\gamma^8)\sigma + (-4800. + 21280.\gamma - 43244.\gamma^2 + 49516.\gamma^3 - 39278.\gamma^4 + 21790.\gamma^5 - 8569.\gamma^6 + 2322.\gamma^7 - 407.\gamma^8 + 42.\gamma^9 - 2.\gamma^{10})\sigma^2 + (51200. - 112640.\gamma + 145024.\gamma^2 - 119296.\gamma^3 + 67136.\gamma^4 - 24892.\gamma^5 + 4798.\gamma^6 + 232.\gamma^7 - 400.\gamma^8 + 96.\gamma^9 - 8.\gamma^{10})\sigma^3 + (25600. - 43520.\gamma + 86272.\gamma^2 - 99040.\gamma^3 + 85724.\gamma^4 - 50328.\gamma^5 + 19775.\gamma^6 - 3934.\gamma^7 - 507.\gamma^8 + 592.\gamma^9 - 187.\gamma^{10} + 30.\gamma^{11} - 2.\gamma^{12})\sigma^4 + (40960. - 98304.\gamma + 164352.\gamma^2 - 177408.\gamma^3 + 158144.\gamma^4 - 107200.\gamma^5 + 59244.\gamma^6 - 24948.\gamma^7 + 7762.\gamma^8 - 1632.\gamma^9 + 188.\gamma^{10} - 8.\gamma^{11})\sigma^5 + \gamma^2(25600. - 48640.\gamma + 66560.\gamma^2 - 59104.\gamma^3 + 42268.\gamma^4 - 22660.\gamma^5 + 9752.\gamma^6 - 3138.\gamma^7 + 712.\gamma^8 - 108.\gamma^9 + 8.\gamma^{10})\sigma^6 + 8(4. - 1.\gamma)^2\gamma^4(4. + 1.\gamma^2)(4 - 2.\gamma + 1.\gamma^2)(2.5 - 1.\gamma + 1.\gamma^2)\sigma^7 + 2.(4. - 1.\gamma)^2\gamma^6(4. + 1.\gamma^2)(2.5 - 1.\gamma + 1.\gamma^2)\sigma^8)^{1/2}$$

$$c = -168. + 144.\gamma - 104.\gamma^2 + 28.\gamma^3 - 3.\gamma^4 + (-64. + 32.\gamma - 160.\gamma^2 + 72.\gamma^3 - 62.\gamma^4 + 16.\gamma^5 - 2.\gamma^6)\sigma + (-128. + 128.\gamma - 240.\gamma^2 + 176.\gamma^3 - 166.\gamma^4 + 78.\gamma^5 - 45.\gamma^6 + 10.\gamma^7 - 1.\gamma^8)\sigma^2 + \gamma^2(-64. + 32.\gamma - 72.\gamma^3 + 44.\gamma^4 - 20.\gamma^5 + 2.\gamma^6)\sigma^3 + (-8.\gamma^4 + 2.\gamma^6 - 8.\gamma^7 + \gamma^8)\sigma^4$$