RESOURCES
for the FUTURE

International Climate Agreements under The Threat of Solar Geoengineering

David M. McEvoy, Matthew McGinty, Todd L. Cherry, and Stephan Kroll

Working Paper 23-36
September 2023

About the Authors

David M. McEvoy is a professor and department chair of the Department of Economics at Appalachian State University.

Matthew McGinty is a professor in the Economics Department at the University of Madison Milwaukee.

Todd L. Cherry is a professor and the John S. Bugas Chair at the University of Wyoming College of Business, where he is also the director of graduate studies.

Stephan Kroll is an associate professor in the Department of Agricultural Resource Economics at Colorado State University.

About the Project

The Resources for the Future Solar Geoengineering research project applies tools from multiple social science research disciplines to better understand the risks, potential benefits, and societal implications of solar geoengineering as a possible approach to help reduce climate risk alongside aggressive and necessary mitigation and adaptation efforts. The project began in 2020 with a series of expert workshops convened under the SRM Trans-Atlantic Dialogue. These meetings resulted in a 2021 article in Science that lays out a set of key social science research questions associated with solar geoengineering research and potential deployment. The Project followed this with additional sponsored research, including a competitive solicitation designed to address research areas highlighted in the Science article. This paper is one of eight research papers resulting from that competition and supported by two author workshops. A key goal of the solicitation and the overall project is to engage with a broader set of researchers from around the globe, a growing number of interested stakeholders, and the public.

About RFF

Resources for the Future (RFF) is an independent, nonprofit research institution in Washington, DC. Its mission is to improve environmental, energy, and natural resource decisions through impartial economic research and policy engagement. RFF is committed to being the most widely trusted source of research insights and policy solutions leading to a healthy environment and a thriving economy.

Working papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review. The views expressed here are those of the individual authors and may differ from those of other RFF experts, its officers, or its directors.

Sharing Our Work

Our work is available for sharing and adaptation under an Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license. You can copy and redistribute our material in any medium or format; you must give appropriate credit, provide a link to the license, and indicate if changes were made, and you may not apply additional restrictions. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. You may not use the material for commercial purposes. If you remix, transform, or build upon the material, you may not distribute the modified material. For more information, visit **https://creativecommons.org/licenses/by-nc-nd/4.0/**

International Climate Agreements under the Threat of Solar Geoengineering

David M. McEvoy, Matthew McGinty, Todd L. Cherry and Stephan Kroll

November 13, 2022

Abstract

The possibility of overshooting global emissions targets has triggered a public debate about the role solar geoengineering (SGE) - using technologies to reflect solar radiation away from Earth - may play in managing climate change. One major concern is that SGE technologies are relatively cheap, and could potentially be deployed by a single nation (the "free driver") that could effectively control the global climate. Another concern is that SGE opportunities may alter countries' incentives to cooperate on abatement. Here we develop a game-theoretic model to analyze how opportunities to deploy SGE impact global abatement and the effectiveness of international environmental agreements (IEAs) on climate change. We show that non-cooperative abatement levels may increase or decrease under the threat of SGE, depending on how damaging the free-driver's level of deployment is on others. We also show the stability of IEAs that govern abatement is challenged by two competing strategic incentives. One is a familiar free-rider incentive, which is the benefit a country earns by leaving an agreement and lowering its abatement. The other incentive is the benefit a country earns by joining an agreement and increasing abatement in order to motivate the free-driver to reduce its level of deployment. We introduce the term *anti-driver* to describe this second incentive. Ultimately, we find that if the anti-driver incentives are high enough, the threat of SGE can expand both the depth (i.e., abatement level) and breadth (i.e., participation level) of stable IEAs compared to a world without SGE.

Keywords: solar geoengineering; solar radiation management; international environmental agreements; self-enforcing agreements; global public goods
JEL classification: C7, D7, F5, H4

1 Introduction

As global emissions of greenhouse gases (GHGs) keep increasing, many are concerned that international efforts focused on mitigation will not be sufficient to avoid excessive warming. The possibility of overshooting emissions targets has triggered a public debate about the role climate interventions - like solar geoengineering - may play in managing global warming. Solar geoengineering (SGE) - also called solar radiation management - describes the process of cooling the planet by reflecting solar radiation away from Earth. There are many concerns with researching and deploying such technologies, and one of the major concerns is international governance. SGE technologies are relatively cheap, and could potentially be deployed by a single nation that could effectively control the global climate. Moreover, the availability of SGE opportunities may alter countries' incentives to cooperate on emissions reductions. In this paper we turn to game theory to help inform us about how SGE opportunities may impact emissions abatement efforts and the overall effectiveness of international environmental agreements (IEAs) on climate change.

The term SGE describes a portfolio of climate intervention technologies that intentionally reflect sunlight away from Earth (Keith et al. 2010). While SGE can take different forms (e.g., marine cloud brightening, space-based mirrors) most attention is focused on stratospheric aerosol scattering (Keith 2000; Crutzen 2006; National Research Council 2015; Smith and Wagner 2018; Wagner 2021), which is the process of injecting aerosols into the upper atmosphere. This form of SGE is estimated to be the least expensive and fastest method of reducing global mean temperatures (Keith et al. 2010). SGE has received growing attention in academic communities (Aldy et al. 2021), policy debates (Keith 2021; Biermann et al. 2022) and recent reports from the intergovernmental panel on climate change (IPCC 2022).

SGE technologies hold the potential to reduce some of the harmful effects of climate change that result from a build up of greenhouse gas emissions in the atmosphere. The technologies also introduce difficult challenges. Stratospheric aerosol scattering is estimated to be inexpensive; enough so that a single wealthy nation could unilaterally deploy an SGE program (Barrett 2008). This means that a single actor has the potential to determine the global average temperature. Wagner and Weitzman (2012) introduced the term "free driver" to describe this possibility. The "free" part is an exaggeration of how cheap the technology is, and "driver" captures the idea that a single actor can drive the global temperature. Since countries will likely have different preferences regarding ideal temperatures, too much or too little SGE could be costly. For this reason SGE is often considered a "good-or-bad" (GoB), depending on the level of deployment (Weitzman 2015; Abatayo et al. 2020; Wagner 2021).

Another potential challenge with SGE is that it could impact incentives to mitigate GHGs (Reynolds 2019). On some level, emissions abatement and SGE are policy substitutes since both activities can reduce the damaging impacts from climate change. The impacts of SGE, however, are temporary and quickly dissipate if deployment stops. Moreover, SGE cannot completely offset damages from a build up of GHGs. For example, SGE would not address (and could exacerbate) problems of ocean acidification associated with GHGs (Williamson and Turley 2012). Since SGE is a new and emerging technology, it may also impose costly and unintended side effects. Given the uncertainty about the effectiveness of these technologies, the most prominent concern is that the availability of SGE may reduce countries' incentives to mitigate GHGs and thereby causing an increase in the stock of GHGs. This concern is often referred to as "moral hazard", but is more accurately defined as "crowding out" incentives to mitigate (Wagner 2021, p. 118).

Our overarching objective is to better understand how the availability of SGE technologies could impact incentives for countries to mitigate GHG emissions and the overall effectiveness of an international agreement on climate change. The model we develop begins with familiar functional forms from the game-theoretic literature on agreements on global greenhouse gas emissions (e.g., Barrett 1994; Finus and McGinty 2019), and we add the availability of SGE to the decision space. The characterization of SGE in the model follows the current understanding in the literature of how the technologies would impact individual and collective welfare. Most importantly, the deployment of SGE by any country will impact temperatures for all countries, and countries are assumed to have heterogeneous preferences for their ideal level of SGE (e.g., Ricke et al. 2013; Weitzman 2015; Heyen et al. 2019).

Other studies have developed game-theoretic models to analyze strategic interactions among countries in the context of SGE. One branch of this literature explores decisions to deploy SGE without including mitigation decisions, and therefore these studies are unable to explore interactions between the two types of investments. Ricke et al. (2013) introduce a model in which SGE can be deployed to decrease damages from impending climate change. Countries have heterogeneous and exogenously determined preferred levels of SGE (it's a GoB), and they explore the effectiveness of coalitions that set SGE levels to maximize joint payoffs of the members. The coalitions can intentionally exclude others from joining, and only members of a coalition can determine how SGE is deployed. They find that large coalitions can be sustained in this environment, but their approach avoids the free-driver problem and associated governance challenges by assuming non-members cannot deploy SGE.

Weitzman (2015) also explores a model in which SGE is deployed to minimize climate damages, and countries have exogenously determined preferred levels. Weitzman shows that the country with the highest preferred level (or lowest preferred temperature) will act as the free-driver to set the global temperature. His model considers a voting architecture that, under certain conditions, can lead to efficient deployment of SGE. Abatayo et al. (2020) test the free-driver hypothesis using a simple model and set of experiments in which countries, like in Weitzman (2015), have exogenously determined preferred levels of SGE. Abatayo et al. find evidence in support of the free-driver hypothesis and find both inefficiencies in SGE and counter-SGE investments. Heyen et al. (2019) also model a world with heterogenous preferences for SGE and the underlying free-driver problem. They show that when countries have the chance to invest in counter-SGE (to offset SGE deployment by those countries with higher preferences), groups of countries can be motivated to cooperate toward a more efficient solution. It is important to reiterate that all of these studies (Ricke et al. 2013; Weitzman 2015; Heyen et al. 2019; Abatayo et al. 2020) focus on SGE decisions without considering abatement opportunities.

Another branch of the literature directly explores the link between emissions abatement and SGE deployment. Moreno-Cruz (2015) propose a two-country model with both mitigation and SGE decisions, and the two investments are imperfect policy substitutes. They also introduce costly side effects from unintended consequences from SGE deployment. Both SGE and mitigation are modeled as global public goods without heterogeneity in preferred levels (i.e., SGE is not modeled as a GoB), but countries can differ by the potential damages they suffer. They show that the availability of SGE can cause aggregate abatement levels to decrease (crowding out) or increase (crowding in) depending on how similar the countries are in terms of the damages they suffer. Cherry et al. (2022) explore the moral hazard conjecture using a laboratory experiment with both mitigation and SGE decisions. On average, they find that the threat of SGE increases mitigation efforts (i.e., crowding in). Note that Moreno-Cruz (2015) and Cherry et al. (2022) do not explore agreements that govern mitigation or SGE decisions.

Millard-Ball (2012) is perhaps the study with overarching research questions closest to ours. They develop a model of global mitigation under the threat of unilateral SGE deployment and use it to explore the stability and effectiveness of IEAs that govern mitigation. Like Moreno-Cruz (2015), abatement and SGE are modeled as imperfect policy substitutes. The Millard-Ball (2012) approach to modeling SGE differs significantly from other studies, and their modeling choices have important implications when interpreting their results. First, they model homogeneous countries with identical preferences (i.e., SGE is not a GoB). Second, deploying SGE is a binary decision and provides net benefits to the deploying country and imposes external costs to all other countries. That is, they take a unique approach and model SGE as a private good with negative externalities. Third, if more than one country decides to deploy SGE, then one randomly selected country is assumed to succeed in deployment. Since all countries want to deploy SGE and all countries are identical, each has a $1/N$ chance of being the lucky deployer. They show that if the external damages are high enough, the threat of SGE deployment can cause all countries to join a cooperative agreement on mitigation in equilibrium.

Our paper is the first to explore international agreements on emissions when SGE is modeled as a good-or-bad (GoB) and governance is complicated by the threat of a free driver (following Weitzman 2015; Wagner 2021). Our approach in this paper is to compare worlds with and without the availability of SGE technologies. We start with a standard model of global emissions abatement from the IEA literature (e.g., Barrett 1994; Barrett 2003; Finus and McGinty 2019) and build in features of SGE. We derive and compare non-cooperative and socially optimal levels, with and without opportunities for SGE. Then we explore the formation of international agreements under the threat of SGE deployment.

We model an IEA with three stages. In the first (*participation stage*), countries decide independently and simultaneously whether or not to join the agreement. In the second stage (*abatement stage*) the agreement members choose abatement levels to maximize collective payoffs while non-members choose abatement independently. In the third stage (*solar geoengineering stage*), both IEA members and non-members make SGE decisions simultaneously and independently.

Our results provide useful insights regarding the relationship between emissions abatement and countries' preferences for SGE. We show that non-cooperative abatement levels may increase or decrease under the threat of SGE, and it depends on how damaging the free-driver's level of SGE deployment is on the other countries. The size of the damages from SGE, in turn, depend on the distribution of preferences for SGE. This result helps us better understand the nuance of the "moral hazard" debate which is centered on the conjecture that SGE opportunities will undermine abatement (e.g., Reynolds 2019; Wagner 2021). Consistent with much of the established literature on IEAs (e.g., Barrett 1994, 1997; McGinty 2007; McEvoy and Stranlund 2009), we find that IEAs under the threat of SGE lead to only marginal improvements in efficiency. In the special case of homogeneous preferences, we show that SGE opportunities do not alter the well-known result in the IEA literature that the largest stable coalitions consist of three members (e.g., Barrett 1994; 2003).

With heterogeneous countries, we find that stability is challenged by two competing incentives. The *free-rider incentive* is the additional payoff a defecting member achieves by leaving the agreement since they lower their abatement responsibilities. On the other hand, the *anti-driver incentive* is the additional payoff achieved by joining an agreement and further dampening the free-driver effect through increased abatement. Ultimately, we find that if the anti-driver incentives are high enough, the threat of SGE can expand both the depth (i.e., abatement level) and breadth (i.e., participation level) of stable IEAs compared to a world without SGE. The threat of a menacing free driver can lead to more cooperation.

In section 2 we model global emissions abatement and SGE deployment. In sections 3 and 4 we derive the non-cooperative and socially optimal abatement and SGE levels. Section 5 introduces the three-stage IEA and derives the stability conditions that define equilibrium agreement sizes. In section 6 we choose parameters and simulate abatement, SGE and stable IEAs to provide further insights. The final section offers a discussion of our main findings and opportunities for future research.

2 Emissions abatement and solar geoengineering

Our approach is to start with a standard model of global emissions abatement from the IEA literature (Barrett 1994; Barrett 2003; Finus and McGinty 2019) and build in features of solar geoengineering. As a baseline and starting point, we first consider a world without the availability of solar geoengineering technologies and characterize non-cooperative and socially optimal abatement levels.

Following (Barrett 1994), we adopt a symmetric model with linear benefits and quadratic abatement costs. Emissions abatement is a pure global public good, individual abatement is denoted as q_i and aggregate abatement is denoted as Q where $Q = \sum_{i \in N} q_i$. Global benefit is $B(Q) = bQ$ and each nation has the same benefit share $\frac{1}{n}$, so each nation's benefit is $B_i(Q) = \frac{bQ}{n}$. Abatement costs, denoted as C, are convex and equal to $C_i(q_i) = \frac{c(q_i)^2}{2}$. We assume identical abatement preferences to clearly highlight the impact of introducing the SGE option and to isolate the role of heterogenous SGE preferences. If both preferences were heterogeneous then there would be countervailing forces that would obfuscate the role of SGE. This allows us to show how countries respond to SGE as an option when they have both identical and heterogeneous SGE preferences. Indeed, we will show that heterogenous SGE preferences results in different non-cooperative abatement levels, even with identical abatement preferences. Interested readers will find analytical solutions for the asymmetric abatement version of this model without SGE in Finus and McGinty (2019).

A country's payoff is $B_i(Q) - C_i(q_i)$, or

$$\pi_i(q_i, Q) = \frac{bQ}{n} - \frac{c\,(q_i)^2}{2}. \tag{1}$$

Country i chooses q_i to maximize equation (1). The non-cooperative equilibrium abatement levels are denoted q^* and Q^* which are

$$\begin{aligned} q^* &= \frac{b}{cn} \\ Q^* &= \frac{b}{c}. \end{aligned} \tag{2}$$

The socially optimal abatement levels maximize aggregate payoff $\Pi = \sum_{i \in N} \pi_i$ and are denoted q^o and Q^o.

$$\begin{aligned} q^o &= \frac{b}{c} \\ Q^o &= \frac{bn}{c}. \end{aligned} \tag{3}$$

The aggregate abatement and payoff differences between the non-cooperative outcome and the social optimum are

$$\begin{aligned} Q^o - Q^* &= \frac{b\,(n-1)}{c} > 0 \\ \Pi^o - \Pi^* &= \frac{b^2\,(n-1)^2}{2cn} > 0. \end{aligned} \tag{4}$$

2.1 Solar geoengineering

Now we consider a world in which countries have an additional channel to manage climate change through the deployment of solar geoengineering (SGE). Deployment of SGE by any country can potentially provide benefits by reducing global temperatures. SGE, however, is an imperfect substitute for emissions abatement in the sense that it does not address the root cause of the problem and cannot entirely offset all of the damages caused by GHG emissions (e.g., does not address ocean acidification problems) (Robock 2008).

Countries are heterogeneous in their preferences for the optimal level of SGE. We modeled the marginal benefit of abatement as homogeneous while introducing heterogeneity in preferences for SGE in order to isolate the individual effect of introducing an SGE option. However, there are multiple reasons why SGE preferences may be heterogeneous. Preferences may be influenced by heterogeneity in the expected unintended consequences or side effects from SGE (e.g., ozone depletion), or by heterogeneity in how climate change impacts are reduced (or exacerbated) depending on geographic location. Different ethical considerations and judgements about the "right" way to management climate change could also result in heterogeneity in preferences. Finally, the heterogeneity in SGE preferences could capture different political implications from acts of deployment.

We assume that a country's preferred level of SGE is a decreasing function of global emissions abatement. Intuitively, as the potential damages from climate change are reduced through emissions abatement, the less a country needs to rely on the new geoengineering technologies. We let the parameter γ_i capture the heterogeneity in preferences for SGE. Country i's preferred level of SGE is denoted as G_i^p and takes the following form

$$G_i^p = \gamma_i(Q^o - Q), \tag{5}$$

where the term in parentheses is the abatement gap - the difference between the socially optimal abatement level and the actual abatement level. Note that if aggregate abatement $Q = Q^o$ then the abatement gap in (5) equals zero and no country prefers positive levels of SGE. We restrict G_i^p to be non negative, and γ_i as a proportion bound between zero and one. If $Q = 0$, then $G_i^p = \gamma_i Q^o$, $\forall i \in N$ which captures that SGE is an imperfect substitute for abatement. The distribution of G_i^p in our model can be likened to the distribution of preferred levels G^* in Abatayo et al. (2020) with the main difference that the preferred levels in our model are endogenous and a decreasing function of aggregate abatement.[1]

Following Weitzman (2015) and Abatayo et al. (2020), increases in SGE benefit a country up to G_i^p, but SGE levels beyond an individual country's preferred point are costly. In this way, SGE can be both a "good" or a "bad" depending on the realized level, referred to as a

---

[1] Abatayo et al. (2020) do not directly consider the link between SGE preferences and emissions abatement. Likewise, Ricke et al. (2013) and Weitzman (2015) start with exogenously determined preferences for solar geoengineering without introducing mitigation.

"GoB" in the literature (Weitzman 2015). We denote the realized level of aggregate SGE for all n countries as G. Let $\beta(G)$ be global benefit from SGE, $\beta_i(G)$ be individual benefit and $\beta(G) = \sum_{i \in N} \beta_i(G)$.

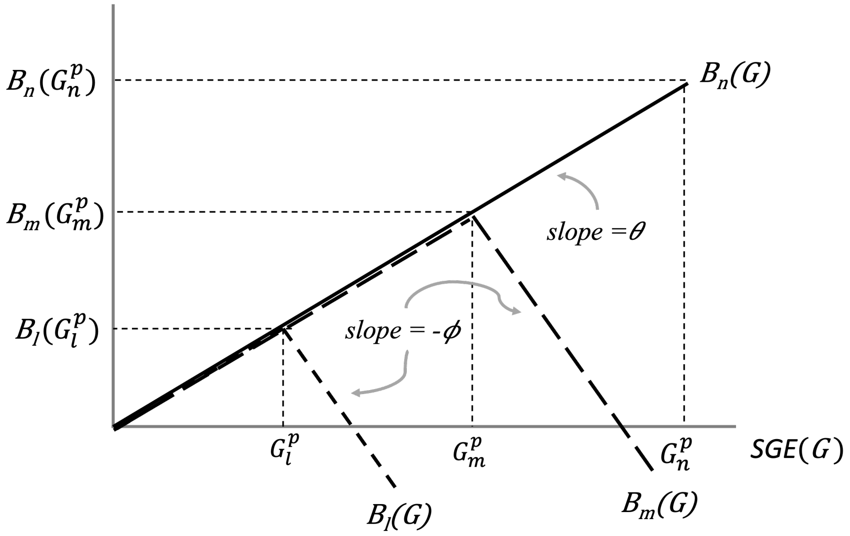
Country i's benefit function from SGE takes the following form

$$\begin{aligned} \beta_i(G) &= \theta G \text{ if } G \leq G_i^p \\ \beta_i(G) &= \theta G_i^p - \phi \left(G - G_i^p\right) \text{ if } G > G_i^p, \end{aligned} \tag{6}$$

where θ is the marginal benefit to SGE when G is less than country i's preferred level, and ϕ is the marginal loss in benefits (i.e., marginal cost) for additional deployment of SGE beyond the preferred level. Each country's benefit function is maximized at G_i^p.

Figure 1 shows example benefit functions for three countries with low (G_l^p), medium (G_m^p) and the highest (G_n^p) preferences for SGE. For SGE levels below G_l^p all n countries increase their benefits by θ for each marginal increase in G. However, increases beyond each country's preferred level results in a marginal reduction in benefits of ϕ for those countries. This can be seen in Figure 1 where the benefit function starts to decrease after peaking at the preferred level.

Figure 1: Benefits to SGE deployment depending on preference levels



We require a parameter restriction regarding the relative impact of marginal abatement and SGE decisions on payoffs that is consistent with SGE being an imperfect substitute for abatement. For all n countries, the maximum individual marginal benefit from SGE is no greater than the individual marginal benefit from emissions abatement. Specifically, this is

$$\frac{b}{n} \geq \theta. \tag{7}$$

The benefit function from SGE, $\beta_i(G)$, can turn negative for sufficiently high levels of SGE beyond a country's preferred level. This can be seen in Figure 1 where the benefit functions for the two countries with low and medium preferences for SGE cross the horizontal axis. The condition for $\beta_i(G) < 0$ is the solution to $\beta_i(G) = \theta G_i^p - \phi\left(G - G_i^p\right) < 0$ which reduces to

$$G \quad > \quad \frac{(\phi + \theta)\, G_i^p}{\phi}, \tag{8}$$

and let G_i^{neg} denote the critical level of SGE beyond which the benefit function turns negative for country i, which yields

$$G_i^{neg} \quad = \quad \frac{(\phi + \theta)\, G_i^p}{\phi}. \tag{9}$$

As an example, if $\theta = \phi$, then $G_i^{neg} = 2G_i^p$, which means that SGE levels more than twice the preferred level for country i result in negative benefits from solar geoengineering. We can use the expression for G_i^p in equation (5) to rewrite equation (9) as

$$G_i^{neg} = \left(\frac{\phi + \theta}{\phi}\right) \gamma_i (Q^o - Q), \tag{10}$$

and clearly G_i^{neg} is decreasing in Q and is equal to zero for all nations when $Q = Q^o$. So at the socially optimal abatement level, any SGE lowers all nations' payoffs. Thus, any positive SGE is strictly a bad for all nations when $Q = Q^o$.

In our model, G is determined as a best-shot technology which means the aggregate level of solar geoengineering is determined by the country that chooses the highest level of deployment. Modeling SGE as a best shot follows the description of SGE technologies in Barrett (2007) and the experiments in Cherry et al. (2022). Others have modeled G using a summation technology (Abatayo et al. 2020; Moreno-Cruz 2015). For simplicity, like Barrett (2008) and Weitzman (2015) we assume the cost of deploying SGE is extremely small and ignore it in our model.[2] Ultimately, a country's payoff function with emissions abatement and SGE opportunities is

$$\begin{aligned} \pi_i(q_i, Q, G) \quad &= \quad \frac{bQ}{n} - \frac{c\,(q_i)^2}{2} + \theta G \text{ if } G \leq G_i^p \\ \pi_i(q_i, Q, G) \quad &= \quad \frac{bQ}{n} - \frac{c\,(q_i)^2}{2} + \theta G_i^p - \phi(G - G_i^p) \text{ if } G > G_i^p. \end{aligned} \tag{11}$$

---

[2] We did explore the implications of adding a fixed cost of SGE deployment to the model. Ultimately, we discovered that introducing fixed costs results in one of two outcomes. If the fixed costs are larger than the SGE benefits for the country with the highest preferred level, then $G = 0$ and we revert back to a world without SGE opportunities. Alternatively, if the fixed costs are less than the SGE benefits for the country with the highest preferred level, there is no impact on SGE deployment, abatement levels and stable coalitions.

3 Non-cooperative abatement and solar geoengineering

In the absence of an agreement, emissions abatement decisions and SGE deployment are modeled as a two-stage game. Countries independently and simultaneously decide on abatement levels in stage 1. Given aggregate emissions Q determined from stage 1, in stage 2 countries independently and simultaneously make their SGE decisions. The game is solved by backward induction and so we begin with stage 2.

3.1 Stage 2 - solar geoengineering

Recall that the distribution of preferred levels of SGE depends on the distribution of γ_i's for all n countries. We denote the highest γ_i as $\gamma^{\max}$, and the highest preferred level as $G^{\max} = \gamma^{\max}(Q^o - Q)$. Since the benefit function $\beta_i(G)$ is maximized at country i's preferred level, and given that solar geoengineering is modeled as a best-shot technology, the country with $G^{\max}$ will maximize payoffs by choosing $G = G^{\max}$ in stage 2.

All other countries with preferred levels less than $G^{\max}$ will anticipate that the chosen G will be above their preferred level and so they will not invest in G. The country or countries with preferred levels equal to $G^{\max}$ are the "free drivers" (Wagner and Weitzman 2012; Weitzman 2015) as their choice of SGE "drives" the global temperature.

Proposition 1: *In stage 2, the level of SGE deployment is determined by the player(s) with the highest preferred level (i.e., the free driver(s)), resulting in* $G = G^{\max}$.

While it is clear that $G = G^{\max}$ after stage 2 and $\beta_i(G^{\max}) > 0$ for the player(s) with the highest preferred level, what is not obvious is whether the aggregated benefits from SGE deployment are positive or negative. We know that SGE levels greater than G_i^{neg} will strictly reduce a player's payoff, while levels less than G_i^{neg} increase payoffs (relative to $G = 0$). The aggregate benefit function from SGE is the sum of (6) and can be expressed as

$$\beta(G^{\max}) = \sum_{i \in N} \left[\theta G_i^p - \phi \left(G^{\max} - G_i^p\right)\right], \tag{12}$$

then collecting G_i^p terms

$$\beta(G^{\max}) = \sum_{i \in N} \left[\left(\theta + \phi\right) G_i^p - \phi G^{\max}\right], \tag{13}$$

and recognizing the last term is a constant summed n times this becomes

$$\beta(G^{\max}) = -\phi n G^{\max} + \left(\theta + \phi\right) \sum_{i \in N} G_i^p. \tag{14}$$

For all players other than the free driver(s), the benefits are increased when $G^{\max}$ decreases. We can now substitute our expressions for $G^{\max}$ and G_i^p, and write this in terms of abatement, which is

$$\beta(G^{\max}) = (Q^o - Q)\left[-n\phi\gamma^{\max} + (\theta+\phi)\sum_{i\in N}\gamma_i\right].$$

The term in [.] is strictly parameters, and Q^o is a constant. The $(Q^o - Q)$ term is non-negative when abatement is less than socially optimal and if the term in [.] is negative then SGE at $G^{\max}$ results in negative aggregate benefits. SGE at $G^{\max}$ results in negative aggregate benefits when the abatement gap is positive if

$$-n\phi\gamma^{\max} + (\theta+\phi)\sum_{i\in N}\gamma_i < 0. \tag{15}$$

Proposition 2: *The aggregate benefits from solar geoengineering are negative if* $\frac{\bar{\gamma}}{\gamma^{\max}} < \frac{\phi}{\theta+\phi}$ *and* $Q < Q^o$.

Proof: Define the average level of γ_i as $\bar{\gamma}$, then $\sum\limits_{i\in N}\gamma_i$ is $n\bar{\gamma}$, and the condition in (15) becomes

$$\begin{aligned}(\theta+\phi)\,n\bar{\gamma} &< n\phi\gamma^{\max}\\ \frac{\bar{\gamma}}{\gamma^{\max}} &< \frac{\phi}{\theta+\phi},\end{aligned} \tag{16}$$

where $\frac{\phi}{\theta+\phi} \in (0,1)$.

Suppose for example that all nations were identical. In this case $\frac{\bar{\gamma}}{\gamma^{\max}} = 1$ and so SGE strictly increases aggregate benefits. However, suppose $\bar{\gamma} = 0.4$ and $\gamma^{\max} = 0.9$, then SGE at the free-driver outcome is welfare reducing if $\theta < \phi$, since $\theta < \phi$ implies $\frac{\phi}{\theta+\phi} \in (0.5, 1)$.[3]

At the individual level, the benefit from SGE at the free-driver outcome is

$$\beta_i(G^{\max}) = (Q^o - Q)\left[(\theta+\phi)\,\gamma_i - \phi\gamma^{\max}\right],$$

and the first term is non-negative and the second is isomorphic to G_i^{neg}. That is,

$$\beta_i(G^{\max}) < 0 \text{ iff } \frac{\gamma_i}{\gamma^{\max}} < \frac{\phi}{\theta+\phi}.$$

3.2 Stage 1 - emissions abatement

Given that $G = G^{\max}$ in stage 2, a country's payoff from (11) in stage 1 takes the following form

$$\pi_i(q_i, Q, G^{\max}) = \frac{bQ}{n} - \frac{c\,(q_i)^2}{2} + (\theta+\phi)\,G_i^p - \phi G^{\max}, \tag{17}$$

---

[3] In our study we do not constrain the relationship between θ and ϕ. Weitzman (2015), however, weighs "overdone" geoengineering (our ϕ) at three times the value for "underdone" geoengineering (our θ).

then in terms of abatement

$$\pi_i(q_i, Q, G^{\max}) = \frac{bQ}{n} - \frac{c\,(q_i)^2}{2} + (Q^o - Q)\left[(\theta + \phi)\,\gamma_i - \phi\gamma^{\max}\right],$$

and the FOC for abatement is

$$\frac{b}{n} - cq_i - \left[(\theta + \phi)\,\gamma_i - \phi\gamma^{\max}\right] = 0, \tag{18}$$

with solution

$$q_i^*(G^{\max}) = \frac{b}{cn} - \frac{(\theta + \phi)\,\gamma_i - \phi\gamma^{\max}}{c}. \tag{19}$$

Proposition 3: *The availability of solar geoengineering technologies causes free drivers to reduce emissions abatement.*

Proof: A free-driver country has $\gamma_i = \gamma^{\max}$ and equation (19) reduces to

$$q_i^*(\gamma^{\max}, G^{\max}) = \frac{b}{cn} - \frac{\theta\gamma^{\max}}{c},$$

which is less than the non-cooperative abatement level in a world without solar geoengineering technologies (from equation (2)). Recall, we introduced the imperfect substitutability parameter restriction $\frac{b}{n} \geq \theta$ in (7), which ensures that abatement q^* is non-negative.

Proposition 4: *The availability of solar geoengineering technologies causes non free-driving countries to increase abatement when* $\frac{\gamma_i}{\gamma^{\max}} < \frac{\phi}{\theta+\phi}$.

Proof: For all countries with preferred SGE levels less than $G^{\max}$, from equation (19) SGE opportunities will increase abatement when

$$(\theta + \phi)\,\gamma_i - \phi\gamma^{\max} < 0 \tag{20}$$

which, when rearranged, is the familiar condition from equation (16)

$$\frac{\gamma_i}{\gamma^{\max}} < \frac{\phi}{\theta + \phi}. \tag{21}$$

Proposition 4 informs us that a country will increase abatement relative to a world without SGE, provided that the free-riding level of solar geoengineering leads to negative payoffs. On the other hand, if the free-driver level of SGE leads to an increase in payoffs, the country will reduce emissions abatement. The intuition is that players can reduce some of their losses from SGE by decreasing the level of $G^{\max}$, which is reduced by an increase in emissions abatement. But if SGE is expected to increase payoffs, the incentive is to reduce abatement to save money on abatement costs. This result helps us better understand the nuance of the "moral hazard" debate which is centered on the conjecture that SGE opportunities will undermine abatement (e.g., Reynolds 2019; Wagner 2021). We show that this is a possibility, but the opposite outcome can also occur - it depends on the distribution of preferences.

We can aggregate $q_i^*(G^{\max})$ to get $Q^*(G^{\max})$ using (19),

$$\begin{aligned}
Q^*(G^{\max}) &= \sum_{i\in N} q_i^*(G^{\max}) \\
Q^*(G^{\max}) &= \sum_{i\in N}\left[\frac{b}{cn} - \frac{(\theta+\phi)\,\gamma_i - \phi\gamma^{\max}}{c}\right] \\
Q^*(G^{\max}) &= \frac{b}{c} - \left(\frac{1}{c}\right)\left[-n\phi\gamma^{\max} + (\theta+\phi)\sum_{i\in N}\gamma_i\right]. \qquad (22)
\end{aligned}$$

Then using the average level $\bar{\gamma}$, the sum $\sum_{i\in N}\gamma_i = n\bar{\gamma}$, and Nash equilibrium abatement without SGE in (2) $Q^* = \frac{b}{c}$ this becomes

$$\begin{aligned}
Q^*(G^{\max}) &= Q^* - \left(\frac{1}{c}\right)\left[-n\phi\gamma^{\max} + (\theta+\phi)\,n\bar{\gamma}\right] \\
Q^*(G^{\max}) &= Q^* - \left(\frac{n}{c}\right)\left[-\phi\gamma^{\max} + (\theta+\phi)\,\bar{\gamma}\right]. \qquad (23)
\end{aligned}$$

This results in the following proposition.

Proposition 5: *The availability of solar geoengineering technologies reduces abatement at the free-driver outcome when* $\frac{\bar{\gamma}}{\gamma^{\max}} > \frac{\phi}{\theta+\phi}$.

Proof: When $\frac{\bar{\gamma}}{\gamma^{\max}} > \frac{\phi}{\theta+\phi}$, $Q^*(G^{\max}) < Q^*$.

For example, if all countries have identical preferred levels of SGE then $\frac{\bar{\gamma}}{\gamma^{\max}} = 1 > \frac{\phi}{\theta+\phi} \in (0,1)$, so the availability of SGE always reduces abatement. Note that Propositions 2 and 5 depend on the same condition. Thus, the aggregate benefits from SGE are positive and hence SGE reduces aggregate abatement at the free-driver outcome when $\frac{\bar{\gamma}}{\gamma^{\max}} > \frac{\phi}{\theta+\phi}$. When the aggregate benefits of SGE are negative SGE increases abatement at the free-driver outcome.

4 Optimal abatement and solar geoengineering

The socially optimal level of abatement and SGE is the solution to choosing Q and G to maximize

$$\Pi \equiv \sum_{i\in N}\pi_i(q_i, Q, G) = \sum_{i\in N}\left[\frac{bQ}{n} - \frac{c\,(q_i)^2}{2}\right] + \beta(G). \qquad (24)$$

If we let m denote the number of countries that would prefer a lower level of SGE compared to the chosen level (i.e., those in which $G > G_i^p$). We can rewrite payoff equation (24) as

$$\Pi \equiv \sum_{i\in N}\pi_i(q_i, Q, G) = \sum_{i\in N}\left[\frac{bQ}{n} - \frac{c\,(q_i)^2}{2}\right] + \sum_{i=m+1}^{n}\theta G + \sum_{i=1}^{m}\left[\theta G_i^p - \phi\,(G - G_i^p)\right], \qquad (25)$$

where the first term after the brackets captures the benefits to those $(n-m)$ countries with preferred SGE levels weakly higher than G. Those are the countries that weakly benefit from SGE. The second term after the brackets captures the m countries that suffer losses from SGE levels above their preferred level.

4.1 Optimal G

When choosing G to maximize $\beta(G)$, the first order condition is

$$(n-m)\theta - m\phi = 0. \tag{26}$$

The first term in (26), $(n-m)\theta$, is the sum of the marginal benefits from another unit of SGE for those countries with preferred levels that are higher than G, and the second term, ϕm, is the sum of the marginal costs imposed on countries that have preferred levels less than G. As long as $(n-m)\theta > \phi m$ it is socially optimal to increase G, and $\beta(G)$ is maximized when $(n-m)\theta = \phi m$. By rearranging terms and solving for m we can define an optimal "stopping rule"

$$m^o = \frac{n\theta}{\theta+\phi}. \tag{27}$$

The stopping rule tells us that it is optimal to increase G up until the point where the number of countries with preferred levels less than G reaches m^o. To illustrate, suppose $\phi = 0$, such that there is no penalty for SGE levels greater than a country's preferred level. In that case $m^o = n$, which means that it is optimal to increase G up until the point that all countries prefer a lower G (given there is no cost to higher levels of G) . On the other hand, if $\theta = 0$, then there is no benefit from G for any country, and so the stopping rule tells us that we shouldn't have any countries with preferred levels lower than G. Finally, if $\theta = \phi$, then $m^o = n/2$, which means it is optimal to stop when half of all countries prefer lower SGE levels. Note that m^o is the integer value equal to $\frac{n\theta}{\theta+\phi}$ or the next lowest integer if $\frac{n\theta}{\theta+\phi}$ is a non-integer.

Without any loss of generality, order nations from lowest to highest γ_i, such that $\gamma^{\min} \equiv \gamma_1 \leq \gamma_2 \leq ... \leq \gamma_n \equiv \gamma^{\max}$. This implies $G^{\min} \equiv G_1^p \leq G_2^p \leq ... \leq G_n^p \equiv G^{\max}$. We can use this ranking and the stopping rule to define the socially optimal level of SGE. Suppose, for example, $m^o = 2$. Then we know that the level of G should be set equal to the preferred level of the country with the second lowest γ_i, which is $G_2^p = \gamma_2(Q^o - Q)$. We can generalize this with the following expression for the optimal level of SGE

$$G^{so} = G_{m^o}^p = \gamma_{m^o}(Q^o - Q). \tag{28}$$

4.2 Optimal Q

Given the expressions for G^{so} and G_i^p, we can rewrite the payoff function in (25) as

$$\begin{aligned}\Pi \equiv \sum_{i\in N}\pi_i(q_i,Q,G^{so}) &= \sum_{i\in N}\left[\frac{bQ}{n}-\frac{c\,(q_i)^2}{2}\right]+\sum_{i=m^o+1}^{n}\theta\gamma_{m^o}(Q^o-Q)+\\ &\sum_{i=1}^{m^o}\left[\theta\gamma_i(Q^o-Q)-\phi\left(\gamma_{m^o}(Q^o-Q)-\gamma_i(Q^o-Q)\right)\right]. \qquad (29)\end{aligned}$$

We can rewrite (29), recognizing that $q_i=\frac{Q}{n}$ in the cost-minimizing solution, as

$$\begin{aligned}\Pi &= bQ-\frac{cQ^2}{2n}+(n-m)\,\theta\gamma_{m^o}(Q^o-Q)+\\ &\quad (Q^o-Q)\sum_{i=1}^{m^o}\left[\theta\gamma_i-\phi\left(\gamma_{m^o}-\gamma_i\right)\right]\\ \Pi &= bQ-\frac{cQ^2}{2n}+(n-m^o)\,\theta\gamma_{m^o}(Q^o-Q)+(Q^o-Q)\sum_{i=1}^{m^o}\left[(\theta+\phi)\,\gamma_i-\phi\gamma_{m^o}\right]\\ \Pi &= bQ-\frac{cQ^2}{2n}+(n-m^o)\,\theta\gamma_{m^o}(Q^o-Q)+(Q^o-Q)\left[-m^o\phi\gamma_{m^o}+(\theta+\phi)\sum_{i=1}^{m^o}\gamma_i\right] \qquad (30)\end{aligned}$$

Then the planner's first-order condition with respect to abatement is

$$\frac{\partial\Pi}{\partial Q}=b-\frac{cQ}{n}-(n-m^o)\,\theta\gamma_{m^o}-\left[-m^o\phi\gamma_{m^o}+(\theta+\phi)\sum_{i=1}^{m^o}\gamma_i\right]=0. \qquad (31)$$

Which results in the socially optimal level of abatement in stage 1, Q^{so}, where

$$Q^{so}=\frac{bn}{c}+\frac{n}{c}\left[m^o\phi\gamma_{m^o}-(n-m^o)\,\theta\gamma_{m^o}-(\theta+\phi)\sum_{i=1}^{m^o}\gamma_i\right]. \qquad (32)$$

Then writing this in terms of the social optimum without SGE Q^o in (3) this is

$$Q^{so}=Q^o+\frac{n}{c}\left[\gamma_{m^o}\left[m^o\left(\theta+\phi\right)-n\theta\right]-(\theta+\phi)\sum_{i=1}^{m^o}\gamma_i\right]. \qquad (33)$$

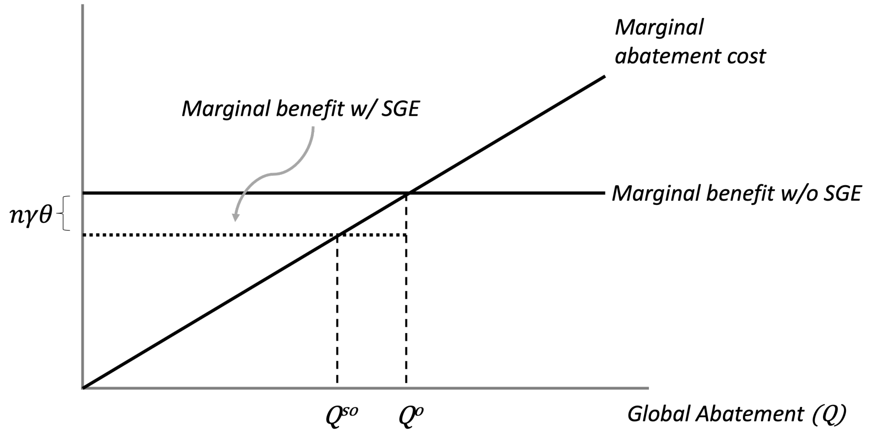
This leads to our next proposition

Proposition 6: *The socially optimal level of emissions abatement is reduced by the availability of solar geoengineering technologies if* $(\theta+\phi)\sum_{i=1}^{m^o}\gamma_i>0$.

Proof: $Q^{so}<Q^o$ if the term in brackets in equation (33) is negative. When substituting the "stopping rule" $m^o=\frac{n\theta}{\theta+\phi}$ from equation (27) and simplifying, the term in brackets reduces to $-(\theta+\phi)\sum_{i=1}^{m^o}\gamma_i$.

The first term of the socially optimal level of abatement in (33) is the familiar optimal abatement level without opportunities for SGE in (3). To illustrate, suppose $\gamma_i = 0$ for all n countries, then $Q^{so} = Q^{o} = \frac{nb}{c}$. If we consider another extreme case in which all countries are identical (i.e., $\gamma_i = \gamma$), then (33) reduces to $Q^{so} = \frac{nb}{c} - \frac{\theta n^2 \gamma}{c}$. It is immediately clear that if all countries have the same preferred level, then the optimal level of SGE is not a "bad" for any country and abatement levels drop relative to a world without SGE opportunities. This scenario with homogeneous countries is depicted in Figure 2. The aggregate marginal abatement cost function is unaffected by SGE, but the aggregate marginal benefit of abatement is reduced by $n\gamma\theta$. When γ and θ are both positive, SGE opportunities cause a decrease in the optimal abatement level as demonstrated in Figure 2.

Figure 2: Social optimum abatement with and without SGE for homogeneous countries



5 International Environmental Agreements

The non-cooperative equilibrium abatement and solar geoengineering levels are inefficient. International environmental agreements are institutions designed to help move countries closer to the social optimum. As a point of departure and to provide a baseline for comparison, we first summarize the stability conditions and equilibrium IEA sizes in a world without SGE opportunities. The fundamental results from abatement models with linear benefits and quadratic costs have been previously published in Barrett (1994, 2003) and Finus and McGinty (2019), and so we get straight to the point.

5.1 IEAs in a world without solar geoengineering

The IEA is a two-stage game. In stage 1, countries independently and simultaneously decide whether to join an agreement. In stage 2, the members of the agreement choose abatement levels to maximize their joint payoffs. Meanwhile, non-members choose their non-cooperative abatement levels. Equilibrium agreement sizes are determined by two stability conditions. An IEA is *internally stable* if no country can earn higher payoffs by leaving the agreement. An IEA is *externally stable* if no country can earn higher payoffs by joining the agreement.

As demonstrated by Barrett (1994, 2003), given the linear-quadratic payoff function, the largest stable coalition size without solar geoengineering opportunities is $s = 3$ members for any values for n, b or c when nations have identical abatement preferences.

5.2 IEAs in a world with solar geoengineering

We now consider agreements to manage global emissions abatement when countries have the opportunity to invest in SGE. The sequence of decision-making follows existing theoretical models and experiments that include both abatement and geoengineering (Moreno-Cruz 2015; Cherry et al. 2022); that is, countries make emissions abatement decisions first and then decide on SGE deployment. The IEA now has three stages. In stage 1 (*participation*), countries decide independently and simultaneously whether to join the IEA. In stage 2 (*abatement*), the IEA members choose abatement levels in order to maximize their joint (i.e., collective) payoffs. Meanwhile, non-members choose their abatement levels unilaterally (non-cooperatively). Finally, in stage 3 (*solar geoengineering*), countries choose their SGE levels. We consider the simplest agreement; one that only governs abatement and all countries are free to choose their optimal SGE levels in stage three. In the final section we discuss alternative institutions that govern both abatement and SGE.

5.2.1 Stage 3 - solar geoengineering

Let the members of an agreement be a coalition S and the set of non-member countries outside an agreement be denoted T with $S \cup T = N$. Denote total abatement for a coalition S as

$Q(S)$. The number of agreement members is denoted as $|S| = s$ and indexed by $i = 1, 2, ..., s$. The number of non-members is $|T| = n - s$ and indexed by $j = 1, 2, ..., (n - s)$.

In an agreement that only governs abatement, the country (member or non-member) with the highest preferred level will act as the free-driver and choose $G(S) = G^{\max}(S) = \gamma^{\max}(Q^o - Q(S))$ in stage three. Since $G(S)$ is decreasing in total abatement, the exact level of SGE is a function of the abatement decisions made by the s members and $n - s$ non-members in stage two.

5.2.2 Stage 2 - abatement

In stage 2, the $n - s$ free-riding non-members each choose their non-cooperative Nash abatement levels $q_i^*(G^{\max})$ from equation (19), which we now label as $q_j^t(G^{\max})$ and include here.

$$q_j^t(G^{\max}) = \frac{b}{cn} - \frac{(\theta + \phi)\gamma_j - \phi\gamma^{\max}}{c}.$$

To help provide intuition regarding how SGE impacts non-member abatement levels, it is convenient to rewrite the function in three terms

$$q_j^t(G^{\max}) \quad = \quad \underbrace{\frac{b}{cn}}_{1} \underbrace{- \frac{\theta\gamma_j}{c}}_{2} \underbrace{+ \frac{\phi(\gamma^{\max} - \gamma_j)}{c}}_{3}$$

The first term is the dominant strategy abatement level in the world without SGE. The second term is the reduction in abatement from SGE being a ‘good’, where γ_j is effectively j’s degree of substitutability between abatement and SGE. The closer γ_j is to one, the closer the substitutability and the greater the reduction in abatement. θ is the marginal benefit from a unit of SGE so $\theta\gamma_j$ is the effective foregone benefit from SGE for a unit of abatement. The third term is the increase in abatement from using abatement as protection against SGE on all units above j’s preferred level. Recall, the marginal cost of SGE in the ‘bad’ region is ϕ.

Meanwhile, the s coalition members choose q_i^s to maximize joint payoffs. The problem is

$$\max_{q_i^s} \sum_{i \in S} \pi_i^s = \sum_{i \in S} \frac{bQ}{n} - \frac{c(q_i^s)^2}{2} + \theta G_i^p - \phi(G^{\max} - G_i^p), \tag{34}$$

when simplifying and expressing in terms of abatement

$$\begin{aligned}
\max_{q_i^s} &= \sum_{i \in S} \frac{bQ}{n} - \frac{c(q_i^s)^2}{2} + (\theta + \phi) G_i^p - \phi G^{\max} \\
\max_{q_i^s} &= \sum_{i \in S} \frac{bQ}{n} - \frac{c(q_i^s)^2}{2} + (\theta + \phi)\gamma_i (Q^o - Q) - \phi\gamma^{\max}(Q^o - Q) \\
\max_{q_i^s} &= \sum_{i \in S} \frac{bQ}{n} - \frac{c(q_i^s)^2}{2} + (Q^o - Q)\left[(\theta + \phi)\gamma_i - \phi\gamma^{\max}\right]. \qquad (35)
\end{aligned}$$

The first order condition is

$$\frac{bs}{n} - cq_i^s - \sum_{i=1}^{s} \left[(\theta + \phi) \gamma_i - \phi \gamma^{\max} \right] = 0, \tag{36}$$

and when solving

$$q_i^s(G^{\max}) = \frac{bs}{nc} - \frac{1}{c} \left[(\theta + \phi) \sum_{i=1}^{s} \gamma_i - s\phi\gamma^{\max} \right]. \tag{37}$$

Total emissions abatement in stage two is therefore

$$Q(G^{\max}) = \sum_{i \in S} q_i^s(G^{\max}) + \sum_{j \in T} q_j^t(G^{\max}). \tag{38}$$

5.2.3 Stage 1 - participation

In the first stage all countries decide independently and simultaneously whether to join the IEA. An equilibrium IEA is one that is both *internally* and *externally* stable. Internal stability requires that no member could increase their payoff by leaving the agreement (i.e., $IS_i(S) = \pi^s(S) - \pi^t(S \backslash i) \geq 0$) and external stability requires that no non-member could increase their payoff by joining (i.e., $ES_i(S) = \pi^t(S) - \pi^s(S \cup i) \geq 0$).

When countries have heterogeneous preferences for SGE, the coalition sizes that satisfy the stability conditions will depend on the specific distribution of γ's. However, in the special case in which countries have homogeneous preferences for SGE, $\gamma_i = \gamma$, signatory, non-signatory and aggregate abatement can be expressed as

$$\begin{aligned} q^t &= \frac{b}{cn} - \frac{(\theta + \phi)\gamma - \phi\gamma}{c} = \frac{b - n\theta\gamma}{cn} \\ q^s &= \frac{bs}{cn} - \frac{1}{c} \left[(\theta + \phi) \sum_{i=1}^{s} \gamma - s\phi\gamma \right] = \frac{bs - ns\theta\gamma}{cn} = \frac{s\,(b - n\theta\gamma)}{cn} \\ Q(s) &= sq^s + (n - s)\,q^t = \frac{\left(s^2 - s + n\right)(b - n\theta\gamma)}{cn} \end{aligned} \tag{39}$$

where, given homogeneous preferences, all subscripts are dropped for clarity. In this special case, the internal stability (IS) condition can be expressed as

$$IS(s) = \left(\frac{b - n\theta\gamma}{n} \right) [Q(s) - Q(s-1)] - \frac{c\left[(q^s)^2 - \left(q^t\right)^2 \right]}{2}. \tag{40}$$

Proposition 7: *When countries have identical preferences for solar geoengineering deployment (i.e., $\gamma_i = \gamma$), the unique stable coalition size is $s^* = 3$.*

Proof: Using $Q(s)$ from (39) and solving $Q(s-1) = \frac{\left[s^2-3s+2+n\right](b-n\theta\gamma)}{cn}$, the first term in (40) reduces to $\frac{2(s-1)(b-n\theta\gamma)^2}{cn^2}$. Given the parameter restriction in (7), $\frac{b}{n} \geq \theta$ stating that the individual marginal benefit from SGE cannot exceed that of abatement, and that $\gamma \in (0,1)$, the second term in (40) can be reduced to $-\frac{(b-n\theta\gamma)^2\left(s^2-1\right)}{2cn^2}$. Combining both terms yields

$$\begin{aligned} IS(s) &= \frac{(b-n\theta\gamma)^2\left[4(s-1)-\left(s^2-1\right)\right]}{2cn^2} \\ IS(s) &= \frac{(b-n\theta\gamma)^2\left(-s^2+4s-3\right)}{2cn^2} \geq 0 \Leftrightarrow s \leq s^* = 3. \end{aligned} \tag{41}$$

From previous studies, we know that with linear-quadratic payoffs and $\gamma = 0$, the largest stable IEA size is $s = 3$. Proposition 7 illustrates that this results holds when $\gamma > 0$ as well. The availability of solar geoengineering opportunities does not alter the size of stable IEAs when countries have the same preferred levels.

The stability conditions are much more complicated when countries have heterogeneous preferences for SGE. The internal stability condition when countries may have different preferences for SGE is

$$IS_i(S) = \left(\frac{b-n\left[(\theta+\phi)\gamma_i - \phi\gamma^{\max}\right]}{n}\right)\left[Q(S)-Q(S\backslash i)\right] - \frac{c\left[\left(q_i^s\right)^2-\left(q_i^t\right)^2\right]}{2}. \tag{42}$$

Note that the IS condition is written in terms of the coalition structure S rather than simply the number of agreement members s (as in (40)). The is because the individual and aggregate abatement levels of the coalition will depend on which particular countries are members. The complexity of the IS condition in (42) does not allow for an analytical solution in the same way that the symmetric case led to Proposition 7. That is, we are not able to explicitly characterize the largest stable agreement size when countries have heterogeneous preferences for SGE. Stability will depend on the distribution of preferences for SGE. This roadblock is common in the IEA literature, and researchers typically turn to simulations to provide further insights (e.g., Barrett 1994; McGinty 2007). In the next section we use numeric examples to help illustrate the relationships between coalition structures, payoffs and stability in both homogeneous and heterogeneous scenarios.

6 Simulations

To help illustrate the payoffs to members and non-members in an IEA that governs emissions abatement, we use numeric examples. Throughout, $b = 1$, $c = 1$, $n = 5$ and $\theta = \phi = 0.1$. We start with the simplest case in which all n countries are homogeneous with respect to their preference for SGE, i.e., $\gamma_i = \gamma$. As a reference point and to frame our analysis relative to the established IEA literature, we first look at the case in which $\gamma = 0$. When $\gamma = 0$, all n countries prefer $G = 0$ and the game effectively reverts to an IEA without SGE opportunities.

Table 1 contains the emissions levels and payoffs for members and non-members when $\gamma = 0$. In a non-cooperative Nash equilibrium, individual countries have a payoff of 0.18 and aggregate payoffs are 0.90. In the social optimum, all n countries join the agreement and each earn 0.5 for a collective payoff of 2.50. From the table it is easy to verify that the largest stable coalition size is $s = 3$, which is consistent with Barrett (1994, 2003). When $s < 3$ the payoff from joining the coalition is higher than remaining outside so the agreement is not externally stable. When $s > 3$ the payoff from leaving the coalition is higher than remaining a member so the agreement is not internally stable.

Table 1: Member and non-member payoffs, homogeneous countries with $\gamma = 0$

s	q^s	q^t	Q	$G^{\max}$	$B(G^{\max})$	π^s	π^t	Π
0	—	0.2	1	0.0000	0.0000	—	0.1800	0.9000
1	0.2	0.2	1	0.0000	0.0000	0.1800	0.1800	0.9000
2	0.4	0.2	1.4	0.0000	0.0000	0.2000	0.2600	1.1800
3^*	0.6	0.2	2.2	0.0000	0.0000	0.2600	0.4200	1.6200
4	0.8	0.2	3.4	0.0000	0.0000	0.3600	0.6600	2.1000
5	1	—	5	0.0000	0.0000	0.5000	—	2.5000

Now consider a scenario in which all n countries have the same positive level of $\gamma = 0.50$. We know from equations (19) and (33) that when all countries have the same preference for SGE, non-cooperative and optimal abatement levels decrease relative to a world without SGE opportunities. From Table 2, it is clear that when $s = 0$, the n non-members have lower abatement levels relative to the scenario in Table 1. Moreover, the socially optimal abatement levels decrease with SGE opportunities and aggregate payoffs increase. Note that, following Proposition 6, since SGE is not a "bad" for any country when they all have the same preferences, countries will deploy a positive level of SGE in the social optimum. Importantly, consistent with Proposition 7, the largest stable IEA remains $s = 3$ with SGE opportunities.

Table 2: Member and non-member payoffs, homogeneous countries with $\gamma = 0.50$

s	q^s	q^t	Q	$G^{\max}$	$B(G^{\max})$	π^s	π^t	Π
0	—	0.15	0.75	2.1250	0.2125	—	0.3513	1.7563
1	0.15	0.15	0.75	2.1250	0.2125	0.3513	0.3513	1.7563
2	0.3	0.15	1.05	1.9750	0.1975	0.3625	0.3963	1.9138
3^*	0.45	0.15	1.65	1.6750	0.1675	0.3963	0.4863	2.1613
4	0.6	0.15	2.55	1.2250	0.1225	0.4525	0.6213	2.4313
5	0.75	—	3.75	0.6250	0.0625	0.5313	—	2.6563

6.1 Heterogeneous countries

Next we consider agreements when countries have heterogeneous preferences for SGE. For ease of comparison, we begin with the same parameter values for n, b, c, θ and ϕ from the previous section. We chose an initial distribution of γ's for the n countries to keep a mean-preserving spread of the homogeneous country example in Table 2. Specifically, we analyze agreements with the following values: $\gamma_1 = 0.2$, $\gamma_2 = 0.4$, $\gamma_3 = 0.5$, $\gamma_4 = 0.6$, and $\gamma_5 = 0.8$. Note that $\bar{\gamma} = 0.5$.

Recall, in an IEA that only governs emissions abatement, the country with the highest preference for SGE - the free driver - will deploy $G(S) = G(S)^{\max} = \gamma^{\max}(Q^o - Q(S))$ in stage three. In Table 3, we show the non-cooperative Nash equilibrium and full participation ($s = n$) abatement and payoff levels for each of the five countries. The superscript f denotes values in an agreement with full participation. The Nash equilibrium abatement and payoffs are equivalent to those under trivial IEA sizes $s < 2$. Note that the non-cooperative Nash and full participation SGE levels are $\mathrm{G}^{\max}(Q^*) = 3.28$ and $\mathrm{G}^{\max}(Q^f) = 0.40$, respectively. It's important to note that the full participation outcome in Table 3 is not the same as the socially optimal outcome. This is because even in an IEA with full participation, the free-driving country (country 5) chooses a level that is higher than the optimal level. Since our distribution of γ's is symmetric, the optimal level of G is the median preferred level (country 3's level, $G = 0.25$).

Table 3: Nash equilibrium and full participation abatement levels and payoffs

		Nash equilibrium		full participation	
country i	γ_i	q_i^*	π_i^*	q_i^f	π_i^f
1	0.2	0.24	-0.013	0.90	0.475
2	0.4	0.20	0.160	0.90	0.495
3	0.5	0.18	0.246	0.90	0.505
4	0.6	0.16	0.331	0.90	0.515
5	0.8	0.12	0.501	0.90	0.535
	$\bar{\gamma} = 0.50$	$Q^* = 0.90$	$\Pi^* = 1.225$	$Q^f = 4.50$	$\Pi^f = 2.525$

In the Nash equilibrium, country 1 takes on the highest level of abatement and earns the lowest payoff. The intuition is that country 1 has the lowest preferred level of SGE, and increasing abatement is a way to protect itself from some of the damages imposed by excessive SGE deployment from the free-driving country. The free-driver, country 5, takes on the lowest abatement and earns the highest payoff since SGE is deployed at their preferred level. In comparison to the scenario without SGE in Table 1, aggregate abatement is lower with SGE (0.90 vs. 1.00) and aggregate payoff is higher with SGE (1.225 vs. 0.90). In particular, countries 3, 4 and 5 abate less and have a higher payoff in the Nash equilibrium with SGE, while countries 1 and 2 abate more (country 1) or the same (country 2) and have lower payoffs.

An IEA with full participation (i.e., the grand coalition) results in a Pareto improvement for all countries. Note that country 1, having the lowest preferred level of SGE, has the most to gain from full cooperation relative to the non-cooperative outcome. However, country 1 still earns the lowest payoff of all n countries. Aggregate abatement under full participation is lower with SGE opportunities compared to without (4.50 vs. 5.0), and aggregate payoffs are higher (2.53 vs. 2.50). For those with SGE levels less than the free driver, the individual gain from an agreement with full participation is increasing in the free-driver's preference for SGE.

6.1.1 Stable coalitions

For the grand coalition to be an equilibrium IEA, it must be the case that no member could be better off by defecting and leaving the agreement. That is, the internal stability condition must be satisfied for all n countries. We can verify the internal stability conditions by comparing payoffs in Table 4 below with those in Table 3. Table 4 includes the abatement levels and payoffs of all possible coalitions of $s = n - 1 = 4$.

Table 4: Member and non-member abatement and payoffs for all $s = 4$

coalition	member abatement	member payoffs	non-member abatement	non-member payoffs	Q	$G^{\max}$	Π
1,2,3,4	q_i^s= 0.78	π_1=0.273; π_2=0.344; π_3=0.379; π_4=0.414	$q_5 = 0.12$	$\pi_5 = 0.782$	3.24	1.41	2.192
1,2,3,5	$q_i^s = 0.74$	π_1=0.275; π_2=0.350; π_3=0.388; π_5=0.501	$q_4 = 0.16$	$\pi_4 = 0.686$	3.12	1.50	2.200
1,2,4,5	$q_i^s = 0.72$	π_1=0.275; π_2=0.353; π_4=0.430; π_5=0.508	$q_3 = 0.18$	$\pi_3 = 0.635$	3.06	1.55	2.201
1,3,4,5	$q_i^s = 0.70$	π_1=0.275; π_3=0.395; π_4=0.435; π_5=0.515	$q_2 = 0.20$	$\pi_2 = 0.580$	3.00	1.60	2.200
2,3,4,5	$q_i^s = 0.66$	π_2=0.358; π_3=0.401; π_4=0.443; π_5=0.528	$q_1 = 0.24$	$\pi_1 = \mathbf{0.462}$	2.88	1.70	2.192

As an example, consider the first coalition in Table 4. This is the coalition in which country 5 is the lone non-member. From the fifth column in Table 4, country 5's payoff as the non-member is 0.782 which exceeds its payoff of 0.535 when inside the grand coalition from Table 3. Clearly, the free-driver country would rather be outside of a coalition of size four, and therefore the internal stability condition is violated.

From this, we immediately verify that the grand coalition is not an equilibrium IEA since it is not internally stable for all n countries. In fact, *only* country 1 would prefer to remain in a coalition of $s = n$ rather than being the single defector in a coalition of size $s = n - 1 = 4$. This is easily confirmed by comparing its payoff as the lone non-member (0.462, in bold) with its payoff as a member of the grand coalition (0.475) from Table 3.

The same procedure is then used to examine the internal stability conditions for agreements of size $s = 4$, by comparing them to agreements when one member defects. There are 10 possible coalitions of size $s = 3$ that must be considered. For any particular coalition of $s = 4$ from Table 4 to be stable, it must be the case that no member could be better off leaving while the others remain in the coalition (now of size $s = 3$). We conduct this procedure and find that no coalition of size $s = 4$ is internally stable for all members.

Ultimately, we are able to show that the largest stable IEA given our distribution of γ_i and chosen parameter values is $s = 2$, although not all coalitions of two members are stable. Table 5 contains abatement and payoffs for all stable IEAs in our example. Note that stable coalitions with members that have lower preferences for SGE will abate more, and achieve higher aggregate payoffs compared to stable coalitions consisting of members with stronger preferences for SGE. In all cases, stable IEAs are only able to marginally improve upon the non-cooperative outcome. This can be confirmed by comparing values from Table 5 with the non-cooperative values from Table 3.

Table 5: Member and non-member abatement and payoffs for all stable IEAs

coalition	member abatement	member payoffs	non-member abatement	non-member payoffs	Q	$G^{\max}$	Π
1,2	$q_i^s = 0.44$	$\pi_1=0.025$; $\pi_2=0.171$	$q_3=0.18$; $q_4=0.16$; $q_5=0.12$	$\pi_3=0.325$; $\pi_4=0.402$; $\pi_5=0.554$	1.34	2.93	1.49
1,3	$q_i^s = 0.43$	$\pi_1=0.029$; $\pi_3=0.249$	$q_2=0.20$; $q_4=0.16$; $q_5=0.12$	$\pi_2=0.244$; $\pi_4=0.398$; $\pi_5=0.551$	1.32	2.94	1.47
2,3	$q_i^s = 0.38$	$\pi_2=0.184$; $\pi_3=0.258$	$q_1=0.24$; $q_4=0.16$; $q_5=0.12$	$\pi_1=0.078$; $\pi_4=0.392$; $\pi_5=0.546$	1.28	2.98	1.46
2,4	$q_i^s = 0.36$	$\pi_2=0.187$; $\pi_4=0.337$	$q_1=0.24$; $q_3=0.18$; $q_5=0.12$	$\pi_1=0.074$; $\pi_3=0.311$; $\pi_5=0.544$	1.26	2.99	1.45
3,4	$q_i^s = 0.34$	$\pi_3=0.265$; $\pi_4=0.341$	$q_1=0.24$; $q_2=0.20$; $q_5=0.12$	$\pi_1=0.069$; $\pi_2=0.228$; $\pi_5=0.542$	1.24	3.01	1.44
4,5	$q_i^s = 0.28$	$\pi_4=0.350$; $\pi_5=0.502$	$q_1=0.24$; $q_2=0.20$; $q_3=0.18$	$\pi_1=0.054$; $\pi_2=0.216$; $\pi_3=0.296$	1.18	3.06	1.42

Of course, the distribution of γ_is and the other parameter choices we just considered is only one example. It alone cannot confirm whether the largest stable IEA consists of two members for *any* heterogeneous distribution of preferences for SGE and any values for n, b, c, ϕ and θ. However, it is important to recognize that the same fundamental tensions remain between countries with different SGE preferences, no matter how those preferences are distributed. Mainly, the free-driving country strictly benefits from SGE levels that match their preferred level, and in an agreement that only governs abatement, there is no way to outright stop the free driver from deploying excessive SGE levels. The best the other countries can do is soften the blow by increasing their abatement. This is why coalitions consisting of members with relatively low preferences for SGE will always abate more than equal size coalitions with members that have relatively high preferences for SGE.

6.1.2 Anti-driver incentives

The stability of IEAs under the threat of solar geoengineering is challenged by two competing incentives. The first incentive, which is a familiar *free-rider incentive*, is the additional payoff a defecting member achieves by leaving the agreement since they lower their abatement responsibilities. The other incentive, which we call the *anti-driver incentive*, is the additional payoff achieved by staying in the agreement and further dampening the free-driver effect. If the free-rider incentive exceeds the anti-driver incentive for any member, the coalition is not stable. We find that this is the case for all coalitions outside of those in Table 5.

We further explore the relationship between these two incentives by examining agreements in situations in which the anti-driver incentive is strong - those in which the free driver can impose significant harm on other countries. Recall, the damages free drivers impose on other countries is increasing with ϕ - the marginal damage for SGE levels above preferred levels - and the gap between $\gamma_i^{\max}$ and γ_i. When ϕ is high and/or the distribution of γ_i is dramatically right-skewed, the free-driver behavior imposes significant costs on other countries, and therefore each has a relatively strong incentive to cooperate within an IEA to increase abatement.

Ultimately we are able to show that the anti-driver incentive could be sufficiently high that it makes membership in an agreement profitable enough to satisfy internal stability while achieving significant abatement beyond non-cooperative levels. To demonstrate, consider an example in which we keep $n = 5$, $b = c = 1$, and $\theta = 0.10$ for ease of comparison with our previous examples, but impose strong anti-driver incentives. In particular the parameter value for ϕ is set relatively high at $\phi = 0.55$, and the distribution of γ_i is: $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.1$, and $\gamma_5 = 0.6$ (i.e., right-skewed distribution). Table 6 contains the Nash equilibrium and full participation abatement levels and payoffs given these example parameters.

Table 6: Nash equilibrium and full participation with strong anti-driver incentives

		Nash equilibrium		full participation	
country i	γ_i	q_i^*	π_i^*	q_i^f	π_i^f
1	0.1	0.465	-0.503	1.0	0.50
2	0.1	0.465	-0.503	1.0	0.50
3	0.1	0.465	-0.503	1.0	0.50
4	0.1	0.465	-0.503	1.0	0.50
5	0.6	0.140	0.570	1.0	0.50
	$\bar{\gamma} = 0.20$	$Q^* = 2.0$	$\Pi^* = -1.44$	$Q^f = 5.00$	$\Pi^f = 2.50$

In the non-cooperative Nash equilibrium, the free-driver chooses an SGE level of $G^{\max} = 1.8$ and earns a payoff of 0.570 while the other countries suffer losses equal to -0.503, leading to an aggregate payoff of $\Pi^* = -1.44$. Note that in the Nash equilibrium, the non free drivers abate significantly more than the free driver (0.47 vs. 0.14), and aggregate abatement is $Q^* = 2.0$. Contrast the Nash equilibrium with full participation in which the coalition of five countries chooses $Q^f = 5$ to maximize their joint payoffs, which leads to a preferred SGE level of zero for all countries and an aggregate payoff of $\Pi^f = 2.50$.

The grand coalition, however, is not internally stable. The free driver has a strong incentive to leave the agreement and reduce their abatement causing the four other countries to increase theirs. By leaving the grand coalition the free driver reduces abatement from 1.0 to 0.14 and earns a payoff of 0.99. We also find that a coalition of size four in which the single non-member is the free driver is also not internally stable; that is, another member has an incentive to leave the coalition, reduce their abatement and increase their payoffs. However, we are able to identify stable coalitions of size $n = 3$ that result in high levels of abatement and increased payoffs relative to the non-cooperative level. The abatement levels and payoffs for an example stable coalition of size three are contained in Table 7.

Table 7: Stable IEA of $n = 3$ with strong anti-driver incentives

country i	γ_i	q_i^s	q_j^t	π_i^s	π_j^t
1	0.1	1.40	—	-0.071	—
2	0.1	1.40	—	-0.071	—
3	0.1	1.40	—	-0.071	—
4	0.1	—	0.47	—	0.794
5	0.6	—	0.14	—	0.961
	$\bar{\gamma} = 0.20$	$Q = 4.79$		$\Pi = 1.54$	

The abatement and payoffs in Table 7 demonstrate it is possible that the threat of SGE can motivate a stable coalition of countries to take on significant abatement responsibilities in order to protect themselves from the free driver. Aggregate abatement with the IEA is $Q = 4.79$, just under the optimal abatement level of 5 in a world without SGE. The three agreement members choose relatively high levels of abatement compared to the two non-members. The free driver abates the least (0.14) and chooses $G^{\max} = 0.126$. The agreement allows the members to reduce their losses compared to the non-cooperative outcome by increasing abatement to protect themselves from the free driver.

For the three IEA members in the stable IEA of size $n = 3$, the anti-driver incentive (i.e., benefit of reducing the level of SGE) is stronger than the free-rider incentive (i.e., avoiding the high abatement cost). Through other examples, we find that the threat of costly SGE can increase stable coalitions well beyond $s = 3$ depending on n and how strong the anti-driver incentives are. In other words, it is possible the threat of SGE expands both the depth (i.e., abatement level) and breadth (i.e., participation level) of stable IEAs compared to a world without SGE. That this is possible, however, does not suggest it is expected. In our example with strong anti-driver incentives in Tables 6 and 7, the marginal damage from SGE beyond the preferred level (ϕ) is over five times the size of the marginal benefit of SGE below the preferred level (θ). While there is some support for assuming the marginal damages exceed the marginal benefits for SGE (e.g., Weitzman 2015), we lack the information to adequately justify the relative values. The important finding is that *if* a free driver can impose significant costs on other countries, SGE opportunities could trigger a subset of countries to form stable agreements that take on significant abatement.

7 Discussion and conclusion

We set out to explore how opportunities to deploy solar geoengineering impact global emissions abatement and the effectiveness of international agreements on climate change. The game-theoretic exercise has led to some new and interesting results. The first is that without cooperation, SGE opportunities can lead to more or less emissions abatement depending on how costly the free-driver's level of SGE is for the other countries. When the aggregate benefits of SGE at the free-driver's preferred level are negative, the threat of SGE causes an increase in aggregate emissions abatement. If the aggregate benefits of SGE are positive, the same threat causes a decrease in abatement. This result helps us better understand the nuance of the "moral hazard" debate which is centered on the conjecture that SGE opportunities will undermine abatement (e.g., Reynolds 2019; Wagner 2021). We show that this is a possibility, but the opposite outcome can also occur - it depends on the distribution of preferences. In the same spirit, the socially optimal level of emissions abatement may increase or decrease under the threat of SGE, it too depends on the distribution of preferences. We show that when the aggregate benefits to SGE under the free-driver are negative (i.e., many countries with relatively low preferences for SGE), SGE opportunities can lead to an increase in emissions abatement.

The IEA we considered is the simplest institution, one that only govern emissions abatement and all countries are free to choose their SGE levels. When countries have homogeneous preferences for SGE, the largest stable IEAs consist of three members, which means that SGE opportunities do not alter the main conclusion from the existing literature using the same underlying global public-goods model (Barrett 2003; Finus and McGinty 2019). With heterogeneous preferences, stability is challenged by two competing incentives. The familiar *free-rider incentive* is the additional payoff a defecting member achieves by leaving the agreement since they lower their abatement responsibilities. The *anti-driver incentive* is the additional payoff achieved by joining an agreement and further dampening the free-driver effect through increased abatement. Ultimately, we find that if the anti-driver incentives are high enough, the threat of SGE can expand both the depth (i.e., abatement level) and breadth (i.e., participation level) of stable IEAs compared to a world without SGE. The threat of a menacing free driver can lead to more cooperation.

What is made clear throughout our analysis is that when an IEA only governs emission abatement, it is very difficult to outright stop the free driver from deploying SGE. An obvious next extension to our analysis is considering different types of IEAs, and in particular institutions that govern both abatement and SGE. One possibility that is informed by the ongoing debate about the acceptability of SGE technologies (e.g., Biermann et al. 2021) is a *non-use agreement*, in which members to an IEA agree not to deploy SGE (Heyen et al. 2019 refer to this a "moratorium agreement"). Of course this type of agreement can only exacerbate the already strong free-rider incentives we observe in the abatement only agreement. Another possible agreement structure inspired by the established IEA literature is a *collective maximization agreement* in which countries jointly determine abatement and SGE levels to maximize the coalition's payoffs. Or a close variant to this, a *no-harm agreement* in which SGE levels are set to match the lowest preferred level of the members (i.e., no member suffers any losses from too-much SGE).

Previous studies have demonstrated the important role financial transfers play at maintaining stability and increasing the effectiveness of IEAs among heterogeneous countries (McGinty 2007). Given the heterogeneity in SGE preferences, one could consider how similar transfer-rules could be designed to better align the would-be free-driver's preferences with the social optimum. The general idea would be that if IEA members earn a significant surplus from jointly cooperating, it may be possible for the coalition to compensate a potential free-rider so that they are weakly better off reducing their SGE deployment to a more efficient level. The challenge with this scheme is that even if it is possible to pay the country with the highest preference to reduce their SGE deployment, then the country with the second highest preference for SGE is the new potential free driver. While a formal analysis is needed, it is easy to see how an agreement with transfers could unravel because of the fluidity of the potential free-driving country.

Our theoretical analysis is a first step in exploring international agreements when solar geoengineering is modeled as a good-or-bad (GoB) and governance is complicated by the threat of a free driver (following Weitzman 2015; Wagner 2021). Of course, our study has many limitations. One limitation is that we only consider one functional form for emissions abatement and SGE benefits/costs. For that reason we cannot generalize our results to other settings, including an often explored case in which both the benefits and costs to abatement are quadratic (e.g., Barrett 1994). Another limitation is that we avoid altogether the potential risks from SGE technologies. SGE technologies are new and untested, which introduces uncertainty about costly side effects, and therefore extending our analysis to include these features is a necessary next step. There is also the related political risk of deploying SGE and becoming a pariah state, which could impact decision making. Finally, our model does not allow countries to invest in counter-SGE, which may help moderate the free-driver problem. Despite these limitations and future opportunities, our research provides new and important insights into the global governance challenge of managing climate change under the threat of solar geoengineering.

Acknowledgments

The authors thank Resources for the Future, the L.A.D. Climate Fund and the National Science Foundation (2033855) for providing financial support for this research. We thank Joe Aldy and Billy Pizer for their detailed feedback and suggestions, and we thank all of the participants that provided comments during the two RFF Solar Geoengineering Research Workshops in May and September, 2022.

References

[1] Abatayo, A. L., Bosetti, V., Casari, M., Ghidoni, R. and Tavoni, M. (2020). "Solar geoengineering may lead to excessive cooling and high strategic uncertainty." *Proceedings of the National Academy of Sciences*, 117(24): 13393-13398.

[2] Aldy, J., T. Felgenhauer, W.A. Pizer, M. Tavoni, M. Belaia, M.E. Borsuk, A. Ghosh, G. Heutel, D. Heyen, J. Horton, D. Keith, C. Merk, J.M. Cruz, J.L. Reynolds, K. Ricke, W. Rickels, S. Shayegh, W. Smith, S. Tilmes, G. Wagner, J.B. Wiener. (2021). "Social science research to inform solar geoengineering; What are the benefits and drawbacks, and for whom?" *Science*, 374(6569): 815-818.

[3] Barrett, S. (1994). "Self-enforcing international environmental agreements. " *Oxford Economic Papers*, 46: 878-894.

[4] Barrett, S. (1999). "A theory of full international cooperation. " *Journal of Theoretical Politics*, 11(4): 519-541.

[5] Barrett, S. (2003). "Environment and statecraft: The strategy of environmental treaty-making: The strategy of environmental treaty-making." Oxford University Press, Oxford, England.

[6] Barrett, S. (2007). "Why cooperate?: the incentive to supply global public goods." Oxford University Press, Oxford, England.

[7] Barrett, S. (2008). "The incredible economics of geoengineering. " *Environmental and Resource Economics*, 39 (1): 45-54.

[8] Biermann, F., J. Oomen, A. Gupta, S.H. Ali, K. Conca, M.A. Hajer, P. Kashwan, L.J. Kotz, M. Leach, D. Messner, C. Okereke, Persson, J. Potocnik, D. Schlosberg, M. Scobie and S.D. VanDeveer. (2022). "Solar geoenegineering: The case for an international non-use agreement. " WIREs Climate Change (754).

[9] Cherry, T., S. Kroll, D.M. McEvoy, D. Campoverde and J.B. Moreno-Cruz (2022). "Climate cooperation in the shadow of solar geoengineering: an experimental investigation of the moral hazard conjecture. " *Environmental Politics*, May: 1-9.

[10] Crutzen, P.J. (2006). "Albedo enhancement by stratospheric sulfur injections: a contribution to resolve a policy dilemma?." *Climatic Change*, 77(3-4): 211.

[11] Finus, M. and M. McGinty. (2019). "The anti-paradox of cooperation: diversity may pay! " *Journal of Economic Behavior & Organization*, 157: 541-559.

[12] Heyen, D., J. Horton and J. Moreno-Cruz (2019). "Strategic implications of counter-geoengineering: Clash or cooperation." *Journal of Environmental Economics and Management*, 95: 153-177.

[13] Intergovernmental Panel on Climate Change. (2022). *Working Group III Contribution to the Sixth Assessment Report.*

[14] Keith, D. (2000). "Geoengineering the climate: History and prospect." *Annual review of energy and the environment*, 25(1): 245-284.

[15] Keith, D., Parson, E. and M. Morgan (2010) "Research on global sun block needed now." *Nature*, 463(7280): 426-427.

[16] Keith, D. (2021). "Toward constructive disagreement about geoengineering." *Science*, 374(6569): 812-815.

[17] McEvoy, D.M. and J.K. Stranlund (2009) "Self-enforcing international environmental agreements with costly monitoring for compliance." *Environmental and Resource Economics*, 42(4): 491-508.

[18] McGinty, M. (2007). "International environmental agreements among asymmetric nations. " *Oxford Economic Papers*, 59(1): 45-62.

[19] Millard-Ball, A. (2012) "The Tuvalu syndrome." *Climatic Change*, 110(3): 1047-1066.

[20] Moreno-Cruz, J.B. (2015) "Mitigation and the geoengineering threat." *Resource and Energy Economics*, 41: 248-263.

[21] NRC [National Research Council]. (2015). "Climate intervention: reflecting sunlight to cool Earth." Washington: National Academies Press.

[22] Reynolds, J.L. (2019). "The governance of solar geoengineering: managing climate change in the Anthropocene." *Oxford University Press*, Oxford, England.

[23] Ricke, K.L., Moreno-Cruz, J.B. and K. Caldeira. (2013) "Strategic incentives for climate geoengineering coalitions to exclude broad participation." *Environmental Research Letters*, 8(1): 014021.

[24] Robock, A. (2008) "20 reasons why geoengineering may be a bad idea." *Bulletin of the Atomic Scientists*, 64(2): 14-18.

[25] Smith, W. and G. Wagner. (2018). "Stratospheric aerosol injection tactics and costs in the first 15 years of deployment " *Environmental Research Letters*, 13(12).

[26] Wagner, G. (2021). *Geoengineering: The Gamble.* Polity Press.

[27] Wagner, G. and M. Weitzman (2012). “Playing God.” *Foreign Policy* (October 24)

[28] Weitzman, M. (2015). “A voting architecture for the governance of free-driver externalities.” *Scandinavian Journal of Economics*, 117: 1049-1068.

[29] Williamson, P. and C. Turley. (2012). “Ocean acidification in a geoengineering context.” *Philosophical Transactions of the Royal Society A*, 370: 4317-4342.