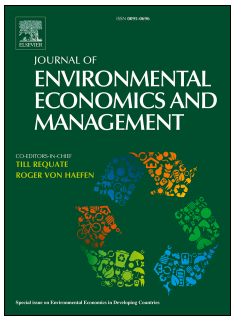

Article:

Heyen, Daniel, Horton, Joshua and Moreno-Cruz, Juan (2019) Strategic implications of counter-geoengineering: clash or cooperation? Journal of Environmental Economics and Management, 95. pp. 153-177. ISSN 0095-0696

https://doi.org/10.1016/j.jeem.2019.03.005

Accepted Manuscript



Strategic implications of counter-geoengineering: Clash or cooperation?

Daniel Heyen, Joshua Horton, Juan Moreno-Cruz

PII: S0095-0696(18)30503-5

DOI: https://doi.org/10.1016/j.jeem.2019.03.005

Reference: YJEEM 2226

To appear in: *Journal of Environmental Economics and Management*

Received Date: 17 July 2018

Revised Date: 13 March 2019

Accepted Date: 15 March 2019

Please cite this article as: Heyen, D., Horton, J., Moreno-Cruz, J., Strategic implications of counter-geoengineering: Clash or cooperation?, *Journal of Environmental Economics and Management* (2019), doi: https://doi.org/10.1016/j.jeem.2019.03.005.

Strategic Implications of Counter-Geoengineering: Clash or Cooperation? *

Daniel Heyen†
Chair of Integr. Risk Manag. and Econ.
ETH Zurich

Joshua Horton
John A. Paulson School of Engineering and Applied Sciences
Harvard University

Juan Moreno-Cruz
School of Environment, Enterprise and Development
University of Waterloo

March 15, 2019

Abstract

Solar geoengineering has received increasing attention as an option to temporarily stabilize global temperatures. A key concern is that heterogeneous preferences over the optimal amount of cooling combined with low deployment costs may allow the country with the strongest incentive for cooling, the so-called free-driver, to impose a substantial externality on the rest of the world. We analyze whether the threat of *counter-geoengineering* technologies capable of negating the climatic effects of solar geoengineering can overcome the free-driver problem and tilt the game in favor of international cooperation. Our game-theoretical model of countries with asymmetric preferences allows for a rigorous analysis of the strategic interaction surrounding solar geoengineering and counter-geoengineering. We find that counter-geoengineering prevents the free-driver outcome, but not always with benign effects. The presence of counter-geoengineering leads to either a climate clash where countries engage in a non-cooperative escalation of opposing climate interventions (negative welfare effect), a moratorium treaty where countries commit to abstain from either type of climate intervention (indeterminate welfare effect), or cooperative deployment of solar geoengineering (positive welfare effect). We show that the outcome depends crucially on the degree of asymmetry in temperature preferences between countries.

Keywords: climate intervention; solar geoengineering; counter-geoengineering; free-driver; strategic conflicts; game theory; cooperation; externality; global warming;

*We gratefully acknowledge funding from the ESRC Centre for Climate Change Economics and Policy and Grantham Foundation for the Protection of the Environment and thank the organisers of the Harvard Solar Geoengineering Research Residency. D.Heyen gratefully acknowledges support from the German Research Foundation (DFG), grant HE 7551/1. Moreno-Cruz acknowledges support from the Canada Research Chairs program. This paper has benefited substantially from discussions and comments at the following events: seminars at FEEM Milan and the Institute for Science, Innovation and Technology Oxford; the EAERE 2017 conference in Athens, Climate Engineering Conference CEC 17 in Berlin, EENR 2018 workshop in Orléans, WCERE 2018 conference in Gothenburg and the GRI workshop at LSE.

†Corresponding author. dheyen@ethz.ch

international environmental agreements
JEL Codes: Q54; H41; D62; D02; D74.

1 Introduction

One option for addressing climate change that is gaining increased attention is Solar Geoengineering (SG), also known as Solar Radiation Management (SRM) (National Research Council 2015). SG aims at (partially) compensating the global warming caused by increased atmospheric levels of greenhouse gases by either releasing cooling particles in the stratosphere (stratospheric aerosol injection) or modifying marine cloud reflectivity (marine cloud brightening). While an optimally designed and implemented SG scheme appears to have the potential to reduce global temperature damages (Moreno-Cruz et al. 2012; Keith and MacMartin 2015; National Research Council 2015), there are concerns that SG's potential benefits are reduced and possibly even reversed in a decentralized world of international 'anarchy'. A key fear is that presumably low deployment costs (McClellan et al. 2012; Smith and Wagner 2018) together with asymmetric preferences over the optimal global temperature change (Heyen et al. 2015) may result in unilateral SG deployment that harms the rest of the world (Horton 2011; National Research Council 2015; Pasztor et al. 2017). This has been termed the "free-driver" problem (Weitzman 2015).[1]

Against this backdrop of potentially welfare deteriorating strategic incentives surrounding a potentially beneficial technology, a recent paper (Parker et al. 2018) explores the idea of counter-geoengineering (CG), or a set of technologies that would give countries threatened by or subject to the free-driver's whims a tool for quickly negating what they regard as harmful SG. While states opposed to unilateral SG could impose a variety of indirect costs (such as trade sanctions) on a free-driver in an effort to halt deployment (Horton 2011), CG would entail a direct response intended to curb such behavior. CG could take one of two forms. 'Neutralizing' CG would entail rendering SG particles inert by, for example, injecting a base to counteract the sulphate aerosols most commonly considered for SG, or employing techniques to accelerate the coagulation and hence atmospheric deposition of SG particles. By contrast, 'countervailing' CG would involve reversing the effects of SG particles by releasing warming agents such as greenhouse gases or specially engineered solid particles to counter the change in radiative forcing caused by SG. Both forms of CG are possible, but neither currently exists; successful development would require achieving adequate forcing efficacy at reasonable cost. The reason why the availability of such CG capabilities might prove beneficial is obviously

[1]The 'free-driver' terminology emphasizes two things: first, the public good nature of interventions in the global climate, i.e. non-excludability and non-rivalry; and second, the potential for a single actor to get in the 'driver seat' (due to low deployment costs) and shape the global climate as she wishes, in contrast to the well-known 'free-rider' problem (Stavins 2011). To emphasize heterogeneous preferences, Weitzman (2015) also refers to SG as a 'public gob', that is, a public good or bad, depending on the amount deployed.

not because further global warming is globally desirable; rather, the very availability of CG might deter the free-driver from unilateral SG deployment and instead promote international cooperation on climate interventions. If CG has this potential to steer climate technology use to overall beneficial levels, then there is a case for countries to invest in CG today as a deterrent to future unilateral SG use.

The present paper provides a first rigorous analysis of the strategic effects of introducing SG and CG into an otherwise standard model of climate economics. We regard SG and CG as two separate and contrasting forms of *climate intervention*. With this understanding, we model climate intervention (via either SG or CG) as a public good game: the operational costs of any climate intervention are borne only by the deploying country, whereas the resulting global temperature change affects all countries. The latter is captured by a non-monotonic benefit function that exhibits an optimal level of global temperature change. A key assumption in our model is that countries disagree about the optimal temperature change and therefore their preferred amount of climate intervention. We give countries two distinct options for cooperation. The first is a *deployment treaty* where countries jointly decide on the climate intervention that maximizes the coalition's overall payoff. The second option, which constitutes one of the novel contributions of the present paper, is a *moratorium treaty*. In a moratorium, an idea often raised in the geoengineering debate (Parker 2014; Victor 2008; Zürn and Schäfer 2013; Parson and Keith 2013), countries commit themselves to abstain from any form of climate intervention. As usual, we assume each country individually determines its willingness to cooperate by comparing payoffs under alternative treaties to the non-cooperative outcome. We study how CG affects the incentives to cooperate by analyzing the game first when only SG is available and hence climate intervention is restricted to cooling; and second when CG is also available and countries are able to cool or warm.

Despite this parsimonious setting, our model delivers a rich set of findings. In the absence of CG, if countries are sufficiently different in their preferred temperature, the non-cooperative outcome is a free-driver equilibrium. If countries have similar preferred temperatures, then the non-cooperative outcome is a free-rider equilibrium. In both cases, cooperation incentives are overall weak: The moratorium treaty is never supported by both countries and therefore unstable, and the deployment treaty is only stable for a relatively small set of parameter constellations. The effect of introducing CG is that the free-driver constellation is not a Nash equilibrium anymore: those who regard the free-driver's cooling as excessive now have a tool to counteract it, and they use it. Absent the opportunity to cooperate, this results in a 'climate clash', an escalation of cooling by SG and warming by CG that typically has no winners and is overall sharply detrimental. If cooperation is an option, however, this bleak outlook of CG in a non-cooperative world may encourage countries to work together. In particular, the free-driver, typically unwilling to cooperate in the absence of CG, may be ready to compromise on the amount of climate intervention. Yet cooperation is not assured, and the outcome might still be

a destructive climate clash. And even if cooperation does occur, it might take the form of a moratorium, which could be worse than the free-driver outcome if climate damages are sufficiently high. The outcome depends crucially on the degree of asymmetry in temperature preferences between countries.

Our paper contributes to two strands of the literature. The first is the emerging literature on strategic interaction and governance surrounding solar geoengineering (Klepper and Rickels 2014; Horton 2011; Barrett et al. 2014; Barrett 2014). We make three specific contributions to this literature. The first pertains to research on non-cooperative geoengineering outcomes under different types of asymmetry (Moreno-Cruz 2015; Manoussi and Xepapadeas 2017; Manoussi et al. 2018; Urpelainen 2012; Weitzman 2015; Heyen 2016). These papers, with the exception of Weitzman (2015) and Heyen (2016), focus on asymmetry in terms of heterogeneous side-effects or different levels of uncertainty but maintain the assumption that countries' preferences regarding the desired climate outcome are perfectly aligned. Our work advances this literature by putting heterogeneous preferences over the global average temperature center-stage. We believe that this source of asymmetry is crucial to capture the idea of excessive SG, frequently referred to as 'free-driving'.[2] Second, we extend Weitzman (2015) and Heyen (2016) in several ways, most importantly by adding the option of CG. The first paper that has put CG center-stage is Parker et al. (2018). We advance their analysis by using a richer and calibrated game-theoretical model (also see Appendix C on the timing of the non-cooperative game) and by studying incentives for cooperation. Third, cooperation incentives surrounding SG have been studied by Millard-Ball (2012) and Ricke et al. (2013). We extend this literature in two ways: first, we study heterogeneous preferences over the global temperature and show that cooperation incentives crucially depend on the degree to which countries disagree about the desired climate. Second, we introduce CG and demonstrate that CG significantly alters countries' incentives to cooperate.

The second strand of the literature we contribute to is the environmental economics literature on public goods, externalities and cooperation, see for instance Barrett (1994) and Finus (2008). Despite the dominant approach of considering symmetric players, the subtle and important role of heterogeneity in strategic environmental settings has been noted and emphasized (Barrett 2001; McGinty 2007; Finus and McGinty 2018). In this context our paper makes three innovations. First, following Weitzman (2015) and Heyen (2016) we allow for the over-provision of a public good by modelling non-monotonic benefit functions with *heterogeneous optimal levels*, a feature not present in other asymmetric public good settings; in contrast to Weitzman (2015), however, we include deployment costs, which gives rise to much richer findings, and situate this discussion in a standard public-good setting with a smooth benefit function. The second

[2]Emmerling and Tavoni (2017) interpret free-driving as over-provision relative to the cooperative (global first-best) solution. This is why free-driving in their sense can also occur in settings with symmetric preferences because countries do not account for the externalities caused by SG.

innovation of our paper is to consider CG, which essentially allows agents to make 'negative' contributions to a public good, an aspect that may be of more general interest in future research beyond geoengineering. Finally, the third contribution of our paper to the environmental economics literature on public goods and cooperation is to introduce a moratorium treaty, i.e. we give agents the option to jointly abstain from contributions to the public good altogether. This form of cooperation, which has not received attention in the literature – unsurprising in light of the focus on symmetric settings – may be of general interest for the analysis of strategic interaction of agents with asymmetric preferences, in particular when side-payments are not available.

We proceed as follows. Sec. 2 presents the model components in detail, with a focus on the case of two countries. Sec. 3 analyzes the deployment stage, in particular the non-cooperative outcomes both with and without CG; these non-cooperative outcomes are the reference points for countries when choosing whether to cooperate, discussed in Sec. 4. Sec. 5 calibrates the model. We then present two robustness checks: Sec. 6 relaxes the assumption that SG and CG have the same cost structure and Sec. 7 generalizes from two to n countries. Sec. 8 concludes.

2 The Model

In our model two countries with asymmetric preferences decide on climate intervention levels, i.e. changes to global temperatures using either SG or CG. Initially we assume SG and CG are symmetric in terms of costs, but we relax that assumption in section 6. The general case with n countries is covered in section 7. Because changes to global temperatures affect every country, we model climate intervention as a public good provision game.

2.1 Timing of Events

In the first stage the two countries can cooperate by forming a climate intervention treaty. The two available options are a *moratorium treaty*, in which the countries commit themselves to deploy neither SG nor CG, and a *deployment treaty*, in which the countries within the coalition commit themselves to choose technology levels so as to maximize the coalition's sum of payoffs. By definition, the deployment treaty implements the climate intervention that maximizes global welfare. If neither treaty comes into effect, countries in the second period choose their climate intervention levels simultaneously and non-cooperatively.[3] In order to assess the game-changing potential of CG we contrast two cases. First, the *'SG only'* case when CG is not available and hence climate interventions are restricted to cooling. We then compare this with the *'CG available'* case in which

[3]We consider the simultaneous game structure to be the most realistic representation of non-cooperative interaction on climate intervention and therefore deviate from the sequential order in Parker et al. (2018). A detailed discussion of the time structure of our model can be found in Appendix C.

countries have the option to increase or decrease global mean temperatures. The non-cooperative outcome depends on whether CG is available or not, and this will in turn have implications for the attractiveness of the treaties.

2.2 Definitions and Assumptions

Climate intervention levels $g_i \in \mathbb{R}$, $i = A, B$, are measured in terms of the resulting temperature change. The global average temperature under climate change T_0 – the *status quo* temperature countries face when making their climate intervention choice – is normalized to zero, $T_0 = 0$.[4] Hence, the *change in global average temperature* T due to climate intervention is

$$T = g_A + g_B. \tag{1}$$

We assume that costs and benefits are quadratic (Barrett 1994; McGinty 2007; Finus and Rübbelke 2013; Diamantoudi and Sartzetakis 2006; Heyen 2016).[5] The costs are

$$C(g_i) = \frac{c}{2} g_i^2 \quad , \; i = A, B \tag{2}$$

with $c > 0$.[6] We assume for simplicity that SG and CG have the same country-independent cost structure. The general case with asymmetric cost structures for SG and CG is covered in Section 6. The climate benefits are

$$B_i(T) = -\frac{b}{2} (T_i - T)^2 \quad , \; i = A, B \tag{3}$$

with $b > 0$.[7] We define the *benefit-cost parameter* $\theta = b/c$. In contrast to operational costs which are private, the benefit function B reflects the public good nature of the climate intervention: Benefits depend on the global average temperature T and hence on the climate intervention levels of both countries. The benefits are highest at $T = T_i$ which justifies calling T_i *country i's preferred temperature*. For a country that suffers from climate change, which is the typical situation, $T_i < 0$.

[4] This temperature includes the effects of any previous mitigation efforts. We do not model mitigation explicitly. The reason is that we are interested in the strategic interaction surrounding SG and CG that can be expected to unfold on a fairly short timescale: climate interventions would have an almost immediate temperature response effect, whereas the effects of mitigation need much longer to materialize.

[5] The calibration in Sec. 5 justifies this assumption.

[6] For the analysis within a public good framework it is crucial to focus on those costs that are borne by each country individually. In the context of a climate intervention these are the direct operational costs of modifying the global climate. Indirect costs that are climate-related are captured within the non-monotonic benefit function B. Indirect costs not related to climate indicators, e.g. health impacts from sulfur particles, are not incorporated in our simple model; including them would likely only strengthen our results as they add another source of external effects, see the discussion in section 8. The cost function (2) captures that the deployment costs a country has to bear are convex in that country's level of geoengineering deployment but is not able to capture that deployment costs may also be affected by deployment by others.

[7] Here we make the simplifying assumption that countries assess climate outcomes using temperature as a proxy for all indicators of climate change. Of course, these simplification comes with limitations and in section 8 we discuss the role of indicators other than temperature.

Country i's *payoff* under the climate intervention profile $g = (g_i)_{i=A,B}$ is

$$\pi_i(g) = B_i(T) - C(g_i) \ . \tag{4}$$

A central component of the model is to allow for different T_i and hence heterogeneous preferences over the optimal amount of climate intervention.[8] Without loss of generality let $T_A \leq T_B$. Accordingly, from now on A is the country that favours relatively strong deployment of SG, whereas country B prefers moderate cooling or no cooling at all. We define the *mean optimal temperature change* $\bar{T} = \frac{T_A+T_B}{2}$ and write

$$T_A = \bar{T} - \Delta \quad , \quad T_B = \bar{T} + \Delta, \quad \text{where} \quad \Delta = \frac{T_B - T_A}{2}. \tag{5}$$

We refer to Δ as the *asymmetry parameter* which equals the standard deviation of the optimal temperature changes T_A and T_B. For $\Delta = 0$, both countries agree on how much the climate ought to change; the higher Δ, the higher the disagreement between the two countries in terms of how to set the global thermostat.[9] One of the advantages of this definition of Δ is that it can easily be extended to the general n country case that we discuss in section 7.

Regarding the overall desirability of some amount of SG, we assume that at the time countries consider a climate intervention through SG (or CG), past efforts at mitigation and 'negative emissions' such as bioenergy with carbon capture and storage (BECCS) have proved insufficient to curb temperatures.

Assumption 1. *The world without any climate intervention is on average too warm,* $\bar{T} < 0$.

In particular, $T_A < 0$. We do not impose assumptions on T_B, so country B might prefer a warmer climate, $T_B > 0$.

2.3 The Decision to Enter a Treaty

We model a climate intervention treaty in line with the literature on international environmental agreements (e.g. Barrett 1994, 2001; Finus 2008). Instead of joint decisions

[8] It is worth emphasizing that our model's approach to capture heterogeneity in terms of different optimal *levels* of a public good is novel. With the exception of Weitzman (2015) and Heyen (2016), the typical focus in the literature has been to assume the same optimal level of the public good but different slopes of the marginal benefit function (e.g. McGinty 2007).

[9] A simple illustrative example provides evidence that countries may prefer different global average temperatures. Assume country A and country B to have pre-industrial temperatures of $16°C$ and $10°C$, respectively. Further assume that climate change increases temperatures in both countries by $3°C$. The climate impact literature suggests that growth rates are maximal for a certain *universal*, i.e. country-independent, temperature; Burke et al. (2015) find growth rates to follow a quadratic inverted U shape with a maximum at around $13°C$. If country A and country B both regard $13°C$ as their optimal temperature, then we have in our notation $T_A = -6°C$ and $T_B = 0°C$, resulting in $\Delta = 3°C$. Such a universal optimal temperature, even if countries' preferences are only partially determined by it, provides a strong argument for heterogeneous preferences over climate intervention in a world of heterogeneous baseline temperatures.

on emission abatement levels, countries in a coalition here jointly decide on *climate intervention levels*. In the first type of treaty, the **deployment treaty**, countries choose the amount of SG that maximizes the coalition's total payoff, i.e. the sum of payoffs across its members. One of the innovations of our paper is to allow for a second type of treaty, the **moratorium treaty**. Here, the countries commit themselves to abstain from climate interventions altogether, $g_i = 0$. One reason to consider this additional type of treaty is the importance of a moratorium in the geoengineering debate (Victor 2008; Parker 2014); furthermore, the aspect of winners and losers is particularly pronounced in the present paper and a moratorium treaty – by definition less appealing than a deployment treaty in terms of the sum of payoffs – might possibly be attractive due to its distributional implications.

In this context it is important to note that we do not include side payments (also known as transfers) in our model. The importance of side payments in increasing the attractiveness of cooperation has often been noted, especially for countries with asymmetric preferences (McGinty 2007; Barrett 2001).[10] Yet we often observe that international treaties designed to overcome domestic interests face strong opposition and that side payments in particular are often seen as politically unacceptable (Gampfer et al. 2014; Diederich and Goeschl 2017). This suggests that studying incentives for cooperation that do not rely on transfers is an important benchmark. The deployment treaty and moratorium treaty are two specific, yet salient, forms of cooperation in the absence of transfers.

As usual in the literature on international environmental agreements we define a treaty to be *stable* if it is a Nash equilibrium in membership strategies. With only two countries this condition reduces to determining whether both countries prefer the treaty in question over the non-cooperative outcome (for the general case see section 7). Because the non-cooperative outcome depends on whether CG is available or not, the stability of a treaty also depends on the availability of CG.[11] We determine stability of the two possible treaties, moratorium treaty and deployment treaty, separately. Therefore it may happen that both treaties are stable. While equilibrium selection is not a focus of our paper, we aim to make the analysis in the $n = 2$ case as easy to follow as possible and hence make the following tie-breaking assumption.

Assumption 2 (Tie-breaking rule)**.** *If both treaties are stable, i.e. if both countries are willing to enter either of the two treaties, then the one most preferred by both countries comes into effect if there is such a clear ordering; if countries disagree on the preferred order, we assume that the moratorium treaty comes into effect.*

[10] Indeed, in the absence of negotiation and transaction costs, it is well known that transfer schemes exist to ensure that the socially optimal configuration makes each party better off (Coase 1960).

[11] Furthermore, with only two countries it does not make a difference whether the coalition is modelled as an *open membership* game, in which a country can enter a coalition without the other members' invitation, or an *exclusive club*, where access to a coalition is conditional on the members' consent (Ricke et al. 2013). See section 7 for a treatment of the case with n countries.

The rationale for this tie-breaking rule is that the status quo of non-deployment may be a focal point, for instance because an error of geoengineering 'commission' is assumed to be worse than an error of geoengineering 'omission' (Weitzman 2015). We will see below that equilibrium selection has a significant impact on the analysis.

We proceed with the equilibrium analysis. We first discuss the non-cooperative equilibria, the fallback option when none of the treaties comes into effect. The relative attractiveness of the non-cooperative case, in turn, determines countries' willingness to enter the moratorium and/or deployment treaty.

3 Optimal Deployment and Non-cooperative Equilibria

We solve the equilibrium via backward induction and thus begin our description with the climate intervention deployment stage. The countries simultaneously choose $g_i \in \mathbb{R}$, $i = A, B$. In the 'SG only' case, deployment is restricted to cooling, $g_i \leq 0$. When CG is available, any temperature level $g_i \in \mathbb{R}$ is feasible.[12]

3.1 Global Optimum

We denote by $(g_i^{**})_{i=A,B}$ the socially optimal configuration that maximizes global welfare $\pi(g) = \pi_A(g) + \pi_B(g)$. The solution to this problem following standard procedure is

$$g_i^{**} = \frac{2\theta}{4\theta + 1}\bar{T} \quad , \; i = A, B \; . \tag{6}$$

It is efficient that both countries deploy the same amount of solar geoengineering due to the homogeneous cost structure. Owing to $\bar{T} < 0$ (Assumption 1), the socially optimal deployment scheme features SG deployment by both countries. Whether CG is available or not has, therefore, no implications for the socially optimal deployment profile. It is straightforward to see that (6) increases in the benefit-cost ratio θ.

3.2 Non-cooperative equilibria

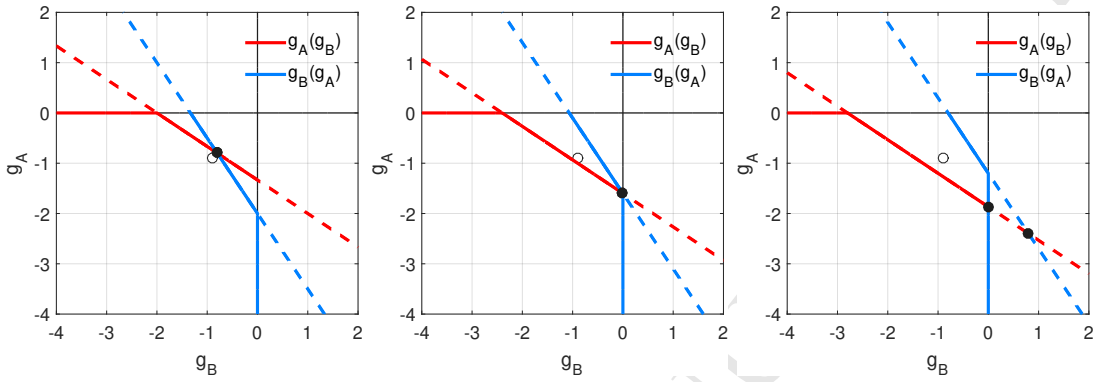
The first step in determining the non-cooperative Nash equilibria is to calculate the best response functions. The conceptually simplest case is when CG is available and hence $g_i \in \mathbb{R}$ unrestricted. In this case, the best response of country i to the other country's climate intervention level g_{-i} is characterized by the first-order condition $d\pi_i(g_i; g_{-i})/dg_i = 0$. In the 'SG only' case, we also need to check whether the non-

[12] The absence of an upper limit on the level of CG corresponds to 'countervailing' CG (Parker et al. 2018), e.g. the release of a potent GHG. The maximal amount of 'neutralizing' CG, in contrast, would be a function of the deployed SG level. We find that CG levels are smaller than SG levels, see below, so that in the context of the present paper it is inconsequential whether we understand CG as countervailing or neutralizing.

positive constraint binds. We get the best response function

$$g_i(g_{-i}) = \begin{cases} \min\left\{\frac{\theta}{\theta+1}\left(T_i - g_{-i}\right), 0\right\} & \text{SG only} \\ \frac{\theta}{\theta+1}\left(T_i - g_{-i}\right) & \text{CG available} \end{cases} \tag{7}$$

Figure 1 shows how the best response functions depend on the asymmetry Δ.



(a) $\Delta = 0\,°C$ (b) $\Delta = 0.4\,°C$ (c) $\Delta = 0.8\,°C$

Figure 1: Best response functions (country A in red, country B in blue) for different asymmetry levels Δ. In all plots $\theta = 2$ and $\bar{T} = -2\,°C$. The solid lines and dashed lines show the best response functions in the 'SG only' and 'CG available' case, respectively. The unfilled circle indicates the socially optimal benchmark (g_A^{**}, g_B^{**}). The asymmetry threshold is $\bar{\Delta} = 0.4\,°C$, see (8). For $\Delta > \bar{\Delta}$ the equilibrium outcome, indicated by a filled black circle, depends on whether CG is available or not.

We now summarize non-cooperative equilibria in the 'SG only' and 'CG available' scenarios and hence determine the game-changing effect of CG in the absence of cooperation possibilities. We define the *asymmetry threshold*

$$\bar{\Delta} := -\frac{1}{2\theta+1}\bar{T}\ . \tag{8}$$

The asymmetry threshold plays an important role in the following discussion, as it helps explain which equilibria obtain under different conditions.

Proposition 1 (Game-changing potential of CG. Non-cooperative equilibria)**.** *There is a unique Nash equilibrium and the outcome depends on parameter settings and whether CG is available:*

(i) The 'SG only' case. For low levels of asymmetry, $\Delta < \bar{\Delta}$, both countries engage in SG. We refer to this outcome as the ***free-rider equilibrium****,*

$$g_A^* = \frac{\theta}{2\theta+1}\bar{T} - \theta\Delta \ < 0 \quad , \quad g_B^* = \frac{\theta}{2\theta+1}\bar{T} + \theta\Delta \ < 0\ . \tag{9}$$

For high levels of asymmetry, $\Delta \geq \bar{\Delta}$, only country A deploys SG. We refer to this

outcome as the ***free-driver equilibrium****,*

$$g_A^* = \frac{\theta}{\theta+1} T_A \quad , \quad g_B^* = 0 \ . \tag{10}$$

(ii) The 'CG available' case. For low levels of asymmetry, $\Delta < \bar{\Delta}$, there is no incentive to deploy CG. The unique equilibrium is therefore the free-rider outcome (9).

For high levels of asymmetry, $\Delta \geq \bar{\Delta}$, country A cools and, simultanously, country B warms. We refer to this outcome as the ***climate clash equilibrium****,*

$$g_A^* = \frac{\theta}{2\theta+1}\bar{T} - \theta\Delta \ < 0 \quad , \quad g_B^* = \frac{\theta}{2\theta+1}\bar{T} + \theta\Delta \ \geq 0 \ .^{13} \tag{11}$$

(iii) For a fixed $\Delta \geq \bar{\Delta}$, switching from the 'SG only' to the 'CG available' case is always bad for country A and makes country B worse off if and only if $\theta > \frac{1+\sqrt{5}}{2}$. Total welfare is unambiguously reduced.

Proof. See appendix A. □

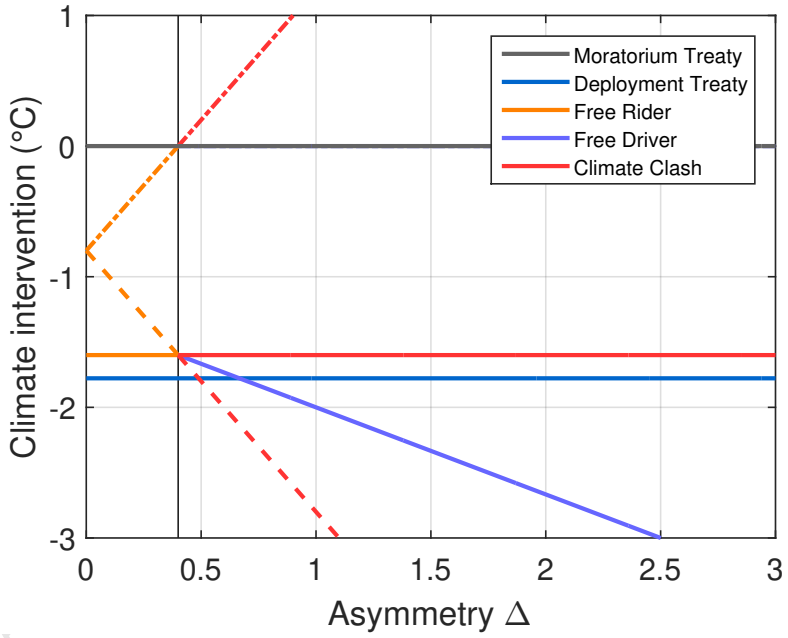


Figure 2: Climate intervention levels of the non-cooperative equilibria as a function of the asymmetry Δ. The parameter settings are as in Figure 1, i.e. $\theta = 2$ and $\bar{T} = -2\ ^\circ C$. The vertical line is at the asymmetry threshold $\bar{\Delta}$. The dashed and dot-dashed lines represent the deployment by country A and country B, respectively, while the solid lines show total levels. For comparison we include the total climate intervention levels under the moratorium treaty (zero) and the deployment treaty (the total level is twice the amount in (6)).

[13] Note that for $\bar{T} = 0$ the climate clash would involve opposite climate interventions by country A and B, $g_A^* = -g_B^*$, with $g^* = g_A^* + g_B^* = 0$, and that deployment treaty and moratorium treaty would coincide when $\bar{T} = 0$. We thank a referee for pointing this out.

Figure 2 shows climate intervention levels for both countries under the non-cooperative equilibria as a function of the asymmetry level Δ. For comparison we include the total SG level under the moratorium treaty (i.e. zero) and the deployment treaty. The free-driver SG level (solid purple line) depends on country A's optimal temperature change T_A but not on T_B and hence the cooling intensifies as the asymmetry level Δ increases. The total temperature change in the climate clash (solid red line) matches the free-rider level (solid orange line) and is independent of the asymmetry level Δ, but is the result of ever diverging SG and CG levels (dashed and dot-dashed red lines) by country A and country B respectively.[14]

The free-rider equilibrium is a well-known outcome in the literature; in particular, the symmetric case $\Delta = 0$ is of this type. The more interesting outcome in the 'SG only' case is the *free-driver* equilibrium. The terminology is from Weitzman (2015) who develops the concept of over-provision of a public good in a setting without deployment costs and with a specific kinked utility function. Our definition coincides with the one in Heyen (2016). The defining characteristic of the free-driver equilibrium is that cooling is excessive from country B's perspective, $T \leq T_B$, and country A is essentially in control of the global thermostat. This excessive cooling does not necessarily imply that country B is worse off relative to a world without any climate intervention. Importantly for our analysis, there is no free-driver equilibrium anymore once CG is available. Country B now has a tool to counter the over-provision of the public good, and due to zero marginal costs (at the point of non-deployment), country B uses this tool. The best response of country A, in turn, is to increase its SG efforts. The only reason why SG and CG levels are bounded in this escalation equilibrium is the convexity of the cost function.

This section has demonstrated that without the possibility of cooperation, CG transforms a free-driver outcome into a climate clash, i.e. those parameter constellations that led to a free-driver equilibrium in the 'SG only' case now lead to a climate clash equilibrium when CG is available. This transformation is overall detrimental as countries waste significant resources on SG and CG. But can CG play a more positive role in the context of cooperation? The next section is dedicated to this question.

4 Incentives for Cooperation

This section analyzes the incentives to cooperate on climate intervention via either a deployment treaty or a moratorium treaty. We begin with the 'SG only' case in section 4.1 and cover the 'CG available' case in section 4.2. All findings are illustrated in Figure 3.

[14]The total deployment level in the climate clash is independent of the level of asymmetry if and only if SG and CG have the same cost structure, see section 6.

4.1 Cooperation incentives when only SG is available

The non-cooperative deployment equilibria derived in the previous section (cf. Proposition 1) are the appropriate reference points when countries are deciding whether they are willing to cooperate by entering a moratorium or deployment treaty. We start with the low asymmetry case where non-cooperation would result in the free-rider outcome.

Proposition 2 (Cooperation incentives in the 'SG only' case. Low asymmetry, $\Delta < \bar{\Delta}$)**.** *Country A prefers the deployment treaty over the free-rider equilibrium irrespective of the level of asymmetry Δ. There is however a value $\Delta_{\text{Max}}^{\text{FreeRider}} \in [0, \bar{\Delta}]$ such that country B prefers the deployment treaty only when $0 \leq \Delta < \Delta_{\text{Max}}^{\text{FreeRider}}$, which is therefore the region where the deployment treaty comes into effect. Both countries prefer the non-cooperative free-rider equilibrium to the moratorium treaty.*

Proof. The algebraic expression for $\Delta_{\text{Max}}^{\text{FreeRider}}$ and derivations are in Appendix A. □

That neither country finds the moratorium treaty attractive is intuitive as both countries engage in SG in the non-cooperative equilibrium, indicating that they find SG valuable even under these non-cooperative conditions; to completely abstain from SG in a moratorium treaty then must be unattractive. The reason why country A prefers the deployment treaty to the non-cooperative free-rider outcome is cost-sharing. The disadvantage from having to compromise with country B on SG deployment levels is, due to the relatively aligned preferences in low asymmetry settings, small compared to the gain from splitting deployment costs. Country B opposes the deployment treaty for asymmetry levels above $\Delta_{\text{Max}}^{\text{FreeRider}}$ since the final temperature outcome in the non-cooperative free-rider equilibrium is close to country B's optimal level T_B (matching this level exactly at $\Delta = \bar{\Delta}$) and country A shoulders the main part of deployment cost. In other words, country B is free-riding on country A's SG deployment. We will see below that country B's opposition to the deployment treaty also extends into the free-driver and climate clash region.

We move on to the case of high asymmetry, where the non-cooperative outcome would be the free-driver equilibrium.

Proposition 3 (Cooperation incentives in the 'SG only' case. High asymmetry, $\Delta \geq \bar{\Delta}$)**.** *There exist values $\Delta_{\text{Min}}^{\text{SG}}$, $\Delta_{\text{Max}}^{\text{SG}}$ and $\Delta_{\text{Morat}}^{\text{SG}}$, all of them larger than $\bar{\Delta}$, such that:*

(i) Country A prefers the free-driver equilibrium to the moratorium treaty throughout and prefers the deployment treaty over the free-driver equilibrium if $\Delta < \Delta_{\text{Max}}^{\text{SG}}$.

(ii) Country B opts for the moratorium treaty when $\Delta > \Delta_{\text{Morat}}^{\text{SG}}$ and prefers the deployment treaty over the free-driver equilibrium if $\Delta > \Delta_{\text{Min}}^{\text{SG}}$. It is $\bar{\Delta} < \Delta_{\text{Min}}^{\text{SG}} < \Delta_{\text{Max}}^{\text{SG}}$.

Therefore, the deployment treaty is stable for $\Delta_{\text{Min}}^{\text{SG}} < \Delta < \Delta_{\text{Max}}^{\text{SG}}$, whereas the moratorium is never stable.

Proof. The algebraic expressions for all relevant levels of the asymmetry parameter Δ and other derivations are in Appendix A. □

Here we see for the first time the appeal of a world without any climate intervention. Country B is willing to enter the moratorium treaty if the disadvantage from being dominated by the free-driver is sufficiently high. However, it is intuitive that the moratorium is not appealing to country A, the free-driver. Therefore, there are no circumstances under which the moratorium treaty can be expected to materialize. The deployment treaty has better chances to form. If the asymmetry exceeds $\Delta^{\text{SG}}_{\text{Min}}$, the free-driver outcome is too harmful for country B which is hence willing to enter the deployment treaty.[15] Country A is also willing to enter the deployment treaty, yet under almost inverse conditions. Specifically, for relatively moderate asymmetry levels, $\Delta < \Delta^{\text{SG}}_{\text{Max}}$, the sharing of deployment costs is attractive enough to justify the compromise in temperature levels. For asymmetry levels higher than $\Delta^{\text{SG}}_{\text{Max}}$, however, the gap between the temperature compromise implicit in the deployment treaty and what country A would like to implement is too wide. But $\Delta^{\text{SG}}_{\text{Min}} < \Delta^{\text{SG}}_{\text{Max}}$, and so there do exist constellations where countries, faced with a looming free-driver outcome, decide to cooperate.

4.2 Cooperation incentives with CG

If asymmetry is low, $\Delta \leq \bar{\Delta}$, cooperation incentives are not changed by the availability of CG as countries have no incentives to deploy CG anyway. We hence focus on the high-asymmetry case where non-cooperation would result in the climate clash.

Proposition 4 (Cooperation incentives in the 'CG available' case. High asymmetry, $\Delta \geq \bar{\Delta}$)**.** *There exist values* $\Delta^{\text{CG}}_{\text{Min}}$, $\Delta^{\text{CG,A}}_{\text{Morat}}$, *and* $\Delta^{\text{CG,B}}_{\text{Morat}}$, *all of them larger than* $\bar{\Delta}$, *and the positive value* $\Delta^{\text{B}}_{\text{Morat,Treaty}}$ *such that:*

(i) Country A unambiguously prefers the deployment treaty over both the climate clash and the moratorium treaty, and prefers the moratorium over the climate clash iff $\Delta > \Delta^{\text{CG,A}}_{\text{Morat}}$.

(ii) Country B prefers the deployment treaty over the climate clash iff $\Delta > \Delta^{\text{CG}}_{\text{Min}} > \bar{\Delta}$, *prefers the moratorium over the climate clash iff* $\Delta > \Delta^{\text{CG,B}}_{\text{Morat}}$, *and prefers the moratorium over the deployment treaty iff* $\Delta > \Delta^{\text{B}}_{\text{Morat,Treaty}}$. *While* $\Delta^{\text{CG,B}}_{\text{Morat}} < \Delta^{\text{CG,A}}_{\text{Morat}}$, *the size of* $\Delta^{\text{B}}_{\text{Morat,Treaty}}$ *relative to other critical levels depends on parameter settings.*

Therefore, the deployment treaty is stable for $\Delta > \Delta^{\text{CG}}_{\text{Min}}$ *and the moratorium treaty is stable for* $\Delta > \Delta^{\text{CG,A}}_{\text{Morat}}$. *Under the tie-breaking Assumption 2, the separating level between deployment treaty and moratorium treaty is* $\Delta^{\text{CG}}_{\text{Max}} := \max(\Delta^{\text{CG,A}}_{\text{Morat}}, \Delta^{\text{B}}_{\text{Morat,Treaty}})$.

[15] The intuition why country B still prefers the free-driver outcome over the deployment treaty for moderate asymmetry levels, $\Delta < \Delta^{\text{SG}}_{\text{Min}}$, is the same as in Proposition 2: The free-driver equilibrium involves no deployment costs for country B, and final temperature changes T, while excessive, are still relatively close to its optimal level T_B.

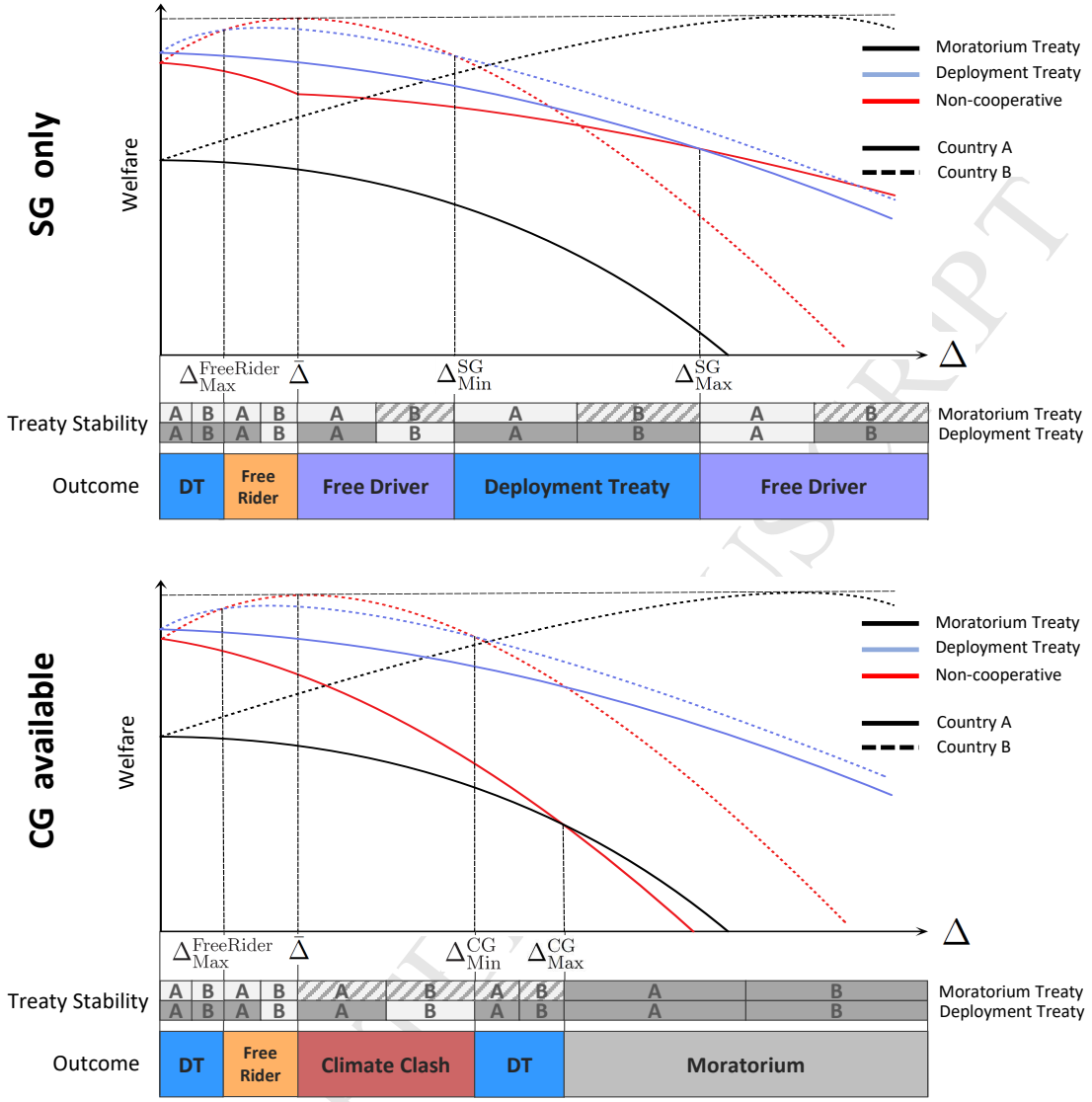


Figure 3: Schematic representation of welfare levels and equilibria as a function of the asymmetry parameter Δ. The upper part shows the 'SG only' case, the lower part the 'CG available' case. The welfare plots are illustrative only. The solid and dashed lines represent country A and B, respectively. The red curves represent all non-cooperative outcomes and are given, depending on asymmetry level Δ and whether CG is available or not, by expressions (9), (10) or (11). The boxes above the equilibrium labels indicate whether countries A and B are willing to join either of the two treaties (dark fill) or not (light fill), respectively. A hatched fill indicates that a country's decision whether to join or not is parameter-dependent but inconsequential for the final outcome. When both treaties are stable (i.e. both treaties are attractive for both countries) and countries disagree about which of the two they prefer, then our tie-breaking rule in Assumption 2 resolves the disagreement. Note that the relative size of the treaty equilibria with and without CG depends on parameter values. See section 5 for a calibration and sensitivity analysis.

Proof. See appendix A. □

The moratorium treaty is stable, i.e. preferred by both countries over the climate

clash, once the asymmetry exceeds $\Delta^{\mathrm{CG,A}}_{\mathrm{Morat}}$. The interest in the moratorium underlines how unattractive the climate clash is. Country B is more interested in the moratorium treaty than country A, which is expressed both by a wider opt-in region ($\Delta^{\mathrm{CG,B}}_{\mathrm{Morat}} < \Delta^{\mathrm{CG,A}}_{\mathrm{Morat}}$) and by a preference for the moratorium over the deployment treaty for levels beyond $\Delta^{\mathrm{B}}_{\mathrm{Morat,Treaty}}$ (a preference that country A never has). This is intuitive when we recall that temperatures under climate change absent any climate intervention (the outcome under the moratorium treaty) are relatively less harmful for country B than for country A. There is a simple intuition why country A is keen to cooperate via the deployment treaty. Not only are deployment costs in the cooperative solution much lower than in the climate clash, the social optimal SG deployment level is also more ambitious and thus closer to T_A. That country B prefers the deployment treaty to the climate clash for moderate asymmetry levels Δ is similar to before: country B's deployment costs are low and the final temperature change is relatively close to B's optimum T_B.

To summarize, we find a rich set of potential outcomes that are depicted in Figure 3. Every outcome (the three non-cooperative as well as the two treaties) materializes under certain conditions, and the boundaries that separate different outcomes are non-trivial. A parameter calibration and sensitivity analysis of the equilibrium boundaries are presented in Section 5. Our findings suggest a substantial potential of CG to change the statics of the global thermostat game: The basic mechanism is CG transforms the game such that the outcome changes from a free-driver in the 'SG only' case to a 'climate clash' when CG is available. This transformation of outcomes is always bad for the free-driver A (and often for country B as well). It is this mechanism that brings the free-driver to the negotiating table when cooperation is possible: the free-driver is now always willing to enter the global optimal deployment treaty. In order to prevent the wasteful climate clash, the free-driver is, under certain conditions, even willing to accept the otherwise very unattractive conditions of a moratorium treaty. We will show in section 7 that this basic mechanism also shapes the general n country case.

4.3 Welfare ranking of outcomes

We have now gained a comprehensive understanding of CG's potential to change the global thermostat game. Are the changes induced by CG for the better or worse? We have partially answered this question above. Proposition 1 shows that, in terms of non-cooperative outcomes, the transformation from free-driver outcome to climate clash induced by CG is detrimental as it decreases global welfare. On the other hand, whenever this bleak outlook induces countries to form a deployment treaty, which by definition implements the global best, then CG's game-changing effect is beneficial. What remains to be understood is how the moratorium treaty ranks in welfare terms. The following result shows that the moratorium, cf. Proposition 4, is only better than the free-driver outcome for high levels of asymmetry. For completeness we also compare the moratorium

treaty to the climate clash. While not important for the welfare impact induced by the presence of CG, this result sheds light on the value of having cooperation options once CG is part of the game.

Proposition 5 (Welfare of Moratorium Treaty)**.**

(i) Global welfare under the moratorium treaty is higher than in the free-driver equilibrium iff $\Delta > \Delta^{\text{Welfare}}_{\text{Morat,Driver}}$*, where* $\Delta^{\text{Welfare}}_{\text{Morat,Driver}} > \bar{\Delta}$

(ii) Global welfare under the moratorium treaty is higher than in the climate clash equilibrium iff $\Delta > \Delta^{\text{Welfare}}_{\text{Morat,Clash}}$*, where* $\Delta^{\text{Welfare}}_{\text{Morat,Clash}} > \bar{\Delta}$*. In relative terms* $\Delta^{\text{Welfare}}_{\text{Morat,Driver}} > \Delta^{\text{Welfare}}_{\text{Morat,Clash}}$.

Proof. See appendix A. □

5 Calibration, sensitivity analysis, and welfare impact

In this section we first calibrate the model parameters b, c and $\bar{T}$. We then determine the sensitivity of equilibrium boundaries to changes in parameters. Finally, we discuss the welfare effect of CG in the calibrated model.

5.1 Parameter calibration

Our calibration of the benefit parameter b rests on Burke et al. (2015) who show that the relationship between (local) temperatures and growth rates follows a universal quadratic relationship. The calibration of the cost parameter c is based on data on stratospheric SG with sulfur aerosols. It combines data on operational cost per kg of load material with the non-linear relation between sulfur load and reduction in radiative forcing. Finally, $\bar{T}$ expresses the amount of atmospheric cooling required to achieve the global optimal temperature at the point of climate intervention. This clearly depends on emissions scenarios. Appendix B provides details on the calibration that results in the following parameter values

$$b = 17.95 \text{ bn}\,\$/^{\circ}C^2 \quad , \quad c = 8.35 \text{ bn}\,\$/^{\circ}C^2 \quad , \quad \bar{T} = -2.1^{\circ}C. \tag{12}$$

We keep asymmetry Δ as an open parameter for two reasons. First, this parameter is the hardest to calibrate as it depends on regional/country-specific preferences over climate outcomes (in contrast to $\bar{T}$ which is a measure of globally aggregated preferences). Second, this provides us with a degree of freedom to describe a variety of interactions between potentially very different agents.

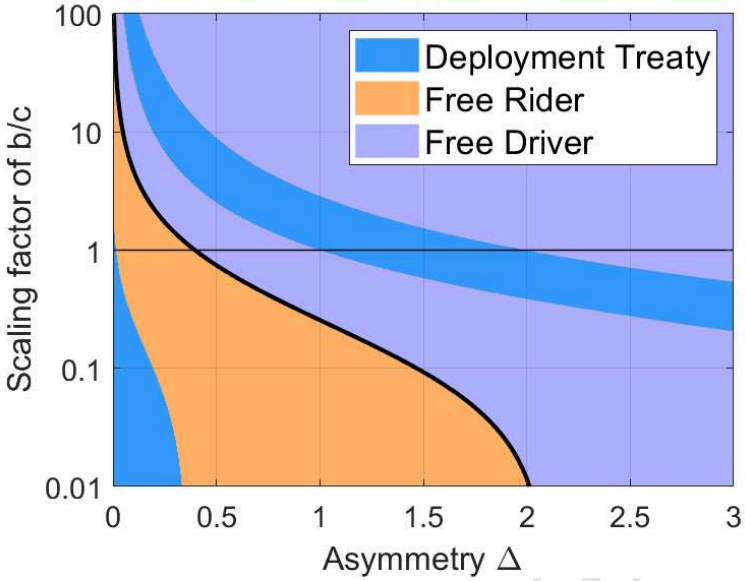
5.2 Outcome boundaries and sensitivity

What outcome can we expect under parameters calibrated as above, and how sensitive are the equilibrium boundaries in Figure 3 to parameter settings? Analysis of the algebraic expression of the equilibrium boundaries (see appendix A) reveals that all boundaries scale linearly with $\bar{T}$ and depend only on the benefit-cost ratio $\theta = b/c$, not b and c separately. The horizontal line in Figure 4 shows the equilibrium boundaries for the best estimate of θ, $b/c = 17.95/8.35$, cf. (12). We check for sensitivity by scaling the benefit-cost ratio upwards and downwards by two orders of magnitude. Figure 4a and 4b depict the 'SG only' and 'CG available' cases, respectively. The solid black line represents $\bar{\Delta}$, the asymmetry threshold from (8) that separates free-rider outcomes to the left from free-driver ('SG only') and climate clash ('CG available') outcomes to the right.

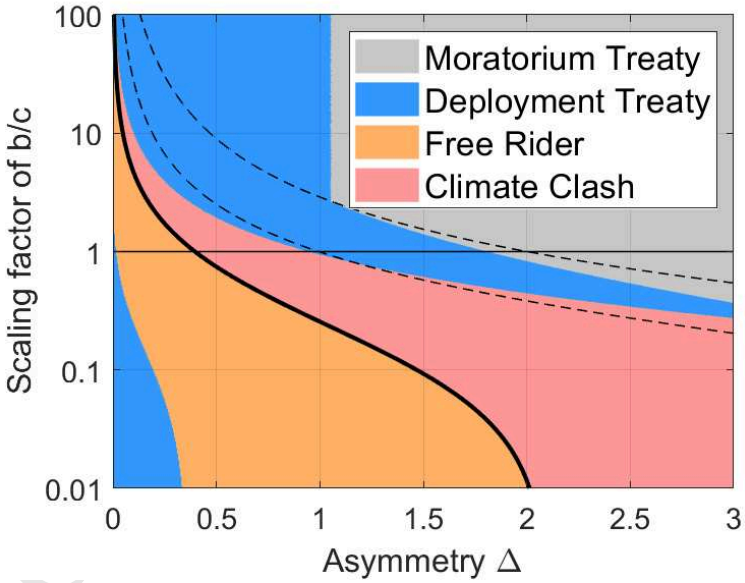
The first observation is that the asymmetry threshold $\bar{\Delta}$ is fairly small for the calibrated parameter values.[16] This means that, in the absence of CG, even a small disagreement over the best use of SG will result in the free-driver outcome. Also note that the deployment treaty is plausible only for the asymmetry range between $1\,°C$ and $2\,°C$. Overall, the free-driver equilibrium is the most likely outcome in the 'SG only' case. We see that, under the calibrated parameter values, CG slightly reduces the set of constellations under which the deployment treaty materializes, whereas the climate clash is the predicted outcome only for a relatively narrow range of asymmetry values. The moratorium treaty, according to our tie-breaking assumption 2, is the predicted outcome for the 'CG available' case for higher values of the asymmetry parameter.

In terms of sensitivity to parameter changes, we have noted above that all boundaries scale linearly with $\bar{T}$. Thus we can focus on the effect of changes in the benefit-cost ratio $\theta = b/c$. Figure 4 demonstrates that all our observations from above are only strengthened if θ gets higher, for instance if operational costs of a climate intervention were significantly lower than current estimates. The free-driver is the typical outcome in the 'SG only' case, and cooperation (through either a deployment or moratorium treaty) in the 'CG available' case almost certain. The outcomes are very different if we consider lower benefit-cost ratios, for instance because climate damages are seen as relatively minor and/or climate interventions much more costly than currently expected. Then, the free-rider outcome becomes much more plausible, in which case the presence of CG would be inconsequential. Interestingly, the likelihood of cooperation (via either

[16] What are large and small values of the asymmetry parameter Δ? Consider the example sketched in footnote 9 where country A and country B have pre-industrial average temperatures of $16°C$ and $10°C$, respectively (this is not an extreme scenario as multiple regions experienced pre-industrial average temperatures beyond $20°C$). If both countries determined their preferences over climate interventions based solely on a universal optimal temperature, e.g. the $13°C$ in Burke et al. (2015), then $\Delta = 3K$. If, less extreme, both countries considered the midpoint between pre-industrial and a certain universal temperature as optimal, then $\Delta = 1.5K$. In this sense it is justified to say that the asymmetry threshold is typically fairly small.



(a) SG only.



(b) CG available.

Figure 4: Sensitivity analysis of the equilibrium boundaries in the $n = 2$ case. The reference benefit-cost ratio $\theta = b/c = 17.95/8.35$, cf. (12), is represented as the horizontal line. To check sensitivity we scale θ upwards and downwards by two orders of magnitude. The solid black curve in both plots represents $\bar{\Delta}$. The dashed lines in (b) represent the deployment treaty boundaries of the 'SG only' case in (a).

deployment or moratorium treaty) decreases as θ decreases, and the climate clash for high levels of asymmetry becomes increasingly plausible.

5.3 Calibrated welfare impacts

In this section we give a calibrated answer to the question whether the game-changing potential of CG is beneficial or detrimental. Figure 5 shows the effect of CG on global welfare.[17] As in Figure 4, the horizontal axis shows the asymmetry between country A and country B, whereas the vertical axis shows a range of benefit-cost ratios; the horizontal line represents the best estimate of $\theta = b/c$ in (12).

There are two regions where CG does not change the outcome of the game and hence leaves global welfare unchanged (indicated by white coloring). First, all asymmetry levels to the left of the asymmetry threshold $\bar{\Delta}$. Here, the non-cooperative outcome is the free-rider equilibrium, and neither country wants to deploy CG in the first place. The second region is where the deployment treaty was the outcome in the 'SG case' and remains the outcome when CG is available.

We find two reasons why CG can be beneficial, indicated by green colors in Figure 5. First, CG can transform a free-driver outcome into a deployment treaty, and our findings suggest that this is likely for intermediate levels of asymmetry and relatively high benefit-cost ratios. The second situation in which CG increases overall welfare is when an extreme free-driver outcome is transformed into a moratorium treaty; in order for the technology-free world to be globally preferable, the asymmetry level must be high so that the free-driver outcome is very problematic.

If the asymmetry is not extreme, however, this transformation from free-driver to moratorium is detrimental (potentially for both countries), represented here in dark red colors. A second (as it turns out relatively rare) scenario in which CG is detrimental is when a deployment treaty (bordered by the dashed lines) in the 'SG only' case is transformed into either a climate clash or a moratorium treaty outcome. Finally, there is a third and important situation in which CG can reduce global welfare: If neither form of cooperation is attractive, then CG transforms what used to be a bad free-driver outcome into an even worse climate clash. This scenario is especially plausible for low benefit-cost ratios.

Appendix D demonstrates that the country-specific effects of CG are fairly clear: typically country A is worse off under CG, whereas country B benefits from the availability of CG. The mixed picture that we see in Figure 5 is hence the superposition of generally contrasting country-specific effects.

[17] The plotted quantity in the contour plot is the welfare difference between the 'CG available' and the 'SG only' case. This difference is, for each separate parameter setting, expressed in terms of the absolute welfare under the social optimal outcome; if, for instance, global welfare under the deployment treaty is -10 units (recall that welfare levels are non-positive by assumption), then a value of -50% in the contour plot means that CG reduces global welfare by 5 units. Note that this plot necessitates reducing one degree of freedom. While the equilibrium boundaries in Figure 4 depend only on the benefit-cost ratio $\theta = b/c$, any welfare analysis depends on both parameters b and c separately. There are different ways to reduce one degree of freedom; here we stipulate that the welfare under the social optimal deployment profile (g_A^{**}, g_B^{**}) is independent of the benefit-cost ratio θ. See Appendix D for more details.

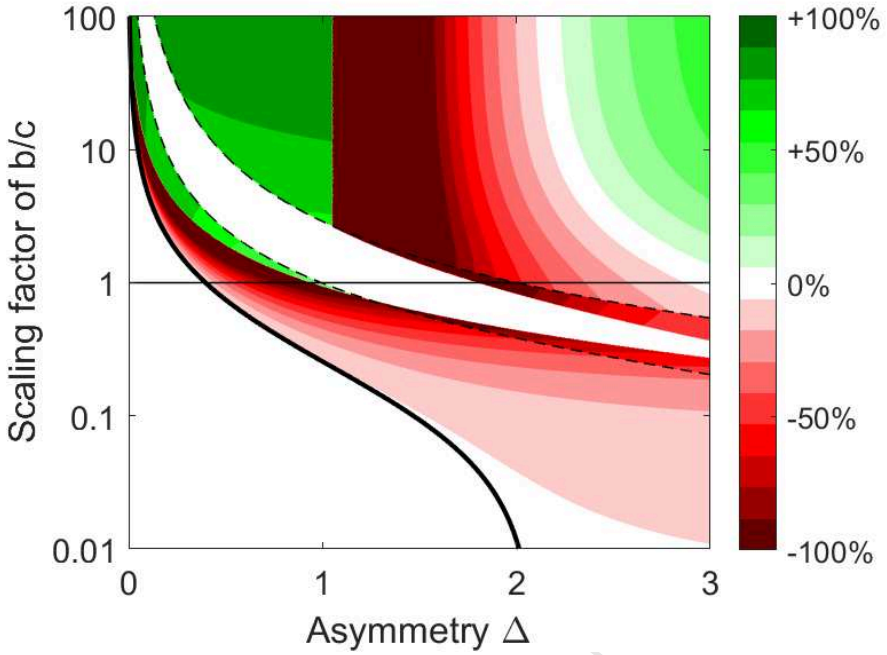


Figure 5: Welfare effect of CG. For every pair of Δ and b/c, the quantity plotted in the contour plot is the difference in welfare with and without CG, normalized by the social optimal welfare (i.e. welfare of the deployment treaty). Green and red colors indicate settings where the impact of CG is positive and negative, respectively, whereas white indicates that CG has no effect on global welfare.

6 Asymmetric costs

This section introduces the possibility that CG has a different cost structure than SG, i.e. the cost function for climate interventions (2) is now differentiated into an SG and CG part,[18]

$$C(g_i) = \begin{cases} \frac{c_{\mathrm{SG}}}{2} g_i^2 & g_i \leq 0 \\ \frac{c_{\mathrm{CG}}}{2} g_i^2 & g_i > 0. \end{cases} \tag{13}$$

To facilitate comparisons with the previous analysis we write

$$c_{\mathrm{SG}} = c \quad , \quad c_{\mathrm{CG}} = \xi \cdot c \tag{14}$$

with $\xi > 0$ as a measure of the asymmetry between SG costs and CG costs. The higher ξ, the costlier is CG relative to SG. The symmetric cost structure studied above is represented by $\xi = 1$. As above, we denote the benefit-cost ratio by $\theta = b/c$; the benefit-cost ratio for CG deployment is θ/ξ.

[18]This cost structure applies to both countries, $i = A, B$. Another extension could be to let different countries have different cost structures. Exploring this generalization is left for future research.

The clash equilibrium is the only one that involves CG. Therefore, all other equilibria remain as before, both in terms of deployment and welfare. With the same methodology used for Proposition 1 we show that the **generalized climate clash equilibrium** is now given by

$$g_A^* = \frac{\theta}{\theta + (1+\theta)\xi} \cdot \left[\xi\bar{T} - (2\theta + \xi)\Delta\right] \quad , \quad g_B^* = \frac{\theta}{\theta + (1+\theta)\xi} \cdot \left[\bar{T} + (2\theta + 1)\Delta\right]. \quad (15)$$

Expression (11) is the special case of (15) when $\xi = 1$. As before we look at the asymmetry level $\bar{\Delta}$ that separates the climate clash equilibrium from the free rider equilibrium, defined by $g_B^* = 0$. From (15) we see that this level is still $\bar{\Delta} = -\frac{\bar{T}}{2\theta+1}$ and therefore in particular not a function of ξ. The total deployment level g^* in the generalized climate clash equilibrium is

$$g^* = \frac{\theta}{\theta + (1+\theta)\xi} \cdot \left[(\xi + 1)\bar{T} + (1 - \xi)\Delta\right]. \quad (16)$$

The deployment levels in the generalized climate clash equilibrium show an intuitive connection to the cost asymmetry parameter ξ. We find

$$\frac{dg_A^*}{d\xi} = \frac{(2\theta+1)\theta^2}{(\theta\xi + \theta + \xi)^2}\left(\Delta - \bar{\Delta}\right) > 0 \quad , \quad \frac{dg_B^*}{d\xi} = -\frac{(2\theta+1)\theta(\theta+1)}{(\theta\xi + \theta + \xi)^2}\left(\Delta - \bar{\Delta}\right) < 0. \quad (17)$$

Due to $\Delta \geq \bar{\Delta}$, the first expression is positive and the second negative. The higher the CG cost, the less CG country B deploys; this implies that country A needs to do less SG. We also see that the overall effect is negative,

$$\frac{dg^*}{d\xi} = -\frac{\theta(2\theta+1)}{(\theta\xi + \theta + \xi)^2}\left(\Delta - \bar{\Delta}\right) < 0. \quad (18)$$

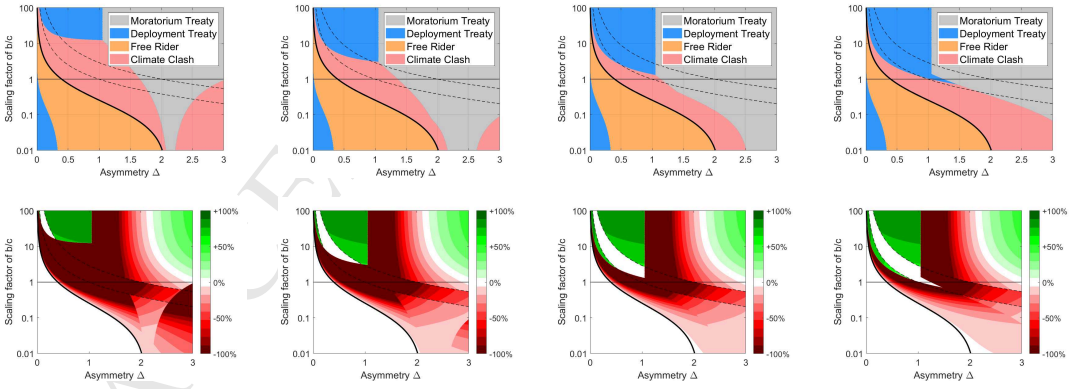
The higher the costs of CG, the more cooling we see in the generalized climate clash equilibrium.

We also immediately observe an intuitive relation between asymmetry Δ and the total temperature change in the generalized climate clash equilibrium. Looking at (16), we see that an increase in asymmetry Δ leads to more cooling if CG is costlier than SG, $\xi > 1$ and to less cooling if CG is cheaper than SG, $\xi < 1$. The insensitivity of the total deployment level g^* that we observed in section 3.2, cf. Figure 2, occurs if and only if CG and SG have the same cost structure, $\xi = 1$.

We now turn to the question how the relative cost structure between SG and CG changes the statics of the game. The 'SG only' case is not affected by a change in CG cost. Figure 4a therefore still describes the set of equilibria without CG, irrespective of ξ. What is affected by a change in relative costs are the equilibria in the presence of CG. Figure 6 shows equilibria in the 'CG available' case (i.e. variations of Figure 4b) together with plots that show the welfare changing effect of CG (i.e. variations of Figure

5) if CG is cheaper than SG, $\xi < 1$. Figure 7 shows the case when CG is more expensive than SG, $\xi > 1$.

We can derive several insights from these figures. First, the climate clash materializes under wider conditions both for very small as well as very large values of ξ; asymmetry in costs of SG and CG favors the climate clash equilibrium. Second, the deployment treaty has the best prospect when CG is costlier than SG and the difference is not too extreme, say when ξ is around 5. Third, the moratorium treaty has the best prospect when CG is cheaper than SG and the difference is not extreme, say ξ around 1/5. The moratorium treaty disappears from the scene of possible outcomes when CG gets very costly relative to SG. Fourth, as CG gets infinitely more costly than SG, the equilibria map converges to the 'SG only' shape, where the free driver equilibrium replaces the climate clash. This makes sense: a very costly CG is similar to no CG at all. Finally, the welfare effect is not straightforward. The simple intuition 'costly CG is similar to no CG' works for low theta values: the CG level deployed by country B here is very low, the free driver accordingly does not have to adjust his level much. The welfare change caused by CG accordingly is small, indicated by white/light red areas for low values of θ in Figure 7. The intuition however is less straightforward for high values of the benefit-cost ratio θ. While the statics of the game determining the equilibrium outcome are still compatible with the 'costly CG is similar to no CG in the first place' logic, the welfare change is significant. The reason seems to be that high benefit-cost ratios θ cause a major CG deployment and accordingly a major adjustment in the free-driver's behavior. But this requires additional research.

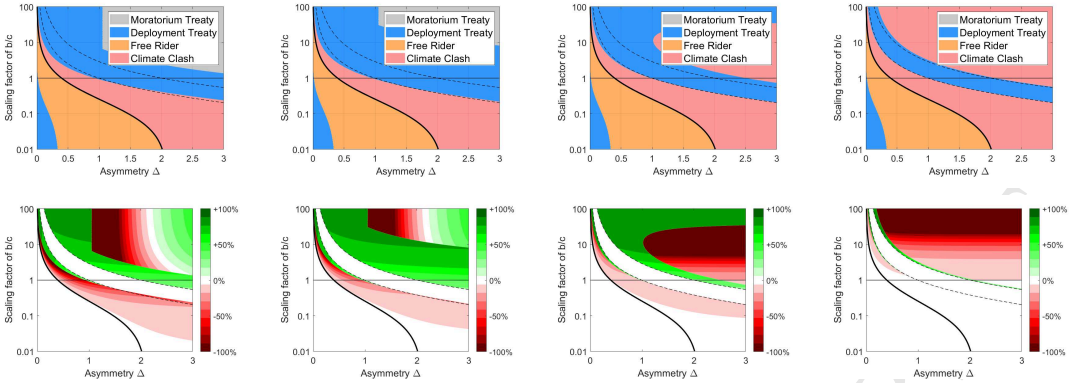


(a) $\xi = 0.01$ (b) $\xi = 0.05$ (c) $\xi = 0.2$ (d) $\xi = 0.5$

Figure 6: Asymmetric costs. CG cheaper than SG, $\xi < 1$.

7 The n countries case

This section extends the setup to the general case of n countries. We derive analytical and numerical results to check the robustness of our results derived under the two-



(a) $\xi = 2$ (b) $\xi = 5$ (c) $\xi = 20$ (d) $\xi = 100$

Figure 7: Asymmetric costs. CG more expensive than SG, $\xi > 1$.

country model. For simplicity we go back to the assumption that SG and CG have the same cost structure, i.e. $\xi = 1$ in this section.

7.1 Model setup

The general structure in terms of benefit function, cost function and timeline of the model remains as before. There are now n countries, each with climate intervention level g_i, $i = 1, \ldots, n$. The *change in global average temperature* T due to climate interventions is $T = \sum_{i=1}^{n} g_i$. The *mean optimal temperature change* is $\bar{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$, where T_i is country i's preferred global average temperature change. We keep assumption 1, i.e. $\bar{T} < 0$. We write

$$T_i = \bar{T} + \Delta\delta_i \ , \tag{19}$$

where $\delta_i \leq \ldots \leq \delta_n$ with $\sum_{i=1}^{n} \delta_i = 0$. We normalize $\frac{1}{n}\sum_{i=1}^{n} \delta_i^2 = 1$, so that Δ is the standard deviation of the optimal temperature change T_i. We call Δ as above the *asymmetry parameter*. For $\Delta = 0$, all countries agree on how much the climate ought to change; increasing Δ represents growing disagreement across countries. Note that the definition for $n = 2$ in section 2 coincides with the definition given here: it is $\delta_A = -1$ and $\delta_B = 1$. We denote by $(g_i^{**})_i$ the socially optimal configuration that maximizes global welfare $\sum_{i=1}^{n} \pi_i(g)$. It is straightforward to show that

$$g_i^{**} = \frac{nb}{n^2 b + c}\bar{T} \qquad , \ i = 1, \ldots, n \ . \tag{20}$$

It is efficient for all countries to deploy the same amount of SG due to the homogeneous cost structure. Owing to $\bar{T} < 0$ (Assumption 1), the socially optimal deployment scheme features SG deployment by all countries. In particular, whether CG is available or not has no implications for the socially optimal deployment profile.

7.2 Non-cooperative equilibria

As before, the asymmetry Δ determines how many countries deploy SG in equilibrium, and the number of countries deploying SG is monotonically decreasing in Δ. The remaining countries, in any case, consider the overall temperature reduction by SG as too high and accordingly either do not deploy SG (in the 'SG only' case) or deploy CG (in the 'CG available' case). In the following we use the notation $\theta = b/c$, $\beta_m = \frac{m\theta}{m\theta+1}$ and denote the average optimal temperature change among the first m countries by $\bar{T}^{(m)} = \frac{1}{m}\sum_{i=1}^{m} T_i$. With these preliminaries, we are ready for our next proposition.

Proposition 6 (Non-cooperative equilibria. General n)**.** *There is a set of values $\Delta^{(m)}$ $(m = 0, \ldots, n)$ that is decreasing in m with $\Delta^{(0)} = \infty$ and $\Delta^{(n)} = 0$. Let the asymmetry parameter be in the interval $\Delta \in [\Delta^{(m)}, \Delta^{(m-1)}]$.*

(i) The 'SG only' case has a unique equilibrium where the m countries with the highest preference for cooling deploy SG

$$g_i^{(m)} = \theta(T_i - \beta_m \bar{T}^{(m)}) \quad i = 1, \ldots m \tag{21}$$

and the remaining countries do not deploy, $g_i^{(m)} = 0$, $i = m+1, \ldots, n$.

(ii) When CG is available, all countries' deployment levels are given by

$$g_i^{(n)} = \theta(T_i - \beta_n \bar{T}) \quad i = 1, \ldots n \tag{22}$$

where the first m are negative (SG deployment) and the remaining $n - m$ positive (CG deployment).

(iii) The transformation induced by CG is typically detrimental, but there are exceptions to this rule.

Proof. See appendix A. □

In the case $n = 2$ we have, as required, $\Delta^{(1)} = \bar{\Delta}$ and the quantities given in Proposition 1 all coincide with (21), evaluated at $m = 2$ (free-rider and climate clash) or $m = 1$ (free-driver).

7.3 Cooperation: assumptions and results

The two forms of treaties, moratorium treaty and deployment treaty, are both modelled as *open-membership games*. Under the moratorium treaty *all* countries bind themselves to abstain from any technology deployment. For a deployment treaty, we stipulate that at most one coalition can form, and this coalition decides on the optimal deployment of SG, where the objective is maximization of the coalition's total payoff. In terms of timing we adopt the Stackelberg leadership assumption: After the coalition has made

its decision, the other countries ('fringe') decide simultaneously and non-cooperatively on optimal SG (in the 'SG only' scenario) or between SG and CG deployment (in the 'CG available' scenario). Note that we allow for at most one coalition and assume that coalitions have at least two members. The reason is simplicity. Furthermore, we rule out CG as the coalition's action. The reason is to allow a good comparison of the 'SG only' and 'CG available' case. One might defend this assumption by saying that further warming the climate through CG clearly has less international justification than SG, so international treaties on CG are less plausible. Nevertheless it would be worthwhile to explore alternative forms of cooperation in future research.

Stability of coalitions. The moratorium treaty is stable if and only if *all* countries prefer the technology-free world over the non-cooperative technology deployment. A deployment treaty (where the size can range between 2 and n) is stable if it is *internally* and *externally* stable. Internal stability means that every coalition member's payoff is higher or the same compared to a scenario in which he leaves the coalition (and the remaining members still form a coalition). External stability of a coalition means that no fringe country can improve her outcome by joining the coalition. Both stability concepts take the decisions of other countries as given. In other words we follow the usual simplifying approach without farsighted players (Mariotti and Xue 2003).

We find that there are stable coalitions that deploy no SG at all. The reason is our simplifying assumption that at most one coalition can form, which results in countries forming non-deploying coalitions in order to prevent coalitions that deploy SG from forming. We regard this as an artifact of our model assumptions and accordingly disregard non-deploying stable coalitions.

Results. To illustrate the general case, Figure 8 shows the results for a setting with $n = 7$ countries (we found similar results for other values of n). We focus on a setting with a clear 'free-driver' in the following sense: One country prefers temperatures lower than $\bar{T}$, all others are symmetric and prefer a moderate cooling. The leftmost vertical line in Figure 8 is at $\Delta = 0$, the next separates two types of non-cooperative outcomes: low asymmetry levels where all countries deploy SG (orange), and levels where only the free-driver deploys SG (purple). The latter outcome is transformed into a climate clash (red) when CG is available. We show two ranges (of different scales) of the asymmetry parameter Δ. The first range (left to the double vertical bar) is $[0, 1.25]$, the second (to the right of the double vertical bar) is the range $[1.25, 5]$. We show the stability of moratorium and deployment treaty, the latter differentiated into full cooperation deployment treaties (all countries are part of the treaty) and partial deployment treaties (only some countries participate). We find that stable coalitions take the form 'free-driver + k others'. For moratorium treaty and full cooperation deployment treaty outcomes, we display individual incentives for cooperation in grey, separated into incentives for the

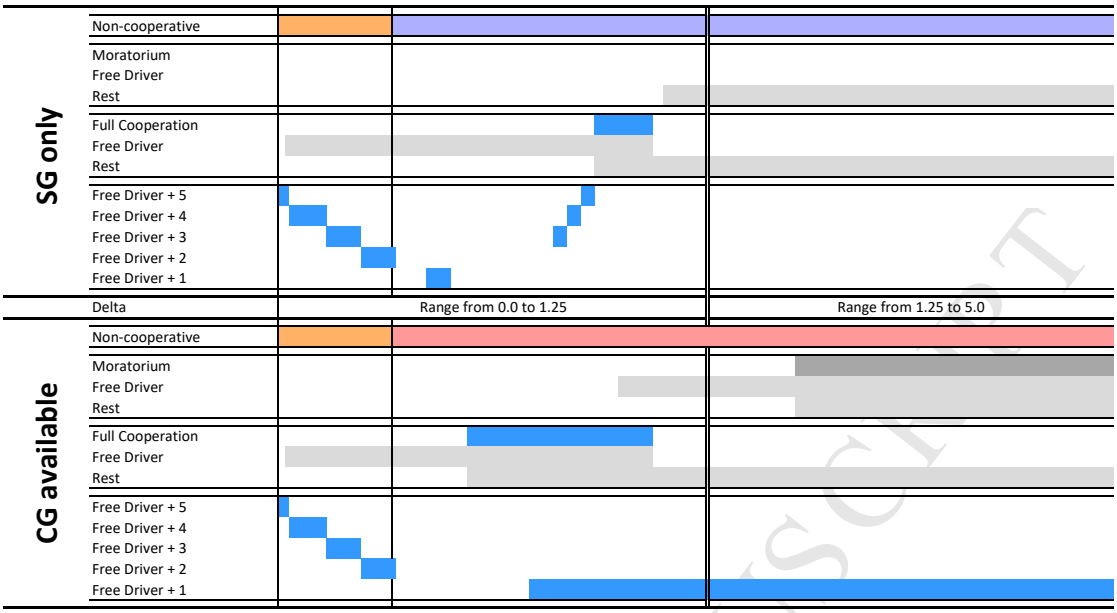


Figure 8: Global thermostat game setting for $n = 7$. One country prefers a strong cooling whereas the other six are symmetric and prefer a moderate cooling. This implies (cf. 19) that $\delta_1 = -2.450$ and $\delta_i = 0.408$ for $i = 2, \ldots, 7$. We keep the parameter calibration from above, cf. (12). The figure presents non-cooperative equilibria and stability of moratorium treaty, full deployment treaty and partial deployment treaties (here always of the form "free-driver + x other countries").

'free-driver' and for the other countries bundled as 'rest' (we code 'rest' as incentivized to cooperate if *all* other countries are willing to cooperate); coalitions are stable where incentives to cooperate overlap.

Our results are in line with the $n = 2$ findings. The non-cooperative case is perfectly analogous. For low asymmetry Δ we get a free-rider outcome (where all countries deploy SG) that is unaffected by the presence of CG; high levels of asymmetry are characterized by a free-driver equilibrium (where only one country deploys), and this free-driver outcome is transformed into a climate clash when CG is available. The cooperation incentives are fairly similar to the $n = 2$ case. Starting with the moratorium treaty, in the 'SG only' case the free-driver is unambiguously opposed to it, while the other countries prefer the technology-free world if asymmetry (and accordingly the gap between other countries' optimal temperatures and what the free-driver implements) is high. Once CG is available, though, the moratorium treaty becomes attractive; in particular, the free-driver – even at intermediate levels of asymmetry – is willing to jointly abstain from deployment in order to prevent the costly climate clash.

Moving on to the deployment treaties (including full and partial cooperation), in the 'SG only' case we find that the deployment treaty can only be stable for low and intermediate levels of asymmetry; for larger levels the free-driver is not willing to compromise. This is to some extent changed when CG is available. Full cooperation remains fairly unattractive for the free-driver (recall that all coalition members count equally

when the coalition's deployment level is determined), but the zone where at least partial cooperation is stable is significantly extended under 'CG available'.

8 Conclusions

We have studied the strategic interaction over fast-acting climate interventions when countries disagree on how much to modify the climate. We have modelled this interaction as a public good game in which countries with asymmetric preferences anticipate the Nash equilibrium of the non-cooperative game and have the option to cooperatively decide on the level of climate intervention. Our main focus has been technological capabilities to quickly counter other countries' (excessive) cooling by means of counter-geoengineering (CG), and in particular the question *how* CG alters the statics of the game and under which circumstances the resulting change in outcomes can prove beneficial.

Our findings are summarized as follows. When climate intervention is restricted to cooling by means of solar geoengineering (SG), then the typical outcome is the 'free-driver' equilibrium. The free-driver, the country that suffers from climate change the most and hence wants to cool the most, may set global temperatures as it pleases; other countries may suffer damages from this excessive cooling but have no measure against it. Cooperation incentives in this case are relatively weak, first and foremost because the free-driver has little reason to compromise. The availability of CG changes this game significantly. We demonstrate that the free-driver outcome is not an equilibrium anymore once dominated countries have CG at their disposal, yet the resulting Nash equilibrium is an even more harmful 'climate clash' in which countries waste significant resources in offsetting SG and CG deployments. This destructive prospect is the very reason why – under certain circumstances – the existence of CG can significantly increase countries' willingness to cooperate. Specifically, the would-be free-driver understands that a climate clash would harm him substantially, and is hence (under a broad set of circumstances) willing to make climate intervention decisions cooperatively. This can enhance collective welfare. Crucially, however, other countries might prefer cooperation in the form of a moratorium that reduces global welfare, or even a climate clash over cooperation altogether.

From a policy perspective the central question is what difference does the existence of a CG capability make to a world where SG is contemplated. Our analysis shows that the answer depends on three key factors. First, the ratio of benefits and costs of climate intervention matters. CG tends to increase cooperation incentives for high benefit-cost ratios but may give rise to a climate clash for low benefit-cost ratios. Second, multiple cooperative agreements can be stable and it matters which of them materializes. Even in the simple, stylized $n = 2$ case, both the moratorium and deployment treaty can be stable, and which one obtains determines how CG affects aggregate welfare. Finally, a

key factor for understanding CG's influence is the level of asymmetry among countries. Where asymmetry is low, the strategic interaction is essentially a free-rider equilibrium and CG makes no difference. Where it is intermediate, a climate clash may ensue. For high levels of asymmetry countries are more willing to cooperate, but our result suggests that extreme levels of asymmetry may favour a welfare-imperfect moratorium.

Given the novelty both of this topic and of our analytical approach to it, we opted to keep the modeling framework as simple as possible, and thus we see various opportunities for extending it. The first possible extension is to allow countries' preferences to incorporate and reflect climate indicators beyond temperature. The most obvious candidate is precipitation, in particular as a major concern surrounding SG is its potential to alter precipitation patterns. An interesting question in this context is whether the inclusion of indicators other than temperature exacerbates the asymmetry between countries or, instead, mitigates the free-driver concern. This points toward linkage with the emerging literature on 'optimal climate states'. A second possible extension is to include indirect effects of geoengineering that are not climate-related, for instance the health effects caused by the particles used for geoengineering (e.g. acid rain and ozone loss from stratospheric SG with sulfur particles). These effects can be captured with a second (negative) 'benefit' function; here, the effects of SG and CG may depend on the sum of *absolute* SG and CG levels and thus not cancel out as in the case of climate related effects. While a thorough analysis is left for future research (building on research into potential SG and CG particles and their possible secondary effects), we can speculate on how this would change our results. It seems plausible that these additional external effects would have little effect on individual choices, but render the climate clash significantly less attractive from a global welfare perspective. In that sense our findings of a problematic climate clash can be interpreted as a lower bound.

Another possible extension is to generalize the time structure of the non-cooperative game. The simultaneous global thermostat game in our model proved rich enough for a variety of equilibria to emerge, but it is worthwhile to study which additional equilibria emerge in a full dynamic game. Importantly, such an extension would allow for a focus on uncertainty and learning. The countries in our model have perfect knowledge of how climate interventions work and how much they cost. It would be interesting to study how uncertainty changes equilibria and welfare, and how learning by doing alters the strategic interaction surrounding climate interventions.

Finally, three more potential extensions revolve specifically around cooperation. First, several valuable robustness checks on our modelling assumptions in the general n country case could be performed: modeling coalition and fringe decisions as simultaneous, in contrast to the Stackelberg leader assumption we have adopted; modeling a coalition as an exclusive club, in contrast to our assumption of an open membership game; and allowing, in addition to the SG coalition, a second CG coalition that deploys CG. Second, a richer set of cooperation possibilities and treaty conditions could be ex-

plored, for example, the potential of transfers and/or sanctions to enhance cooperation, the implications of transaction costs or exit options, or other forms of cooperation treaties that go beyond moratorium treaty and deployment treaty. One particularly interesting question to consider would be how CG might interact with other indirect costs imposed on unilateral SG and thereby affect prospects for cooperation. Lastly, the subject of equilibrium selection is ripe for further research. Our results have demonstrated that multiple stable cooperation equilibria are possible, and which one obtains determines the ultimate desirability of climate interventions such as CG. Our assessment of CG would be much more positive if we were sure that, where deployment and moratorium treaties are both stable, the former materializes. In this sense it is of central interest to understand which of multiple stable treaties is more likely to emerge.

References

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

Barrett, S. (2001). International cooperation for sale, *European Economic Review* **45**(10): 1835–1850.

Barrett, S. (2014). Solar geoengineering's brave new world: Thoughts on the governance of an unprecedented technology, *Review of Environmental Economics and Policy*.

Barrett, S., Lenton, T. M., Millner, A., Tavoni, A., Carpenter, S., Anderies, J. M., Chapin III, F. S., Crépin, A.-S., Daily, G., Ehrlich, P., Folke, C., Galaz, V., Hughes, T., Kautsky, N., Lambin, E. F., Naylor, R., Nyborg, K., Polasky, S., Scheffer, M., Wilen, J., Xepapadeas, A. and de Zeeuw, A. (2014). Climate engineering reconsidered, *Nature Climate Change* **4**(7): 527–529.

Burke, M., Hsiang, S. M. and Miguel, E. (2015). Global non-linear effect of temperature on economic production, *Nature* **527**(7577): 235–239.

Coase, R. H. (1960). The Problem of Social Cost, *The Journal of Law and Economics* **3**: 1–44.

Diamantoudi, E. and Sartzetakis, E. S. (2006). Stable international environmental agreements: An analytical approach, *Journal of Public Economic Theory* **8**(2): 247–263.

Diederich, J. and Goeschl, T. (2017). Does Mitigation Begin At Home?, *Technical report*, Discussion Paper Series, University of Heidelberg, Department of Economics.

Emmerling, J. and Tavoni, M. (2017). Quantifying Non-Cooperative Climate Engineering, *Fondazione Eni Enrico Mattei Working Paper*.

Finus, M. (2008). Game Theoretic Research on the Design of International Environmental Agreements: Insights, Critical Remarks, and Future Challenges, *International Review of Environmental and Resource Economics* **2**(1): 29–67.

Finus, M. and McGinty, M. (2018). The anti-paradox of cooperation: Diversity may pay!, *Journal of Economic Behavior & Organization*.

Finus, M. and Rübbelke, D. T. G. (2013). Public good provision and ancillary benefits: The case of climate agreements, *Environmental and Resource Economics* **56**(2): 211–226.

Gampfer, R., Bernauer, T. and Kachi, A. (2014). Obtaining public support for North-South climate funding: Evidence from conjoint experiments in donor countries, *Global Environmental Change* **29**: 118–126.

Heyen, D. (2016). Strategic Conflicts on the Horizon: R&D Incentives for Environmental Technologies, *Climate Change Economics* **7**(4): 1650013.

Heyen, D., Wiertz, T. and Irvine, P. J. (2015). Regional disparities in SRM impacts: the challenge of diverging preferences, *Climatic Change* **133**(4): 557–563.

Horton, J. B. (2011). Geoengineering and the myth of unilateralism: pressures and prospects for international cooperation, *Stanford Journal of Law, Science & Policy* **4**: 56–69.

Keith, D. W. and MacMartin, D. G. (2015). A temporary, moderate and responsive scenario for solar geoengineering, *Nature Climate Change* **5**: 201–206.

Klepper, G. and Rickels, W. (2014). Climate Engineering: Economic Considerations and Research Challenges, *Review of Environmental Economics and Policy*.

Manoussi, V. and Xepapadeas, A. (2017). Cooperation and Competition in Climate Change Policies: Mitigation and Climate Engineering when Countries are Asymmetric, *Environmental and Resource Economics* **66**(4): 605–627.

Manoussi, V., Xepapadeas, A. and Emmerling, J. (2018). Climate engineering under deep uncertainty, *Journal of Economic Dynamics and Control* **94**: 207–224.

Mariotti, M. and Xue, L. (2003). *Farsightedness in coalition formation*, Cheltenham: Edward Elgar.

McClellan, J., Keith, D. W. and Apt, J. (2012). Cost analysis of stratospheric albedo modification delivery systems, *Environmental Research Letters* **7**(3): 034019.

McGinty, M. (2007). International environmental agreements among asymmetric nations, *Oxford Economic Papers*.

Millard-Ball, A. (2012). The Tuvalu Syndrome, *Climatic Change* **110**(3-4): 1047–1066.

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

Moreno-Cruz, J. B., Ricke, K. L. and Keith, D. W. (2012). A simple model to account for regional inequalities in the effectiveness of solar radiation management, *Climatic Change* **110**(3-4): 649–668.

National Research Council (2015). Climate Intervention: Reflecting Sunlight to Cool Earth, *Technical report*.

Parker, A. (2014). Governing solar geoengineering research as it leaves the laboratory, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **372**(2031): 20140173.

Parker, A., Horton, J. B. and Keith, D. W. (2018). Stopping Solar Geoengineering Through Technical Means: A Preliminary Assessment of Counter Geoengineering, *Earth's Future*.

Parson, E. A. and Keith, D. W. (2013). End the deadlock on governance of geoengineering research, **339**(6125): 1278–1279.

Pasztor, J., Scharf, C. and Schmidt, K.-U. (2017). How to govern geoengineering?, *Science* **357**(6348): 231–231. 00000.

Pierce, J. R., Weisenstein, D. K., Heckendorn, P., Peter, T. and Keith, D. W. (2010). Efficient formation of stratospheric aerosol for climate engineering by emission of condensible vapor from aircraft, *Geophysical Research Letters*.

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

Shaviv, N. J. (2005). On climate response to changes in the cosmic ray flux and radiative budget, *Journal of Geophysical Research: Space Physics* **110**: A08105. 00095.

Smith, W. and Wagner, G. (2018). Stratospheric aerosol injection tactics and costs in the first 15 years of deployment, *Environmental Research Letters* **13**(12): 124001.

Stavins, R. N. (2011). The problem of the commons: Still unsettled after 100 years, **101**(1): 81–108.

Urpelainen, J. (2012). Geoengineering and global warming: a strategic perspective, *International Environmental Agreements: Politics, Law and Economics* **12**(4): 375–389.

Victor, D. (2008). On the regulation of geoengineering, *Oxford Review of Economic Policy* **24**(2): 322–336.

Weitzman, M. L. (2015). A Voting Architecture for the Governance of Free-Driver Externalities, with Application to Geoengineering, *The Scandinavian Journal of Economics* **117**(4): 1049–1068.

Zürn, M. and Schäfer, S. (2013). The paradox of climate engineering, *Global Policy* **4**(3): 266–277.

Appendix A Proof of the Propositions

Proof of Proposition 1. To prove (i) and (ii), begin with the 'CG available' case. The best response functions lead to $g_A^* = \frac{\theta}{2\theta+1}\bar{T} - \theta\Delta$ and $g_B^* = \frac{\theta}{2\theta+1}\bar{T} + \theta\Delta$. It is straightforward to see that $g_B^* < 0$ iff $\Delta < \bar{\Delta}$. This is accordingly the free-rider region where it is inconsequential whether CG is available or not. For $\Delta \geq \bar{\Delta}$ we have a climate clash with $g_B^* \geq 0$ in the 'CG available' case. In the 'SG only' case it is necessarily $g_B^* = 0$ when $\Delta \geq \bar{\Delta}$. Country A's best response is the free-driver level $g_A^* = \frac{\theta}{\theta+1}T_A$. To show (iii) we look at the difference in payoffs between the free-driver and climate clash. This difference is

$$\begin{cases} \frac{b\theta}{(\theta+1)(2\theta+1)}\left[(\theta^2+\frac{5}{2}\theta+1)(\Delta-\bar{\Delta})^2 + 2(\theta+1)(-\bar{T})(\Delta-\bar{\Delta})\right] & \text{country A} \\ \frac{1}{2}\frac{b\theta(\theta^2-\theta-1)}{(\theta+1)^2}(\Delta-\bar{\Delta})^2 & \text{country B} \\ \frac{1}{2}\frac{b\theta(2\theta^2+2\theta+1)}{(\theta+1)^2}(\Delta-\bar{\Delta})^2 + \frac{1}{2}\frac{b\theta}{2\theta+1}(-\bar{T})(\Delta-\bar{\Delta}) & \text{total welfare} \end{cases} \tag{23}$$

Because of $-\bar{T} > 0$ and $\Delta \geq \bar{\Delta}$, these quantities are always positive for country A and total welfare, and positive for country B iff $\theta^2 - \theta - 1 > 0$. The latter is equivalent with $\theta > \frac{1+\sqrt{5}}{2}$.

Proof of Proposition 2.

Moratorium Treaty. For country A,

$$\pi_A(0,0) - \pi_A(g_A^*, g_B^*) = \frac{2b\theta}{(2\theta+1)^2}\Big[-\bar{T}^2(\theta+\tfrac{3}{4}) + \bar{T}(\theta+\tfrac{1}{2})\Delta + (\theta+\tfrac{1}{2})^2\Delta^2\Big]$$

This is negative at $\Delta = 0$. The only positive root is at

$$\Delta_{\text{Morat}}^{\text{CG,A}} = -\frac{1+2\sqrt{\theta+1}}{2\theta+1}\bar{T}\ , \tag{24}$$

which is larger than $\bar{\Delta}$. The label indicates that we make use of this quantity in the 'CG available' case. See proposition 4. For country B,

$$\pi_B(0,0) - \pi_B(g_A^*, g_B^*) = \frac{2b\theta}{(2\theta+1)^2}\Big[-\bar{T}^2(\theta+\tfrac{3}{4}) - \bar{T}(\theta+\tfrac{1}{2})\Delta + (\theta+\tfrac{1}{2})^2\Delta^2\Big]$$

This is negative at $\Delta = 0$. The only positive root is at

$$\Delta_{\text{Morat}}^{\text{CG,B}} = -\frac{-1+2\sqrt{\theta+1}}{2\theta+1}\bar{T}\ , \tag{25}$$

which is again larger than $\bar{\Delta}$. So neither country prefers the moratorium treaty over the free-rider outcome.

Deployment Treaty. We begin with country A. It is

$$\pi_A(g_A^{**}, g_B^{**}) - \pi_A(g_A^*, g_B^*) = \frac{b\theta}{2(2\theta+1)^2(4\theta+1)}\Big[\bar{T}^2 - 2\bar{T}(8\,\theta^2+10\,\theta+3)\Delta + (16\theta^3+20\,\theta^2+8\theta+1)\Delta^2\Big]$$

which is positive at $\Delta = 0$. Because the expression has no positive root we see that country A always prefers the treaty over the free-rider equilibrium. For country B, it is

$$\pi_B(g_A^{**}, g_B^{**}) - \pi_B(g_A^*, g_B^*) = \frac{b\theta}{2(2\theta+1)^2(4\theta+1)}\left[\bar{T}^2 + 2\bar{T}(8\theta^2+10\theta+3)\Delta + (16\,\theta^3+20\,\theta^2+8\,\theta+1)\Delta^2\right]$$

which is positive at $\Delta = 0$. The unique root smaller than $\bar{\Delta}$ is

$$\Delta_{\text{Max}}^{\text{FreeRider}} := \frac{-3-4\theta+2\sqrt{4\theta^2+5\theta+2}}{(4\theta+1)(2\theta+1)}\bar{T} \tag{26}$$

So country B prefers the deployment treaty over the free-rider iff $\Delta < \Delta_{\text{Max}}^{\text{FreeRider}}$. It is straightforward to show that $-3-4\theta+2\sqrt{4\theta^2+5\theta+2} < 0$ and therefore $\Delta_{\text{Max}}^{\text{FreeRider}} > 0$.

Proof of Proposition 3. We begin with the comparison of moratorium treaty and free-driver. For country A,

$$\pi_A(0,0) - \pi_A\left(\tfrac{\theta}{\theta+1}T_A, 0\right) = -\frac{b\theta T_A^2}{2\theta+2} < 0\,,$$

so country A always prefers the free-driver. For country B,

$$\pi_B(0,0) - \pi_B\left(\tfrac{\theta}{\theta+1}T_A, 0\right) = \frac{b\theta\,(3\theta+2)}{2(\theta+1)^2}\left(\Delta - \bar{T}\right)\left(\Delta + \frac{\theta+2}{3\theta+2}\bar{T}\right)\,,$$

which is positive for $\Delta > \Delta_{\text{Morat}}^{\text{SG}} := -\frac{\theta+2}{3\theta+2}\bar{T} > 0$.
We continue with the comparison of deployment treaty and free-driver. We begin with country A. It is

$$\pi_A(g_A^{**}, g_B^{**}) - \pi_A\left(\tfrac{\theta}{\theta+1}T_A, 0\right) = \frac{b\theta}{2(\theta+1)(4\theta+1)}\left[3\bar{T}^2 - 6\bar{T}\Delta - (4\theta+1)\Delta^2\right]$$

which is positive at $\Delta = 0$. The unique positive root is

$$\Delta_{\text{Max}}^{\text{SG}} := -\frac{3+2\sqrt{3\theta+3}}{4\theta+1}\,\bar{T} \tag{27}$$

and $\Delta_{\text{Max}}^{\text{SG}} > \bar{\Delta}$. This means that country A prefers the deployment treaty to the free-driver outcome iff $\bar{\Delta} \le \Delta < \Delta_{\text{Max}}^{\text{SG}}$. We continue with country B. It is

$$\pi_B(g_A^{**}, g_B^{**}) - \pi_B\left(\tfrac{\theta}{\theta+1}T_A, 0\right) = \frac{b\theta}{2(\theta+1)^2(4\theta+1)}\left[\bar{T}^2(2-\theta) + 2\bar{T}(4+7\theta) + (12\,\theta^2+11\,\theta+2)\Delta^2\right].$$

The root larger than $\bar{\Delta}$ is

$$\Delta_{\text{Min}}^{\text{SG}} := -\frac{7\,\theta+4+2\sqrt{3\,\theta^3+9\,\theta^2+9\,\theta+3}}{12\,\theta^2+11\,\theta+2}\,\bar{T} \tag{28}$$

At $\Delta = \bar{\Delta}$, the above expression is

$$-\frac{2b\theta(\theta+1)\bar{T}^2}{(4\theta+1)(2\theta+1)^2} < 0$$

so that country B prefers the free-driver outcome to the deployment treaty iff $\Delta < \Delta_{\text{Min}}^{\text{SG}}$. It is $\Delta_{\text{Max}}^{\text{SG}} - \Delta_{\text{Min}}^{\text{SG}} = -2\frac{\sqrt{(\theta+1)}}{12\theta^2+11\theta+2}\bar{T}\cdot\left[\sqrt{3}(3\theta+2) + \sqrt{(\theta+1)} - \sqrt{3}(\theta+1)\right]$, which is clearly positive. The relative size of $\Delta_{\text{Morat}}^{\text{SG}}$ on the one hand and $\Delta_{\text{Min}}^{\text{SG}}$ and $\Delta_{\text{Max}}^{\text{SG}}$ on the other hand is dependent on θ.

Proof of Proposition 4. The algebraic expressions for climate clash and free-rider equilibrium are the same; because of that some relevant quantities have already been defined in Proposition 2.

(i) That country A prefers the moratorium treaty over the climate clash iff $\Delta > \Delta^{\mathrm{CG,A}}_{\mathrm{Morat}}$ has been demonstrated in the proof of Proposition 2. To see that country A always prefers the deployment treaty over the moratorium, note that

$$\pi_A(g_A^{**}, g_B^{**}) - \pi_A(0,0) = -\frac{2b\theta}{4\theta+1}\bar{T}(2\Delta - \bar{T})$$

which is positive due to $-\bar{T} > 0$.

(ii) That country B prefers the moratorium treaty over the climate clash iff $\Delta > \Delta^{\mathrm{CG,B}}_{\mathrm{Morat}}$ has been demonstrated in the proof of Proposition 2. It is immediately clear that $\Delta^{\mathrm{CG,A}}_{\mathrm{Morat}} > \Delta^{\mathrm{CG,B}}_{\mathrm{Morat}}$. Comparing deployment treaty and climate clash, country B prefers the former iff Δ is larger than

$$\Delta^{\mathrm{CG}}_{\mathrm{Min}} := -\frac{\left(3+4\theta+2\sqrt{4\theta^2+5\theta+2}\right)}{(4\theta+1)(2\theta+1)}\bar{T}\ . \tag{29}$$

In terms of deployment treaty vs. moratorium we have

$$\pi_B(g_A^{**}, g_B^{**}) - \pi_B(0,0) = \frac{2b\theta}{4\theta+1}\bar{T}(\bar{T}+2\Delta)\ .$$

This means that country B prefers the moratorium treaty to the deployment treaty iff

$$\Delta > -\frac{1}{2}\bar{T} =: \Delta^{\mathrm{B}}_{\mathrm{Morat,Treaty}}\ . \tag{30}$$

(iii) For the moratorium treaty to be stable it is necessary that both countries prefer it over the climate clash; this is equivalent with $\Delta > \Delta^{\mathrm{CG,A}}_{\mathrm{Morat}}$. In addition, because of assumption 2, only one of the two countries needs to prefer the moratorium over the deployment treaty. From (i) we know that country A never prefers the moratorium, from (ii) we know that country B prefers the moratorium treaty over the deployment treaty iff $\Delta > \Delta^{\mathrm{B}}_{\mathrm{Morat,Treaty}}$. Under assumption 2 the moratorium treaty hence realizes for all asymmetry levels above $\Delta^{\mathrm{CG}}_{\mathrm{Max}} := \max(\Delta^{\mathrm{CG,A}}_{\mathrm{Morat}}, \Delta^{\mathrm{B}}_{\mathrm{Morat,Treaty}})$.

Proof of Proposition 5.

(i) It is

$$\pi(0,0) - \pi\left(\tfrac{\theta}{\theta+1}T_A, 0\right) = \frac{b\theta}{2(\theta+1)^2}(\Delta - \bar{T})\left(\Delta(2\theta+1) + \bar{T}(2\theta+3)\right)$$

which is negative at $\Delta = 0$. The unique positive root is

$$\Delta^{\mathrm{Welfare}}_{\mathrm{Morat,Driver}} := -\frac{2\theta+3}{2\theta+1}\bar{T} \tag{31}$$

and it is straightforward to show that $\Delta^{\mathrm{Welfare}}_{\mathrm{Morat,Driver}}$ is larger than $\bar{\Delta}$.

(ii) We have

$$\pi(0,0) - \pi(g_A^*, g_B^*) = \frac{b\theta}{(2\theta+1)^2}\Big[-\bar{T}^2(4\theta+3) + (4\theta^2+4\theta+1)\Delta^2\Big]\ ,$$

which is negative at $\Delta = 0$. The unique positive root is at

$$\Delta^{\mathrm{Welfare}}_{\mathrm{Morat,Clash}} := -\frac{\sqrt{4\theta+3}}{2\theta+1}\bar{T}$$

and it is straightforward to show that $\Delta^{\text{Welfare}}_{\text{Morat,Clash}}$ is larger than $\bar{\Delta}$ and $\Delta^{\text{Welfare}}_{\text{Morat,Driver}} > \Delta^{\text{Welfare}}_{\text{Morat,Clash}}$.

Proof of Proposition 6. We prove part (i) and (ii) together. Consider the general n country case. We use again $\theta = b/c$ and define $\beta_m = \frac{m\theta}{m\theta+1}$ and the average optimal temperature change among the first m countries $\bar{T}^{(m)} = \frac{1}{m}\sum_{i=1}^{m} T_i$. The best response of country i to the other countries' geoengineering deployment level $T_{-i} = \sum_{j\neq i}^{n} g_j$ is characterized by the first order condition $\frac{d\pi_i(g_i;T_{-i})}{dg_i} = 0$. In the 'SG only' world it is necessary to check whether the non-positive constraint binds. We calculate the best response function

$$g_i(T_{-i}) = \begin{cases} \min\left\{\frac{\theta}{\theta+1}\left(T_i - T_{-i}\right), 0\right\} & \text{SG only} \\ \frac{\theta}{\theta+1}\left(T_i - T_{-i}\right) & \text{CG available} \end{cases} \tag{32}$$

The game consisting only of the first m countries, i.e. the m countries with the highest preferences for cooling, has the equilibrium

$$g_i^{(m)} = \theta(T_i - \beta_m \bar{T}^{(m)}) \ . \tag{33}$$

The overall temperature change in this equilibrium is $\sum_{i=1}^{m} g_i^{(m)} = \beta_m \bar{T}^{(m)}$. This is the equilibrium of the 'SG only' case if and only if country $m+1$ considers the temperature reduction as too much (and hence is unwilling to deploy more SG) and country m is willing to contribute SG (i.e. the game of the first $m-1$ countries results in a total temperature reduction that does not exceed country m's optimal reduction so that country m, due to vanishing marginal costs at the point of non-contribution, is willing to deploy SG). This is the case iff

$$\beta_{m-1}\bar{T}^{(m-1)} > T_m \geq \beta_m \bar{T}^{(m)} \ , \tag{34}$$

which is equivalent to

$$\Delta\left(\beta_{m-1}\bar{\delta}^{(m-1)} - \delta_m\right) > (1-\beta_{m-1})\bar{T} \quad \text{and} \quad \Delta\left(\beta_m \bar{\delta}^{(m)} - \delta_{m+1}\right) \leq (1-\beta_m)\bar{T} \tag{35}$$

Define $\Delta^{(m)} = \frac{1-\beta_m}{min(0,\beta_m\bar{\delta}^{(m)} - \delta_{m+1})}\bar{T} \in [0,\infty]$ for $m = 1,\ldots,n-1$ and set $\Delta^{(n)} = 0$ and $\Delta^{(0)} = \infty$. It is easy to see that $\Delta^{(m)}$ decreases in m. That (33) is the equilibrium of the SG only game then is equivalent with $\Delta^{(m)} \leq \Delta < \Delta^{(m-1)}$. The equilibrium when CG is available is always characterized by (33) with $m = n$ the first m contributions being negative and the remaining $n-m$ positive. In the case $n = 2$ we have, as required, $\Delta^{(1)} = \bar{T}$ and the quantities given in Proposition 1 all coincide with (33), evaluated at $m = 2$ (free-rider and climate clash) or $m = 1$ (free-driver).

We turn to part (iii). Let $m \leq n$ be such that in the 'SG only' case exactly m countries deploy SG, $\Delta^{(m)} \leq \Delta < \Delta^{(m-1)}$. It is straightforward to see that the availability of CG decreases welfare relative to the 'SG only' case iff

$$E := (1+\theta)\sum_{k=1}^{n}(T_k - \beta_n\bar{T})^2 - (1+\theta)\sum_{k=1}^{m}(T_k - \beta_m\bar{T}^{(m)})^2 - \sum_{k=m+1}^{n}(T_k - \beta_m\bar{T}^{(m)})^2 \ > 0 \tag{36}$$

We use (19) to write expression E as a quadratic function in Δ, $E = C_0 + C_1\Delta + C_2\Delta^2$. We find

$$C_0 = (1+\theta)n\bar{T}^2(1-\beta_n)^2 - (1+\theta)m\bar{T}^2(1-\beta_m)^2 - (n-m)\bar{T}^2(1-\beta_m)^2 \tag{37}$$

$$C_1 = 2\bar{T}\bar{\delta}^{(m)}(1-\beta_m)^2\left(\tfrac{\beta_m}{1-\beta_m}n - m\theta\right) \tag{38}$$

$$C_2 = \theta\sum_{k=m+1}^{n}\delta_k^2 + \beta_m(\bar{\delta}^{(m)})^2\big(2m\theta - \beta_m(m\theta+n)\big) \tag{39}$$

From Proposition 1 (iii) we know that $E > 0$ for all parameter constellations in the case $n = 2$. It is cumbersome to analytically determine the parameter constellations for which $E > 0$ for general n. We instead did a numerical analysis to get a sense of the conditions. The first observation from this analysis is that the extreme free-driver setting, $\delta_1 < 0$ and $\delta_k = -\delta_i/(n-1)$, seems to have the highest potential to result in $E < 0$, i.e. exceptions to the rule of welfare-decreasing CG. In this extreme free-driver setting, we find constellations with $E < 0$ for all $n \geq 5$ (whereas an equidistant δ-profile has $E < 0$ constellations only for $n \geq 9$). Constellations with $E < 0$ are characterized by high levels of asymmetry Δ and low benefit-cost ratios θ. Future research is needed to analytically determine the conditions under which CG decreases/increases welfare in the non-cooperative case.

Appendix B Calibration

For our parameter calibration we assume that countries base their decisions on benefits and costs in a certain year. The reason is twofold. First, we have modeled climate intervention as a one-shot (timeless) game in the first place, leaving more realistic models featuring a dynamic game structure for future research. The second reason is that we focus on the short-term interaction between SG and CG, leaving aside decisions with a long-term time profile such as choices on mitigation and R&D; because costs and benefits of climate interventions in this model have the same time profile, discounting does not affect the relative size of benefits and costs. We therefore focus on benefits and costs in a certain year, all expressed in terms of USD in 2015, the year of the most recent assessment of SG costs.

Benefit parameter b. Let g denote the growth rate. Burke et al. (2015) finds a quadratic relation between temperature and growth, $g = const + b_1 T + b_2 T^2$ with $b_1 = 0.0127\,°C^{-1}$ and $b_2 = -0.0005\,°C^{-2}$ (Extended Data Table 1). We can write this as

$$g = const - \frac{\tilde{b}}{2}(T - T^*)^2 \tag{40}$$

with $\tilde{b} = -2b_2 = 1/1000\,°C^{-2}$ and $T^* = -\frac{b_1}{2b_2} = 12.7\,°C$.

We now turn to the link between growth rate g and benefit function B, where we assume that a country's benefit function is given by GDP. The GDP as a function of temperature is $B(T) = Y_0(1+g)$, where Y_0 is the GDP at the beginning of the period. We can hence write $B(T) = const - Y_0 \frac{\tilde{b}}{2}(T - T_i)^2$. For our analysis we assume that the countries are of the size of the US. As explained above, we express all monetary quantities in terms of 2015 values. The GDP of the US in 2015 was $Y_0 = 17.95$ trillion \$. The quadratic coefficient of the benefit function $b = \tilde{b} \cdot Y_0$ thus reads

$$b = 17.95 \text{ bn}\,\$/°C^2\ . \tag{41}$$

Cost parameter c. The following table reflects the best currently available estimates of annual costs of stratospheric geoengineering with sulfur. The range of stratospheric sulfur load is taken from Pierce et al. (2010). We assume a linear relation between sulfur load and cost, and choose the mid-point of the range 2 to 8 billion \$ for 5 Mt of sulfur load in National Research Council (2015), referring to McClellan et al. (2012). A recent study by Smith and Wagner (2018) shows numbers in the same ballpark. The effect of stratospheric load on changes in radiative forcing is read from the SO2 scenario in Figure 4 in Pierce et al. (2010). The associated change in temperatures is based on the climate sensitivity lambda=$0.54\,°C\,m^2/W$, which corresponds to an equilibrium temperature change of 2.1 $°C$ (Shaviv 2005). We can then fit the model $C(T) = \frac{c}{2}(\Delta T)^2$ to the relationship between costs and temperature change and get

$$c = 8.35 \text{ bn}\,\$/°C^2\ . \tag{42}$$

Variable						Source
Sulfur load (Mt)	0	2	5	10	20	McClellan et al. (2012)
Costs (bn $)	0	2	5	10	20	National Research Council (2015)
Δ RF (Wm^{-2})	0	0.9	1.8	2.8	4.1	Pierce et al. (2010)
Δ T (K)	0	0.486	0.972	1.512	2.214	Shaviv (2005)

Table 1: Available data for cost estimates of stratospheric geoengineering with sulfur.

Temperature parameter $\bar{T}$. The parameter $\bar{T}$ captures by how much the average temperature exceeds the optimum at the beginning of the global thermostat game. We assume that preindustrial temperatures were on average optimal, and use for our numerical illustration Shaviv (2005) with an equilibrium temperature change of $2.1°C$. This corresponds to $\bar{T} = -2.1°C$.

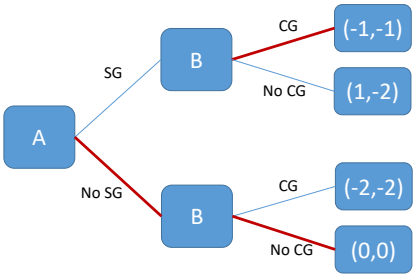
Appendix C The timing of the global thermostat game

This section discusses the time structure of our model. There are two separate modelling assumptions pertaining to the time structure: (i) the time structure of the non-cooperative global thermostat game, in particular the temporal order of SG and CG, and (ii), in the context of cooperation possibilities, the temporal deployment order of treaty members ('coalition') and non-members ('fringe'). We discuss both aspects separately.

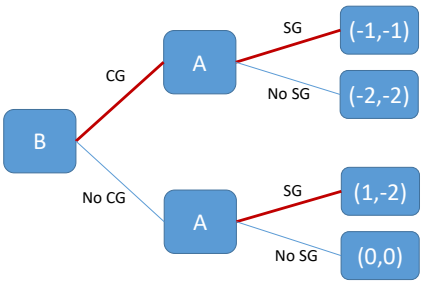
Time structure of the non-cooperative game. In particular we here discuss the relation between our model and Parker et al. (2018). Both our model and Parker et al. (2018) fall into the class of models that abstain from modelling climate interventions as a full dynamic game with repeated interaction. Within this class there are essentially three ways to model the relative timing of SG and CG, illustrated in Figure 9 with the specific payoff structure from Parker et al. (2018). The time structure of Parker et al. (2018) is depicted in Figure 9a: Country A first decides whether to deploy SG, then country B chooses whether to deploy CG or not. The unique Nash equilibrium in this setting is (No SG, No CG): The threat of CG deters country A's use of SG. Figure 9b shows the alternative sequential timing in which country B decides on CG first, followed by country A's SG decision (this timing obviously is only meaningful for countervailing CG as neutralizing CG requires a previous SG deployment). The unique Nash equilibrium with the reversed sequential order is (CG,SG). The use of SG makes sense for country A irrespective of B's decision; this, in turn, makes CG the only reasonable decision for B. So we see that CG's deterrence effect in Parker et al. (2018) crucially depend on the assumption of country B 'having the last word' on climate intervention. Note that the results also depend on the payoff structure: one important case is if SG and CG are symmetric, for instance when 'no SG / CG' results in $(-2, 1)$ instead of $(-2, -2)$. Then the climate clash with a joint deployment of SG and CG is the unique equilbrium of the game prediction, irrespective of the temporal order.[19]

There is no logical reason why (countervailing) CG could not precede SG. In this sense, none of the two sequential settings seems to be a plausible representation of the global thermostat game. Therefore, the simultaneous game (represented in Figure 9c), which gives no technology an advantage over the other in terms of the time structure, is the most plausible among the three. The assumption of simultaneous SG and CG deployment (together with a richer action space and payoff structure) has been adopted in our model, see section 2. We see that the unique Nash equilibrium in the simultaneous variant of the Parker et al. (2018) game is (SG,CG), and this is in line with our finding of the climate clash equilibrium. To summarize: The deterrence effect in Parker et al. (2018) rests on the specific time structure. Other temporal orders of SG and CG give rise to the analog of our climate clash equilbrium.

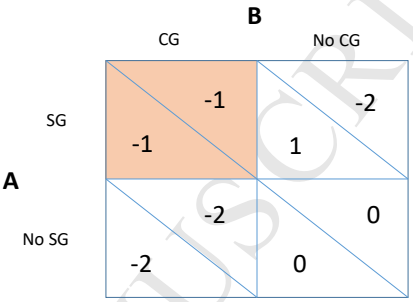
[19] We thank a reviewer for pointing this out.



(a) The timing in Parker et al. (2018).



(b) Reversed sequential timing.



(c) Simultaneous game.

Figure 9: Alternative time structures of the game in Parker et al. (2018). The payoff structure is the same in all subfigures.

Temporal deployment order of treaty members and non-members. A separate question regarding timing arises when we include the option of cooperation. The literature has distinguished two assumptions on the relative timing of treaty members ('coalition') and non-members ('fringe'), see e.g. Finus (2008). Either members and non-members make their decisions, here on climate interventions, simultaneously ('Cournot assumption'), or the coalition moves first followed by simultaneous decisions by non-members ('Stackelberg leader assumption'). For the case $n = 2$ both assumptions are equivalent; for our analysis with a general n in section 7 we adopted the Stackelberg leader assumption, leaving the comparison with the Cournot case for future research.

Appendix D Welfare change induced by CG

Choices for the welfare-changing effect of CG. While the type of equilibrium only depends on the benefit-cost ratio $\theta = b/c$, the comparison of welfare levels, and therefore also statements on the welfare changing effect of CG in Figure 5, is not determined by the choice of θ alone. Two choices are needed in this context. The first choice is on b and c for a given θ. Options include

(i) Keep b fixed. Then only $c = b/\theta$ varies with θ.

(ii) Keep c fixed. Then only $b = \theta c$ varies with θ.

(iii) Choose b and c such that social optimal welfare does not depend on θ. Total welfare is $-\frac{4b^2\Delta^2+bc(\Delta^2+\bar{T}^2)}{4b+c}$. That this equals total welfare evaluated at reference values b_0 and c_0 (in our case the calibrated values in (12)) implies

$$c = b_0 \cdot \frac{4\theta+1}{4\theta_0+1} \cdot \frac{4\Delta^2\theta_0 + \Delta^2 + \bar{T}^2}{4\Delta^2\theta + \Delta^2 + \bar{T}^2} \tag{43}$$

Obviously, b here is θc.

The second choice we need to make is how to measure changes in welfare levels. Options include

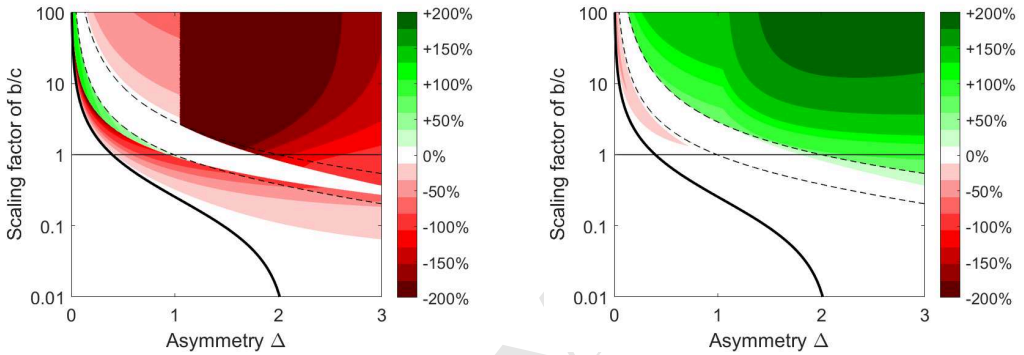
(i) Focus on absolute welfare changes

(ii) Measure welfare changes in terms of the respective outcome in the 'SG only' case

(iii) Measure welfare changes in terms of the respective social optimal outcome

On both questions we have chosen option (iii).

Country-specific welfare change. Figure 10 shows the country-specific welfare impact of CG for $n = 2$; this effectively disaggregates the aggregate effect shown in Figure 5. As before, red and green colors indicate a harmful and beneficial impact of CG, respectively. The plots suggest that country A is typically worse off under CG, while country B benefits from the availability of CG.



(a) Country A. (b) Country B.

Figure 10: The welfare impact of CG, differentiated into effects on country A and country B. As in Figure 5, the welfare differences between CG and SG are normalized by the *total* welfare under the deployment treaty. Note that the scale here is different from the $[-100\%, 100\%]$ range in Figure 5.