Mitigation strategies under the threat of solar radiation management

Fabien Prieur,* Martin Quaas,† Ingmar Schumacher‡

Abstract

We develop a two-player, two-period game of climate change. In the first period, players non-cooperatively choose their emission level, which yields a private benefit. Aggregate emissions then determines the extent of the environmental damage that is incurred in the second period. In the second period, depending on aggregate emissions, players can decide to undertake solar radiation management (SRM) at some cost. SRM is a means to reduce temperature, but at the same time it is perceived as a risky activity since it potentially involves a series of negative environmental impacts. This is captured by assuming that positive SRM induces a shift from a certain damage to a (larger) expected damage. Solving for the Nash equilibrium in SRM strategies allows us to identify a critical emission threshold that triggers (unilateral) SRM by the player that is the most vulnerable to climate change. Moving back to the first period, we then investigate how the potential deployment of SRM in the future affects current emissions. In particular, we wonder if the threat of SRM may act as a coordination device. For that purpose we compare two alternative scenarios. Either the players commit to meeting the threshold and face the resulting coupled constraint on aggregate emissions. Or they get rid of it and

*EconomiX, Université Paris Nanterre, 200 avenue de la République, 92001 Nanterre Cedex, France. E-mail: fabien.prieur@u-paris10.fr.

†Department of Economics, University of Kiel, Wilhelm-Seelig-Platz 1, 24118 Kiel, Germany. E-mail: quaas@economics.uni-kiel.de.

‡IPAG Business School, 184 boulevard Saint-Germain, 75006 Paris, France, and Department of Economics, Ecole Polytechnique Paris, France. Tel.: +352 621264575. E-mail: ingmar.schumacher@ipag.fr.

non-cooperatively choose their emissions knowing that this will trigger (unilateral) SRM. Solving for the Nash equilibrium corresponding to the latter situation and the coupled constraint Nash equilibrium for the former, we show that under some conditions players may individually find it optimal to refrain from emitting too much in order to avoid SRM. This conclusion holds true even in the worst scenario in which one player – the one that won't undertake SRM – bears alone the responsibility of meeting (or not) the constraint.

Keywords: climate change, solar geoengineering, heterogenous damages, strategic interaction, commitment

JEL classification: C72, Q54.

1 Motivation

After the semi-success of the Paris agreement – or total failure, it depends on whether one regards a glass as either half full or half empty – and the election of Trump as president of the US, it seems pretty clear that the international community will not be able to promote a coordination of mitigation efforts in the short and even medium term. Yet, global warming is beginning to take place before our eyes and negative impacts of climate change are already felt and expected to intensify. In such a situation of coordination failure and insufficient world mitigation, and given the inertia of the climate system, it is now urgent to find an alternative, if any, to at least hinder climate change. Then, the obvious question is: what can be done?

Some in the scientific community argues that we should engineer the climate. Geoengineering is a term that encompasses a series of measures intended to cope with warming. This ranges from building special plants to remove carbon from the atmosphere, through reforestation, to solar radiation management (SRM). SRM can be broadly defined as the process consisting in spraying sulfate aerosol into the stratosphere to block the sunlight and cool the earth.

In this paper, as in most of the literature, we will focus on solar geoengineering, or SRM. The reason for this choice is that these techniques correspond to the most promising but also the most challenging and controversial solutions to affect the climate system. Most promising should be understood as effective and affordable. To understand how it works, we can simply refer to a natural experiment, the eruption of Mount Pinatubo in 1991 in the Philippines. This eruption was accompanied with the release of 20 million metric tons of sulfur dioxide which then reached the upper atmosphere and reacted with water vapor to form a global haze. As a result, scientists observed a decrease in direct solar radiation of about 30% whereas the average temperature on earth dropped by 0.5 Celcius for more than a year. So the cooling potential of these measures seem well established. They are also affordable. In a recent report, the Royal Society (2009) emphasized that SRM and marine clouds generation were the cheapest solutions to control temperature. In addition, it would be possible thanks to SRM to bring back temperature to any desired

level at a much lower cost than with mitigation only.

The other side of the coin is that SRM involves a series of new (specific) risks that must be accounted for. There is evidence that SRM may (substantially) affect regional precipitation patterns and volumes worldwide (due to modified atmospheric and ocean circulation). Referring again to the Pinatubo eruption, this event caused large hydrological responses, including reduced precipitation, soil moisture, and river flow in many world regions (NCAR, 2007). The same kind of evidence can be found in the more distant past with the eruption of the Laki volcano that took place in the eighteen century in Iceland and lasted for eight months. This contributed to famine in Africa, India, and Japan (Robock, 2008). Besides, one important specific risk with SRM is referred to as the termination effect (see Goes et al. 2011). The basic principle underlying this effect is the following: if for any reason one stops SRM - after several decades of use - then serious troubles are to be expected as a result of the abrupt climate response. This is because it tackles the symptom and not the cause of warming, which is the continuous increase in CO2 concentration. Finally, there are many other potentially negative impacts (not directly related to the climate) ranging from the increase in ocean acidity and ozone hole to the impact on biodiversity etc.

Last but not least, two important characteristic of SRM are worth noticing. First, if the deployment of large scale (S)RM will likely be a means to reduce average temperature, its impact will feature spatial heterogeneity. As perfectly emphasized by Quaas and Rickels (2016): "The climate system (temperature, precipitation, wind etc.) is expected to react to the application of RM measures differently from region to region. Consequently, application of RM measures would not only imply an uneven distribution of costs and benefits but also great differences in the desired levels of RM." Second, because it is cheap, unilateral deployment of SRM will be possible. The combination of these two characteristics obviously raises many other issues.

Economists have started to address some of these issues for the last ten years or so. A non-exhaustive list of implications involved by the use of SRM includes among others, moral hazard (lower mitigation once SRM becomes a valuable option, Moreno-Cruz and

Smulders, 2010), desirability and consequences of R&D in the SRM domain (Quaas et al. 2017), optimal climate policy and carbon tax in the presence of SRM (Moreno-Cruz and Keith, 2013, Heutel et al., 2015), governance (as SRM may be undertaken unilaterally, the questions are: who is gonna to do it? what will be the target given the heterogeneity in terms of temperature preferences etc., see Barrett, 2008, and Weitzman, 2015).

In this paper, we are more particularly interested in the new strategic interaction that may come out of the availability and affordability of SRM measures. Our contribution is very close to Moreno-Cruz (2015). The author develops a simple two-period game where players (countries or regions) non-cooperatively choose their mitigation effort in the first period and then may undertake SRM during the second one. Within this framework, he looks at the impact of unilateral SRM on mitigation efforts at the subgame perfect Nash equilibrium by comparing these efforts with those that would arise at the first best or cooperative solution. He concludes that unilateral SRM deployment (in the future) may induce inefficiently high levels of mitigation (present) by the threatened region when asymmetries in terms of climate change and SRM damages are sufficiently large. He finally observes that, and we quote, the latter region "would be better off negotiating a treaty where both countries jointly implement mitigation." However, Moreno-Cruz does not explain how this coordination might take place and we all know now that coordination is very unlikely to occur in order to deal with climate change. So he barely addresses this idea that some region may prefer something else than uncoordinated and too high emissions then followed by SRM by others.

This is the motivation of the current paper: our aim is to properly deal with the "something else," that is the alternative scenario to SRM deployment. But we believe that the analysis should still be conducted in non cooperative setting and instead involves a commitment problem. Indeed, the main question we want to ask is under which conditions is it optimal for one region to refrain from emitting too much to avoid the threat of SRM?

To address this issue, we develop a two-player, two-period game similar to Moreno-Cruz (2015). As to the impact of SRM, we want to convey the idea that if SRM may be good for some player, there is also a lot of uncertainty surrounding its impact (think

about the termination effect). So in the end, we don't really know what may happen once it will be deployed. The simplest way to capture this idea is to assume that positive SRM induces a shift from a certain damage to a (larger) expected damage. The analysis of the second period problem reveals the existence a threshold level for aggregate emissions that triggers (unilateral) SRM deployment by the player who is the most vulnerable to climate change. This implies that, as seen from the first period, players' payoffs are defined piecewise depending on whether first period aggregate emissions exceed or remain below the threshold. Players – particularly the one who is subject to the threat of SRM – may naturally want to take care of the existence of the threshold because it affects the payoffs. This is where our approach differs from the one of Moreno-Cruz. To model this situation, we adopt a commitment perspective. This consists in endowing the threatened player (only) with another discrete decision. Indeed, this player now has to choose whether or not he/she will commit to the constraint on aggregate emissions imposed by the threshold, and refrain from emitting too much (to avoid SRM). This means that we should add a third stage in our game that comes first and during which this decision is taken.

To form this decision, the player has to determine what can be the possible outcomes of the interaction in the two next stages (periods). In fact, there are two possible equilibria. Either this player does not care about the constraint on emissions and chooses his/her emission level freely and non-cooperatively given the future SRM reaction of the other player. This is referred to as the equilibrium without commitment. Or, he may want to take into account the constraint in his/her first period problem and then simply releases a level of emissions exactly equal to the difference between the threshold and the other player optimal emission level. This characterizes the equilibrium with commitment.

The last important question to be addressed is: what is the best strategy? Our results show that even in the worst scenario in which one player bears alone the responsibility of meeting (or not) the constraint on aggregate emissions, this player may prefer to commit to it. This conclusion holds under some conditions that all point to the same requirement that the equilibrium with commitment should not be too painful to this player (compared to the alternative). These conditions involve the cost and benefit of commitment. On

the cost side, given that commitment requires to reduce emissions, the threshold should be high enough so that there is an incentive to make the effort. Moreover, the expected damage under SRM should be high enough compared to the certain damage with commitment. This ensures that the benefit from commitment, in terms of avoided damage, is sizable.

The paper is organized as follows. Section 2 presents the basic model and a benchmark situation. Then Section 3 characterizes the possible equilibria and provides some general existence results. In Section 4, we develop a simple linear quadratic game which allows us to address the main question raised by the present analysis. Section 5 concludes.

2 Game setting and benchmark

Two-period, two-player game. Players, indexed by $j = i, -i$, release emissions, e_j, during the first stage for a private benefit, $F_j(e_j)$. Aggregate emissions $e = \sum_j e_j$ affect some climate variables, measured at the regional scale, $x_j = x_j(e)$ (temperature), which in turn determines the extent of the damage, $D_j(x_j(e))$. In the second stage, player $-i$ (the South, she) may unilaterally decide to implement RM measures, g_{-i}, by incurring a monetary cost $C(g_{-i})$. If player $-i$ undertakes RM this is because it is globally beneficial for her. For simplicity, we assume that $x_j = x$ for all j, that x is linearly dependent on aggregate emissions so that player $-i$ damage with positive RM amounts to $D_{-i}(e - g_{-i})$. Our first assumption summarizes the conditions imposed on these functional forms.

Assumption 1 *Players share the same private benefit from emissions: $F_i = F_{-i} = F$, with $F'' < 0$. Damage functions in the absence of RM, D_j, satisfy: $D'_j > 0$, $D''_j > 0$ for $j = i, -i$. RM measures if implemented impose a cost C, with $C' > 0$, $C'' > 0$ for all $g_{-i} \geq 0$.*

The main ingredient of the model concerns the way we model the intrinsic heterogeneity between players and the specific heterogeneity in the impact of RM measures by player $-i$. The starting point is that player $-i$ is the most exposed to the (negative) impact of

climate change. That is why we consider that this player may find the level of aggregate emissions, and related damage, so high that she prefers to deploy RM techniques in order to keep the damage under control. There won't be any particular problem surrounding this action if it was innocuous to player i (the North, he). However, we argue that RM by player $-i$ makes player i feel that climate damage becomes more variable, or riskier, and can potentially be costly to him. To account for the existing threat of RM, we replace the sure damage (in the absence of RM) with an expected damage, $E[D_i]$, that is faced by player i because of the deployment of RM techniques. In short, we impose:

Assumption 2 *In the absence of RM, player $-i$ is more vulnerable to climate change impacts than player i. That is, a marginal increase in aggregate emissions is more damaging to player $-i$: $D_i' < D_{-i}'$ for all e. The impact of RM measures features spatial heterogeneity. Player $-i$ damage function is unchanged whereas player i damage becomes more variable or risky, i.e., D_i should be replaced with $E[D_i]$. In addition, the implementation of RM techniques represents a threat to player i: $E[D_i] > D_i$.*

Timing of the game: players first choose their emission levels. Then player $-i$ decides whether or not to undertake RM and its RM effort (if any).

Before going any further, we may want to characterize the benchmark situation in which SRM is not an available option. This is the standard static emission game. Players basically choose their emission levels by solving:

$$\max_{e_j} F(e_j) - D_j(e) \tag{1}$$

Then we can establish that

Proposition 1

- *In the absence of SRM measures, there exists a unique Nash equilibrium.*
- *Aggregate emissions, e^u, are implicitly defined by*

$$e^u = M(e^u) \text{ with } M(e^u) = \sum_j (F')^{-1}(D_j'(e^u)) \text{ and } M'(e^u) = \frac{\sum_j D_j''}{F''} < 0. \tag{2}$$

Player j emissions, e_j^u, are then given by

$$F'(e_j^u) = D'_j(e^u) \text{ for } j = i, -i. \tag{3}$$

In the remainder of the analysis, this Nash equilibrium will be referred to as the *unconstrained equilibrium*. There is nothing much to say about it except that we have $e_i^u > e_{-i}^u$.

Now, we can move to the most interesting analysis of the original game in which the threat of SRM is operating. The key point will be to show the existence of two different types of regimes and equilibria and to understand what does it imply in terms of player i's decisions.

3 Equilibrium outcomes

As it will be apparent soon, our two-stage game is not standard but it can be solved, at least partly, using standard techniques. In particular, we proceed backward for the resolution.

During the second stage the South chooses 1/ whether or not to undertake SRM and 2/ the level of deployment (if any). Given the aggregate emission level, e, player $-i$ solves:

$$\min_{g_{-i} \geq 0} D_{-i}(e - g_{-i}) + C(g_{-i}). \tag{4}$$

From the first order condition

$$C'(g_{-i}) \geq D'_{-i}(e - g_{-i}) \text{ with equality if } g_{-i} > 0,$$

and under Assumption 1, we can define $\bar{e}$ such that $C'(0) = D'_{-i}(\bar{e}) \Leftrightarrow \bar{e} = (D'_{-i})^{-1}(C'(0))$. For all $e \leq \bar{e}$, $g_{-i} = 0$ (corner regime w.r.t SRM) whereas $g_{-i} > 0$ when $e > \bar{e}$. In the latter case, the RM effort is given by (with a slight abuse of notation):

$$g_{-i} = g_{-i}(e) \text{ with } g'_{-i} = \frac{D''_{-i}}{C'' + D''_{-i}} \in (0, 1). \tag{5}$$

So at this stage, we get the expression of a threshold level of aggregate emissions, which will play a crucial role in the first stage because with this information in mind, player(s) will have to answer the following question: should we want to exceed this threshold or not, given the cost and benefit associated with one or the other decision.

We observe that the higher $C'(0)$, the cost of the first unit of SRM, the higher $\bar{e}$. In addition, the lower the marginal damage D'_{-i}, the higher $\bar{e}$.

In order to have an interesting problem, in what follows, we take as an assumption the following condition:

Assumption 3 *We impose* $\sum_j (F')^{-1}(D'_j(\bar{e})) > \bar{e} \Leftrightarrow e^u > \bar{e}$.

This basically implies that the threshold is going to represent a potentially binding constraint on players' optimization program during the first stage. Indeed, we want to be sure that under some circumstances, the threat of SRM is operative.

Let us now examine the problem that players face during the first stage. We should start by giving the expression of their payoffs taking as given $\bar{e}$ and the reaction function $g_{-i} = g_{-i}(e)$. For player i

$$\Pi_i(e_i, e) = F(e_i) - \begin{cases} D_i(e) \text{ if } e \leq \bar{e} \\ E[D_i(e - g_{-i}(e))] \text{ else} \end{cases} \tag{6}$$

For player $-i$:

$$\Pi_{-i}(e_{-i}, e) = F(e_{-i}) - \begin{cases} D_{-i}(e) \text{ if } e \leq \bar{e} \\ D_{-i}(e - g_{-i}(e)) + C(g_{-i}(e)) \text{ else} \end{cases} \tag{7}$$

We can observe that there exist two possible regimes (situations) and two different – but related – problems to study depending on $e \gtreqless \bar{e}$. This is the purpose of the next subsections to analyze the outcome of the specific interaction between the players.

3.1 Equilibrium without commitment

First, suppose that players do get rid of the threshold, meaning that emissions will go above the threshold. The players' optimization program can be written as

$$\begin{array}{l} \max_{e_i} F(e_i) - E[D_i(e - g_{-i}(e))], \\ \max_{e_{-i}} F(e_{-i}) - D_{-i}(e - g_{-i}(e)) - C_{-i}(g_{-i}(e)), \end{array} \tag{8}$$

we can solve these problems, that basically form a simultaneous emission game conditional on the reaction function of player $-i$. The corresponding equilibrium is index by n. Then we can check that the solution is admissible, i.e. $e^n > \bar{e}$.

More precisely, players' optimality conditions read as follows:

$$\begin{array}{l} F'(e_i) = (1 - g'_{-i}(e))E[D'_i(e - g_{-i}(e))] \\ F'(e_{-i}) = D'_{-i}(e - g_{-i}(e)) \end{array} \tag{9}$$

Aggregate emissions are implicitly given by:

$$\begin{array}{l} e = N(e) \text{ with } N(e) = (F')^{-1}(D'_{-i}(e - g_{-i}(e))) + (F')^{-1}((1 - g'_{-i}(e))E[D'_i(e - g_{-i}(e))]) \\ \text{and } N'(e) = \frac{(1-g'_{-i})D''_{-i} + (1-g'_{-i})^2 E[D''_i] - g''_{-i}E[D'_i]}{F''} \gtreqless 0, \end{array} \tag{10}$$

Then, we can claim that

Proposition 2

- *There exists an interior solution if*

$$\frac{E[D'_i(\bar{e})]}{D'_i(\bar{e})} \leq \frac{1}{1 - g'_{-i}(\bar{e})} \tag{11}$$

- *Aggregate emissions, e^n, satisfy $e^n > \bar{e}$. But they can be higher or lower than aggregate emissions at the unconstrained equilibrium, e^u.*

Of course, in order for the North to be ready to take the risk of positive SRM, the expected damage shouldn't be too high compared to the certain damage incurred in

the absence of SRM. Add discussion: things can get really worse in terms of aggregate emissions, with positive SRM, as it's possible that countries release more emissions than in the benchmark. This is actually the pessimistic scenario some politicians and scholars have in mind when they think about the impact of having the SRM option available. But, this is not the entire story since we can also characterize another equilibrium with very distinct features.

3.2 Equilibrium with commitment

Second, players – in particular player i who is subject to the threat of unilateral SRM - may want to take into account the threshold as it affects their payoffs. A natural scenario is the one in which the North alone takes care of the constraint $e \leq \bar{e}$: since we endow the South with the (additional) RM strategy, we may also allow the North to adapt to it by deciding whether or not to meet the constraint, i.e. to maintain aggregate emissions below or at $\bar{e}$. After all, once we consider that RM represents a threat for the North, then it comes naturally the question of how to deal with it. This is actually the main point of the paper. We want to determine under which conditions the North may decide to refrain from emitting too much given the threat of positive RM. We'll also investigate the repercussions in terms of individual and aggregate emissions.

Another justification of our approach is that it is logical to start with the analysis of the worst situation for the North, i.e., the situation in which the North bears alone the responsibility of protecting the world from RM. In fact, if we find conditions under which the North prefers to comply with the constraint in the worst scenario then, for sure, the results will continue to hold in more balanced situations where the burden is shared by the two players (argument: negotiation powers).

In any case, player i cannot control the SRM effort directly. But, as it is apparent from the analysis above, its emission strategy in the first stage (period) of the game is a means to influence player $-i$'s decision to implement SRM measures in the second stage. In order to account for this opportunity, we add the constraint $e \leq \bar{e}$ to player i's first

stage optimization when characterizing a solution with $g_{-i} = 0$ in the second stage. Of course, we implicitly consider that only player i (the North) has the capacity to release so many emissions as $\bar{e}$, or to set emissions to a level close to that threshold. On the contrary, player $-i$'s emission potential is limited (the South; so it cannot except to saturate the constraint by itself).

The problems to solve for the two players reduce to: for player i

$$\max_{e_i} F(e_i) - D_i(e) \text{ s.t. } e \leq \bar{e} \tag{12}$$

while player $-i$'s problem is unchanged, still characterized by (1)-(3). Player i optimality condition now reads as follows:

$$F'(e_i) \geqq D_i'(e) \text{ with equality if } e < \bar{e}. \tag{13}$$

One can check that as long as the constraint doesn't bite, the solution of the game coincides with the unconstrained equilibrium. However, whenever the constraint is binding, i.e. as long as Assumption 3 holds, aggregate emissions set at the ceiling $\bar{e}$. So we obtain:

Proposition 3

- *There exists a unique constrained equilibrium if and only if*

$$(F')^{-1}(D_{-i}'(\bar{e})) \leq \bar{e}. \tag{14}$$

- *Player $-i$'s emission level is given by*

$$F'(e_{-i}^c) = D_2'(\bar{e}) \Leftrightarrow e_{-i}^c = (F')^{-1}(D_{-i}'(\bar{e})), \tag{15}$$

- *player i releases the difference to achieve $\bar{e}$:*

$$e_i^c = \bar{e} - (F')^{-1}(D_{-i}'(\bar{e})). \tag{16}$$

Remark. Player i's emissions won't satisfy in general $F'(e_i^c) = D_i'(\bar{e})$. It would only be by chance that when player $-i$ emits e_{-i}^c given in (15) and aggregate emissions set up to the level $\bar{e}$, player i could choose its optimal (unconstrained) emission level.

Finally, players' payoffs are equal to:

$$\begin{aligned}\Pi_i^c &= F(\bar{e} - (F')^{-1}(D'_{-i}(\bar{e}))) - D_i(\bar{e}),\\ \Pi_{-i}^c &= F((F')^{-1}(D'_{-i}(\bar{e}))) - D_{-i}(\bar{e}).\end{aligned} \tag{17}$$

Once the possible outcomes have been characterized, the last step of the analysis consists in answering the North's question: which way to go? This requires to compare the emission levels in the different scenarios and ultimately the payoffs. If it is relatively easy to compute the payoffs in the constrained case, the same exercise is tricky at the equilibrium without compliance because emission strategies are defined implicitly only. So we have no other option but to resort to specific functional forms. This is what we do in the next section.

4 Application

In order to develop further the analysis, let us make use of the following functional forms:

$$\begin{aligned}&C(g_2) = c_1 g_2 + \tfrac{c_2 g_2^2}{2},\\ &D_2(e) = \tfrac{d_2 e^2}{2},\\ &\underline{D}_1(e) = \tfrac{\underline{d}_1 e^2}{2}\\ &F(e_i) = a e_i (b - \tfrac{e_i}{2})\\ &\text{Binary risk: } E[D_1] = p\underline{D}_1(e) + (1-p)\overline{D}_1(e) \text{ with } \overline{D}_1(e) = \tfrac{\overline{d}_1 e^2}{2} \text{ and } \overline{d}_1 > \underline{d}_1.\end{aligned} \tag{18}$$

These functional forms satisfy Assumption 1. We choose the simplest type of risk but this is just a example. Actually we simply want $E[D_1] = Ed_1 > \underline{d}_1$. To account for the heterogeneity between the two players we should also impose: $\underline{d}_1 < d_2$. To sum up, to be consistent with Assumption 2, we should impose:

$$\underline{d}_1 < \min\{d_2, Ed_1\}. \tag{19}$$

The main steps of the analysis are relegated into the Appendix or omitted when straightforward. The South (resp. North) is now referred to as player 2 (resp. 1).

Aggregate emissions e^u, at the unconstrained Nash equilibrium, are:

$$e^u = \frac{2ab}{a + \underline{d}_1 + d_2}$$

The RM strategy by player 2 and net emissions for a given e (solution of the second stage) are respectively given by:

$$\begin{array}{l} g_2(e) = \frac{d_2 e - c_1}{c_2 + d_2} \geq 0 \Leftrightarrow e \geq \bar{e} = \frac{c_1}{d_2} \\ \tilde{e} = e - g_2(e) = \frac{c_1 + c_2 e}{c_2 + d_2} \end{array} \tag{20}$$

The counterpart of Assumption 3 for the general problem now is:

$$e^u > \bar{e} \Leftrightarrow \frac{2ab}{a + \underline{d}_1 + d_2} > \bar{e} \Leftrightarrow \frac{2ab}{a + \underline{d}_1 + d_2} > \frac{c_1}{d_2}. \tag{21}$$

Under condition (21), player 1 has to choose between two opposite strategies: 1/ meeting the constraint imposed on aggregate emissions in order to avoid potentially costly RM effort by the South in the second stage of the game or 2/ getting rid of this constraint and facing positive RM. To understand what is the best option, we have to characterize and compare the two possible outcomes, starting with the equilibrium without compliance.

First period: players' and aggregate emission levels at the interior equilibrium

$$\begin{array}{l} e_1^n = \frac{b(c_2+d_2)(a(c_2+d_2)+c_2 d_2) - c_2 E d_1 (c_1 + b c_2)}{(c_2+d_2)(a(c_2+d_2)+c_2 d_2) + c_2^2 E d_1}, \\ e_2^n = \frac{(c_2+d_2)(ab(c_2+d_2) - d_2(c_1+bc_2)) + b c_2^2 E d_1}{(c_2+d_2)(a(c_2+d_2)+c_2 d_2) + c_2^2 E d_1}, \\ e^n = \frac{(c_2+d_2)(2ab(c_2+d_2) - c_1 d_2) - c_1 c_2 E d_1}{(c_2+d_2)(a(c_2+d_2)+c_2 d_2) + c_2^2 E d_1}, \end{array} \tag{22}$$

and of course we want these levels to be non-negative (which will be ensured later).

This in turn gives the equilibrium level of SRM and net emissions, $\tilde{e}$:

$$\begin{array}{l} g_2(e^n) = \frac{d_2}{c_2+d_2}(e^n - \bar{e}) \\ \tilde{e}^n = \frac{a(c_2+d_2)(c_1+2bc_2)}{(c_2+d_2)(a(c_2+d_2)+c_2 d_2) + c_2^2 E d_1} \end{array} \tag{23}$$

Next we can characterize the corner equilibrium in the particular (and simplest) situation where player 1 is the only one who bears the burden of meeting the constraint. Other

situations, in which players share the burden based on negotiations, will be discussed later but to get our point it's enough at this stage to show that there exist circumstances under which player 1 may actually prefer to refrain from emitting too much even in the worst scenario considered.

Corner solution: aggregate emissions by construction amount to $e^c = \bar{e}$ and are split between players as follows:

$$\begin{aligned} e_1^c &= \tfrac{(a+d_2)\bar{e}-ab}{a}, \\ e_2^c &= \tfrac{ab-d_2\bar{e}}{a}, \\ e^c &= \bar{e}(= \tfrac{c_1}{d_2}). \end{aligned} \tag{24}$$

Again, for the problem to be non-trivial, one must have $e_j^c \geq 0$ for $j = 1, 2$. This boils down to imposing: $\bar{e} \in [\frac{ab}{a+d_2}, \frac{ab}{d_2}]$. Combining this condition with (21), we impose the following condition on the threshold level of emission:

$$\frac{ab}{a+d_2} \leq \bar{e} \leq \min\left\{\frac{2ab}{a+\underline{d}_1+d_2}, \frac{ab}{d_2}\right\}, \tag{25}$$

what is the true upper bound is mainly an issue of how heterogenous are the two players in terms of damages. If player 1 is much less exposed to the damage than player 2 ($\underline{d}_1 < d_2 - a$) then the upper bound is ab/d_2. To avoid unnecessary discussion, let us consider that this is the case.

We can now proceed to the comparison between the two solutions. As to the emission levels, we can establish that:

Proposition 4

- *The ranking between emission levels satisfies $e^n > e^c$ and $e_1^n > e_1^c$ if and only if*

$$(c_2+d_2)(2abd_2 - c_1(a+d_2)) - c_1c_2Ed_1 > 0. \tag{26}$$

- *Under condition (25), this inequality holds if*

$$\frac{Ed_1}{\underline{d}_1} < \frac{c_2+d_2}{d_2}. \tag{27}$$

- *Under this ranking, we also have* $\tilde{e}^n > e^c$.

Discussion, interpretation of (27). Note that (26) can also be expressed in such a way that it yields an upper bound on the expected damage of player 1.

Before going to main point – the comparison between player 1's payoffs obtained in the two scenarios – we should say a word about player 2's situation and how it changes when one switches from the interior to the corner solution. At first glance, one would expect that player 2 is better off at the corner equilibrium because in this case this is player 1 who does most of the job of emission control and player 2 doesn't have to undertake costly SRM. Moreover, we may want to get rid of the paradoxical situation where player 2 prefers the interior solution and player 1 the corner.

Given that under (27), we have $\tilde{e}^n > e^c$, a sufficient condition for $\Pi_2^c > \Pi_2^n$ is obviously $e_2^n < e_2^c$. This is equivalent to:

$$(c_2 + d_2)(c_2(2abd_2 - c_1(a + d_2)) - 2ac_1d_2) - c_1c_2^2Ed_1 > 0, \tag{28}$$

which is nothing more than a slightly stronger version of (26). In order to avoid unnecessary complications, let's assume hereafter that this condition holds and focus on the more interesting problem faced by player 1.

Noticing that $e_1^n = b - \frac{c_2Ed_1}{a(c_2+d_2)}\tilde{e}^n$, direct manipulations give the expression of player 1's payoffs at the two possible equilibria:

$$\begin{array}{l} \Pi_1^c = \frac{1}{2a}\left(-3(ab)^2 + 4ab(a + d_2)\bar{e} - ((a + d_2)^2 + a\underline{d}_1)\bar{e}^2\right), \\ \Pi_1^n = \frac{1}{2a(c_2+d_2)^2}\left((ab(c_2 + d_2))^2 - (c_2^2Ed_1 + a(c_2 + d_2)^2)Ed_1(\tilde{e}^n)^2\right). \end{array} \tag{29}$$

Of course, it would be extremely messy to directly compare these two payoffs. However, defining Δ_k, $k = c, n$, as

$$\Delta_c = 4(ab)^2[(a + d_2)^2 - 3a\underline{d}_1],$$
$$\Delta_n = (c_2 + d_2)^2\left[(a(c_1 + 2bc_2)^2 + 2c_2^2(a(c_2 + d_2) + c_2d_2))^2 + 4c_2^2((c_1 + 2bc_2)^2 + c_2^2)(a(c_2 + d_2) + c_2d_2)^2\right],$$

we can easily show that:

Proposition 5

- *Suppose that*

$$\underline{d}_1 < \frac{(a+d_2)^2}{3a}, \tag{30}$$

then player 1's payoffs are non-negative at the corner equilibrium if and only if

$$\bar{e} \geq \tilde{e} = \frac{4ab(a+d_2) - \sqrt{(}\Delta_c)}{2((a+d_2)^2 + a\underline{d}_1)} \tag{31}$$

- *Player 1's payoffs are negative at the interior equilibrium if and only if*

$$Ed_1 > \tilde{E}d_1 = \frac{\sqrt{(}\Delta_n) - Ed_1(c_2+d_2)[a(c_1+2bc_2)^2 + 2c_2^2(a(c_2+d_2)+c_2d_2)]}{2Ed_1^2c_2^2[(c_1+2bc_2)^2+c_2^2]}. \tag{32}$$

- *Conditions (30) and (31) together imply that* $\Pi_1^c > \Pi_1^n$.

Interpretation of the conditions and discussion. Basically the conditions above all point to the same requirement: the corner equilibrium not to be too painful to player 1. This is more likely the case when the threshold emission level is not too low, the expected damage under SRM is high enough (in order for player 1 to have an incentive to refrain from emitting too much) the certain damage is low enough etc. Conclusions are quite expected but this deserves further discussion. At least, the results are meaningful.

Finally, we can conclude the analysis by displaying a numerical example that illustrates all of the results above (this has been done already). And we can discuss the implications of the RM option/threat on the outcome of climate negotiations.

If one believes that the situation considered so far, where the North bears all the burden of meeting the constraint $e \leq \bar{e}$, is not relevant because for instance the North has a leader position or higher negotiation power, then we can develop further the analysis by investigating more balanced cases where both regions share the burden, with (un)equal weights. In order to study the corresponding game, we need to switch to the concept of coupled constraint Nash equilibrium (Rosen, 1965) etc.

5 Conclusion

We develop a two-player, two-period game of climate change. In the first period, players non-cooperatively choose their emission level, which yields a private benefit. Aggregate emissions then determines the extent of the environmental damage that is incurred in the second period. In the second period, depending on aggregate emissions, players can decide to undertake solar radiation management (SRM) at some cost. SRM is a means to reduce temperature, but at the same time it is perceived as a risky activity since it potentially involves a series of negative environmental impacts. This is captured by assuming that positive SRM induces a shift from a certain damage to a (larger) expected damage. Solving for the Nash equilibrium in SRM strategies allows us to identify a critical emission threshold that triggers (unilateral) SRM by the player that is the most vulnerable to climate change. Moving back to the first period, we then investigate how the potential deployment of SRM in the future affects current emissions. In particular, we wonder if the threat of SRM may act as a coordination device. For that purpose we compare two alternative scenarios. Either the players commit to meeting the threshold and face the resulting coupled constraint on aggregate emissions. Or they get rid of it and non-cooperatively choose their emissions knowing that this will trigger (unilateral) SRM. Solving for the Nash equilibrium corresponding to the latter situation and the coupled constraint Nash equilibrium for the former, we show that under some conditions players may individually find it optimal to refrain from emitting too much in order to avoid SRM. This conclusion holds true even in the worst scenario in which one player – the one that won't undertake SRM – bears alone the responsibility of meeting (or not) the constraint.

A Appendix

A.1 Proof of proposition 2

The equation $e = N(e)$ must be studied for $e > \bar{e}$. Then, we have that $M(\bar{e}) \leq \lim_{e \to \bar{e}} N(e) = (F')^{-1}(D'_{-i}(\bar{e})) + (F')^{-1}((1 - g'_{-i}(\bar{e}))E[D'_i(\bar{e})])$ is equivalent to (11). And we already imposed $\bar{e} < M(\bar{e})$. Moreover it's straightforward to check that $\lim_{e \to \infty} N(e) < \infty$. This establishes existence. We cannot at first glance conclude that the solution is unique because it's difficult to sign the last term at the numerator of the derivative $N'(e)$ (it involves the third derivative of C and D_{-i}). However, one may note that the derivative is negative if one takes a quadratic cost function, which ensures uniqueness.

A.2 Proof of proposition 5

One can note that Π_1^k, $k = c, n$, is a simple second order polynomial in $\bar{e}$ and $\tilde{e}^n$ respectively. It is then possible to solve the polynomial in $\bar{e}$ to provide a necessary, and a necessary and sufficient, condition for $\Pi_1^c \geq 0$. We basically find an interval of variation of $\bar{e}$, denoted by $[\bar{e}^-, \bar{e}^+]$, with $\bar{e}^- = \tilde{e} > 0$, such that $\Pi_1^c \geq 0$ for all $\bar{e} \in [\bar{e}^-, \bar{e}^+]$. Then it remains to show that there exists a non-empty intersection between this interval and the one imposed in (25), which is easy to do. This means that we can find a combination of parameters such that the statement in Proposition 5 is true. For the interior payoffs, the second polynomial in $\tilde{e}^n$ can be rewritten as a second order polynomial in Ed_1 and we can also identify a necessary and sufficient condition for $\Pi_1^i \geq 0$ etc.

References

[1] Barrett, S. 2008. The incredible economics of geoengineering. Environmental and Resource Economics 39, p. 45-54.

[2] Goes, N, Tuana, M., and Keller, K. 2011. The economics (or lack thereof) of aerosol geoengineering. Climatic Change 109, p. 719-744.

[3] Heutel, G. Moreno-Cruz, J., and Shayegh, S. 2015. Solar geoengineering, uncertainty and the price of carbon. NBER working paper # 21355.

[4] National Center for Atmospheric Research. 2007.

[5] Moreno-Cruz, J. 2015. Mitigation and the geoengineering threat. Resource and Energy Economics 41, p. 248-263.

[6] Robock, A. 2008. 20 reasons why geoengineering may be a bad idea. Bulletin of the Atomic Scientists 64, p. 14-18.

[7] Rosen, J. 1965. Existence and uniqueness of equilibrium point for concave N-person games. Econometrica 33, p. 520-534.

[8] Royal Society (2009)

[9] Moreno-Cruz, J. and Smulders, S. 2010.

[10] Moreno-Cruz , J. and Keith, D. 2013. Climate policy under uncertainty: a case for solar geoengineering. Climatic Change 121, p. 431-444.

[11] Quaas, M., and Rickels, W. 2017. Regulation of Radiation Management: Uncertainty and Incentives. Mimeo.

[12] Quaas, M, Quaas, J., Rickels, W., and Boucher, O. 2017. Are there reasons against open-ended research into solar radiation management? A model of intergenerational decision-making under uncertainty. Journal of Environmental Economics and Management 84, p. 1-17.

[13] Weitzman, M. 2015. A Voting architecture for the governance of free-driver externalities, with application to geoengineering. Scandinavian Journal of Economics 117, p. 1049-1068.