Geoengineering in climate negotiations

Adrien Fabre*& Gernot Wagner†

March 7, 2020

Abstract Geoengineering is the intentional modification of the Earth's climate (to counteract global warming), and little is known about the entire consequences that it entails, which might prove dramatic. Our game theoretic model shows that, if some countries face moderate impacts of a low level of mitigation but fear the uncertain consequences of geoengineering, they could agree for a high mitigation policy in order to avoid the deployment of geoengineering by countries facing disastrous damages of climate change. Thus, the (unused) option of geoengineering helps reaching a welfare-improving subgame perfect equilibrium as compared to the low mitigation one that would prevail without this possibility. Consequently, if research on geoengineering reduces uncertainty and dismisses (resp. confirms) the catastrophic scenario, so that no country fears geoengineering anymore, the equilibrium that prevails is geoengineering (resp. low mitigation), even if this could be suboptimal for aggregate welfare. Thus, our model brings an original argument against research on geoengineering, as the uncertainty that surrounds it can act as a commitment device in climate negotiations.

*Paris School of Economics, Université Paris 1 Panthéon - Sorbonne. adrien.fabre@psemail.eu.
†New York University. gwagner@nyu.edu.

1 Risky geoengineering option can make an ambitious climate mitigation agreement more likely

Game theory helps frame the strategic interactions among countries (Schelling, 1990), with global climate negotiations as a prominent example (Barrett, 2006; MacKay et al., 2015). That holds for multilateral negotiations under the umbrella of the United Nations Framework Convention on Climate Change as well as for bilateral negotiations and those among smaller groups of countries. Science points to the need for ambitious climate action (IPCC, 2018). However, assuming symmetric players in a weakest-link game, as long as one country, or group of countries, prefers high mitigation (H) to low (L) while the other prefers L to H, an ambitious agreement (H) is unlikely.

1.1 Simple game theory explains lack of ambitious climate mitigation

Consider two countries or groups of countries, 1 and 2. Both countries would have to prefer H to L for H to prevail (Table 1) (Nash, 1951). This holds as long as the aggregate level of mitigation is characterized by the minimum of each country's mitigation policy, a stylized representation of the arithmetic of mitigation focused on strategic interaction among symmetric players (MacKay et al., 2015).

Table 1: Climate resulting from each pair of moves. High mitigation agreement (H) is only possible if both players choose H over low (L) mitigation.

Moves by players 1 \ 2	H_2	L_2
H_2	H	L
L_2	L	L

This weakest-link game representation stands in contrast to models of total mitigation as the sum of each country's contribution (Barrett, 2010; Sandler, 2018). It also assumes that mitigation is costly, with H costlier than L. Mitigation might not be as costly as often assumed (Gillingham & Stock, 2018), leading, for example, to some countries pursuing ambitious climate action on their own (Wagner & Weitzman, 2016), or it might be in some countries' self-interest to pursue more ambitious mitigation policies to coax others into doing the same (Kotchen, 2018; Nordhaus, 2015). Nordhaus (2015), in particular, points to how powerful countries taking ambitious climate mitigation action could force others into following suit. Either possibility may indeed help break through the global climate policy logjam. Failing that, Table 1 presents the simplest possible explanation for the woeful lack of ambitious climate action to date. Any player would only play H, if H is also the preferred outcome of the other player(s). Otherwise the one player choosing H would incur the extra mitigation cost in vain. This is often described as the "free-rider" problem, the direct result of climate mitigation's public goods nature (Baumol, 1952).

1.2 Geoengineering is cheap and imperfect, and all-too tempting

In contrast to mitigation, geoengineering (G) is often characterized as fast, cheap, and imperfect (Keith, 2000). Here we drop "fast" and focus on the latter two. In fact, while G is often synonymous with solar radiation management or modification (SRM), with carbon dioxide removal (CDR) instead best thought as expensive mitigation (Moreno-Cruz et al., 2018), here some CDR techniques could count, too. For example, ocean fertilization in an attempt to remove carbon from the atmosphere might be "slow" yet is also both cheap and imperfect. The combination makes it risky, widening the list of possible G techniques relevant to this analysis.

Modelling G necessitates extending our game to two periods: first, players choose mitigation levels H or L; second, they decide on G. The latter is simply a decision of whether to pursue G or not. G is so cheap that the largest hurdle is the willingness to engage in it in the first place. G is also so powerful that one country's choice determines the overall outcome. While for mitigation, global ambition here equals the minimum level of ambition across the two players, for G it only takes one player to pursue it for the entire world to experience the outcome—both its potential (net) benefits, and the potentially large risks. All of it figures into the decision to pursue G. Moreover, the aggregate level of mitigation is irrelevant to a geoengineered climate, so that the climate resulting from the players' moves across both periods can be characterized by H, L or G. Conversely, the preferences of each player can be reduced to their rankings over possible resulting climates H, L, or G.

Table 2 summarizes the impact of G on the negotiated outcome: As in Table 1, H only happens if both players prefer it. Meanwhile, either player resorting to G implies the entire world facing its consequences. These properties not least explain G's "free-driver" properties (Schelling, 1992; Victor, 2008; Wagner & Weitzman, 2012; Weitzman, 2015). G is at once so cheap and powerful that a single player's preferences dominate the outcome, regardless of how many other players there are.

Table 2: Climate resulting from each pair of moves. Low mitigation (L) without geoengineering (G) dominates high mitigation (H) (Table 1), while G dominates all.

Moves by players 1 \ 2	H_2	L_2	G_2
H_2	H	L	G
G_2	L	L	G
L_2	G	G	G

1.3 Simple game theory of how availability of G can induce H

If all players agreed on either of the three possible negotiation positions as their top choice, the outcome would be just that, shown in grey in Table 3. If any player preferred G to all other choices, G would win (shown in italics). That leaves four cases.

Table 3: Climatic outcomes depending on each player's full set of preferences. Availability of geoengineering (G) could lead to high mitigation agreement (H, in bold). Symmetric outcomes are omitted for simplicity.

1 \ 2	H > L > G	H > G > L	L > G > H	L > H > G	G > L > H	G > H > L
H > L > G	H					
H > G > L	H	H				
L > G > H	L	G	L			
L > H > G	L	**H**	L	L		
G > L > H	*G*	*G*	*G*	*G*	*G*	
G > H > L	*G*	*G*	*G*	*G*	*G*	*G*

In all four, one player ranks L as the top choice, while the other prefers H above all else. That situation mimics the "simple" negotiation, shown in Table 1, save for the addition of G. If G were ranked last by both, the game would collapse to the simple negotiation, where one ranks L > H and the other ranks H > L. G does not interfere. L wins. Conversely, if G were to be ranked second by both players, G would win. Two cases remain.

If the player who prefers L ranked G > H while the other player who prefers H ranked L > G, L would win. G does not alter the outcome, as both players prefer L to G. That, however, would imply that those most hurt by climate change, and thus most interested in pursuing H, would rank L > G. That might be a justifiable ranking, given G's inherent risks. But desperation begets risky choices.

Here this willingness on the part of those preferring H > L to engage in risky choices to avoid the worst implies that those same players might also rank G > L. G is not the first choice, but it might be better than L: H > G > L, for those most vulnerable to a low mitigation climate. The fundamental assumption here—and this is indeed an assumption—is that H precludes G. Ambitious mitigation is sufficient to avoid having any player want to play G: H > G for any player. That assumes G would only be used in an emergency setting, as a last-ditch effort to avoid bad climatic outcomes due to low mitigation efforts L (Victor et al., 2009; Weitzman, 2015). While not unusual to assume as much, it does preclude using moderate G as part of an overall climate policy portfolio (Moreno-Cruz et al., 2018). This assumption immediately leads to H > G > L for those ranking H > L. It also leads to L > H > G for the second type of player who ranks L > H.

These rankings, in turn, immediately lead to the final case, where the availability of G induces H (in bold, in Table 3). It is also the only situation with H as the outcome despite only one player, or group of players, ranking H first. The threat of G leads both players to undertake H, although one apparently faces higher mitigation costs than direct benefits from avoided climate damages.

1.4 Discussion

We are not the first to analyze the impact of the mere availability of G on mitigation outcomes. Some have focused on G crowding out ambitious mitigation, necessitating complex governance regimes in their own right (Quaas et al., 2017; Weitzman, 2015). This often comes under the heading of "moral hazard" (Keith, 2000; Wagner & Zizzamia, 2019). Indeed mere mention of G might lessen a single actor's willingness to mitigate, or it might do the opposite (Merk et al., 2016). Others have indeed argued how the threat of G, coupled with G's inherent uncertainties, can lead to more ambitious mitigation action within and across countries (Heyen, 2015; Moreno-Cruz, 2015; Victor et al., 2009). (The possibility of "counter-geoengineering," too, might play this role of helping catalyze H (Bas & Mahajan, 2019; Heyen et al., 2019).) Here we boil this reasoning down to its bare essentials, focusing on simple, relative preference ordering of each player over climate outcomes H, L, and G. These players could be singular actors in bilateral negotiations. They could also stand in for groups of players. G's fundamental characteristics might lead players previously stuck in a low-mitigation equilibrium L to consider H.

For G to induce H, the threat of deploying G must be credible. And it must indeed be a threat—something opting for H can avoid. All that relies on G being risky and shrouded in uncertainty. The more research discovers about G, decreasing its potential risks and uncertainties, the less it might be able to induce H. Conversely, the possibility of G deployed as a supplement to H might have similar effects of helping induce H, but it may well be too tempting to trade off some mitigation action for moderate geoengineering (Moreno-Cruz et al., 2018).

Lastly, if any player—a singular state, for example, tempted by G to maintain the high-emitting status quo—were to prefer G altogether, the free-driver effect kicks in and G prevails. The only possible solution in such a scenario to prevent the free driver from forcing G upon everyone else: strong governance at the global level, effectively leading to a ban on unilateral G.

The availability of G is no guarantee of inducing H. Viewed in the right light, and with the bare minimum of oft-invoked assumptions, it might help break the decades-long logjam of climate negotiations—both formally at the international level, and informally among smaller groups of state and possibly non-state actors.

A The formal model

A.1 Description of the model

Let us consider the following two-period two-player game, where it is common knowledge that both players are rational. In first period, each player i chooses a level of mitigation which is high (H) or low (L, with $L < H$), while in second period they choose whether to undertake geoengineering (G) or not ($\varnothing$). Let $a_i = (m_i, g_i) \in \mathcal{A} := \{L; H\} \times \{\varnothing; G\}$ be the action of player i. The resulting climate (at the end of period 2) $C \in \mathcal{C} := \{L; H; G\}$ is set equal to G if $\cup_i g_i = G$ and to the aggregate level of mitigation $\min_i \{m_i\}$ otherwise. In other terms, the climate is characterized by geoengineering as soon as one player deploys it, and by the level of mitigation otherwise, given that a high aggregate level of mitigation can only be attained if both players undertake it. Thus, an action of type $(\cdot, G)$ can be identified with G while an action of the form $(m_i, \varnothing)$ can be identified with m_i. Table 4 summarizes this characterization, identifying an action with H, L or G accordingly.

Table 4: Climate resulting from each state.

$a_1 \backslash a_2$	H	L	G
H	H	L	G
L	L	L	G
G	G	G	G

The estimated probability of a catastrophic impact of geoengineering is common knowledge, and can take three values: $p \in \mathcal{P} := \{0; \pi; 1\}$, with $\pi \in (0; 1)$ a small probability. Depending on whether p equals 0, π, or 1, we say that the *situation* is one where geoengineering is *secure*, *risky* or *catastrophic*. Furthermore, we denote by $V_i : \mathcal{A} \times \mathcal{C} \to \mathbb{R}$ player i's payoff, which is common knowledge. Finally, let us denote by $W = \sum_i n_i V_i$ the social welfare, where $n_i > 0$ is the weight given to player i's welfare.

For simplicity, we assume that the players are risk-neutral (as this does not affect the results), and that costs add up so that $V_i = H_i 1_{a_i=H} + \epsilon 1_{g_i=G} + L_i 1_{C=L} + G_i^p 1_{C=G}$. With this specification, we express the payoffs as a departure from the reference climate H. In addition to (player-specific) payoffs relative to the climate, we consider that two actions are costly: high mitigation (whose cost is player-specific) and geoengineering (with a common cost $-\epsilon$ which should be viewed as small). We interpret all the parameters (i.e. H_i, L_i, G_i^p and ϵ) as negative, although in principle our model allows them to take any value. We simply impose that V_i is injective (i.e. that payoffs are distinct) to avoid multiple equilibria. In particular, this implies $\epsilon < 0$ and allows to know what player is responsible for G. Otherwise with $\epsilon = 0$, one player knowing that the other will play G would be indifferent in playing G as well, and new non-intuitive equilibria would appear.

A.2 Simplification of the problem

In the following, we take the situation p as given and thus forget the exponent p in G_i^p.

A player j would play H only if $H_i > G_i > L_i$ for all i (otherwise it is clear that some player will not play H and that j will incur the cost of mitigation in vain). But then, if $H_i > G_i > L_i$ for all i, it is an optimal and stable strategy for all players to play H. Each player then obtains their maximal payoff by playing $\varnothing$ in period 2. Hence, no player ever plays (H, G) in our setting with complete information. Thus, we ignore such actions in the following and identify one's action with H, L or G.

Table 5summarizes the payoffs of both players. As it is suboptimal that a player plays H while the other does not play H, such actions are put in parentheses and can be dismissed for the derivation of the equilibrium. To the extent that $-\epsilon$ is small enough,[1] Tables 4 and 5 show that i's payoffs can be characterized by the climate (as actions in parentheses will never been played), meaning that i's preference ordering between climates H, L and G is the same as the ordering between payoffs H_i, L_i and G_i, and does not depend on i's action.

Table 5: Payoffs of player i. Payoffs in parentheses correspond to suboptimal actions.

$a_i \backslash a_{-i}$	H	L	G
H	H_i	$(H_i + L_i)$	$(H_i + G_i)$
L	L_i	L_i	G_i
G	$G_i + \epsilon$	$G_i + \epsilon$	$G_i + \epsilon$

B Solution

Now that the problem has been simplified, let us derive the subgame perfect equilibrium, depending on the payoffs.

Table 5 shows that when $G_i + \epsilon > H_i, L_i$, it is a dominant strategy for i to play G, unless $-i$ already plays G. In the latter case, both players want the geoengineered climate G yet prefer that the other pays for the cost of deployment. This is a chicken game where the optimal strategy is that both players play a correlated equilibrium in period 2: they agree on the player who will deploy geoengineering (and incur ϵ) by flipping a coin. In any case, $G_i + \epsilon > H_i, L_i$ leads to the climate G. Thus, mitigation cannot occur if geoengineering is the best alternative for one player.

Putting all this together, we can infer the subgame perfect equilibrium in function of the payoffs; we report the resulting climates in Table 6. If one's player's preferred climate is G, they imposes G to the other player (these cases correspond to the red cells). If both players have the same preferred climate, they have no incentive to deviate from the the state where they both play this preferred action (green cells). If both players prefer L to G, none of them have any interest in playing G, and L will be played by both as long as one of them does not prefer H to L (italic cells). In the case where one player prefers both L and G to H, they will not play H; so that G will be the outcome if at least one of the player prefers G to L.

Only one last case remains: $L_i > H_i > G_i$ while $H_{-i} > G_{-i} > L_{-i}$ (bold cells). Both players prefer H to G, but differ in that they rank L first or last. G is preferred to L by one player (and H is not unanimous), but the other player can avoid G (its worst outcome) by playing the best outcome of the other player: H. In this case, the fear of geoengineering leads both player to undertake a high mitigation, although one of them faces higher mitigation costs than benefits from avoided damages.

Table 6: Best equilibrium climate depending on each player's preferences.

$1 \backslash 2$	$G_2 > L_2, H_2$	$L_2 > H_2 > G_2$	$L_2 > G_2 > H_2$	$H_2 > G_2 > L_2$	$H_2 > L_2 > G_2$
$G_1 > L_1, H_1$	G	G	G	G	G
$L_1 > H_1 > G_1$	G	*L*	*L*	**H**	*L*
$L_1 > G_1 > H_1$	G	*L*	*L*	G	*L*
$H_1 > G_1 > L_1$	G	**H**	G	H	H
$H_1 > L_1 > G_1$	G	*L*	*L*	H	H

[1]More precisely, this reasoning fails when $-\epsilon > \min_i \{H_i, L_i\} > 0$ as this implies $G_i + \epsilon < L_i < G_i$ or $G_i + \epsilon < H_i < G_i$.

References

S. Barrett. *Environment and Statecraft: The Strategy of Environmental Treaty-Making*. OUP Oxford, Oxford ; New York, 2006. ISBN 978-0-19-928609-6. 2

S. Barrett. *Why Cooperate?: The Incentive to Supply Global Public Goods*. OUP Oxford, Oxford ; New York, 2010. ISBN 978-0-19-958521-2. 2

M. Bas & A. Mahajan. Security implications of geoengineering and counter-geoengineering. 2019. 4

W. J. Baumol. Welfare Economics and the Theory of the State. In C. K. Rowley & F. Schneider, editors, *The Encyclopedia of Public Choice*. Springer US, Boston, MA, 1952. ISBN 978-0-306-47828-4. 2

K. Gillingham & J. H. Stock. The Cost of Reducing Greenhouse Gas Emissions. *Journal of Economic Perspectives*, 2018. 2

D. Heyen. *Five Essays in the Economics of Climate Engineering, Research, and Regulation under Uncertainty*. Dissertation, 2015. 4

D. Heyen, J. Horton, & J. Moreno-Cruz. Strategic implications of counter-geoengineering: Clash or cooperation? *Journal of Environmental Economics and Management*, 2019. 4

IPCC. Global Warming of 1.5 °C. Technical report, 2018. 2

D. W. Keith. Geoengineering and the climate: History and Prospect. *Annual Review of Energy and the Environment*, 2000. 2, 4

M. J. Kotchen. Which Social Cost of Carbon? A Theoretical Perspective. *Journal of the Association of Environmental and Resource Economists*, 2018. 2

D. J. C. MacKay, P. Cramton, A. Ockenfels, & S. Stoft. Price carbon - I will if you will. *Nature*, 2015. 2

C. Merk, G. Pönitzsch, & K. Rehdanz. Knowledge about aerosol injection does not reduce individual mitigation efforts. *Environmental Research Letters*, 2016. 4

J. Moreno-Cruz, G. Wagner, & D. Keith. An Economic Anatomy of Optimal Climate Policy. SSRN Scholarly Paper ID 3001221, Social Science Research Network, Rochester, NY, 2018. 2, 4

J. B. Moreno-Cruz. Mitigation and the geoengineering threat. *Resource and Energy Economics*, 2015. 4

J. Nash. Non-Cooperative Games. *Annals of Mathematics*, 1951. 2

W. Nordhaus. Climate Clubs: Overcoming Free-Riding in International Climate Policy. *American Economic Review*, 2015. 2

M. F. Quaas, J. Quaas, W. Rickels, & O. Boucher. 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*, 2017. 4

T. Sandler. Collective action and geoengineering. *The Review of International Organizations*, 2018. 2

T. Schelling. *The Strategy of Conflict*. Harvard University Press, Cambridge, Mass., 2nd edition edition, 1990. ISBN 978-0-674-84031-7. 2

T. C. Schelling. Some Economics of Global Warming. *The American Economic Review*, 1992. 3

D. G. Victor. On the regulation of geoengineering. *Oxford Review of Economic Policy*, 2008. 3

D. G. Victor, M. G. Morgan, F. Apt, & J. Steinbruner. The Geoengineering Option - A Last Resort against Global Warming. *Foreign Affairs*, 2009. 4

G. Wagner & M. L. Weitzman. Playing God. *Foreign Policy*, 2012. 3

G. Wagner & M. L. Weitzman. *Climate Shock: The Economic Consequences of a Hotter Planet*. Princeton University Press, reprint edition edition, 2016. 2

G. Wagner & D. Zizzamia. Green Moral Hazards. SSRN Scholarly Paper ID 3486990, Social Science Research Network, Rochester, NY, 2019. 4

M. L. Weitzman. A Voting Architecture for the Governance of Free-Driver Externalities, with Application to Geoengineering: A voting architecture for the governance of free-driver externalities. *The Scandinavian Journal of Economics*, 2015. 3, 4