PhD Programme:

Science and Management of Climate Change

PhD Dissertation

Adaptation and Mitigation in the Context of International Environmental Agreements

Strategic Interactions and Effects on Negotiations' Outcome

SSD: SECS-P/01

PhD Coordinator
Prof. Carlo Carraro

Supervisor
Prof. Francesco Bosello

PhD Candidate
Francesco Furini
Student ID: 956308

Abstract

This research focuses on the analysis, from a game theoretical perspective, of International environmental agreements in the presence of adaptation. Despite its private good nature, adaptation plays a crucial role in climate agreements negotiation because of its strategic relation with mitigation. For this reason, it is very important to include both strategies in IEAs modeling.

The dissertation is a collection of three papers which expand the existing literature on IEAs in different directions: 1) the standard mitigation-adaptation game (M+A-Game) is analysed in a Stackelberg scenario; 2) the strategic relation between mitigation and adaptation and its effect on climate negotiation is analysed assuming that mitigation, attenuating climate change damages, can also affect the effectiveness of adaptation; 3) the existing theoretical results are tested through an Integrated Assessment Model (IAM) application.

The strategic relation between mitigation and adaptation, the effect of adaptation on mitigation strategies and on negotiation's outcome are analysed. Successful climate cooperation requires both large stable coalitions and high welfare improvements with respect to non-cooperation. The paradox of cooperation persists in most of the game configurations considered. Optimistic results arise only in a situation in which strategic complementarity holds both in mitigation and mitigation-adaptation space.

Abstract (Italiano)

L'obiettivo di questa ricerca è analizzare, applicando la teoria dei giochi, la formazione di accordi internazionali sul clima in presenza di adattamento. Nonostante la natura di bene privato dell'adattamento, il suo ruolo è fondamentale nell'ambito della negoziazione climatica per via della interdipendenza strategica con la mitigazione. Per questo motivo, è molto importante includere entrambe le strategie nella modellizzazione di accordi internazionali sul clima.

La tesi è composta da tre articoli che espandono lo stato dell'arte della letteratura sugli accordi internazionali sul clima in queste direzioni: 1) il gioco con mitigazione e adattamento (M+A-Game) è analizzato in un contesto di Stackelberg leadership; 2) la relazione strategica tra mitigazione e adattamento e le conseguenze sulla negoziazione climatica sono analizzate assumendo che la mitigazione, limitando gli impatti ambientali futuri, è in grado di influenzare l'efficacia dell'adattamento; 3) i risultati teorici esistenti sono testati attraverso l'applicazione di un *Integrated Assessment Model (IAM).*

Il focus dell'analisi è sulla relazione strategica tre mitigazione e adattamento, sull'effetto dell'adattamento sulle strategie di mitigazione e sul risultato della negoziazione Il successo della cooperazione climatica richiede un'ampia partecipazione che porti a sostanziali miglioramenti nel benessere sociale rispetto ad una situazione non cooperativa. Il paradosso della cooperazione persiste nelle molteplici configurazioni analizzate, e solo in una situazione in cui sia le strategie di mitigazione sia mitigazione e adattamento sono complementi strategici le previsioni sui risultati della negoziazione sono ottimiste.

Table of Content

EXECUTIVE SUMMARY

Public debate on climate change and on the need for immediate and effective solutions is becoming increasingly central. International negotiations for the reduction of greenhouse gas (GHG) emissions, within the annual UNFCCC Conferences of Parties (COPs), are attracting the interest of medias and public opinion. Despite the increasing pressures for the implementation of effective measures against climate change, and despite decades of negotiations, the solution still appears far and difficult to achieve.

To support and explain the reasons behind climate negotiation failures, economic science applies a game theoretical approach to model International Environmental Agreements (IEAs) games. Starting from the early 90's, IEAs games have largely explained the obstacles to the formation of a large and effective agreement. The first strand of this literature has modeled climate negotiation considering only mitigation, i.e., the reduction of GHG emissions to limit future climate change, while omitting the possibility for players to implement adaptation, i.e., measures that combat and reduce the impacts of climate change. The reason behind this modeling choice is that mitigation represents the objective of cooperation: it is a public good which generates a positive externality. With cooperation, members of the coalition internalize the positive externality and provide the optimal abatement level. These models have widely shown the difficulties in reaching an effective agreement. This result is known as the paradox of cooperation: stable coalitions for the reduction of GHG emissions are either small or, if large participation is achieved, it only brings small welfare improvements from non-cooperation. Hence, cooperation is reached only when it is least needed.

Over the last decade, a new strand of IEAs games literature has emerged. In addition to mitigation, adaptation is introduced as a policy strategy to combat climate change. Despite its private good nature, adaptation plays a crucial role in climate agreements negotiation because of its strategic relation with mitigation. The few existing models of IEAs in the presence of adaptation find optimistic conclusions in terms of cooperation. After the introduction of adaptation, players' mitigation strategies can become strategic complements. If the strategic relation between adaptation and mitigation is strong enough, mitigation strategies are no longer substitutes. This fact neutralizes the leakage effect, reduces free riding incentives and increases participation.

This work has the intent of further exploring IEAs in the presence of adaptation. On the one hand it aims to extend existing theoretical results, on the other hand these results are tested through an application of an Integrated Assessment Model (IAM).

With the standard game theoretic approach, the analyses performed in a pure mitigation context are conducted in a game considering both mitigation and adaptation. The paradox of cooperation is tested both in a Nash-Cournot and in a Stackelberg scenario. Furthermore, the strategic relation between mitigation and adaptation is further analysed considering conditions that could lead to strategic complementarity between these two policy options.

With an Integrated assessment approach, the existing theoretical results are tested considering real aggregate data on 6 different macro regions. Strategic interactions and coalition stability are analysed in a complex and asymmetric context.

The dissertation is a collection of three papers (chapters) which expand the existing literature on IEAs in different directions: 1) the standard mitigation-adaptation game (M+A-Game) is analysed in a Stackelberg scenario; 2) the strategic relation between mitigation and adaptation and its effect on climate negotiation is analysed assuming that mitigation, attenuating climate change damages, can also affect the effectiveness of adaptation; 3) the existing theoretical results are tested through an integrated assessment model application.

The first Chapter is a paper co-authored with Professor Michael Finus. It analyses whether and how Stackelberg leadership of signatories affects the formation of stable climate agreements in a mitigation (M-) and in a mitigation-adaptation (M+A-) game with symmetric players. Larger coalitions are obtained when mitigation reaction functions are downward sloping (always the case in the M-Game, and one possibility in the M+A-Game). However, wide participation only leads to low welfare improvements. When mitigation reaction functions are upward sloping (only possible in the M+A-Game), Stackelberg leadership turns to be an obstacle for cooperation. Stackelberg leadership is not able to improve upon the Nash-Cournot scenario, and the paradox of cooperation is persisting also in the M+A-Game.

Chapter two is a paper co-authored with Professor Francesco Bosello. It considers a standard IEAs game with symmetrical players and assumes a double mitigation-adaptation relation. The common assumption that higher mitigation decreases the marginal benefit of adaptation and vice versa is enriched allowing for the possibility that mitigation, leading to lower and more manageable damages, determines a greater effectiveness of adaptive measures. With the additional assumption, complementarity between mitigation and adaptation is possible. Upward sloping mitigation reaction functions are less likely to occur, but still possible and, if this is the case, the grand coalition can form. Nonetheless, large participation can induce substantive welfare gains, avoiding the paradox of cooperation, only if adaptation and mitigation are strategic complements.

In the third and last chapter, an updated version of the RICE model is applied to test the main theoretical results on the introduction of adaptation in IEAs. Adaptation and mitigation are found to be strategic substitutes and, after the introduction of adaptation, individual mitigation (emissions) levels can be complements. Adaptation also favours cooperation, but it does not lead to large participation. While in the M-Game no internally stable coalitions form, the M+A-Game leads to 2-players (out of 6) internally stable coalitions. Allowing for optimal transfers between coalition members, in both games full cooperation is achieved. However, the positive effect of adaptation on coalition stability can still be found looking at free riding incentives. Adaptation reduces (or eliminate) the leakage effect and weakens the incentives to deviate from the agreement.

Further extensions are possible, and already scheduled for future research work, both from a game theoretic and an integrated assessment perspective. The standard adaptation-mitigation game can be analysed by assuming that adaptation is chosen before mitigation to test the strategic role of anticipatory adaptation. For what concerns the IAM part, the last chapter of this dissertation still requires additional work to lead to more robust results. Furthermore, IAMs could be also applied to test the theoretical results of the first two papers of this dissertation. Stackelberg leadership and more complex adaptation-mitigation interconnections can be modeled to see their effects in presence of asymmetries and intertemporal maximization processes.

CHAPTER ONE

The (Un)importance of Stackelberg Leadership for the Formation of (Un)successful International Climate Agreements*

Abstract

We analyze in a simple game-theoretic model whether and how Stackelberg leadership of signatories affects the formation of stable climate agreements in a mitigation (M-Game) and in a mitigation-adaptation game with symmetric players. We show generally that stable coalitions are larger under the Stackelberg scenario than under the Nash-Cournot scenario in the M-Game and in the M+A-Game if reaction functions in mitigation space are downward sloping. In the M+A-Game, if reaction functions are upward sloping, this relation is reversed. In order to evaluate outcomes, we contrast the total potential gains from cooperation with the gains achieved by stable coalitions. This allows testing for Barrett's paradox of cooperation as established for the M-Game in Barrett (1994), and later reiterated by many others: stable coalitions are either small or if they are large, the potential gains from cooperation are small. We show that this paradox generally carries over to the M+A-Game under the Stackelberg scenario. This is also true in the M+A-Game under the Nash-Cournot assumption, except if, apart from upward sloping reaction functions in mitigation space, mitigation and adaptation are complements and not as commonly assumed substitutes. Thus, our results neither support the expectation that Stackelberg leadership nor the inclusion of adaptation in climate change negotiations as emerges from Bayramoglu et al. (2018) will cut through the Gordon node of unsuccessful climate agreements.

Keywords: Climate change, mitigation-adaptation game, international environmental agreements, Stackelberg leadership

* This paper, co-authored with Professor Michael Finus, is submitted, and under revision, in The Journal of Environmental Economics and Management (JEEM).

1. Introduction

Mitigation and adaptation are two strategies to combat climate change. Mitigation directly targets the cause of the problem, reducing greenhouse gases emissions, causing global warming. Instead, adaptation aims at ameliorating the negative consequence of global warming. Whereas mitigation is typically viewed as a pure public good, adaptation is seen as a private good (reducing only damages of the party conducting adaptation). Addressing global warming requires international cooperation: isolated actions will not make a big difference if other countries do not follow suit. However, the signature and ratification of effective international climate agreements have proved to be difficult in the past. There is a widespread consensus that the Kyoto Protocol has not been able to curb the increase of greenhouse gases in the past, and also most scholars have doubts about the effectiveness of the Paris Accord signed in 2015 as highlighted by the latest IPCC 1.5 degrees report (IPCC 2018). As the effects of global warming become more and more visible, adaptation becomes increasingly important as a policy option. This is not only evident by the increasing literature on the costs and effectiveness of adaptation as well as about the practical and technical obstacles of implementation, in particular, in developing countries (IEG 2013 and World Bank 2010), but adaptation is also an integral part of almost all recent climate change negotiations (UNFCCC 2014 and 2016). The main obstacle of addressing the cause of global warming is the public good nature of mitigation. Reducing emissions comes at a cost that is borne by individual countries, but the benefits are enjoyed by all countries worldwide. This free-rider incentive structure is certainly amplified by policy makers' myopia, focusing on the short-term cost of mitigation and discounting the future benefits of reduced climate damages.

International climate negotiation failures have been largely explained by game-theoretic models of international environmental agreements (IEAs).[1] In the standard workhorse model with only

---

[1] The first models go back to Barrett (1994), Carraro and Siniscalco (1993) and Hoel (1992). This literature on IEAs has grown substantially over recent years. A collection of the most influential

mitigation and symmetric players, i.e., models in which mitigation is the only strategy to address global warming and in which all countries have the same (strictly concave) payoff function, only small agreements are stable if signatories and non-signatories choose their mitigation levels simultaneously, which has been called the Nash-Cournot scenario. For Stackelberg leadership of signatories, more optimistic results have been obtained in terms of the size of stable agreements (Barrett 1994; Diamantoudi and Sartzetakis 2006; Rubio and Ulph 2006). However, as Barrett (1994) coined it, the paradox of cooperation persists: stable coalitions are either small or if they are large, the potential gains from cooperation are small. Recently, Bayramoglu et al. (2018) showed for the Nash-Cournot scenario that more optimistic results may be obtained if countries have a second strategy at their avail, namely adaptation, which they coined the mitigation-adaptation game. Based on insights from Ebert and Welsch (2011 and 2012) in the context of two countries, they show that in such an extended game with *n* players and the possibility to form coalitions, mitigation levels in different countries may no longer be strategic substitutes but may become complements if the cross effect between mitigation and adaptation is sufficiently strong. They demonstrate that with a complementary relationship of mitigation levels, reaction functions are no longer downward sloping but are upward sloping in mitigation space. Most importantly, with upward sloping reaction functions, larger agreements are stable, irrespective whether mitigation and adaptation are substitutes (as commonly believed) or complements (as an unlikely but possible option according to Ingham et al. 2013). Overall, it appears that Bayramoglu et al. (2018) derive a more optimistic conclusion regarding the prospects of cooperation if adaptation is added as a second strategy to an IEA-game, provided cross effects between mitigation and adaptation are sufficiently strong.

---

articles have been collected in a volume by Finus and Caparros (2015), including a survey in the introduction to this volume. In this volume, various extensions of the standard model are included for which in some cases more positive results are obtained. The importance of this topic is also highlighted by some of the finest recent papers, e.g., Battaglini and Harstad (2016) and Harstad (2012).

In this paper, we introduce Stackelberg leadership in a mitigation-adaptation game as Eisenack and Kähler (2016) have done for two players, but extend the analysis to n players and coalition formation.[2] In order to allow for a direct comparison between the mitigation and the mitigation-adaptation game as well as across the two scenarios, Nash-Cournot and Stackelberg leadership, we employ the setting of Bayramoglu et al. (2018). We address two research questions in this paper.

1) Does Stackelberg leadership improve over the Nash-Cournot scenario? We provide a general proof that stable coalitions are larger if reaction functions in mitigation space are downward sloping, which is always the case in the mitigation game and is one option in the mitigation and adaptation game. However, the reverse is true if reaction functions are upward sloping, which is another option in the mitigation-adaptation game. Importantly, if stable coalitions are larger under the Stackelberg than under the Nash-Cournot scenario, improvements in terms of global welfare are (very) small. If this relationship is reversed, i.e., stable coalitions are smaller under Stackelberg leadership than under the Nash-Cournot scenario, Nash-Cournot leads usually to better outcomes.

2) Does the paradox of cooperation as established by Barrett (1994) for the M-Game and later iterated by many others also hold for the M+A-Game? We show that for the Stackelberg scenario this paradox directly carries over without qualification in the M+A-Game. For the Nash-Cournot scenario we come to a less positive conclusion than Bayramoglu et al. (2018). Only if, apart from upward sloping reaction functions in mitigation space due to strong cross effects between mitigation and adaptation, mitigation and adaptation are complements, the paradox will disappear, otherwise it persists.

In what follows, we lay out the model in section 2, derive our results in section 3 and conclude in section 4 with some hints about future research. Section 3 derives first some general results that help

---

2 There is a long tradition of economic applications of Stackelberg leadership. See for instance Basu and Singh (1990), Endres (1992), Gal-Or (1985) and Vickers (1985). In particular, including the possibility of Stackelberg leadership in IEAs modeling is crucial to understand some of the mechanisms behind climate negotiations.

to understand the basic driving forces and incentive structure for coalition formation across the two games and two scenarios and then discusses some further interesting properties based on simulations, which are reported in Appendix A.6. All proofs are contained in the Appendix, A.1 to A.5.

2. The Model

2.1 Payoff Functions

We consider n symmetric countries $i = 1, 2, ..., n$, with N the set of all countries. We compare two different games. In the Mitigation Game (M-Game), countries have only mitigation as a strategy to combat climate change, whereas in the Mitigation-Adaptation Game (M+A-Game), they also have adaptation as a second strategy.

Following Bayramoglu et al. (2018), the payoff function of every country i in the M-Game is given by:

$$w_i(M, m_i) = B_i(M) - C_i(m_i) \tag{1.a}$$

whereas in M+A-Game it is given by:

$$w_i(M, m_i, a_i) = B_i(M, a_i) - C_i(m_i) - D_i(a_i). \tag{1.b}$$

In the M-Game, the individual payoff comprises benefits B_i, which are a function of total mitigation, $M = \sum_{i=1}^{n} m_i$, minus the cost C_i, which is a function of individual mitigation m_i. In the M+A-Game, benefits are a function of both strategies, total mitigation M and individual adaptation a_i. Costs comprise mitigation cost $C_i(m_i)$ and adaptation cost $D_i(a_i)$ where the latter cost is a function of individual adaptation a_i. Both, mitigation, the pure public good, as well as adaptation, the pure private good, contribute to benefits.[3]

The strategy space of country i is given by $m_i \in \left[0, \overline{m_i}\right]$ and $a_i \in \left[0, \overline{a_i}\right]$. If we assume $w_i\left(M, m_i, a_i = 0\right) = w_i\left(M, m_i\right)$, both games are directly comparable. Moreover, we assume that all countries have the same payoff function, i.e., all countries are assumed to be ex-ante symmetric. Hence, we can drop index i, whenever no misunderstanding is possible. However, as will become clear below, countries may nevertheless be ex-post asymmetric as in our model countries endogenously choose whether they join an agreement and become signatories (S) or remain outside and become non-signatories (NS), as these groups choose different mitigation levels. If we want to stress this difference, we use subscript S and NS, respectively.

All welfare functions, as well as their first and second derivatives, are assumed to be continuous. Following Bayramoglu et al. (2018), we introduce the following assumptions, with the understanding that assumptions a) and b) apply to both games whereas the remaining assumptions apply only to the M+A-Game. In terms of notation, we denote for instance $B_M = \partial B / \partial M$, $B_{MM} = \partial^2 B / \partial^2 M$ and $B_{Ma} = B_{aM} = \partial^2 B / \partial M \partial a_i$.

[3] It is generally known that the public good provision game can be alternatively framed as an emission game; they are dual problems. In the context of mitigation and adaptation, this is evident by comparing Bayramoglu et al. (2018) and Rubio (2018). In the emission game, the equivalent to the benefit function in the public good game is the damage function with aggregate emissions and adaptation being the arguments in this function.

General Assumptions I

Both Games

a) $B_M > 0$, $B_{MM} < 0$, $C_m > 0$, $C_{mm} > 0$.

b) $\lim_{M \to 0} B_M > \lim_{m \to 0} C_m > 0$.

M+A-Game

c) $B_a > 0$, $B_{aa} \le 0$, $D_a > 0$, $D_{aa} \ge 0$.

If $B_{aa} = 0$, *then* $D_{aa} > 0$ *and vice versa: if* $D_{aa} = 0$, *then* $B_{aa} < 0$.

d) $\lim_{a \to 0} B_a > \lim_{a \to 0} D_a > 0$

e) i) $B_{aM} = B_{Ma} < 0$ *or ii)* $B_{aM} = B_{Ma} > 0$.

These assumptions and their implications are discussed in Bayramoglu et al. (2018). Mitigation and adaptation are substitutes as commonly assumed for assumption e) i), but are complements for assumption e) ii). It will become apparent that for most results, the sign of the cross derivative does not matter, though the absolute size of this derivative will turn out to be important. In order to reduce the complexity of some of the subsequent proofs, we assume that third derivatives are equal to zero, which implies linear reaction functions. In the Appendix, we mention whenever we need this assumption, though it will no longer be mentioned in the text.

2.2 The Coalition Formation Game

We consider the workhorse model of international environmental agreements, which is a two-stage cartel formation game. In the first stage, countries decide on their membership. Those countries, which join coalition P, $P \subseteq N$, are called signatories and those which remain outside are called non-signatories. In the second stage, signatories act as a single player, choosing their economic strategies by maximizing the aggregate payoff over all signatories. Non-signatories act as single players, maximizing their own payoff. The solution of the second stage leads to an economic strategy vector for every coalition P of size p, $1 \leq p \leq n$. If this strategy vector is unique, notation simplifies and we can write $w_i^*(p)$. As we will see below, as all signatories $i \in P$ choose the same strategy vector and the same applies to all non-signatories $j \notin P$ (though signatories and non-signatories will choose different strategy vectors) we can also write $w_S^*(p)$ and $w_{NS}^*(p)$, with the understanding that $w_{NS}^*(p)$ does not exist if $p = n$ and $w_S^*(p) = w_{NS}^*(p)$ if $p = 1$.[4] Below, we derive sufficient conditions (see General Assumptions II below), which guarantee the existence and uniqueness of second stage equilibria.

For the second stage, we need to distinguish between the two games, the M- and M+A-Game. Moreover, we distinguish between the Nash-Cournot (NC) and the Stackelberg (ST) scenario. Under the NC-scenario, signatories and non-signatories choose their economic strategies simultaneously, and under the ST-scenario they do so sequentially, with signatories being the Stackelberg leader and non-signatories the followers, in line with the assumptions in the literature on IEAs (e.g., Barrett 1994 and Rubio and Ulph 2006), but with a small modification: we do not allow for a singleton player to act as a leader in the case in which no coalition forms.

If coalition P is empty or, which is equivalent, if it consists of only one player, the equilibrium economic strategy vector will be the same as in the Nash equilibrium in games without coalition formation. Conversely, if coalition $P = N$, i.e., the grand coalition has formed, this corresponds to the social optimum. In both extreme cases, there are no leaders and followers and the NC- and ST-scenario coincide. Hence, difference in equilibrium strategies between the two scenarios in the second stage arise when there is partial cooperation, i.e., $1 < p < n$.

---

4 Strictly speaking, $p = 0$ and $p = 1$ imply the same coalition structure. For notational simplicity, we assume $1 \leq p \leq n$.

In the first stage, and making already use of the symmetry assumption, and the simplified notation because of a unique economic strategy vector for every coalition of size p , $1 \leq p \leq n$, a coalition of size p is stable if it is internally and externally stable.

$$\text{Internal stability:} \qquad w_S^*(p) \geq w_{NS}^*(p-1) \tag{2}$$

$$\text{External stability:} \qquad w_{NS}^*(p) \geq w_S^*(p+1)$$

Internal stability requires that a signatory has no incentive to leave a coalition of size p . External stability requires that a non-signatory has no incentive to join a coalition of size p . A coalition which is internally and externally stable is called stable and the size of such a coalition is denoted by p^*. It is important to note that despite second stage equilibria for $p = 1$ and $p = n$ are the same for the NC- and ST-scenario, internal stability for $p = n$ and external stability for $p = 1$ will be different.

2.3 Assumptions in the Second Stage

Under the NC-scenario, we assume in line with Bayramoglu et al. (2018) that all countries choose their mitigation levels in the M-Game and their mitigation and adaptation levels in the M+A-Game simultaneously. As shown by Bayramoglu et al. (2018), in the M+A-Game, this is equivalent to all countries choosing first their mitigation levels and then all countries choosing their adaptation levels.

Under the ST-scenario, we assume signatories choose first their economic strategies (mitigation in the M-Game and mitigation and adaptation in the M+A-Game) as leaders and then non-signatories do the same as followers. In the M+A-Game, this is equivalent to any alternative sequence as long as signatories choose their mitigation levels first and each group does not choose adaptation before mitigation.[5]

---

[5] If adaptation was chosen before mitigation, the strategic role of adaptation would change and would lead to different outcomes (see Breton and Sbragia 2019 Eisenack & Kähler 2016, Masoudi and Zaccour 2017 and 2018 and Zehaie 2009).

Below, we list the first order conditions in an interior equilibrium (which follows from the General Assumptions I above) in the two games under the two alternative scenarios.

Consider first the NC-scenario. In the M-Game, signatories internalize the externality among its p members whereas non-signatories just maximize their own welfare. Hence, (3.a) and (3.b) imply $\frac{C_m(m_S)}{p} = C_m(m_{NS})$ and therefore $m_S > m_{NS}$ due to the strict convexity of the mitigation cost function. In the M+A-Game, the same is true considering (4.a) and (4.b) and the fact that signatories and non-signatories will choose the same adaptation level according to (5), i.e., $a_i = a_S = a_{NS}$, as adaptation is a private good.

Table 1: First Order Conditions under the NC- and ST-Scenario in the Two Games*

	M-Game	
	NC-scenario	ST-scenario
Signatories	$p \cdot B_M(M) = C_m(m_S)$ (3.a)	$p \cdot \left[B_M(M)(1 + R'_{NS}) \right] = C_m(m_S)$ (6.a)
Non-signatories	$B_M(M) = C_m(m_{NS})$ (3.b)	$B_M(M) = C_m(m_{NS})$ (6.b)
	M+A-Game	
	NC-scenario	ST-scenario
Signatories	$p \cdot B_M(M, a_i) = C_m(m_S)$ (4.a)	$p \cdot \left[B_M(M, a_i)(1 + R'_{NS}) \right] = C_m(m_S)$ (7.a)
Non-Signatories	$B_M(M, a_i) = C_m(m_{NS})$ (4.b)	$B_M(M, a_i) = C_m(m_{NS})$ (7.b)
Both	$B_a(M, a_i) = D_a(a_i)$ (5)	$B_a(M, a_i) = D_a(a_i)$ (8)

* Let $M_{NS} = R_{NS}(M_S)$. Then, $R'_{NS} = \partial M_{NS} / \partial M_S$ with $M_S = p \cdot m_S$ and $M_{NS} = (n - p) \cdot m_{NS}$.

Let us now consider the ST-scenario. Firstly, compared to the NC-scenario, it is evident from Table 1 that only the first order conditions of signatories regarding mitigation have changed. Secondly, we notice that the Stackelberg leaders choose their economic strategies such as to find the point on the followers' reaction function associated with the highest possible welfare for the leaders. That is, signatories as leaders, take into consideration how non-signatories will react. Thirdly, if we let

$m_{NS} = r_{NS}(M_{-j})$ be the best response of one non-signatory, given the mitigation level of all other players M_{-j}, or, using the symmetry assumption, which implies that all non-signatories de facto behave equally, we can define an aggregate best response function of all non-signatories $M_{NS} = R_{NS}(M_S)$ with M_{NS} being the aggregate mitigation level of all non-signatories and M_S the aggregate mitigation level of all signatories (and hence $M = M_S + M_{NS}$). Accordingly, $r_{NS}^{'}(M_{-j})$ and $R_{NS}^{'}(M_S)$ are the respective slopes of these best response or reaction functions. Similarly, we can derive the slopes of individual and aggregate best response functions of signatories, $r_S^{'}(M_{-i})$ and $R_S^{'}(M_{NS})$. Fourthly, these slopes are derived by totally differentiating the first order conditions for mitigation. Following Bayramoglu et al. (2018), in the M+A-Game, this takes into account that equilibrium mitigation and adaptation are linked. That is, before total differentiation of (4.a) and (4.b), respectively, we notice that (5) implicitly defines equilibrium adaptation as a function of total mitigation, i.e., $a_i^*(M)$. For convenience, we reproduce the result of Bayramoglu et al. (2018) in Proposition 1 below.

Proposition 1: Slopes of Reaction Functions in Mitigation and Adaptation Space

Let $\Psi^M = B_{MM}$ *in the M-Game and* $\Psi^{M+A} = B_{MM} + \dfrac{(B_{aM})^2}{D_{aa} - B_{aa}}$ *in the M+A-Game. The slopes of individual and aggregate reaction functions of signatories and non-signatories in mitigation space are given by* $r_S^{'}\left(M_{-i\in P}\right) = \dfrac{p \cdot \Psi}{C_{mm}\left(m_S\right) - p \cdot \Psi}$, $R_S^{'}\left(M_{NS}\right) = \dfrac{p^2 \cdot \Psi}{C_{mm}\left(m_S\right) - p^2 \cdot \Psi}$, $r_{NS}^{'}\left(M_{-j\notin P}\right) = \dfrac{\Psi}{C_{mm}\left(m_{NS}\right) - \Psi}$ *and* $R_{NS}^{'}\left(M_S\right) = \dfrac{(n-p)\cdot\Psi}{C_{mm}\left(m_{NS}\right) - (n-p)\cdot\Psi}$, *respectively. That is, reaction functions are downward sloping in the M-Game because* $\Psi^M < 0$ *and the same is true in the M+A-Game if* $\Psi^{M+A} < 0$ *with a slope strictly larger than* -1 *and strictly smaller than* 0. *In the M+A-Game, if* $\Psi^{M+A} > 0$ *reaction functions are upward sloping.*

In the mitigation-adaptation space, given each country's reaction function $a_i = f(M)$*, the slope of this function is given by* $f^{'}(M) = \frac{\partial a_i}{\partial M} = \frac{B_{aM}}{D_{aa} - B_{aa}}$ *and hence the reaction function is downward sloping if* $B_{aM} < 0$ *and upward sloping if* $B_{aM} > 0$.

Proof: See Bayramoglu et al. (2018), Proposition 2.

The most interesting part of Proposition 1 is that reaction functions in mitigation space do not have to be downward sloping in the M+A-Game, as this is always the case in the M-Game, but can be upward sloping in the M+A-Game. Thus, the leakage effect in terms of mitigation, due to mitigation levels in different countries being strategic substitutes, may turn into an anti-leakage effect such that mitigation levels become strategic complements. The latter possibility arises if the cross effects between mitigation and adaptation are strong, i.e., B_{aM} and $\frac{\partial a_i}{\partial M}$ are large in absolute terms because $\frac{(B_{aM})^2}{D_{aa} - B_{aa}} = B_{aM} \cdot \frac{\partial a_i}{\partial M}$, even though, interestingly, the sign of B_{aM} does not matter (as B_{aM} is squared in Ψ^{M+A}). That is, it does not matter whether mitigation and adaptation are strategic substitutes $(B_{aM} < 0)$ or complements ($B_{aM} > 0$) but only that this cross effect is sufficiently large (compared to the direct effect B_{MM} such that $\Psi^{M+A} > 0$ is possible).[6]

With reference to Table 1, under the ST-scenario, comparing the first order conditions of signatories and non-signatories in the M-Game ((6.a) and (6.b)) and in the M+A-Game ((7.a) and (7.b)), we have $\frac{C_m(m_S)}{p \cdot (1 + R_{NS}^{'})} = C_m(m_{NS})$. Hence, only if $R_{NS}^{'} > 0$ (which is only possible in the M+A-Game), can we conclude $m_S(p) > m_{NS}(p)$, given the convexity of the mitigation cost function. In contrast, if

6 In the following, we rule out the uninteresting and special case of $\Psi^{M+A} = 0$.

$-1 < R_{NS}^{'} < 0$, which is always the case in the M-Game and is one possibility in the M+A-Game, $m_S(p) < m_{NS}(p)$ is possible if p is small.

In terms of the existence and uniqueness of second stage equilibria, it turns out that our General Assumptions I are sufficient when reaction functions are downward sloping but additional General Assumptions II need to be imposed in case reaction functions are upward sloping. That is, we need further assumptions in the M+A-Game if $\Psi^{M+A} > 0$ as explained in Appendix A.1.

General Assumptions II

In the M+A-Game let $\Psi^{M+A} = B_{MM} + \frac{(B_{aM})^2}{D_{aa} - B_{aa}}$*. If* $\Psi^{M+A} > 0$*, for any coalition size* p*, a sufficient condition for the existence of a unique second stage equilibrium is:*

$$\Psi^{M+A} \cdot \left[\frac{p^2}{C_{mm}(m_S)} + \frac{(n-p)}{C_{mm}(m_{NS})} \right] < 1$$ *under the Nash-Cournot scenario and*

$$\Psi^{M+A} \cdot \left[\frac{p^2 \cdot (1 + R_{NS}^{'})}{C_{mm}(m_S)} + \frac{(n-p)}{C_{mm}(m_{NS})} \right] < 1$$ *under the Stackelberg scenario.*

We note that if $\Psi^{M+A} > 0$, $1 + R_{NS}^{'} > 0$.

3. Results

3.1 Preliminaries and Definitions

In the following analysis, we focus on comparing the sizes and success of stable agreements under the NC- and ST-scenario. In order to explain differences, it will be helpful to consider some general properties of coalition formation under the two scenarios. Moreover, it will turn out to be useful to work with a specific welfare function in order to illustrate a couple of general interesting points. On the one hand, and as it is well-known from the literature on IEAs, only this allows to make sharp predictions about first stage equilibria (i.e., the size of stable agreements). On the other hand, this allows running simulations for those results, which cannot be obtained analytically; again, a feature quite common in the literature on IEAs. Nevertheless, it will be apparent from subsection 3.2 that we are able to derive a couple of very general and useful results, which are complemented in subsection 3.3 with further details based on simulations.

Definition 1: Positive Externality, Superadditivity and Cohesiveness

Let $n \geq p \geq 2$.

i) *PEP: The expansion of coalition* $p-1$ *to* p *exhibits a positive (negative) externality if:*

$$w_{NS}^{*}(p) > (<)\, w_{NS}^{*}(p-1).$$

If this holds for all p *,* $n \geq p \geq 2$*, the game is a positive (negative) externality game.*

ii) *SAD: The expansion of coalition* $p-1$ *to* p *is superadditive if:*

$$p \cdot w_{S}^{*}(p) \geq (>) [p-1] \cdot w_{S}^{*}(p-1) + w_{NS}^{*}(p-1).$$

If this holds for all p *,* $n \geq p \geq 2$*, the game is superadditive.*

iii) *WCOH: The expansion of coalition* $p-1$ *to* p *is welfare cohesive if:*

$$p \cdot w_{S}^{*}(p) + [n-p] \cdot w_{NS}^{*}(p) > [p-1] \cdot w_{S}^{*}(p-1) + [n-p+1] \cdot w_{NS}^{*}(p-1)$$

If this holds for all p *,* $n \geq p \geq 2$*, the game is welfare cohesive.*

iv) *MCOH: The expansion of coalition* $p-1$ *to* p *is mitigation cohesive if:*

$$p \cdot M_{S}^{*}(p) + [n-p] \cdot M_{NS}^{*}(p) \geq (>) [p-1] \cdot M_{S}^{*}(p-1) + [n-p+1] \cdot M_{NS}^{*}(p-1)$$

If this holds for all p *,* $n \geq p \geq 2$*, the game is mitigation cohesive.*

The first two properties may be viewed as positive properties in that they help to explain whether stable coalitions will be small or large. Positive externality makes it attractive to stay outside a coalition whereas for negative externalities just the opposite holds. Superadditivity can be viewed as a necessary condition to make joining a coalition attractive. In a superadditive and negative externality game, the grand coalition is the unique stable agreement (Weikard 2009). Thus, cooperation does not pose a problem. In contrast, in positive externalities games, typically, stable coalitions are small. This is evident if superadditivity fails, but even if it holds, the positive externality effect may be stronger than the superadditivity effect such that only small coalitions are stable.

The third and the fourth property can be viewed as normative properties. Clearly, in the grand coalition, total welfare and total mitigation levels are higher than in any other coalition (see Bayramoglu et al. 2018). However, it may not always be true that these levels increase with every enlargement of a coalition, irrespective of its size, as we will illustrate and explain in more detail below. Note that a sufficient condition for welfare cohesiveness is superadditivity and positive externalities.

In line with the literature on IEAs and following Bayramoglu et al. (2018), we consider a welfare function with quadratic benefits and quadratic costs in order to illustrate some results. In the M-Game, we assume:

$$w_i^M = \left(bM - \frac{g}{2} M^2 \right) - \frac{c}{2} m_i^2 \tag{9}$$

and in the M+A-Game we consider:

$$w_i^{M+A} = \left(bM - \frac{g}{2} M^2 \right) + a_i \left(\beta - fM \right) - \frac{c}{2} m_i^2 - \frac{d}{2} a_i^2 \text{ such that } B_{aM} < 0 \tag{10.a}$$

and

$$w_i^{M+A} = \left(bM - \frac{g}{2} M^2 \right) + a_i \left(\beta + fM \right) - \frac{c}{2} m_i^2 - \frac{d}{2} a_i^2 \text{ such that } B_{aM} > 0 \tag{10.b}$$

assuming that all parameters b, g, c β, f, and d are strictly positive. If we were to set $g = 0$, we could retrieve the linear-quadratic welfare function, also frequently considered in the literature on IEAs. However, in this case, in the M-Game, countries would have a dominant strategy ($\Psi^M = 0$ and reaction functions would be orthogonal), implying that the NC- and ST-scenario are identical. For expositional clarity, we ignore this case.

It is also clear that by setting $a_i = 0$ in the M+A-Game, we are back in the M-Game. In Appendix A.2, we derive conditions for the parameters, which ensure that the sufficient conditions for existence and uniqueness are satisfied plus additional conditions, which ensure interior equilibria.

For welfare function (9), $\Psi^M = -g$ and $r_{NS}^{'} = -\frac{g}{c+g}$ and hence reaction functions are downward sloping. We notice that the absolute value of this slope increases in the benefit parameter g and decreases in the cost parameter c. For welfare functions (10.a) and (10.b), $\Psi^{M+A} = \frac{f^2 - g \cdot d}{d}$ which is negative if $f^2 - g \cdot d < 0$ and positive if $f^2 - g \cdot d > 0$. Accordingly, the slope of the reaction function, $r_{NS}^{'} = \frac{\left(f^2 - g \cdot d\right)}{c \cdot d + \left(f^2 - g \cdot d\right)}$, may either be negative or positive. The difference between (10.a) and (10.b) is just the sign of the cross derivative B_{aM}, which does neither affect Ψ^{M+A} nor $r_{NS}^{'}$.

In our simulations, which are reported in Appendix A.6, we consider five runs, covering a wide range of parameter values, displaying results for the NC- and ST-scenario. In Table A.1, we consider the M-Game (and hence $\Psi^M < 0$), whereas in Table A.2 to A.5 we consider the M+A. In Tables A.2 and A.3 $\Psi^{M+A} < 0$ is assumed whereas in Tables A.4 and A.5 $\Psi^{M+A} > 0$. The difference is that the first table of each set (i.e., Table A.2 and A.4), assumes $B_{aM} < 0$ and the second (i.e., Tables A.3 and A.5) assumes $B_{aM} > 0$. Hence, we cover all possible interesting parameter constellations.

In order to evaluate stable coalitions, we consider two indices in our simulations, restricting ourselves to the welfare dimension, even though similar indices could be defined in terms of mitigation. We recall that no-cooperation with $p = 1$ corresponds to the classical Nash equilibrium without coalition formation and full cooperation with $p = n$ corresponds to the social optimum. We denote total welfare with W, $W = \sum_{i=1}^{n} w_i$, and use superscripts to refer to the social optimum, SO, Nash equilibrium, NE, and stable coalitions in the NC- and ST-scenario, respectively.

Definition 2: Importance of Cooperation and Improvement upon the Nash Equilibrium

- *The Importance of Cooperation Index (ICI) measures the percentage global welfare improvement from moving from no-cooperation (NE) to the social optimum (SO):*

$$ICI = \frac{W^{SO} - W^{NE}}{W^{NE}} \cdot 100$$

- *The Improvement upon the Nash equilibrium Index (INI) measures the percentage global welfare improvement obtained in a stable equilibrium under the NC- and ST-scenario, respectively:*

$$INI^{NC} = \frac{W^{NC}\left(p^{NC*}\right) - W^{NE}}{W^{NE}} \cdot 100,$$

$$INI^{ST} = \frac{W^{ST}\left(p^{ST*}\right) - W^{NE}}{W^{NE}} \cdot 100.$$

Both indices are relative measures as absolute values are meaningless without any benchmark. Index *ICI* measures the potential gains from cooperation or what Barrett (1994) called the “need for cooperation”. Index *INI* measure the performance of stable coalitions. Clearly, if *ICI* is small, also *INI* must be small, even stable coalitions may be large. If *ICI* is large, *INI* may be small because only small coalitions are stable. Hence, cooperation is interesting and successful if *ICI* and *INI* is large because the potential gains from cooperation are large and these gains are reaped because large coalitions are stable. Relating Barrett’s paradox of cooperation to the above indices means that either only small coalitions are stable in which case *INI* is small, or large coalitions are stable, but then *ICI* and hence *INI* are small. That is whenever cooperation would be needed most, stable coalitions achieve little. It is also for this reason that focusing only on the size of stable coalitions p^* is not sufficient for an evaluation, we also need to evaluate outcomes in terms of global welfare gains.

3.2 Propositions

In this subsection, we derive some general results, which are summarized in Proposition 2 below. In the M-game, reaction functions are downward sloping (Proposition 2.a). Consequently, signatories having a strategic advantage (i.e., a first mover advantage) under the ST-scenario, will lower their mitigation level compared to the NC-scenario, knowing that non-signatories will partly make up for this by mitigating more. Overall, for any given coalition size p, total mitigation will be lower under the ST- than under the NC-scenario. The Stackelberg leader will be better off and the reverse is true for the follower. It is for this reason that stable coalitions under the ST-scenario will be at least as large than under the NC-scenario. Hence, we provide a general proof of this relation which has been found in many papers on IEAs and which will also be illustrated for our specific welfare functions below. Moreover, as it is evident from Proposition 2.a, this result extends to the M+A-Game, provided reaction functions are downward sloping.

Proposition 2: Comparison of NC- and ST-Scenario, Mitigation, Payoffs and Stable Coalitions

Consider a generic coalition of size p, $n > p > 1$.

a) *In the M-Game with* $\Psi^M < 0$ *and in the M+A-Game if* $\Psi^{M+A} < 0$, *and hence reaction functions are downward sloping in mitigation space, the following relations hold for every* p, $n > p > 1$*:*

- $M^{NC}(p) > M^{ST}(p)$, $m_S^{NC}(p) > m_S^{ST}(p)$ *and* $m_{NS}^{NC}(p) < m_{NS}^{ST}(p)$;
- $w_S^{NC}(p) < w_S^{ST}(p)$ *and* $w_{NS}^{NC}(p) > w_{NS}^{ST}(p)$. *It follows that*
- $p^{ST*} \geq p^{NC*}$.

b) *In the M+A-Game with* $\Psi^{M+A} > 0$*, implying upward sloping reaction functions in mitigation space, the following relations hold for every* p*,* $n > p > 1$*:*

- $M^{NC}(p) < M^{ST}(p)$, $m_S^{NC}(p) < m_S^{ST}(p)$ *and* $m_{NS}^{NC}(p) < m_{NS}^{ST}(p)$*;*
- $w_S^{NC}(p) < w_S^{ST}(p)$, $w_{NS}^{NC}(p) < w_{NS}^{ST}(p)$ *and* $W^{NC}(p) < W^{ST}(p)$
- $m_S^{ST}(p) - m_S^{NC}(p) > m_{NS}^{ST}(p) - m_{NS}^{NC}(p)$ *if the mitigation cost function is a strictly convex polynomial function such that*

  $w_S^{ST}(p) - w_S^{NC}(p) < w_{NS}^{ST}(p) - w_{NS}^{NC}(p)$*. It follows that*
- $p^{ST*} \leq p^{NC*}$.

Proof: See Appendix A.3.

It is also evident from Proposition 2.a why it is not possible to draw any general conclusion about total mitigation levels and global welfare for stable coalitions under the two scenarios. In terms of global welfare, we do not know whether $W^{NC}(p^{NC}) > W^{ST}(p^{ST})$ or the reverse is true for a given p as signatories are better off but non-signatories worse off under the ST- than under the NC-scenario. Hence, we also do not know generally whether $W^{NC}(p^{NC*}) < W^{ST}(p^{ST*})$ or the opposite is true in equilibrium. In terms of global mitigation, we know that $M^{NC}(p) > M^{ST}(p)$ but $p^{NC*} \leq p^{ST*}$ and hence, generally, $M^{NC}(p^{NC*}) <,> M^{ST}(p^{ST*})$.

Finally, Proposition 2.b stresses that the intuition the ST-scenario always leads to larger stable coalitions is wrong if reaction functions in mitigation space are upward sloping, which is possible in the M+A-Game if cross effects are strong enough such that $\Psi^{M+A} > 0$. In such a matching game, both, signatories and non-signatories, increase their mitigation levels under the ST- compared to the NC-scenario. This also translates into a Pareto-improvement for all countries and hence in higher total welfare. However, compared to the NC-scenario, non-signatories gain more than signatories under the ST-scenario, i.e., there is a second mover advantage.[7] The reason is that signatories increase their mitigation levels more than non-signatories and hence carry higher additional mitigation costs. This explains why the size of stable coalitions are generally weakly smaller under the ST- than NC-scenario. Again, this makes it impossible to conclude generally whether $M^{NC}\left(p^{NC*}\right)<M^{ST}\left(p^{ST*}\right)$ and $W^{NC}\left(p^{NC*}\right)<W^{ST}\left(p^{ST*}\right)$ hold or the reverse is true.

In order to illustrate the relation between stable coalitions under the two scenarios in the two games, we determine stable coalitions for our specific welfare functions as introduced above.

Proposition 3: Stable Coalitions in the M- and M+A-Game for Specific Welfare Functions

Consider payoff function (9) in the M-Game and (10.a) and (10.b) in the M+A-Game and assume the conditions on parameters in Appendix A.2 to hold. The size of stable coalitions p^* under the CN- and ST-scenario are given by (assuming that $n \geq 7$):

	M-GAME		**M+A-GAME**			
	$\Psi<0$		$\Psi<0$		$\Psi>0$	
	NC	ST	NC	ST	NC	ST
p^*	$p^{NC*}\in[1,2]$	$p^{ST*}\in[2,n]$	$p^{NC*}\in[1,2]$	$p^{ST*}\in[2,n]$	$p^{NC*}=\{3,n\}$	$p^{ST*}=\{2,3\}$

Proof: See Appendix A.4.

It is evident that for downward sloping reaction functions, under the ST-scenario, even the grand coalition could form. In contrast, under the NC-scenario, only small coalitions are stable. For upward

[7] This is in line with the literature on Stackelberg games with symmetric players (though usually confined to two players). There is a first (second) mover advantage in the presence of downward (upward) sloping reaction functions (Gal-Or 1985).

sloping reaction functions, the reverse is true. Under the NC-scenario, large coalitions can be stable, whereas under the ST-scenario only small coalitions are stable.[8]

In order to rationalize different equilibrium coalition sizes, we consider the general properties in the two games under the two scenarios.

Proposition 4: Properties in the M- and M+A-Game under the CN- and ST-scenario

Consider the general welfare function (1.a) in the M-Game and welfare function (1.b) in the M+A-Game. Further assume the General Assumptions I and II to hold. Then the following conclusion can be drawn:

	M-GAME		M+A-GAME			
	$\Psi<0$		$\Psi<0$		$\Psi>0$	
	NC	ST	NC	ST	NC	ST
PEP	✓	fails when MCOH fails	✓	fails when MCOH fails	✓	✓
SAD	may fail for small p	✓	may fail for small p	✓	✓	✓
WCOH	may fail for small p	may fail for small p	may fail for small p	may fail for small p	✓	✓
MCOH	✓	may fail for small p	✓	may fail for small p	✓	✓

Properties as defined in Definition 1; ✓ = property holds for all expansion $p-1$ to p, $2\leq p\leq n$, except for PEP for which $2\leq p<n$.

Proof: See Appendix A.5.

Under the NC-scenario, the game is a positive externality game. Total mitigation increases steadily with an expansion of the coalition from which also non-signatories benefit due to the non-exclusiveness of the public good. Non-signatories reduce their contribution to this public good if

8 Bayramoglu et al. (2018) show that for welfare function (10.a) and (10.b) and $\Psi>0$ in the M+A-Game, $p^{NC*}\in[3,n]$. We find that if $n\geq 7$ (as assumed in our simulations), this leads to $p^{NC*}=\{3,n\}$. See Appendix A.4.

reaction functions are downward sloping (and hence have not only higher benefits but also lower mitigation costs). However, even if $\Psi > 0$, non-signatories contribute less than proportionally to the total increase in total mitigation and hence also enjoy a positive externality from the expansion of the coalitions. Therefore, with positive externalities, there is an incentive to remain a non-signatory.

Moreover, under the NC-scenario, if $\Psi < 0$, it is also evident that superadditivity may fail due to the leakage effect, which is also an obstacle to form large stable coalitions. Together, this explains why only small coalitions are stable if reaction functions are downward sloping. In contrast, if reaction functions are upward sloping, superadditivity always holds, as the game has turned into a matching game with anti-leakage. This allows to form larger stable coalitions, including the grand coalition in the M+A-Game if $\Psi > 0$. It is also evident that if the leakage effect is present (i.e., $\Psi < 0$), welfare cohesiveness may fail (as a result of a failure of superadditivity).

Under the ST-scenario, the negative conclusion about the size of stable coalitions if reaction functions are downward sloping (i.e., $\Psi < 0$) is just reversed. Roughly speaking, and as our simulations will confirm, the steeper the reaction function, the larger is the strategic advantage of the leader over the follower and hence the larger will be stable coalitions. Moreover, superadditivity always holds, and at least for not too large coalitions, the enlargement of coalitions may not be associated with positive but with negative externalities, making it attractive for non-signatories to join the coalition. The fact that larger coalitions may not necessarily lead to substantially better outcomes, as will be confirmed in subsection 3.3 based on simulations, is already apparent by the fact that welfare and mitigation cohesiveness does not generally hold if $\Psi < 0$.[9] In other words, larger stable coalitions under Stackelberg leadership comes at a price.

---

9 Welfare cohesiveness fails whenever the superadditivity effect is dominated by the negative externality effect. Mitigation cohesiveness may fail as the Stackelberg leaders use their strategic advantage to reduce their contribution to the public good, which may not be compensated by the followers' additional mitigation effort.

Such a price also needs to be paid under the ST-scenario if reaction functions are upward sloping (i.e., $\Psi > 0$). Even though welfare and mitigation cohesiveness hold throughout, stable coalitions are small and smaller than in the NC-scenario. The small coalitions are due to the fact that positive externalities hold throughout and non-signatories benefit more than signatories from Stackelberg leadership of signatories.

3.3 Further Results and Discussion

The discussion in this subsection is supported by our simulations, which are summarized in Tables A.1 to A.5 in Appendix A.6. We address two research questions. 1) Does the ST-scenario improve over the NC-scenario? In order to answer this question, Table A.1 to A.3 will be helpful. That is, we focus on downward sloping reactions in mitigation space, i.e., $\Psi < 0$, as for upward sloping reaction functions, i.e., $\Psi > 0$, we already know that $p^{*NC} \geq p^{*ST}$. 2) Does the paradox of cooperation as established by Barrett (1994) for the M-Game and later iterated by many others also hold for the M+A-Game? In order to answer this question, Table A.2 to A.5 will be helpful.

3.3.1 Does the ST-scenario improve over the NC-scenario?

In the M-Game, we know $p^{NC*} \in [1,2]$ and $p^{ST*} \in [2,n]$ from Proposition 3. Hence, under the NC-scenario, whenever n is sufficiently large ($n = 100$ in our simulations), stable coalitions cannot achieve much. Under the ST-scenario, we find that the steeper the reaction functions (implying a high ratio of the parameters g/c for welfare function (9)), the larger will be p^{ST*}. However, it is easily proved that $ICI = \frac{W^{SO} - W^{NE}}{W^{NE}} \cdot 100$ decreases in the ratio g/c. As Table A.1 confirms if p^{ST*} approaches the grand coalition, the value of *ICI* is very small. Accordingly, also $INI^{ST} = \frac{W^{ST}\left(p^{ST*}\right) - W^{NE}}{W^{NE}} \cdot 100$ must be very small. Also the reverse is true, if reaction functions are flat (low value of g/c), *ICI* is large but p^{ST*} is small and hence *INI* is small. Hence, overall, *INI* is generally small under both scenarios, and the ST-scenario only marginally improves upon the NC-scenario.[10]

In the M+A-Game and downward sloping reaction functions in mitigation space, we know $p^{NC*} \in [1,2]$ and $p^{ST*} \in [2,n]$ from Proposition 3, irrespective of the sign of the cross derivative B_{aM}. As in M-Game, the same conclusion conclusions are valid with reference to Tables A.2 and A.3. The NC-scenario allows only for small stable coalitions anyway, and the ST-scenario, though it may generate larger stable coalitions, makes hardly any difference to no cooperation: *INI* shows low values for all parameter constellations and the ST-scenario only marginally improves upon the NC-scenario. For completeness, it is worthwhile mentioning that for upward sloping reaction functions (Tables A.4 and A.5), the ST-scenario implies usually lower welfare gains than the NC-scenario as stable coalitions tend to be smaller. Only in a few cases, when $p^{NC*} = 3$ will the ST-scenario marginally improve upon the NC-equilibrium, but this is when *INI* is anyway small.

3.3.2 Does the paradox of cooperation also hold for the M+A-Game?

For the NC-scenario and downward sloping reaction functions, the paradox holds because we know $p^{NC*} \in [1,2]$ from Proposition 3 and hence *INI* is low (see Tables A.2 and A.3), irrespective of the sign of the cross derivative B_{aM}. We know this may change for upward sloping reaction functions, as then $p^{CN*} = \{3,n\}$ from Proposition 3, and hence the grand coalition may be stable. However, as is evident from Table A.4 for $B_{aM} < 0$, even if $p^{NC*} = n$, *ICI* and hence *INI* are small, and if $p^{NC*} = 3$ *ICI* may be large but *INI* is small (because $p^{NC*} = 3 << n$), the classical paradox of cooperation. Only if $B_{aM} > 0$ (Table A.5), we find that $p^{NC*} = n$ as well as *ICI* and *INI* may be large. This appears to be the only "anti-paradox constellation". However, this rests on the rather unlikely assumption that

---

10 For one parameter constellation in Table A.1, the sizes of stable coalitions under both scenarios are identical, $p^{NC*} = p^{ST*} = 2$, and hence $INI^{NC} > INI^{ST}$.

mitigation and adaptation are complements apart from upward sloping reaction functions in mitigation space. This may be seen as qualifying the positive conclusion as derived by Bayramoglu et al. (2018).

Finally, for the ST-scenario, *INI* is always small. For downward sloping reaction functions, as in the M-Game, also in the M+A-Game large stable coalitions go along with small *ICI* and hence *INI*. For upward sloping reaction functions, stable coalitions are always small and hence *INI* is also small. Hence, the paradox of cooperation holds for all parameter constellations for the ST-scenario.

Looking to both stable coalition size and welfare performances in the NC-scenario, is interesting to compare our results with Heugues (2014). The paper exogenously introduce complementarity between abatement strategies in a pure emissions game, i.e., without modeling adaptation. It is found that large coalitions, up to half of the players, can form in the cases where cooperation is most needed. Hence, differently from our results, upward sloping reaction functions are not able to lead to full cooperation. Nonetheless large improvements from non-cooperation are achieved. We find instead that upward sloping reaction functions can lead to full cooperation but, under the standard assumption of substitutability between mitigation and adaptation, only small welfare improvements from non-cooperation are achieved.

4. Summary and Conclusion

In this paper, we considered the standard two-stage coalition formation game with symmetric players. We explored four different settings: a) mitigation game (M-Game), b) mitigation-adaptation game (M+A-Game), c) Nash-Cournot scenario (NC-scenario) and d) Stackelberg scenario (ST-scenario). In the first stage of the game, players choose whether to sign an agreement and be part of a climate agreement or to remain outside as a singleton. In the second stage, signatories choose their economic strategies (mitigation or mitigation and adaptation) by maximizing their aggregate welfare, while non-signatories maximize their individual welfare. The sequence of these decisions differed between the NC- and the ST-scenario.

Our analysis combined the features of two contributions. The first contribution by Barrett (1994), Diamantoudi and Sartzetakis (2006) and Rubio and Ulph (2006) studied the effect of ST-scenario on the size of stable agreements in the M-Game. The second contribution by Bayramoglu et al. (2018) studied the effect of moving from the M- to the M+A-Game under the NC-scenario, i.e., when all players simultaneously choose their economic strategies.

We complemented these studies by considering Stackelberg leadership in the M+A-Game. This allowed us to address two research questions. 1) Does the ST-scenario improve over the NC-scenario? 2) Does the paradox of cooperation as established by Barrett (1994) for the M-Game and later iterated by many others also hold for the M+A-Game?

We found that the ST-scenario leads to larger stable coalitions if reaction functions in mitigation space are downward sloping, i.e., mitigation levels in different countries are strategic substitutes. This happens because signatories reduce their mitigation efforts, forcing followers to mitigate more compared to the NC-scenario. Therefore, participation is more attractive in the ST- than in the NC-scenario. However, we found that whenever the difference in stable coalition sizes is large between the two scenarios, the potential gains from cooperation are small. Hence, the ST-scenario only marginally improves upon the NC-scenario. In contrast, if reaction functions in mitigation space are upward sloping in the M+A-Game, stable coalitions are even smaller in the ST- than in the NC-scenario, which is also reflected in lower equilibrium total welfare. Thus, taken together, the ST-scenario does not always lead to larger stable coalitions and larger global welfare than the NC-scenario, but if this is the case, the welfare improvements are very marginal.

The results for the ST-scenario confirmed Barrett's paradox of cooperation: either coalitions are small or, if they are large, the potential gains from cooperation are small. This is also true for the NC-scenario, with one exception: reaction functions in mitigation space need to be upward sloping, and, additionally, mitigation and adaptation need to be complements and not substitutes. Hence, the paradox of cooperation extends to a richer coalition game, which includes adaptation as an additional strategy to mitigation for the widespread assumption that mitigation and adaptation are substitutes.

For future research, two obvious extensions come to mind. Firstly, we assumed that adaptation is either chosen simultaneously with mitigation or after mitigation. In other words, we considered “reactive adaptation”. However, in a dynamic game in which negotiations spread over some time and in which contracts are renegotiated, like for instance in Battaglini and Harstad (2016) and Harstad (2012), one can easily perceive that adaptation becomes “active” as considered for instance by Buob and Stephan (2011) and Heuson et al. (2015). Coalition formation games have been analysed under this assumption by Masoudi and Zaccour (2017 and 2018) considering cooperation on adaptation with R&D spillovers and by Breton and Sbragia (2019) using a specific climate cost function that considers vulnerability. Secondly, we assumed symmetric players. In order to capture the current interesting discussion whether industrialized countries should support developing countries by providing adaptation because of their high vulnerability to climate change and their lack of adaptation capacity, the model would need to be extended to allow for asymmetry in terms of benefit and cost functions like this is considered in Eyckmans et al. (2016), Lazkano et al. (2016) and Li and Rus (2018).

References:

Barrett, S., (1994), Self-Enforcing International Environmental Agreements. "Oxford Economic Papers", 46, pp.878–894.

Battaglini, M. and B. Harstad (2016), Participation and Duration of Environmental Agreements. “Journal of Political Economy”, vol. 124(1), pp. 160-204.

Basu, K. and N. Singh (1990), Entry-Deterrence in Stackelberg Perfect Equilibria. "International Economic Review", vol. 31(1), pp. 61-71.

Bayramoglu, B., M. Finus and J.-F. Jacques (2018), Climate Agreements in Mitigation-Adaptation

Game. “Journal of Public Economics”, vol. 165, pp. 101-113.

Breton, M. and L. Sbragia (2019), The Impact of Adaptation on the Stability of International Environmental Agreements. "Environmental and Resource Economics", vol. 74(2), pp. 697-725.

Buob, S. and G. Stephan (2011), To Mitigate or to Adapt: How to Confront Global Climate Change. “European Journal of Political Economy”, vol. 27(1), pp. 1–16.

Carraro, C. and D. Siniscalco (1993), Strategies for the International Protection of the Environment. "Journal of Public Economics", vol. 52(3), pp. 309–328.

Diamantoudi, E. and E.S. Sartzetakis (2006), Stable International Environmental Agreements: An Analytical Approach. “Journal of Public Economic Theory”, vol. 8(2), pp. 247-263.

Ebert, U. and H. Welsch (2011), Optimal Response Functions in Global Pollution Problems Can be Upward-sloping: Accounting for Adaptation. “Environmental Economics and Policy Studies”, vol. 13(2), pp. 129–138.

Ebert, U. and H. Welsch (2012), Adaptation and Mitigation in Global Pollution Problems: Economic Impacts of Productivity, Sensitivity, and Adaptive Capacity. “Environmental and Resource Economics”, vol. 52, pp. 49–64.

Eisenack, K. and L. Kähler (2016), Adaptation to Climate Change Can Support Unilateral Emission Reductions. "Oxford Economic Papers", vol. 68(1), pp. 258–278.

Endres, A. (1992), Strategic Behavior Under Tort Law. "International Reviw of Law and Economics", vol. 12, pp. 377-380.

Eyckmans, J., S. Fankhauser and S. Kverndokk (2016), Development Aid and Climate Finance. “Environmental and Resource Economics”, vol. 63 (2), pp. 429–450.

Finus, M. and A. Caparrós (2015), Game Theory and International Environmental Cooperation. The International Library of Critical Writings in Economics. Edward Elgar, Cheltenham, UK.

Gal-Or, E., (1985), First Mover and Second Mover Advantages. "International Economic Review", vol. 26(3), pp. 649-653.

Harstad, B. (2012), Climate Contracts: a Game of Emissions, Investments, Negotiations and Renegotiations. "Review of Econonomic Studies", vol. 79 (4), pp. 1527–1557.

Heugues, M. (2014), International Environmental Cooperation: a New Eye on the Greenhouse Gas Emissions' Control. "Annals of Operations Research", vol. 220 (1), pp. 239–262.

Heuson, C., W. Peters, R. Schwarze and A.-K. Topp (2015), Investment and Adaptation as Commitment Devices. “Environmental and Resource Economics”, vol. 62, pp. 769-790.

Hoel, M. (1992), International Environment Conventions: The Case of Uniform Reductions of Emissions. “Environmental and Resource Economics”, vol. 2(2), pp. 141-159.

Independent Evaluation Group (IEG) (2013). Adapting to Climate Change: Assessing World Bank Group Experience: Phase III of the World Bank Group and Climate Change. Washington DC. http://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/cc3_full_eval.pdf

Ingham, A., Ma, J. and Ulph, A.M. (2013), Can Adaptation and Mitigation be Complements? "Climatic Change", vol. 120(1–2), pp. 39–53.

IPCC (2018), Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. World Meteorological Organization, Geneva, Switzerland, pp. 1-32.

Lazkano, I., W. Marrouch and B. Nkuiya (2016), Adaptation to Climate Change: How Does Heterogeneity in Adaptation Costs Affect Climate Coalitions? “Environment and Development Economics”, vol. 21(06), pp. 812–838.

Li, H. and Rus, H (2018), Climate Change Adaptation and International Mitigation Agreements with Heterogeneous Countries. “Journal of the Association of Environmental and Resource Economists”, vol. 6(3), pp. 503-530.

Masoudi, N. and G. Zaccour (2017), Adapting to climate change: Is cooperation good for the environment? "Economic Letters", vol. 153(C), pp. 1-5.

Masoudi, N. and G. Zaccour (2018), Adaptation and International Environmental Agreements. "Environmental and Resource Economics", vol. 71(1), pp. 1-21.

Rubio, S.J. (2018), Self-Enforcing International Environmental Agreements: Adaptation and Complementarity. Working Paper, 029.2018, Fondazione Eni Enrico Mattei.

Rubio, S.J. and A. Ulph (2006), Self-enforcing International Environmental Agreements Revisited.

“Oxford Economic Papers”, vol. 58(2), pp. 233-263.

UNFCCC (2014), Report of the Adaptation Committee to the Subsidiary Body for Scientific and Technological Advice. Forty-first session of COP20, Lima, Peru, FCCC/SB/2014/2.

UNFCCC (2016), Adaptation under the UNFCCC after the Paris Agreement. https://unfccc.int/news/adaptation-under-the-unfccc-after-the-paris-agreement

Vickers, J. (1985), Delegation and the Theory of the Firm. "The Economic Journal", Vol. 95, pp. 138-147.

Weikard, H.-P. (2009), Cartel Stability under an Optimal Sharing Rule. “Manchester School”, vol. 77(5), pp. 575-593.

World Bank. (2010). Economics of Adaptation to Climate Change: Synthesis Report. Washington DC. http://documents.worldbank.org/curated/en/646291468171244256/pdf/702670ESW0P10800EACCSynthesisReport.pdf

Zehaie, F. (2009) The Timing and Strategic Role of Self-Protection. "Environmental and Resource Economics", vol. 44(3), pp. 337–350.

Appendix

A.1 Derivation of the General Assumptions II

The procedure to derive sufficient conditions for the existence and uniqueness of mitigation and adaptation equilibria for every coalition of size p follows Bayramoglu et al. (2018). The procedure is based on the concept of replacement functions. Let $m_S = g_S(M)$ be the individual replacement function of a signatory and $m_{NS} = g_{NS}(M)$ be the replacement function of a non-signatory. The aggregate replacement function $G(M)$ is derived by summing over all replacement functions, which for symmetry is

$$\sum_{i=1}^{n} m_i = p \cdot m_S + (n-p) \cdot m_{NS} = M = G(M) = \sum_{i=1}^{n} g_i(M) = p \cdot g_S(M) + (n-p) \cdot g_{NS}(M) .$$

If every replacement function is downward sloping over the entire mitigation space, the aggregate replacement function will be downward sloping as well (which is the vertical aggregation of individual replacement functions) and hence will intersect with the 45-degree line once. In other words, the level of M, which satisfies the equality above is the equilibrium M^*, which upon substitution into individual replacement functions gives m_S^* and m_{NS}^*. As we will see below, replacement functions are downward sloping (like reaction functions, see Proposition 1) if $\Psi < 0$. In the case of upward sloping replacement functions ($\Psi > 0$), a sufficient condition for uniqueness is that the aggregate replacement function has a slope of less than 1 over the entire domain such that it intersects with the 45-degree line and only once. Finally, as reaction functions of adaptation as function of total mitigation (see Proposition 1) are continuous and single valued, also equilibrium adaptation levels will be unique. Below, we derive the sufficient conditions in the case of the ST-scenario, which are those in the NC-scenario as derived by Bayramoglu et al. (2018) if we set $R_{NS}^{'} = 0$.

The first order conditions of signatories in the M-Game and M+A-Game (6.a) and (7.a) in Table 1, respectively, using the concept of individual replacement functions, read:

$$p \cdot \left[B_M(M)\left(1 + R_{NS}^{'}\right) \right] = C_m\left(m_S(M)\right)$$

$$p \cdot \left[B_M\left(M, a_i(M)\right) \cdot \left(1 + R_{NS}^{'}\right) \right] = C_m\left(m_S(M)\right)$$

Total differentiation with respect to M, and ignoring third derivatives for simplicity, gives the slope of the individual replacement function of signatories, keeping in mind the different values of Ψ in the M- and M+A-Game:

$$g_S^{'}(M) = \frac{p \cdot \left[\Psi \cdot \left(1 + R_{NS}^{'}\right)\right]}{C_{mm}(m_S)}.$$

For non-signatories, we find, using the first order conditions (6.b) and (7.b) in Table 1, respectively:

$$B_M(M) = C_m\left(m_{NS}(M)\right)$$

$$B_M\left(M, a_i(M)\right) = C_m\left(m_{NS}(M)\right)$$

and hence we derive the slope of the individual replacement of non-signatories:

$$g_{NS}^{'}(M) = \frac{\Psi}{C_{mm}(m_{NS})}.$$

Accordingly, the slope of the aggregate replacement function is given by:

$$G^{'}(M) = \Psi \cdot \left[\frac{p^2 \cdot \left[\left(1 + R_{NS}^{'}\right)\right]}{C_{mm}(m_S)} + \frac{(n-p)}{C_{mm}(m_{NS})}\right]$$

which is negative if $\Psi < 0$, but is positive if $\Psi > 0$, and hence we need that $G^{'}(M) < 1$ holds, which is the sufficient condition we state in the General Assumptions II.

A.2 Existence, Uniqueness Conditions for an Interior Second Stage Equilibrium for Welfare Function (9), (10a.) and (10.b)

In the M-Game for welfare function (9), we have $B_M = b - g \cdot M$, $B_{MM} = -g < 0$, $C_m = c \cdot m_i$, $C_{mm} = c$ and hence $\Psi^M = B_{MM} = -g$. Hence, we have: $r_S^{'}(M_{-i}) = -\frac{p \cdot g}{c + p \cdot g}$, $R_S^{'}(M_{NS}) = -\frac{p^2 \cdot g}{c + p^2 \cdot g}$, $r_{NS}^{'}\left(M_{-j}\right) = -\frac{g}{c+g}$ and $R_{NS}^{'}(M_S) = -\frac{(n-p) \cdot g}{c + (n-p) \cdot g}$. Moreover, $m_S^{NC} = \frac{p \cdot b}{\left(p^2 + n - p\right) \cdot g + c}$, $m_{NS}^{NC} = \frac{m_S^{NC}}{p}$, $m_S^{ST} = \frac{p \cdot b \cdot c}{(n-p)^2 \cdot g^2 + 2(n-p) \cdot c \cdot g + c \cdot g \cdot p^2 + c^2}$ and

$$m_{NS}^{ST} = \frac{\left(g\cdot(n-p)+c\right)\cdot b}{(n-p)^2\cdot g^2+2(n-p)\cdot c\cdot g+c\cdot g\cdot p^2+c^2}.$$

For both scenarios, the existence and uniqueness condition is always satisfied because $\Psi^M < 0$. No further conditions for an interior equilibrium need to be imposed as mitigation levels are always positive for any (positive) value of parameters.

In the M+A-Game, considering payoff function (10.a) for which $B_{aM} < 0$, we have $B_M = b - g\cdot M - f\cdot a_i$, $B_{MM} = -g < 0$, $B_{Ma} = -f < 0$, $B_a = \beta - f\cdot M$, $B_{aa} = 0$, $C_m = c\cdot m_i$, $C_{mm} = c$, $D_a = d\cdot a_i$, $D_{aa} = d$ and $\Psi^{M+A} = -g + \frac{(-f)^2}{d} = \frac{f^2 - g\cdot d}{d}$. The sign of Ψ depends on the sign of $f^2 - g\cdot d$. From the existence and uniqueness condition under the NC-scenario $\Psi\cdot\left[\frac{p^2}{C_{mm}(m_S)} + \frac{(n-p)}{C_{mm}(m_{NS})}\right] < 1$, noticing that $C_{mm}(m_S) = C_{mm}(m_{NS}) = c$ as well as Ψ are constants, the left-hand side of this inequality increases in p. Hence, using $p = n$, we derive for payoff function (10.a) $c\cdot d - n^2\cdot\left(f^2 - g\cdot d\right) > 0$ for this condition. We notice that this condition is not binding if $f^2 - g\cdot d < 0$ as expected.

For reaction functions, we derive:

$$r_S'(M_{-i}) = \frac{p\cdot\left(f^2-d\cdot g\right)}{c\cdot d - p\cdot\left(f^2-d\cdot g\right)}.,\ R_S'(M_{NS}) = \frac{p^2\cdot\left(f^2-d\cdot g\right)}{c\cdot d - p^2\cdot\left(f^2-d\cdot g\right)},\ r_{NS}'\left(M_{-j}\right) = \frac{f^2-d\cdot g}{c\cdot d - \left(f^2-d\cdot g\right)},$$

$$R_{NS}'(M_S) = \frac{(n-p)\cdot\left(f^2-d\cdot g\right)}{c\cdot d-(n-p)\cdot\left(f^2-d\cdot g\right)} \text{ and } f'(M) = \frac{-f}{d}.$$

For the NC-scenario, we have:

$$m_S^{NC} = \frac{p\cdot(b\cdot d-\beta f)}{c\cdot d - \left(p^2+n-p\right)\cdot\left(f^2-d\cdot g\right)},\ m_{NS}^{NC} = \frac{m_S^{NC}}{p} \text{ and } a_i^{NC} = \frac{\beta\cdot c - \left(n-p+p^2\right)\cdot\left(b\cdot f - g\cdot\beta\right)}{c\cdot d - \left(n-p+p^2\right)\cdot\left(f^2-d\cdot g\right)}.$$

Five conditions, as identified in Bayramoglu et al. (2018), need to hold. We will state all the conditions below, after the analysis of the Stackelberg scenario conditions.

In the ST-scenario, we find:

$m_S^{ST} = \frac{c \cdot p \cdot d \cdot (b \cdot d - \beta \cdot f)}{Z}$, $m_{NS}^{ST} = \frac{\left(c \cdot d - (n-p) \cdot \left(f^2 - g \cdot d\right)\right) \cdot (b \cdot d - \beta \cdot f)}{Z}$ and

$$a_i^{ST} = \frac{(b \cdot f - \beta \cdot g) \cdot (f^2 - d \cdot g) \cdot (n-p)^2 + c \cdot \beta \cdot (c \cdot d + (n-p) f^2) - X}{Z}$$

with $Z := (f^2 - d \cdot g) \cdot \left(-c \cdot d \cdot (p^2 + 2n - 2p)\right) + (f^2 - d \cdot g)^2 \cdot (n-p)^2 + c^2 \cdot d^2$ and

$X := b \cdot f \cdot c \cdot d \cdot (n - p + p^2) + g \cdot \beta \cdot c \cdot d \cdot (2(n-p) + p^2)$.

The numerators of m_S^{ST} and of m_{NS}^{ST} are greater than zero if $(b \cdot d - \beta \cdot f) > 0$ which is exactly the same condition than in the NC-scenario ($C4^{ST} = C4^{NC}$). The remaining term in m_{NS}^{ST} is positive due to the existence and uniqueness condition $C3^{ST} = C3^{NC}$ as stated below. The denominator Z , as shown in detail in the following, is also positive due to the existence and uniqueness condition $C3^{ST} = C3^{NC}$ below. Finally, for adaptation level, a_i^{ST} , the additional condition $C5^{ST}$ is needed to guarantee positive individual adaptation levels, which implies that the numerator of a_i^{ST} is positive.

Recalling that non-signatories' aggregate mitigation reaction function is $R_{NS}'(M_S) = \frac{(n-p) \cdot \Psi}{C_{mm}(m_{NS}) - (n-p) \cdot \Psi}$, the existence and uniqueness condition of the ST-scenario is $\Psi \cdot \left[\frac{p^2 \cdot (1 + R_{NS}')}{C_{mm}(m_S)} + \frac{(n-p)}{C_{mm}(m_{NS})} \right] < 1$. For welfare function (10.a), this condition reads:

$\frac{(f^2 - d \cdot g) \cdot \left(-c \cdot d \cdot (p^2 + 2n - 2p)\right) + (f^2 - d \cdot g)^2 \cdot (n-p)^2 + c^2 \cdot d^2}{c \cdot d \cdot \left(c \cdot d - (n-p) \cdot \left(f^2 - d \cdot g\right)\right)} > 0$. We can show that this condition holds due to the existence and uniqueness condition $C3^{ST} = C3^{NC}$ below. Looking at the numerator, we note that it is identical to the term Z , which is in the denominator of equilibrium mitigation and adaptation levels as stated above. The second term $(f^2 - d \cdot g)^2 \cdot (n-p)^2$ is always positive. Hence, we have to sign $c^2 \cdot d^2 - \left(c \cdot d \cdot (p^2 + 2n - 2p) \cdot (f^2 - d \cdot g)\right)$. Dividing by $c \cdot d$, we obtain $c \cdot d - (p^2 + 2n - 2p) \cdot (f^2 - d \cdot g)$, which is always greater than 0 if $C3^{ST} = C3^{NC}$ as stated below holds. $\left(c \cdot d - (p^2 + 2n - 2p) \cdot (f^2 - d \cdot g)\right.$ takes on the lowest value for $p = n$. Replacing $p = n$, we obtain $\left. C3^{ST} = C3^{NC} \right)$. With this step, we have also proved that $Z > 0$. Looking at the denominator of the condition above, it is also clear that it is positive because of $C3^{ST}=C3^{NC}$ $\left(c\cdot d-(n-p)\cdot\left(f^2-d\cdot g\right)>0 \text{ if } c\cdot d-n^2\cdot\left(f^2-g\cdot d\right)>0\right)$.

Remark: An alternative existence and uniqueness condition in the Stackelberg game is obtained by deriving the second order condition for signatories, which needs to be negative for a maximum. We obtain the following condition:

$(f^2-d\cdot g)\cdot\left(-c\cdot d\cdot(p^2+2n-2p)\right)+(f^2-d\cdot g)^2\cdot(n-p)^2+c^2\cdot d^2>0$ which is the numerator of the condition above.

Taken together, the conditions that need to be satisfied in the M+A-Game for the NC- and the ST-scenario are the following:

$$C1^{ST}=C1^{NC}:\ b-g\cdot M-f\cdot a>0$$

$$C2^{ST}=C2^{NC}:\ \beta-f\cdot M>0$$

$$C3^{ST}=C3^{NC}:\ c\cdot d-n^2\cdot\left(f^2-g\cdot d\right)>0$$

$$C4^{ST}=C4^{NC}:\ b\cdot d-\beta\cdot f>0$$

$$C5^{NC}:\ \beta\cdot c-n^2\cdot\left(b\cdot f-g\cdot\beta\right)>0$$

$$C5^{ST}:\ (b\cdot f-\beta\cdot g)\cdot(f^2-d\cdot g)\cdot(n-p)^2+c\cdot\beta\cdot(f^2\cdot p-f^2\cdot n+c\cdot d)-b\cdot f\cdot c\cdot d\cdot(p^2+n-p)+ \\ g\cdot\beta\cdot c\cdot d\cdot(p^2+2n-2p)>0$$

where $C1$ and $C2$ are required for the General Assumptions I to hold; $C3$ is the existence and uniqueness condition; $C4$ and $C5$ are the mitigation and adaptation non-negativity conditions, respectively. Substituting the highest possible equilibrium mitigation and adaptation levels for given p in $C1$ and $C2$, it turns out that these two conditions are captured by the non-negativity conditions $C4$ and $C5$. Therefore, for both scenarios, only condition $C3$ to $C5$ are relevant, with $C3$ being only relevant if $f^2-g\cdot d>0$, i.e., if $\Psi^{M+A}>0$.

Moving now to the case of $B_{aM}>0$ i.e., considering explicit payoff function (10.b), it turns out that some of the conditions above can be dropped and no additional conditions need to be imposed.

A.3 Proof of Proposition 2

Mitigation Game

We want to prove $M^{NC}(p) > M^{ST}(p)$. Let us assume the opposite, namely: $M^{NC}(p) < M^{ST}(p)$. In the pure mitigation game, $\Psi^{M} < 0$, and therefore $R_{NS}^{'} < 0$. Then from the first order conditions in Table 1 and the General Assumptions I, we have:

$$C_m\left(m_S^{ST}\right) = p \cdot \left[B_M\left(M^{ST}\right) \cdot \left(1 + R_{NS}^{'}\right)\right] < p \cdot \left[B_M\left(M^{NC}\right) \cdot \left(1 + R_{NS}^{'}\right)\right] < p \cdot \left[B_M\left(M^{NC}\right)\right] = C_m\left(m_S^{NC}\right)$$

for signatories and

$$C_m\left(m_{NS}^{ST}\right) = B_M\left(M^{ST}\right) < B_M\left(M^{NC}\right) = C_m\left(m_{NS}^{NC}\right)$$

for non-signatories, assuming $n > p > 1$. It follows that $C_m\left(m_S^{ST}\right) < C_m\left(m_S^{NC}\right)$, $C_m\left(m_{NS}^{ST}\right) < C_m\left(m_{NS}^{NC}\right)$. Therefore, given the convexity of cost functions, $m_S^{ST} < m_{NS}^{NC}$ and $m_{NS}^{ST} < m_{NS}^{NC}$ must hold. Hence, $M^{NC}(p) > M^{ST}(p)$, which contradicts our initial assumption $M^{NC}(p) < M^{ST}(p)$. Thus, we have: $M^{NC}(p) > M^{ST}(p)$. Consequently, $m_{NS}^{NC}(p) < m_{NS}^{ST}(p)$ must hold from the first order conditions of non-signatories and for $M^{NC}(p) > M^{ST}(p)$ it must be that $m_S^{NC}(p) > m_S^{ST}(p)$ holds.

Signatories, as Stackelberg leaders, will be better off (or equally well off) than in the simultaneous move game by axiomatic reasoning, i.e., $w_{NS}^{NC}(p) \geq w_{NS}^{ST}(p)$. Non-signatories, as followers, will have lower benefits due to lower M and higher costs due to higher m_{NS}. Therefore, we have $w_{NS}^{NC}(p) > w_{NS}^{ST}(p)$. Taken together, $p^{*ST} \geq p^{*NC}$ follows from the condition of internal stability (2).

Mitigation-Adaptation Game

In a first step, we differentiate the left-hand side of signatories' first order conditions in mitigation space (7.a) under the ST-scenario with respect to M:

$$\frac{\partial\left[p \cdot \left(B_M\left(M, a_i(M)\right) \cdot \left(1 + R_{NS}^{'}\right)\right)\right]}{\partial M} = p \cdot \left[\left[B_{MM} + B_{Ma} \cdot \frac{\partial a_i}{\partial M}\right] \cdot \left(1 + R_{NS}^{'}\right)\right].$$

assuming third derivatives to be zero. Knowing that $\frac{\partial a_i}{\partial M} = \frac{B_{aM}}{D_{aa} - B_{aa}}$ and rearranging terms, we obtain:

$$\frac{\partial\left[p\cdot\left(B_M\left(M,a_i\left(M\right)\right)\cdot\left(1+R_{NS}^{'}\right)\right)\right]}{\partial M}=p\cdot\left[\Psi\cdot\left(1+R_{NS}^{'}\right)\right].$$

Then, differentiating the benefit side of non-signatories' first order conditions (7.b), we obtain:

$$\frac{\partial\left[\left(B_M\left(M,a_i\left(M\right)\right)\right)\right]}{\partial M}=\left[B_{MM}+B_{Ma}\cdot\frac{\partial a_i}{\partial M}\right]=\Psi\,.$$

The signs of these derivatives depend on the sign of Ψ (as $1+R_{NS}^{'}>0$ is always true). Therefore, for both, signatories and non-signatories, the left-hand side of marginal benefits in their respective first order conditions will decrease (increase) in the level of total mitigation M if $\Psi<(>)0$.

1) Let us assume $\Psi<0$. We want to show $M^{NC}\left(p\right)>M^{ST}\left(p\right)$ but assume the opposite: $M^{NC}\left(p\right)<M^{ST}\left(p\right)$.

From signatories' first order conditions under the NC-scenario (4.a) and under the ST-scenario (7.a), keeping in mind that with $\Psi<0$ the marginal benefits in the first order conditions decreases in total mitigation M, the following holds:

$$C_m\left(m_S^{ST}\right)=p\cdot\left[B_M\left(M^{ST},a_i^{ST}\left(M^{ST}\right)\right)\cdot\left(1+R_{NS}^{'}\right)\right]<p\cdot\left[B_M\left(M^{NC},a_i^{NC}\left(M^{NC}\right)\right)\cdot\left(1+R_{NS}^{'}\right)\right]<$$

$$p\cdot\left[B_M\left(M^{NC},a_i^{NC}\left(M^{NC}\right)\right)\right]=C_m\left(m_S^{NC}\right)$$

For non-signatories, using (4.b) and (7.b) accordingly, we have:

$$C_m\left(m_{NS}^{ST}\right)=B_M\left(M^{ST},a_i^{ST}\left(M^{ST}\right)\right)<B_M\left(M^{NC},a_i^{NC}\left(M^{NC}\right)\right)=C_m\left(m_{NS}^{NC}\right).$$

It follows that $C_m\left(m_S^{ST}\right)<C_m\left(m_S^{NC}\right)$ and $C_m\left(m_{NS}^{ST}\right)<C_m\left(m_{NS}^{NC}\right)$ hold and, therefore, given the convexity of cost functions, $m_S^{ST}<m_{NS}^{NC}$ and $m_{NS}^{ST}<m_{NS}^{NC}$ must hold. These inequalities contradict the assumption $M^{NC}\left(p\right)<M^{ST}\left(p\right)$ so that $M^{NC}\left(p\right)>M^{ST}\left(p\right)$ must hold. Consequently, $m_{NS}^{NC}\left(p\right)<m_{NS}^{ST}\left(p\right)$ must hold from the first order conditions of non-signatories and hence for $M^{NC}\left(p\right)>M^{ST}\left(p\right)$ we must have $m_S^{NC}\left(p\right)>m_S^{ST}\left(p\right)$.

Stackelberg leaders will be better off (or equal well off) than in the simultaneous game by axiomatic reasoning. For non-signatories, the variables that affect their welfare by going from the Nash-Cournot to the Stackelberg scenario are total mitigation (that also affects equilibrium adaptation levels) and individual mitigation. We know that mitigation costs will increase due to higher m_{NS}. In order to evaluate the overall effect, we totally differentiate non-signatories' welfare function:

$$\Delta w_{NS} = \frac{\partial B(M, a_i)}{\partial M} \cdot \Delta M + \frac{\partial B(M, a_i)}{\partial a_i} \cdot \frac{\partial a_i}{\partial M} \cdot \Delta M - \frac{\partial C\left(m_{NS}^{NC}\right)}{\partial m_{NS}} \cdot \Delta m_{NS} - \frac{\partial D\left(a_i^{NC}\right)}{\partial M} \cdot \frac{\partial a_i}{\partial M} \cdot \Delta M$$

and, using the first order conditions in terms of adaptation, $B_a = D_a$, we get:

$$\Delta w_{NS} = B_M \cdot \Delta M - C_m\left(m_{NS}\right) \cdot \Delta m_{NS} .$$

As we know from above that $\Delta M < 0$ and $\Delta m_{NS} > 0$, it follows that non-signatories' welfare will drop when moving from the NC- to the ST-scenario. Therefore, pulling results together for $\Psi < 0$, it holds that $w_S^{CN}(p) < w_{NS}^{ST}(p)$ and $w_{NS}^{CN}(p) > w_{NS}^{ST}(p)$, though nothing can be said about aggregate welfare $W(p)$. From the last two inequalities and considering the internal stability condition (2), it follows that $p^{*ST} \geq p^{*NC}$.

2) We now consider $\Psi > 0$. We want to show $M^{NC}(p) < M^{ST}(p)$.

Due to upward-sloping mitigation reaction functions, we need to consider two possibilities:

$M^{NC}(p) < M^{ST}(p)$ would be compatible only with $m_S^{NC}(p) < m_S^{ST}(p)$ and $m_{NS}^{NC}(p) < m_{NS}^{ST}(p)$;

$M^{NC}(p) > M^{ST}(p)$ would be compatible only with $m_S^{NC}(p) > m_S^{ST}(p)$ and $m_{NS}^{NC}(p) > m_{NS}^{ST}(p)$.

We note that, axiomatically, the Stackelberg leader will receive a higher (or equal) welfare compared to the simultaneous game. To see how signatories' welfare will change when moving from the NC- to the ST-scenario, we total differentiate welfare function (1.b). The result would be the same for non-signatories, except for individual mitigation levels (as done below). We have:

$$\Delta w_S = \frac{\partial B(M, a_i)}{\partial M} \cdot \Delta M + \frac{\partial B(M, a_i)}{\partial a_i} \cdot \frac{\partial a_i}{\partial M} \cdot \Delta M - \frac{\partial C\left(m_S^{NC}\right)}{\partial m_S} \cdot \Delta m_S - \frac{\partial D\left(a_i^{NC}\right)}{\partial M} \cdot \frac{\partial a_i}{\partial M} \cdot \Delta M$$

and, using the information $B_a = D_a$ from the first order conditions with respect to adaptation, we get:

$$\Delta w_S = B_M \cdot \Delta M - C_m\left(m_S\right) \cdot \Delta m_S .$$

From the first order conditions of signatories under the NC-scenario (4.a) in Table 1, we know that $p \cdot B_M = C_m(m_S)$. We also know that in case of upward sloping mitigation reaction functions, $|\Delta M| > |p \cdot \Delta m_s|$ as also non-signatories change their mitigation levels in the same direction as signatories. Therefore, $|B_M \cdot \Delta M| > |C_m(m_S) \cdot \Delta m_s|$ must be true, implying that the benefit effect dominates the cost effect. Consequently, signatories can only increase their welfare by becoming Stackelberg leaders by increasing their mitigation level compared to the NC-scenario. Therefore for $\Psi > 0$, we will have: $M^{NC}(p) < M^{ST}(p)$, $m_S^{NC}(p) < m_S^{ST}(p)$ and $m_{NS}^{NC}(p) < m_{NS}^{ST}(p)$.

For non-signatories, we have:

$$\Delta w_{NS} = B_M \cdot \Delta M - C_m(m_{NS}) \cdot \Delta m_{NS} .$$

From the first order conditions of non-signatories under the NC-scenario (4.b) in Table 1, we know that $B_M = C_m$. We also know that because of upward sloping mitigation reaction functions $|\Delta M| > |\Delta m_{NS}|$ holds and hence $|B_M \cdot \Delta M| > |C_m(m_{NS}) \cdot \Delta m_{NS}|$. Hence, taken together, $w_S^{NC}(p) < w_S^{ST}(p)$ and $w_{NS}^{NC}(p) < w_{NS}^{ST}(p)$ and hence $W^{NC}(p) < W^{ST}(p)$ if $\Psi > 0$.

Finally, we need to show $m_S^{ST}(p) - m_S^{NC}(p) > m_{NS}^{ST}(p) - m_{NS}^{NC}(p)$ and $w_S^{ST}(p) - w_S^{NC}(p)$ $< w_{NS}^{ST}(p) - w_{NS}^{NC}(p)$ which results in $p^{*ST} \leq p^{*NC}$. Looking at signatories' and non-signatories' welfare functions, we can rewrite those as follows: $w_{NS}^{NC} = w_S^{NC} + \left(C\left(m_S^{NC}\right) - C\left(m_{NS}^{NC}\right)\right)$ and $w_{NS}^{ST} = w_S^{ST} + \left(C\left(m_S^{ST}\right) - C\left(m_{NS}^{ST}\right)\right)$. Using this, $w_S^{ST}(p) - w_S^{NC}(p) < w_{NS}^{ST}(p) - w_{NS}^{NC}(p)$ translates into $C\left(m_S^{ST}\right) - C\left(m_S^{NC}\right) > C\left(m_{NS}^{ST}\right) - C\left(m_{NS}^{NC}\right)$. This will be true provided $m_S^{ST}(p) - m_S^{NC}(p) >$ $m_{NS}^{ST}(p) - m_{NS}^{NC}(p)$ holds, which we need to prove. Assume mitigation cost functions to have the following form: $C(m_S) = \frac{c}{\varepsilon} \cdot m_S{}^{\varepsilon}$, $C(m_{NS}) = \frac{c}{\varepsilon} \cdot m_{NS}{}^{\varepsilon}$ with $\varepsilon > 1$ and hence $C_m(m_S) = c \cdot m_S{}^{\varepsilon-1}$, $C_m(m_{NS}) = c \cdot m_{NS}{}^{\varepsilon-1}$. From the first order conditions with respect to mitigation in the NC-scenario we know that $\frac{C_m(m_S)}{p} = C_m(m_{NS})$ and hence $\frac{c \cdot m_S{}^{\varepsilon-1}}{p} = c \cdot m_{NS}{}^{\varepsilon-1}$ and consequently $m_S^{NC\,\varepsilon-1} = p \cdot m_{NS}^{NC\,\varepsilon-1}$ so that $m_S^{NC} = \sqrt[\varepsilon-1]{p} \cdot m_{NS}^{NC}$. From the first order conditions under the ST-scenario for mitigation we know that $\frac{C_m(m_S)}{p\cdot(1+R_{NS}^{'})}=C_m(m_{NS})$. For our polynomial cost function, we obtain $m_S^{ST\,\varepsilon-1}=p\cdot(1+R_{NS}^{'})\cdot m_{NS}^{ST\,\varepsilon-1}$ so that $m_S^{ST}=\sqrt[\varepsilon-1]{p\cdot(1+R_{NS}^{'})}\cdot m_{NS}^{ST}$. Basic algebraic manipulation delivers: $\Delta m^{NC}=m_S^{NC}-m_{NS}^{NC}=\left(\sqrt[\varepsilon-1]{p}-1\right)m_{NS}^{NC}$ and $\Delta m^{ST}=m_S^{ST}-m_{NS}^{ST}=\left(\sqrt[\varepsilon-1]{p\cdot(1+R_{NS}^{'})}-1\right)m_{NS}^{ST}$. Now because of $\Psi>0$, $R_{NS}^{'}>0$ and, therefore, $m_S^{CN}-m_{NS}^{CN}<m_S^{ST}-m_{NS}^{ST}$. Rearranging this inequality, we have: $m_S^{ST}(p)-m_S^{NC}(p)>m_{NS}^{ST}(p)-m_{NS}^{NC}(p)$.

A.4 Proof of Proposition 3

For the NC-scenario, Bayramoglu et al. (2018) demonstrated that in M-Game and in M+A-Game with $\Psi<0$ stable coalition size can be either $p^*=1$ or $p^*=2$. In the M+A-Game with $\Psi>0$ they have shown that $p^*\geq 3$ as internal stability holds for all smaller p but external stability does not. Now, cumbersome calculations (which are available upon request) show that if $n\geq 7$, either $p^*=3$ or $p^*=n$ as confirmed by our simulations.

For the ST-Scenario, in the M- and M+A-Game, we know from Proposition 2 $p^{NC*}\leq p^{ST*}$ if $\Psi<0$. Hence, we need to show that $p^{ST}=2$ is always internally stable as this implies $p^{ST*}\geq 2$. Hence, we compute $IS(p):=w_S^*(p)-w_{NS}^*(p-1)$ in the M- and M+A-Game, substitute $p=2$ and show that $IS(p=2)\geq 0$. As $IS(p)$ is a large term, in particular in the M+A-game, we do not reproduce it here, though results are available upon request. In order to show that $p^{ST*}\in[2,n]$, it suffices to run simulations which delivers p^{ST*} in the entire interval. We have conducted such simulations of which Tables A.1, A2 and A.3 provide a (small) sample. Again, all simulations are available upon request. Finally, in the M+A-game and $\Psi>0$, we know from Proposition 3 that $p^{NC*}\geq p^{ST*}$ and for welfare function (10.a) and (10.b) that $p^{NC*}=\{3,n\}$. Hence, it suffices to produce examples which deliver $p^{ST*}\in\{2,3\}$ provided we can show that $p^{ST*}\neq 1$. This is indeed the case because for $p=2$ the internal stability condition $IS(p=2):=w_S^*(2)-w_{NS}^*(1)\geq 0$ holds and $p^{ST}=1$ is externally unstable. Further notice that the internal stability condition at $p=2$ is identical to the condition of superadditivity, which we know holds from Proposition 4 for any expansion $p-1$ to p, $n \geq p \geq 2$ in the Stackelberg scenario.

Remark: In the M-Game, knowing that the slope of the reaction function increases in g and decreases in c, one can calculate the following limits: $\lim_{g\to\infty} IS(p)=0$ and $\lim_{c\to 0} IS(p)=0$ which proves that any coalition p is internally stable for those limits, including the grand coalition, in which case all smaller coalitions will be externally unstable. A detailed proof is available upon request.

A.5 Proof of Proposition 4

Mitigation Cohesiveness (MCOH)

The difference $M(p)-M(p-1)$ can also be investigated by considering $\frac{\partial M}{\partial p}$, treating p as a continuous variable. Bayramoglu et al. (2018) have shown that in the NC-scenario in the M- and M+A-Game. Following their approach, only minor modifications for the ST-scenario are necessary. Total differentiation of the first order conditions of signatories and non-signatories in the M- and M+A-Game, as provided in Table 1, delivers after rearranging terms, and recalling the difference of the term in the two games (and setting third derivatives to zero):

$$\frac{\partial m_S}{\partial p} = \frac{p \cdot \Psi \cdot \frac{\partial M}{\partial p} \cdot \left(1+R_{NS}^{'}\right)}{C_{mm}\left(m_S\right)} + \frac{B_M \cdot \left(1+R_{NS}^{'}\right)}{C_{mm}\left(m_S\right)}$$

$$\frac{\partial m_{NS}}{\partial p} = \frac{\Psi \cdot \frac{\partial M}{\partial p}}{C_{mm}\left(m_{NS}\right)}.$$

We know that $\frac{\partial M}{\partial p} = m_S + p \cdot \frac{\partial m_S}{\partial p} - m_{NS} + (n-p) \cdot \frac{\partial m_{NS}}{\partial p}$. Substituting $\frac{\partial m_S}{\partial p}$ and $\frac{\partial m_{NS}}{\partial p}$ from above and rearranging terms, we obtain:

$$\frac{\partial M}{\partial p} = \frac{m_S - m_{NS} + \frac{p \cdot B_M \cdot \left(1+R_{NS}^{'}\right)}{C_{mm}\left(m_S\right)}}{1-\Psi \cdot \left[\frac{p^2 \cdot \left(1+R_{NS}^{'}\right)}{C_{mm}\left(m_S\right)} + \frac{(n-p)}{C_{mm}\left(m_{NS}\right)}\right]}$$

The term $\frac{p \cdot B_M \cdot \left(1+R_{NS}^{'}\right)}{C_{mm}\left(m_S\right)}$ is always positive and the denominator is always positive by the General Assumptions II. Hence, if $m_S - m_{NS} \geq 0$, we can conclude $\frac{\partial M}{\partial p} > 0$. We know that $m_S - m_{NS} \geq 0$ if $\Psi > 0$ in which case we can also conclude $\frac{\partial m_{NS}}{\partial p} > 0$ and $\frac{\partial m_S}{\partial p} > 0$ from above. If $\Psi < 0$, $m_S - m_{NS} < 0$ is possible and hence nothing can be generally concluded. In order to show that is possible for some the examples provided in Appendix A.6 are sufficient.

Positive Externality (PEP)

In the context of the NC-scenario, see Bayramoglu et al. (2018). In the ST-scenario, we derive exactly the same condition:

$$\frac{\partial w_{NS}}{\partial p} = B_M \cdot \left[\frac{\partial M}{\partial p} \cdot \left(1 - \frac{\Psi}{C_{mm}\left(m_{NS}\right)}\right)\right]$$

noting that $B_M > 0$ from the General Assumptions I and $\left(1 - \frac{\Psi}{C_{mm}\left(m_{NS}\right)}\right) > 0$ from the sufficient condition of existence and uniqueness as stated in the General Assumptions II. Therefore, $\frac{\partial w_{NS}}{\partial p}$ depends on the sign of $\frac{\partial M}{\partial p}$. Whereas $\frac{\partial M}{\partial p} > 0$ always holds in the NC-scenario, and this is also true in the ST-scenario if $\Psi > 0$ as we know from above, we also know that in the ST-scenario $\frac{\partial M}{\partial p} < 0$ is possible provided $\Psi < 0$ in which case non-signatories do not enjoy a positive but suffer from a negative externality if the coalition is expanded.

Superadditivity (SAD)

We need to show: $p \cdot w_S^*\left(p\right) \geq (>) \left[p-1\right] \cdot w_S^*\left(p-1\right) + w_{NS}^*\left(p-1\right)$ for all p, $2 \leq p \leq n$. For the NC-scenario Bayramoglu et al. (2018) established in both games that a sufficient condition for SAD to hold are (weakly) upward sloping reaction functions, i.e., $\Psi \geq 0$. For the ST-scenario, SAD must hold by axiomatic reasoning. Step 1: Any move from $p-1$ to p implies one more signatory. Keeping total mitigation of the p signatories at the same level than at $p-1$ $\big(p \cdot m_S(p) = \left[p-1\right] m_S(p-1) + m_{NS}(p-1)\big)$, total mitigation cost will have decreased among the p signatories as the first order conditions of mitigation imply cost-effectiveness among signatories. The $n-p$ non-signatories will not have changed their strategies in Step 1. Step 2: The p Stackelberg leaders choose their equilibrium strategies by maximizing their aggregate payoff, controlling the best-response of non-signatories. If they choose different strategies in step 2 compared to step 1, the aggregate welfare of the p signatories must have further increased. For the final move from $p-1=n-1$ to $p=n$, when there are no outsiders left after the move, the SAD-condition is equal to welfare cohesiveness (WCOH) and WCOH for this last move does generally hold because total welfare in the grand coalition is strictly larger than in any other coalition in an externality game by axiomatic reasoning.

Welfare Cohesiveness (WCOH)

If a game is superadditive and exhibits a positive externality throughout, this is sufficient that WCOH holds. Both conditions hold in both scenarios for $\Psi > 0$. In order to prove that WCOH may fail to hold, the examples provided in Appendix A.6 are sufficient.

TABLES

Table A.1: Mitigation Game*

PARAMETERS	r'_{NS}	ICI	NASH-COURNOT						STACKELBERG					
			PEP	SAD	WCOH	MCOH	p*	INI	PEP	SAD	WCOH	MCOH	p*	INI
b=10, g=1, c=1.	-0.5000	0.01	✓	p>17	P>16	✓	1	0	p>30	✓	p>27	p>30	51	0
b=10, g=5, c=1.	-0.8333	0	✓	p>17	P>16	✓	1	0	p>59	✓	p>53	p>59	84	0
b=10, g=100, c=1	-0.9901	0	✓	p>17	P>16	✓	1	0	p>90	✓	p>85	p>90	100	0
b=10, g=0.01, c=1.	-0.0099	32.37	✓	p>14	✓	✓	1	0	✓	✓	✓	✓	3	0.51
b=10, g=0.001, c=1.	-0.0001	426.11	✓	✓	✓	✓	2	1.71	✓	✓	✓	✓	3	4.41
b=10, g=1, c=300.	-0.0033	122.87	✓	✓	✓	✓	2	1.26	✓	✓	✓	✓	3	2.37
b=10, g=1, c=0.1.	-0.9901	0	✓	p>17	P>16	✓	1	0	p>70	✓	p>62	p>70	92	0
b=10, g=0.1, c=350.	-0.0003	1258.81	✓	✓	✓	✓	2	1.89	✓	✓	✓	✓	2	1.79
b=30, g=1, c=1.	-0.5000	0.01	✓	p>17	P>16	✓	1	0	p>30	✓	p>27	p>30	51	0

Table A.2: Mitigation-Adaptation Game, $\Psi < 0$, $B_{aM} < 0$ *

PARAMETERS	r'_{NS}	$f'(M)$	Ψ	ICI	NASH-COURNOT						STACKELBERG					
					PEP	SAD	WCOH	MCOH	p*	INI	PEP	SAD	WCOH	MCOH	p*	INI
Base	-0.4444	-0.20	-0.80	0.01	✓	p>17	p>15	✓	1	0	p>27	✓	p>24	p>27	52	0
b=3	-0.4444	-0.20	-0.80	0	✓	p>17	p>15	✓	1	0	p>32	✓	p>24	p>32	100	0
β=11	-0.4444	-0.20	-0.80	0.01	✓	p>17	p>15	✓	1	0	p>30	✓	p>24	p>30	83	0.01
g=2	-0.6429	-0.20	-1.80	0	✓	p>17	p>16	✓	1	0	p>43	✓	p>28	p>43	99	0
f=0.5	-0.4872	-0.10	-0.95	0.01	✓	p>17	p>15	✓	1	0	p>33	✓	p>25	p>33	92	0.01
c=0.5	-0.6154	-0.20	-0.80	0	✓	p>17	p>16	✓	1	0	p>39	✓	p>34	p>39	66	0
c=50	0.0157	-0.20	-0.80	13.10	✓	p>15	✓	✓	1	0	p>2	✓	p>2	p>2	7	1.05
c=300	-0.0027	-0.20	-0.8	94.19	✓	✓	✓	✓	2	0.82	✓	✓	✓	✓	3	1.50
c=500	-0.0016	-0.20	-0.80	135.57	✓	✓	✓	✓	2	0.79	✓	✓	✓	✓	3	1.77
d=2	-0.3333	-0.50	-0.50	0.02	✓	p>17	p>15	✓	1	0	p>23	✓	p>20	p>23	54	0
d=50	-0.4949	-0.02	-0.98	0.01	✓	p>17	p>15	✓	1	0	p>29	✓	p>27	p>29	51	0

Base simulation parameters: b=10, β=10, g=1, f=1, c=1, d=5. The other simulations analyze the change of one parameter value

Table A.3: Mitigation-Adaptation Game, $\Psi < 0$, $B_{aM} > 0$ *

PARAMETERS	$r_{NS}^{'}$	$f^{'}(M)$	Ψ	ICI	NASH-COURNOT						STACKELBERG					
					PEP	**SAD**	**WCOH**	**MCOH**	**p***	**INI**	**PEP**	**SAD**	**WCOH**	**MCOH**	**p***	**INI**
Base	-0.4444	0.20	-0.80	0.01	✓	p>17	p>15	✓	1	0	p>26	✓	p>24	p>26	46	0
b=3	-0.4444	0.20	-0.80	0.01	✓	p>17	p>15	✓	1	0	p>26	✓	p>24	p>26	46	0
g=2	-0.6429	0.20	-1.80	0	✓	p>17	p>16	✓	1	0	p>40	✓	p>37	p>40	65	0
g=0.21	-0.0099	0.20	-0.01	32.39	✓	p>14	✓	✓	1	0	✓	✓	✓	✓	3	0.51
f=2.23	-0.0054	0.45	-0.01	70.06	✓	p>10	✓	✓	1	0	✓	✓	✓	✓	3	1.44
f=-0.5	-0.4872	0.10	-0.95	0.01	✓	p>17	p>15	✓	1	0	p>29	✓	P>27	p>29	50	0
c=0.5	-0.6154	0.20	-0.8	0	✓	p>17	p>16	✓	1	0	p>38	✓	P>35	p>38	62	0
c=300	-0.0027	0.20	-0.8	121.10	✓	✓	✓	✓	2	1.06	✓	✓	✓	✓	3	2.17
c=500	-0.0016	0.20	-0.8	186.91	✓	✓	✓	✓	2	1.09	✓	✓	✓	✓	3	2.60
d=1.00001	0	1	0	4469.51	✓	✓	✓	✓	2	1.97	✓	✓	✓	✓	2	1.97
d=50	-0.4949	0.02	-0.98	0.01	✓	p>17	p>15	✓	1	0	p>29	✓	P>27	p>29	51	0

Base simulation parameters: b=10, β=10, g=1, f=1, c=1, d=5. The other simulations analyze the change of one parameter value

Table A.4: Mitigation-Adaptation Game, $\Psi > 0$, $B_{aM} < 0$ *

PARAMETERS	$r_{NS}^{'}$	$f^{'}(M)$	Ψ	ICI	NASH-COURNOT						STACKELBERG					
					PEP	**SAD**	**WCOH**	**MCOH**	**p***	**INI**	**PEP**	**SAD**	**WCOH**	**MCOH**	**p***	**INI**
Base	0	-0.33	0.11	176.42	✓	✓	✓	✓	3	0.01	✓	✓	✓	✓	2	0.07
beta=9.286	0	-0.33	0.11	217.07	✓	✓	✓	✓	3	0.63	✓	✓	✓	✓	2	0.09
g=2.11	0	-0.33	0	172.46	✓	✓	✓	✓	3	0.21	✓	✓	✓	✓	2	0.07
c=45001	0	-0.33	0.11	195771.75	✓	✓	✓	✓	3	0.23	✓	✓	✓	✓	2	0.08
d=21.1	0	-0.31	0	190.47	✓	✓	✓	✓	3	0.23	✓	✓	✓	✓	2	0.08
CASE 1	0.0001	-0.33	0.29	0	✓	✓	✓	✓	100	0	✓	✓	✓	✓	2	0
CASE 2	0.0001	-0.99	4.99	0	✓	✓	✓	✓	100	0	✓	✓	✓	✓	2	0
CASE 3	0.0001	-0.99	9.99	0	✓	✓	✓	✓	100	0	✓	✓	✓	✓	2	0

Base simulation parameters: b=10, β=10, g=2, f= 6.5, c=50000 d=20. The other simulations analyze the change of one parameter value.

CASE 1: b=10, β=30, g=1.9, f=6.5999, c=3000, d=19.8. CASE 2: b=10, β=10, g=2 f=6.999999, c=50000, d=7. CASE 3: b=100, β=100, g=5 f=14.999999, c=100000, d=15.

Table A.5: Mitigation-Adaptation Game, $\Psi > 0$, $B_{aM} > 0$ *

PARAMETERS	r'_{NS}	$f'(M)$	Ψ	ICI	NASH-COURNOT						STACKELBERG					
					PEP	SAD	WCOH	MCOH	p*	INI	PEP	SAD	WCOH	MCOH	p*	INI
Base	0	0.33	0.11	18.64	✓	✓	✓	✓	3	0.01	✓	✓	✓	✓	3	0.01
b=1	0	0.33	0.11	1154.69	✓	✓	✓	✓	3	0.87	✓	✓	✓	✓	3	0.87
beta=1	0	0.33	0.11	7402.91	✓	✓	✓	✓	3	5.57	✓	✓	✓	✓	3	5.60
g=2.11	0	0.33	0	1775.05	✓	✓	✓	✓	3	2.11	✓	✓	✓	✓	3	2.11
c=1126	0.0001	0.33	0.11	33.92×10^5	✓	✓	✓	✓	3	3.66	✓	✓	✓	✓	3	3.71
d=18.37	0.0001	0.35	0.29	10.06×10^6	✓	✓	✓	✓	100	10.06×10^6	✓	✓	✓	✓	3	2.25
d=21.1	0	0.31	0	1814.15	✓	✓	✓	✓	3	2.16	✓	✓	✓	✓	3	2.16
CASE 2	0.0001	0.33	0.29	12.58×10^8	✓	✓	✓	✓	100	12.58×10^8	✓	✓	✓	✓	3	0.62
CASE 3	0.0001	0.49	0.49	87.41×10^5	✓	✓	✓	✓	100	87.41×10^5	✓	✓	✓	✓	3	2.09

Base simulation parameters: b=10, β=30, g=2, f= 6.5, c=3000 d=20. The other simulations analyze the change of one parameter value.
CASE 2: b=10, β=10, g=2 f=6.999999, c=50000, d=7. CASE 3: b=1, β=1, g=1 f=2.9999, c=5000, d=6.

* For the general properties of the game (PEP, SAD, WCOH and MCOH), ✓ means that they hold for every coalition of size p. If this is not the case, p values indicated refer to intervals or specific values for which a given condition holds. For any other interval or values of p, the condition fails. If SAD holds for a given p, it means that the move from $p-1$ to p is superadditive. For $f'(M)$, Ψ, ICI and INI we round to two digits and for r'_{NS} we round to 4 digits.

CHAPTER TWO

Accounting for adaptation and its effectiveness in International Environmental Agreements*

Abstract

This paper analyses, within a standard International Environmental Agreement game, the effect of the introduction of adaptation on climate negotiation. The model expands the existing literature by considering a double relation between the two strategies. The common assumption that higher mitigation decreases the marginal benefit of adaptation and vice versa is enriched allowing for the possibility that mitigation, leading to lower and more manageable damages, determines a greater effectiveness of adaptive measures. We show that the general results from the literature still hold. In particular, that the presence of adaptation can determine upward sloping mitigation reaction functions and that, in this case, the grand coalition can form. Nonetheless, large participation can induce substantive welfare gains only if adaptation and mitigation are strategic complements. We show that, co*mplementarity is facilitated when adaptation effectiveness is linked to mitigation levels. At the same time, this condition also shrinks the possibility to observe upward sloping mitigation reaction function. This suggests a key role played also by the "nature" of complementarity between mitigation and adaptation that in some cases can reduce the room for the formation of large and welfare improving climate change agreements.*

Keywords: Climate change, adaptation effectiveness, mitigation-adaptation strategic relation, International Environmental Agreements game.

---

* This paper, co-authored with Professor Francesco Bosello, is submitted, and under revision, in Games and Economic Behavior.

1. Introduction

Contrasting climate change is increasingly recognised as one of the key challenges that our society has to address. The two pillars of climate change policy are mitigation and adaptation. The former acts directly on the cause of the problem, reducing emissions and lowering future climate change. The latter acts on its consequences tackling directly climate impacts. Recognizing the global public bad nature of climate change, the international community started since the beginning of the 90's a complex negotiation process under the umbrella of the United Nation Framework Convention on Climate Change to set and coordinate an equitable, effective and efficient climate action.

These negotiations rounds offered a natural and extremely fertile ground to apply game-and coalition theory models. These have been amply used to predict the potential outcomes of international negotiations on climate change and, in particular, to explain the very reason behind the unsatisfactory achievement of "large" cooperation or of "substantive" environmental effectiveness. The first stream of literature on International Environmental Agreements (IEAs), however, has largely focused on mitigation, while adaptation has been somewhat overlooked. The general results of this literature is that a large (with many countries) and stable (without internal defeaters or external entrants) coalition is not possible to obtain (Hoel 1992, Carraro and Siniscalco 1993, Barrett 1994)[11]. The driver of the outcome is the incentive to free ride. Intuitively: when, or exactly because, the agreement produces large environmental benefits (it is stringent), then the incentive to free ride by single players is too strong to be possibly offset by the remaining members of the coalition. Thus, the coalition shrinks until the incentive to free ride becomes sufficiently small. This typically occurs when few members (two/three) are left. There is a "dual" interpretation of this exit: when large coalitions do form, it is

---

11 The literature on IEA games has extended over the years. For a collection of some of the most influential papers see Finus and Caparros (2015). Many aspects of climate negotiation are analyzed and, in some cases, more optimistic conclusions are achieved (see for instance Barrett and Dannenberg (2012), Eyckmans and Finus (2007) and Finus and Maus (2018)).

because the underlying agreements entail a low incentive to free ride, i.e., they are close to a "do nothing" business as usual.

In the last decade, a more recent stream of literature emerged studying the interplay between mitigation and adaptation in IEAs. This crop of papers is mostly based on empirical Integrated Assessment Modes with only few proposing theoretical framework (Zehaie 2009, Ebert and Welsch 2011 and 2012, Ingham et al. 2013, Eisenack and Kahler 2016). Even fewer have analyzed how the presence of adaptation could affect the size and stability of an IEA (Benchekroun et al. 2011, Buob and Siegenthaler 2011, Marrouch and Chaudhuri 2011, Auerswald et al. 2018, Bayramoglu et al. 2018).

The presence of adaptation might indeed change the nature of the emission reduction game, acting on the incentive to free ride. A first suggestion in this direction is offered by Auerswald et al 2018 showing that in a leader-follower game the commitment to adapt by a group of signatory countries can be a credible signal of a low willingness to mitigate. This can induce outsiders to increase their mitigation effort respect to a "no adaptation" case. Then, total abatement in the presence of adaptation would depend on the shape of the respective mitigation reaction functions. A further refinement in this direction, is proposed by Marrouch and Chaudhuri (2011). They show that, if adaptation is possible, the optimal reply to a potential free rider a climate agreement is more adaptation and not more mitigation. This could reduce the free riding benefit and thus foster the stability of a climate coalition. But there is more. Under given conditions, the presence of adaptation can make mitigation reaction functions upward sloping, therefore an abating coalition acting as Stackelberg leader can induce higher abatement in non-signatories.

The upward sloping nature of mitigation reaction function induced by adaptation is thus a key property for the stability of climate coalition. This issue has been extensively examined by Bayramoglu et al. (2018). One of the interesting points of the paper is that upward sloping mitigation reaction functions occur when the adaptation-mitigation interaction is "sufficiently large", regardless of whether the two strategies are complements or substitutes. Nonetheless, complementarity and substitutability do play a role in determining abatement and welfare levels of the M+A-Game.

The strategic relation between mitigation and adaptation has been examined by a parallel stream of literature, concluding that whether the two are complements or substitutes is mostly an empirical matter. Adaptation and mitigation are commonly seen and modeled as economic substitutes: if the cost of mitigation falls (rises), then the optimal response is to increase (decrease) the level of mitigation and decrease (increase) the level adaptation. However, for instance, Ingham et al. (2013) show that, when adaptation costs depend on the amount of mitigation, the two strategies can be complements. This can occur if adaptation were harder to implement under faster rates of climate change. By reducing emissions, countries not only reduce the rate of climate change, but also facilitate (buy time for) adaptation. As assumed by Ingham et al. (2013), adaptation and mitigation may be linked by more than one relationship and as a result the standard assumption of strategic substitutability may be reversed.

Starting from this idea, the purpose of this paper is to further investigate the interconnections between mitigation-adaptation and the effects that they could have on climate negotiation outcomes.

Our starting point is Bayramoglu et al. (2018), whose analysis we enrich inserting a double connection between adaptation and mitigation[12]. Differently from Ingham et al. (2013), we introduce the second relation in the benefit side of the payoff function, and not in the cost side. In this way, we keep the two strategies interconnected only through benefit effects, while their costs are independent as in Bayramoglu et al. (2018).

The first relation we consider is the standard one: higher levels of adaptation, reducing the marginal damage from climate change, weaken the benefits from mitigation. Conversely, higher levels of

[12] Our modification is therefore introduced in the M+A-Game, while the pure mitigation game does not change from Bayramoglu et al. (2018).

mitigation, by generating less damages to be attenuated, reduce the incentive for protection. The second, is similar in spirit to Ingham et al. (2013). Specifically, we include an adaptation effectiveness parameter that depends on total mitigation level so that higher mitigation determines higher productivity of adaptation. This assumption finds support in the scientific literature. The IPCC AR4 (IPCC 2007) states that as climate change increases, not only do the costs increase but also the options for successful (we read effective) adaptation diminish. Adaptation effectiveness appears closely linked to the rate and magnitude of climate change according to Adger et al. (2007) while Romero-Lankao et al. (2014) state that "several lines of evidence indicate that effective adaptation requires changes in approach and becomes much more difficult if warming exceeds 2°C above preindustrial levels".

In summary, the present paper: i) enriches the interaction between adaptation and mitigation linking the effectiveness of adaptation to mitigation levels, ii) finds how this enriched interaction influences the complementary or substitutability of mitigation and adaptation, iii) studies the effect of this more complex interaction on stable coalition size, mitigation level and welfare performance.

In what follows: section 2 introduces the game theoretical model and its general assumptions, section 3 presents the game, section 4 solves the two stages of the game presenting major results, section 5 concludes.

2. The model

We consider n symmetric players (countries) $i = 1, 2, ..., n$, and 2 different games. In the pure mitigation game (M-Game), our reference, players can only use emissions reduction as a strategy to combat climate change. In the mitigation-adaptation game (M+A-Game), adaptation is introduced. In the M-Game we adopt the same general payoff function of Bayramoglu et al. (2018) given by:

$$w_i(M, m_i) = B_i(M) - C_i(m_i) \tag{1.a}$$

In the M+A-game we introduce a substantial modification, with the payoff function given by:

$$w_i\left(M, m_i, a_i\right) = B\left(M, a_i, \varepsilon\left(M\right)\right) - C_i\left(m_i\right) - D_i\left(a_i\right) \tag{1.b}$$

The total welfare will be the sum of all individual payoffs $W = \sum_{i=1}^{n} w_i$.

In the M-Game (function 1.a), individual payoff is given by the benefit B_i from total mitigation $M = \sum_{i=1}^{n} m_i$, minus the cost C_i of individual mitigation m_i. In the M+A-Game (function 1.b), benefits depend on total mitigation M , individual adaptation a_i and its effectiveness ε that is a function of total mitigation. Adaptation effectiveness $\varepsilon\left(M\right)$ also interacts with adaptation as it will be clear from the general assumptions that characterize the model. The cost functions $C_i\left(m_i\right)$ and $D_i\left(a_i\right)$ depend on the individual mitigation and adaptation levels. Every player i in the M-Game will decide its level of individual mitigation $m_i \in \left[0, \overline{m_i}\right]$, while in the M+A-Game it will also set its individual adaptation level $a_i \in \left[0, \overline{a_i}\right]$. Players have identical individual payoff functions, and thus are ex-ante symmetric. For notation simplicity in the following we drop the index i. However, after the first stage of the game, players can have different objective functions, and therefore mitigation levels, depending on whether they are part of the climate coalition or not. Where appropriate, we stress this difference using the index S for signatories and NS for non-signatories.

All payoff functions are assumed to be continuous with continuous first and second derivatives. Then, the following general assumptions on payoff functions are introduced[13]. The subscripts refer to the kind of the derivative, e.g. $B_M = \frac{\partial B}{\partial M}$, $B_{MM} = \frac{\partial^2 B}{\partial^2 M}$ and $B_{Ma} = \frac{\partial^2 B}{\partial M \cdot \partial a}$.

[13] Those not related to adaptation effectiveness properties are the same of Bayramoglu et al. (2018) .

General assumptions

M-Game:

a) $B_M > 0$, $B_{MM} < 0$, $C_m > 0$, $C_{mm} > 0$.

b) $\lim_{M\to 0} B_M > \lim_{m\to 0} C_m > 0$

M+A-Game:

ai) $\overline{B_M} = B_M + B_\varepsilon \cdot \varepsilon_M > 0$, $\overline{B_{MM}} < 0$, $C_m > 0$, $C_{mm} > 0$.

bi) $\lim_{M\to 0} \overline{B_M} > \lim_{m\to 0} C_m > 0$

c) $B_a > 0$, $B_{aa} \leq 0$, $D_a > 0$, $D_{aa} > 0$.

If $B_{aa} = 0$, then $D_{aa} > 0$ and vice versa: if $D_{aa} = 0$, then $B_{aa} < 0$.

d) $\lim_{a\to 0} B_a > \lim_{a\to 0} D_a > 0$.

e) $\varepsilon_M > 0$, $\varepsilon_{MM} \leq 0$.

f) $B_\varepsilon > 0$, $B_{\varepsilon\varepsilon} \leq 0$, $B_{\varepsilon M} = B_{\varepsilon\varepsilon} \cdot \varepsilon_M \leq 0$.

g) $B_{aM} = B_{Ma} < 0$.

h) $B_{a\varepsilon} = B_{\varepsilon a} > 0$.

i) $\overline{B_{aM}} = \dfrac{\partial B_a\left(M, a, \varepsilon(M)\right)}{\partial M} = \overline{B_{Ma}} = \dfrac{\partial \overline{B_M}}{\partial a} = B_{Ma} + B_{\varepsilon a} \cdot \varepsilon_M >< 0$

With, in assumption ai):

$$\overline{B_{MM}} = \frac{\partial\left(B_M\left(M, a_i, \varepsilon(M)\right) + B_\varepsilon\left(M, a_i, \varepsilon(M)\right) \cdot \varepsilon_M(M)\right)}{\partial M} = B_{MM} + B_{M\varepsilon} \cdot \varepsilon_M + \varepsilon_{MM} \cdot B_\varepsilon + \overline{B_{\varepsilon M}} \cdot \varepsilon_M < 0$$

with $\overline{B_{\varepsilon M}} = \dfrac{\partial B_\varepsilon\left(M, a, \varepsilon(M)\right)}{\partial M} = B_{\varepsilon M} + B_{\varepsilon\varepsilon} \cdot \varepsilon_M = 2B_{\varepsilon\varepsilon} \cdot \varepsilon_M$

Assumptions a), ai), c) and f) set the standard properties of concave benefit and convex cost functions. This configuration guarantees, together with assumptions b), bi) and d), strictly positive mitigation and adaptation equilibrium levels in both games.

Assumptions a), ai) and c) also describe the nature of mitigation and adaptation. Mitigation is a pure public good, as the marginal benefits are affected by total mitigation and not only by individual one. Adaptation is instead a private good: its marginal benefits depend only on individual adaptation levels. Assumption e) represents the new element of this theoretical work. Adaptation effectiveness is defined as an increasing concave (or linear) function of total mitigation level. The logic behind this assumption is the idea that adaptive measures cannot be equally applied regardless of the damage level. In case of catastrophe, adaptation can't be applied: even with an extreme effort, the damage can't be substantively attenuated. Mitigation allows to avoid extreme damages and, as emissions decrease and climate change slows, adaptation starts to be increasingly effective. In this context, ε sets the amount of damage that can be avoided through adaptive measures. It can be conceived as a 0 to 1 variable: for example, value 1 would mean that the damage could be completely absorbed by adaptation, while with a value of 0.5 damage could be reduced by the 50%. The smaller the severity of damages the larger the portion that can be eliminated by adaptation. This is what assumption e) captures.

The relation between mitigation and adaptation is described by assumptions g), e) and h). Assumption g) confirms their standard trade-off: higher levels of adaptation reduce the marginal benefit of mitigation and vice versa. For assumption e), higher mitigation determines higher effectiveness of adaptive measures. This in turn, through assumption h) that describes the relation between adaptation and its effectiveness, increases marginal benefits from adaptation. The overall resulting cross derivative sign is uncertain. It is given by the sum of these two effects as described in assumption i).

3. The game

The International Environmental Agreement game is typically solved as a 2-stage game where countries choose whether to participate or not in a mitigation coalition in the first stage and set their level of mitigation (in the M-game) or of both mitigation and adaptation (in the M+A-game) in the second stage.

Countries who join the coalition P in the first stage are the signatories (S), other countries are the non-signatories (NS). The size of coalition P is indicated by p. If $p = n$ the grand coalition (full cooperation) forms. In an externality game this always represents the social optimum (SO). If $p = 1$ the non-cooperative Nash equilibrium (NE) forms. We refer to all other coalition sizes $1 < p < n$ as partial cooperation.

In a game with a positive externality, a necessary condition for coalition formation is profitability. The coalition of size p is profitable if each signatory gets a higher payoff inside the coalition rather than in a non-cooperative Nash equilibrium. Formally, profitability can be written:

$$w_S(p) > w^{NE}.$$

In the presence of free riding incentives, profitability is not a sufficient condition for the formation of a climate agreement. Stability is also needed. According to the majority of IEAs literature, here we use the open membership rule (D'Aspremont et al. 1983). i.e., players can join and leave the coalition without the consensus of others. Consequently, stability should be both internal and external.

Internal stability requires that none of the signatories would be better off leaving the coalition P while other players stay in the coalition. External stability requires that none of the non-signatories would be better off joining the coalition. Formally, internal stability can be written:

$$w_S^*(p) \geq w_{NS}^*(p-1) \tag{2}$$

while external stability is:

$$w_{NS}^{*}\left(p\right)\geq w_{NS}^{*}\left(p+1\right)$$

When profitability, internal and external stability are jointly satisfied the coalition can be formed. For every coalition P, we allow for a unique equilibrium vector of mitigation (mitigation and adaptation) decisions to exist. We introduce the uniqueness and existence condition in the next section.

In the second stage of the game, the coalition acts as a single player maximizing its payoff and internalizing the positive externality arising from mitigation. In the M-Game, signatories will choose the level of mitigation that maximizes their aggregate payoff, while each non-signatory will instead choose the mitigation level maximizing its individual payoff. In the M+A-Game signatories and non-signatories simultaneously set their mitigation and adaptation levels or, equivalently, they first set their mitigation levels and then, in a second step, adaptation levels[14].

4. Solving the game

The game is solved by backward induction. We start from the second stage analysing mitigation and adaptation first order conditions, equilibrium levels and interdependencies, and then we move to the analysis of the first stage coalition formation.

4.1 Second stage: mitigation and adaptation decisions

4.1.1 Preliminaries

In the second stage, after a coalition P has formed in the first stage, signatories and non-signatories choose their optimal strategies. The first order condition for mitigation levels are given by:

$$p\cdot B_{M}\left(M\right)=C_{m}\left(m\right) \tag{3.a}$$

---

[14] The equivalence of the two games can be easily derived from the FOCs of the M+A-Game, and the demonstration is identical to Bayramoglu et al. (2018), thus we omit it. Final equilibrium levels will be the same as long as adaptation is not chosen before mitigation. The strategic role of anticipatory adaptation has been studied, in a 2-players context, by Zehaie (2009).

in the M-Game and:

$$p \cdot \left[B_M \left(M, a, \varepsilon(M) \right) + B_\varepsilon \left(M, a, \varepsilon(M) \right) \cdot \varepsilon_M \left(M \right) \right] = C_m \left(m \right) \tag{3.b}$$

in the M+A-Game. $p = 1$ for non-signatories and $p \geq 2$ for signatories if a coalition of at least 2 players has formed.

In the M+A-Game we have the additional FOC for adaptation:

$$B_a \left(M, a \right) = D_a \left(a \right) \tag{4}$$

These FOCs, enable to identify some relations between signatories and non-signatories in terms of mitigation, adaptation and welfare levels.

Lemma 1 (Mitigation, adaptation and payoff levels' relations between signatories and non-signatories) If a coalition of size $p \geq 2$ has formed in the first stage, then the following holds:

- $m_S^* \left(p \right) > m_{NS}^* \left(p \right)$
- $a_S^* \left(p \right) = a_{NS}^* \left(p \right)$
- $w_S^* \left(p \right) < w_{NS}^* \left(p \right)$

Proof: see Appendix A.1

The first and last statements hold for both games, while the second one, involving adaptation, only refers to the M+A-Game. The free ride incentive is well explained by the relations of Lemma 1 as non-signatories are better off than signatories in both M- and M+A-Game.

Mitigation FOCs (3.a) and (3.b) define players' reaction functions in the mitigation space. Total mitigation is the individual mitigation level of player *i* plus the mitigation of all other players: $M = m_i + M_{-i}$. In this way, every mitigation FOC defines m_i as a function of M_{-i}. This is the individual reaction function of player i in the mitigation space. In both M- and M+A-Game we can identify the individual mitigation best response functions of signatories and non-signatories respectively as $r_S\left(M_{-i\in P}\right)$ and $r_{NS}\left(M_{-j\notin P}\right)$. Aggregate mitigation reaction of signatories $R_S\left(M_{NS}\right)$ and of non-signatories $R_{NS}\left(M_S\right)$ capture the strategic interaction between the two groups of players.

Moving to adaptation, FOC (4) is equal for all players. Individual adaptation can be expressed as a function of total mitigation. This indicates the reaction functions in the adaptation space $f\left(M\right)=a^*\left(M\right)$ that is equal for signatories and non-signatories.

A preliminary step of the analysis is to define the condition that guarantees to have a unique second stage equilibrium. The following assumption needs to be satisfied.

Additional assumption (Existence and uniqueness conditions of a second stage equilibrium). *In the M+A-Game, let* $\Psi^{M+A}=\overline{B_{MM}}+\frac{\left(B_{aM}+B_{a\varepsilon}\cdot\varepsilon_M\right)^2}{D_{aa}-B_{aa}}$. *If* $\Psi^{M+A}>0$, *then a unique second stage equilibrium exists if:* $\Psi^{M+A}\cdot\left[\frac{p^2}{C_{mm}\left(m_S\right)}+\frac{\left(n-p\right)}{C_{mm}\left(m_{NS}\right)}\right]<1$

Proof: see Appendix A.1

4.1.2 General results

Proposition 1 (slopes of reaction functions in the mitigation space). *Let* $\Psi^{M} = B_{MM}$ *for the M-Game and* $\Psi^{M+A} = \overline{B_{MM}} + \frac{\left(B_{aM} + B_{a\varepsilon} \cdot \varepsilon_M\right)^2}{D_{aa} - B_{aa}}$ *for the M+A-Game. Slopes of individual and aggregate signatories' mitigation reaction function are given respectively by* $r'_S\left(M_{-i\in P}\right) = \frac{p \cdot \Psi}{C_{mm}\left(m_S\right) - p \cdot \Psi}$ *and* $R'_S\left(M_{NS}\right) = \frac{p^2 \cdot \Psi}{C_{mm}\left(m_S\right) - p^2 \cdot \Psi}$. *Slopes of individual and aggregate non-signatories' reaction functions are given respectively by* $r'_{NS}\left(M_{-j\notin P}\right) = \frac{\Psi}{C_{mm}\left(m_{NS}\right) - \Psi}$ *and* $R'_{NS}\left(M_S\right) = \frac{\left(n-p\right)\cdot\Psi}{C_{mm}\left(m_{NS}\right) - \left(n-p\right)\cdot\Psi}$.

The proof follows the same lines of the derivation of existence and uniqueness condition in Appendix A.1 and is therefore omitted. The slopes are derived by totally differentiating the first order condition for mitigation (3.a) and (3.b).

In the M+A-Game, reaction functions can be upward sloping, and this will happen in our model when $\Psi > 0$.

The substantive difference from Bayramoglu et al. (2018) stems from the term determining the slope of mitigation reaction functions in the M+A-Game. In Bayramoglu et al. (2018) it is defined as: $A^{M+A} = B_{MM} + \frac{\left(B_{aM}\right)^2}{D_{aa} - B_{aa}}$.

We can notice that "our" $\Psi^{M+A} < A^{M+A}$. In fact, the positive squared term of Ψ^{M+A} is lower than the one of A^{M+A} as it is composed by two compensating effects. The negative term is also smaller ($\overline{B_{MM}} < B_{MM}$) by the definition of $\overline{B_{MM}}$. Therefore, in our configuration, introducing adaptation effectiveness dependence on total mitigation level, increases the stringency of the condition to have upward sloping mitigation reaction functions in the M+A-game. Accordingly, if a country increases its mitigation commitment it will be easier to be in the case where all other countries react by reducing their mitigation levels. Intuitively, in our M+A-game, mitigation entails a double positive externality. On the one hand all the players will receive a higher direct benefit from total mitigation, on the other hand countries will receive higher benefits from their adaptive measures as they will be more effective. Having a double positive externality, the free riding incentive will be higher and hence mitigation levels will likely be strategic substitutes.

Now, endowed with our new assumption on adaptation effectiveness, we explicit the strategic relations between adaptation and mitigation, i.e., conditions for complementarity or substitutability.

Proposition 2 (Adaptation-mitigation strategic relation) *In the mitigation-adaptation space, the slope of each player's reaction function* $f(M)$ *is given by* $f'(M)=\frac{\partial a}{\partial M}=\frac{B_{aM}+B_{a\varepsilon}\cdot\varepsilon_M}{D_{aa}-B_{aa}}$. *Then, mitigation and adaptation will be substitutes or complements if* $B_{aM}+B_{a\varepsilon}\cdot\varepsilon_M<0$ *or* $B_{aM}+B_{a\varepsilon}\cdot\varepsilon_M>0$ *respectively.*

Proof: see Appendix A.2

Proposition 2 sets the possibility to have strategic complementarity between mitigation and adaptation. Strategic complementarity occurs when the positive term $B_{a\varepsilon}\cdot\varepsilon_M$, originated by the dependence of adaptation effectiveness on mitigation, dominates the standard negative interdependency B_{aM}. This outcome is an alternative formalization of the findings of Ingham et al. (2013) in which complementarity could arise in the special case where adaptation costs were depending on the amount of mitigation[15].

---

15 Other studies in which adaptation and mitigation are found to be strategic complements are Yohe and Strzepeck (2004 and 2007). They focus on tipping points saying that, when impacts from climate

It is worth stressing, an important implication of Propositions 1 and 2. As in Bayramoglu et al. 2018, it is the interaction between mitigation and adaptation that can determine upward sloping mitigation reaction functions, but not the nature of the relation. In Ψ, the term $B_{aM} + B_{a\varepsilon} \cdot \varepsilon_M = \overline{B_{aM}}$ (which determine the nature of interaction between mitigation and adaptation) is squared. Thus, upward sloping mitigation reaction functions can occur either with complementary or substitute mitigation and adaptation. What is needed is that the strategic relation is "sufficiently" strong (large value of $B_{aM} + B_{a\varepsilon} \cdot \varepsilon_M$ in absolute terms).

This said, the strategic relation between adaptation and mitigation does play an important role on the final equilibrium levels of the M+A-Game. Compared to the pure mitigation game, we can conclude, in line with Bayramoglu et al. (2018):

Corollary (Mitigation levels in the M-Game and in the M+A-Game) *Consider an arbitrary coalition of size p formed at the first stage of the game. At the second stage, if adaptation and mitigation are strategic substitutes (complements) then we will have* $m_S^{M+A}(p) < (>) m_S^M(p)$, $m_{NS}^{M+A}(p) < (>) m_{NS}^M(p)$ *and* $M^{M+A}(p) < (>) M^M(p)$.

Proof: see Appendix A.2

Complementarity or substitutability between mitigation and adaptation determines the change in mitigation levels moving from the M-Game to the M+A-Game. When the two strategies are complements, for any given coalition size p, individual and total mitigation levels will be higher in the M+A-Game compared to the pure mitigation game. If they are substitute, mitigation levels will be lower.

change are ''not smooth, non-monotonic and not manageable'', adaptation-mitigation complementarity should be the rule and not the exception.

Compared to Bayramoglu et al. (2018) we can also claim that, under the standard assumption of substitutability, the introduction of the dependence of adaptation effectiveness upon mitigation, induces a lower decrease of mitigation levels in the M+A-Game. This is true for both signatories and non-signatories, and hence also applies to total mitigation (Appendix A.2).

4.2. First stage of the game

In the first stage players choose whether to join the mitigation coalition or not. This is the crucial stage of the game in which cooperation takes form.

4.2.1. General properties

To characterize the incentives to join a coalition P and to analyze the effect on coalition size, on second stage mitigation and welfare levels we first introduce three properties of the game: positive externality property (PEP), superadditivity (SAD) and cohesiveness (COH). We refer to two dimensions of cohesiveness. The standard one that is the welfare dimension (WCOH), and the mitigation dimension (MCOH).

Definition 2: Superadditivity, Positive externality, and Cohesiveness

i) *Superadditivity holds if, for every coalition size* $p \geq 2$ *and for every* $i \in P$ *:*

$$p \cdot w_S^*(p) \geq [p-1] \cdot w_S^*(p-1) + w_{NS}^*(p-1)$$

ii) *Positive externality property holds if, for every* $j \notin P$ *:*

$$w_{NS}^*(p) \geq w_{NS}^*(p-1)$$

iii) *Mitigation cohesiveness holds if, for every coalition size* $p \geq 2$ *and for every* $i \in P$ *:*

$$p \cdot M_S^*(p) + (n-p) \cdot M_{NS}^*(p) \geq [p-1] \cdot M_S^*(p-1) + [n-p+1] \cdot M_{NS}^*(p-1)$$

iv) *Welfare cohesiveness holds if, for every coalition size* $p \geq 2$ *and for every* $i \in P$ *:*

$$p \cdot w_S^*(p) + (n-p) \cdot w_{NS}^*(p) \geq [p-1] \cdot w_S^*(p-1) + [n-p+1] \cdot w_{NS}^*(p-1)$$

Superadditivity and positive externality are linked to stable coalition size. Looking to the internal stability condition (2), it is clear that superadditivity is a necessary condition for coalition stability. If coalition of size p is stable, then the move from $p-1$ to p is superadditive. If we consider a coalition of size $p=2$, then SAD is a sufficient condition for its stability. If the move from $p=1$ to $p=2$ is superadditive, then the coalition $p=2$ is internally stable. PEP refers instead at the positive externality generated by the coalition. It holds when the welfare of players outside the coalition benefit from an enlargement of participation. This property is an obstacle to the stability of large coalitions. As the coalition gets larger, the incentives to stay outside are bigger and therefore the internal stability condition is more difficult to be satisfied.

If both SAD and PEP hold, then welfare cohesiveness holds as well. However, neither of the two cohesiveness properties are associated with coalition stability. They refer instead to the positive effect that higher participation to a climate agreement would have on total mitigation and welfare levels. When cohesiveness holds, larger coalitions would bring higher total mitigation (MCOH) and/or higher total welfare (WCOH). However, if they fail, larger coalitions would bring a loss in terms of total mitigation and/or welfare. For this reason, we should not only look at the number of participants in a mitigation agreement, but also which are the final mitigation and welfare levels[16].

[16] For cohesiveness properties, we don't focus on adaptation. In our model it is directly related to total mitigation level and, therefore, we could link it to mitigation cohesiveness property. When total mitigation increases in the size of coalition p (MCOH holds), adaptation will decrease (increase) if the two strategies are substitutes (complements).

Proposition 3 (general properties of the games) *For what concerns the properties of our pure mitigation and mitigation-adaptation game:*

a) *SAD always holds when* $\Psi > 0$
b) *PEP always holds in both games*
c) *MCOH always holds in both games*

Proof: see Appendix A.3

The general properties of the game do not differ from Bayramoglu et al. (2018). We now check how the enrichment of the relationship adaptation-mitigation, affects three aspects: stable coalition size, mitigation levels and the welfare of coalitions[17].

4.2.2. Explicit payoff functions

To investigate these aspects, we introduce explicit functional forms for the payoff of the two games. That of the M-Game is taken from Bayramoglu et al. (2018):

$$w_i^M = \left(bM - \frac{g}{2} M^2 \right) - \frac{c}{2} m_i^2 \tag{5}$$

In the M+A-Game the payoff function is instead given by:

$$w_i^{M+A} = \left(bM - \frac{g}{2} M^2 \right) + a\left(\beta + \varepsilon - fM\right) - \frac{c}{2} m_i^2 - \frac{d}{2} a^2$$

Where $\varepsilon\left(M\right) = e \cdot M$. Therefore, the extended payoff for the mitigation adaptation game is:

$$w_i^{M+A} = \left(bM - \frac{g}{2} M^2 \right) + a\left(\beta + e \cdot M - fM\right) - \frac{c}{2} m_i^2 - \frac{d}{2} a^2 \tag{6}$$

[17] Adaptation levels can be directly related to total mitigation level given the strategic relations between the two strategies. For coalition evaluation we only focus on global welfare though the same analysis, with very similar conclusions, could be conducted also in total mitigation terms.

The payoff functions are quadratic-quadratic. Parameters b, g, c, β, e, f and d are assumed to be strictly positive.[18] The M- and M+A-Game are directly comparable as the former can be obtained setting adaptation level to zero in the latter.

Appendix A.4 analyzes the payoff functions verifying that all the general assumptions of the model are satisfied, that the existence and uniqueness additional assumption holds and that mitigation and adaptation levels are non-negative.

To quantify welfare impacts we use the indices described in:

Definition 3 (Importance of Cooperation Index and Improvement upon the Nash Equilibrium)

- the Importance of Cooperation Index (ICI) is the percentage welfare improvement moving from the Nash equilibrium (NE) to the social optimum (SO):

$$ICI = \frac{W^{SO} - W^{NE}}{W^{NE}} \cdot 100$$

- the Improvement upon the Nash equilibrium Index (INI) is the percentage welfare improvement brought by the equilibrium coalition compared to non-cooperation:

$$INI = \frac{W^{*}\left(p^{*}\right) - W^{NE}}{W^{NE}} \cdot 100$$

The index ICI measures the need for cooperation. High (low) ICI values indicate large (small) gains from full cooperation compared to the Nash equilibrium. The INI index measures the achievement of a given stable coalition comparing the welfare of that coalition with the non-cooperative situation. The two indices are related as the highest possible value of INI is ICI: they will have the same value in the cases where the social optimum is reached.

---

18 The perhaps less interesting case of linear-quadratic payoff function could be easily obtained setting g=0.

4.2.3. Simulation results

The following results are highlighted.

Result 1 (climate cooperation in a pure mitigation setting)

In the mitigation game, at most a 2 players stable coalition forms. This occurs when the need for cooperation (ICI) is large i.e. when mitigation reaction functions are very flat. The 2-players stable coalition brings just small welfare improvements compared to non-cooperation (small INI) and remains far from the welfare improvements potentially achievable in full cooperation (see Table 1).

The addition of adaptation highlights different results depending on $\Psi < 0$ or $\Psi > 0$ (Tables 2, 3, 4 and 5 in Appendix A.4), namely:

Result 2 (Effects of adaptation in case of $\Psi < 0$)

With $\Psi < 0$, mitigation reaction functions remain downward sloping also in the M+A-Game. Stable coalition size can be at most 2. Thus, there is not a detectable effect of adaptation on the stable coalition size. Nonetheless, when, in some parameterization, the M-Game does not allow stable coalition to form, the M+A-Game allows 2 players stable coalitions. This requires flat mitigation reaction functions and high need for cooperation.

The M+A-Game generally leads to higher welfare equilibrium level than the M-Game. This happens because there are two instruments that can be used to cut climate change costs and, in a first best, two instruments can never perform worse than one. The gains are however small and the INI index is always small. Finally, if the strategies are substitutes (complements), then the total mitigation level are lower (higher) in the M+A-Game than in the M-Game. The mitigation gap increases with the slope of the reaction function in the mitigation-adaptation space i.e., $f'(M)$ is large in absolute terms.

Result 3 (Effects of adaptation in case of $\Psi > 0$)

With $\Psi > 0$ the M+A-Game can present upward sloping mitigation reaction functions. In our configuration, $\Psi > 0$ can be obtained without violating any condition only in those parameterizations where in the M-Game reaction functions are extremely flat and the 2 players stable coalition forms. In the M+A-Game this leads to stable coalitions formed by 3 or all n players. Adaptation thus increases the size of stable coalitions. Looking to the game properties and Proposition 3, the enlargement of stable coalitions can be explained by more favorable condition for cooperation as, with upward sloping mitigation reaction functions, superadditivity always holds. In turn, as MCOH always holds, WCOH follows. Therefore, in this case, enlarging the coalition is always good for total welfare.

When the stable coalition size is 3, total mitigation level in the stable coalitions is higher (lower) when adaptation and mitigation are complements (substitutes) compared to the M-Game. The need of cooperation ICI is high in case of complementarity and low in the case of substitutability.

When the grand coalition forms, the social optimum is reached. Equilibrium mitigation levels will always be higher in the M+A-Game than in the M-Game regardless of the relationship between the two strategies. However, when mitigation and adaptation are substitute, the grand coalition is obtained only in cases where the need for cooperation is very low and, therefore, the improvement from non-cooperation is also small. When the two strategies are complements and the grand coalition forms, the social optimum is reached, and brings very high welfare improvement from the Nash equilibrium.

5. Summary and Conclusion

This paper investigates how the presence of adaptation can influence the size and stability of international climate change agreements, their mitigation and welfare levels. It does so introducing a richer interaction across mitigation and adaptation. Namely, following the suggestions from the empirical and theoretical literature on adaptation, the possibility that adaptation effectiveness depends on the level of mitigation. In the light of this enrichment, the paper also re-examines the nature of the strategic interaction between mitigation and adaptation.

Our analysis confirms that the presence of adaptation can make mitigation reaction functions upward sloping. The interesting point is that this can occur when mitigation and adaptation are either complement or substitute. What is needed is a “sufficiently large” interdependence across the two strategies. However, when adaptation effectiveness is made dependent upon mitigation levels, the possibility to observe this outcome reduces. Counterintuitive it may seem, this is explained by the fact that, with that modification, the positive externality produced by one’s mitigation on others increases: mitigation acts now not only reducing others’ climate change damages directly, but also indirectly, improving their adaptation effectiveness. This reinforces the tendency to reply with less mitigation by one player to more mitigation by another player.

When mitigation reaction functions remain downward sloping, then the presence of adaptation does not enlarge the size of stable coalitions compared to the pure mitigation game. However, there are more stable coalitions. When mitigation reaction functions are upward sloping, adaptation increases the size of stable coalition and can lead to the formation of the grand coalition.

Complementarity or substitutability across mitigation and adaptation, on their turn, impact the abatement and the welfare level of the stable coalitions. Respect to this point, we show formally that complementarity can be originated when adaptation effectiveness depends upon mitigation levels.

Then, when the two strategies are complements, for any given coalition size, individual and total mitigation levels will be higher in the M+A-Game compared to the M-Game. If they are substitute, mitigation levels will be lower. Nonetheless, under the standard assumption of substitutability, the introduction of the dependence of adaptation effectiveness upon mitigation, induces a lower decrease of mitigation level in the M+A-Game.

The M+A-Game generally leads to higher welfare equilibrium levels than the M-Game. This intuitively because players are endowed with an additional instrument to maximize their objective function. With downward sloping mitigation reaction functions and small coalition size, the welfare gains are however small either compared to non-cooperation or the pure mitigation game. With upward sloping mitigation reaction functions and larger stable coalition two situations can emerge. When there is mitigation-adaptation substitutability, large, and possibly, the grand coalition are obtained only in cases where the need for cooperation is very low. Therefore, the welfare improvement from non-cooperation is also small. With mitigation-adaptation complementarity, when the social optimum is reached, it brings very high welfare improvement from non-cooperation.

We can derive two major policy implications from our work. The first is that a joint negotiation on mitigation and adaptation seems always welfare improving. Also when mitigation and adaptation are substitute, and mitigation reaction functions are downward sloping, adaptation increases the number of stable coalitions. This is a potentially positive message in the context of a fragmented regime or a bottom-up approach to climate negotiations like that endorsed by the Paris agreement. For instance, by supporting adaptation in developing countries, developed countries could spur the formation of abating "clubs", that could be a starting point to then achieve further mitigation goals. Moreover if, as it seems possible, mitigation and adaptation are complements and mitigation reaction function are upward sloping, joint negotiation on mitigation and adaptation can lead to the formation of a stable grand coalition. Here however, we flag a second insight which is less positive. Indeed, we showed that complementarity is facilitated when adaptation effectiveness is linked to mitigation level as suggested by many authors. Nonetheless, this same condition also shrinks the possibility to observe upward sloping mitigation reaction function, which is crucial for large abating coalitions. This suggests that also the "nature" of complementarity matters, and that in some cases this can reduce the room for the formation of large and welfare improving mitigation coalition.

Possible extensions of this work could take in account on the one hand players heterogeneity, as considered for instance in Eyckmans et al. (2016) and Lazkano et al. (2016), and on the other hand a more precise specification of the adaptation effectiveness function. In our explicit function, in order to satisfy the general assumptions and the non-negativity conditions, we considered adaptation effectiveness as a linearly increasing function of total mitigation level. A more accurate representation could consider evidences from the literature to specify the concavity of this function. It could be also interesting to test our theoretical outcomes with an empirical application of an Integrated Assessment Model (IAM). IAMs have been already applied for coalition formation analysis in a pure mitigation context[19]. However, despite some models have been extended including adaptation (Agrawala et al. 2011, De Bruin et al 2009), coalition formation has not been analysed.

References:

Agrawala, S., F. Bosello, C. Carraro, K. De Bruin, R. Dellink and E. Lanzi (2011), Plan or React? Analysis of Adaptation Costs and Benefits Using Integrated Assessment Models. “Climate Change Economics”, vol. 2(3), pp. 175–208.

Adger,W.N., S. Agrawala, M.M.Q. Mirza, C. Conde, K. O’Brien, J. Pulhin, R. Pulwarty, B. Smit and K. Takahashi (2007), Assessment of adaptation practices, options, constraints and capacity. “Climate Change 2007: Impacts, Adaptation and Vulnerability”. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, pp. 717-743.

Altamirano-Cabrera, J.C. and M. Finus (2006), Permit Trading and Stability of International Climate Agreements. “Journal of Applied Economics”, vol. 9(1), pp. 19-47.

Auerswald H., K.A. Konrad, and M. Thum (2018), Adaptation, Mitigation and Risk-taking in Climate Policy. Journal of Economics, vol. 124(3), pp. 269-287.

[19] See for instance Altamirano-Cabrera and Finus 2006 and Eyckmans and Finus 2003.

Barrett, S., (1994), Self-Enforcing International Environmental Agreements. "Oxford Economic Papers", 46, pp.878–894.

Barrett, S., A. Dannenberg (2012), Climate negotiations under scientific uncertainty. “Proceedings of the National Academy of Sciences of the United States of America”, vol. 109(43), pp. 17372-17376.

Bayramoglu, B., M. Finus and J.-F. Jacques (2018), Climate Agreements in Mitigation-Adaptation Game. “Journal of Public Economics”, vol. 165, pp. 101-113.

Benchekroun, H., W. Marrouch and A. Ray Chaudhuri (2011), Adaptation Effectiveness and Free-Riding Incentives in International Environmental Agreements. “CentER Discussion Paper Series”, No. 2011-120.

Buob S. and S. Siegenthaler (2011), Does Adaptation Hinder Self-enforcing International Environmental Agreements?, “Department of Economics and Oeschger Centre for Climate Change Research”, University of Bern, Bern.

Carraro, C. and D. Siniscalco (1993), Strategies for the International Protection of the Environment. "Journal of Public Economics", vol. 52(3), pp. 309–328.

D’Aspremont, C., A. Jacquemin, J. Gabszewicz, J. Weymark (1983), On the Stability of Collusive Price Leadership. "The Canadian Journal of Economics", vol. 16(1), pp. 17-25.

De Bruin, K., R. Dellink, R.S.J. Tol (2009), AD-DICE: an Implementation of Adaptation in the DICE Model. “Climatic Change”, vol. 95(1-2), pp. 63-81.

Ebert, U. and H. Welsch (2011), Optimal Response Functions in Global Pollution Problems Can be Upward-sloping: Accounting for Adaptation. “Environmental Economics and Policy Studies”, vol. 13(2), pp. 129–138.

Ebert, U. and H. Welsch (2012), Adaptation and Mitigation in Global Pollution Problems: Economic Impacts of Productivity, Sensitivity, and Adaptive Capacity. “Environmental and Resource Economics”, vol. 52, pp. 49–64.

Eisenack, K. and L. Kähler (2016), Adaptation to Climate Change Can Support Unilateral Emission Reductions. "Oxford Economic Papers", vol. 68(1), pp. 258–278.

Eyckmans, J., S. Fankhauser and S. Kverndokk (2016), Development Aid and Climate Finance. “Environmental and Resource Economics”, vol. 63(2), pp. 429–450.

Eyckmans, J. and M. Finus (2006), Coalition Formation in a Global Warming Game: How the Design of Protocols Affects the Success of Environmental Treaty-making. “Natural Resource Modeling”, vol. 19(3), pp. 323-358.

Eyckmans, J. and M. Finus. (2007), Measures to Enhance the Success of Global Climate Treaties. “International Environmental Agreements”, vol. 7, pp. 73–97.

Finus, M. and A. Caparrós (2015), Game Theory and International Environmental Cooperation. The International Library of Critical Writings in Economics. Edward Elgar, Cheltenham, UK.

Finus, M. and S. Maus. (2008), Modesty May Pay! “Journal of Public Economic Theory, vol. 10, pp. 801–26.

Hoel, M. (1992), International Environment Conventions: The Case of Uniform Reductions of Emissions. “Environmental and Resource Economics”, vol. 2(2), pp. 141-159.

Ingham, A., Ma, J. and Ulph, A.M. (2013), Can Adaptation and Mitigation be Complements? "Climatic Change", vol. 120(1–2), pp. 39–53.

IPCC (2007), Summary for Policymakers. “Climate Change 2007: Impacts, Adaptation and Vulnerability”. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 7-22.

Lazkano, I., W. Marrouch and B. Nkuiya (2016), Adaptation to Climate Change: How Does Heterogeneity in Adaptation Costs Affect Climate Coalitions? “Environment and Development Economics”, vol. 21(06), pp. 812–838.

Marrouch W. and A.R. Chaudhuri (2011), International Environmental Agreements in the Presence of Adaptation. FEEM Working Paper No. 35.2011.

Romero-Lankao, P., J.B. Smith, D.J. Davidson, N.S. Diffenbaugh, P.L. Kinney, P. Kirshen, P. Kovacs, and L. Villers Ruiz (2014), North America. “Climate Change 2014: Impacts, Adaptation, and Vulnerability”. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1439-1498.

Yohe, G. W. and K. Strzepeck. (2004), Climate Change and Water Resources Assessment in South Asia: Addressing Uncertainties. "Climate Change and Water Resources in South Asia", M. M. Q. Mizra, ed., Taylor and Francis, The Netherlands.

Yohe, G. W. and K. Strzepeck. (2007), Adaptation and Mitigation as Complementary Tools for Reducing the Risk of Climate Impacts, "Mitigation and Adaptation Strategies for Global Change", vol. 12, pp. 727–739.

Zehaie, F. (2009) The Timing and Strategic Role of Self-Protection. "Environmental and Resource Economics", vol. 44(3), pp. 337–350.

Appendix

A.1 Proof of Lemma 1 and Additional Assumption

A.1.1 Lemma 1

To analyse the relation between mitigation levels of signatories and non-signatories in the 2 games, we look at mitigation FOCs (3.a) and (3.b). We consider $p=1$ for non-signatories and $p \geq 2$ for signatories. Both in the mitigation and in the mitigation-adaptation game, we can identify the following relation: $\frac{C_m(m_S)}{p} = C_m(m_{NS})$. Therefore, for every coalition size $p \geq 2$, given the convexity of cost functions, each signatory will mitigate more than each non-signatory in both games.

For what concerns adaptation levels in the M+A-Game, from FOC (4) it is clear that all players will choose the same adaptation level that will depend on total mitigation.

From these conclusions it follows that, for both M and M+A-Game (because of symmetry and equal adaptation levels): $w_S^*(p) > w_{NS}^*(p)$ for every $p \in [2, n)$. Signatories will face higher mitigation costs than non-signatories, while benefits will be identical.

A.1.2 Additional Assumption: Existence and Uniqueness Condition

The procedure to derive the sufficient condition for existence and uniqueness of a second stage equilibrium follows Bayramoglu et al. (2018). It is based on the concept of replacement functions, that are useful to check for the existence and uniqueness of the Nash equilibrium in public aggregative games (Cornes & Hartley 2007). Individual replacement functions for signatories and non-signatories are defined as $m_S = g_S(M)$ and $m_{NS} = g_{NS}(M)$ respectively. The aggregate replacement function $G(M)$ consists in the summation of individual replacement functions. Because of players symmetry, we have:

$$\sum_{i=1}^{n} m_i = p \cdot m_S + (n-p) \cdot m_{NS} = M = G(M) = \sum_{i=1}^{n} g_i(M) = p \cdot g_S(M) + (n-p) \cdot g_{NS}(M)$$

If the aggregate replacement function is downward sloping, then it will intersect with the 45-degree line once. This point represents the unique equilibrium mitigation level M^* which, substituting into individual replacement functions, gives the individual equilibrium mitigation levels m_S^* and m_{NS}^*. As we show below, replacement functions are downward sloping (like reaction functions, see Proposition 1) if $\Psi < 0$. With negative slope of replacement functions, then the existence of a unique equilibrium is guaranteed. When the slope is positive ($\Psi > 0$) a sufficient condition to have a unique equilibrium is a slope of the aggregate replacement function lower than 1 for every possible coalition size p , so that the 45-degree line is intersected once. For what concerns adaptation levels, as they can be expressed as continuous and single valued functions of total mitigation, they will be unique whenever total mitigation level is unique.

Now, to derive the sufficient condition for existence and uniqueness equilibrium, we rewrite the mitigation FOCs of the two games considering individual mitigation as a function of total mitigation. In this way, from FOCs (3.a) and (3.b), we obtain the individual replacement functions for the two games:

$$p \cdot \left[B_M \left(M \right) \right] = C_m \left(m_i \left(M \right) \right)$$

$$p \cdot \left[B_M \left(M, a\left(M \right), \varepsilon\left(M \right) \right) + B_\varepsilon \left(M, a\left(M \right), \varepsilon\left(M \right) \right) \cdot \varepsilon_M \left(M \right) \right] = C_m \left(m_i \left(M \right) \right)$$

Total differentiation of these two conditions with respect to M , gives the slope of individual replacement functions for signatories (considering $p \geq 2$) and for signatories (considering $p = 1$). The individual slopes of replacement functions are given by:

$$g'_S \left(M \right) = \frac{p \cdot \Psi}{C_{mm} \left(m_S \right)}$$

$$g'_S \left(M \right) = \frac{\Psi}{C_{mm} \left(m_{NS} \right)}$$

The aggregate replacement function $G\left(M \right)$, is obtained by summing up all individual replacement functions: $G'\left(M \right) = p \cdot g'_S \left(M \right) + \left[n - p \right] \cdot g'_{NS} \left(M \right)$:

$$G'\left(M \right) = \frac{p^2 \cdot \Psi}{C_{mm} \left(m_S \right)} + \frac{\left(n - p \right) \cdot \Psi}{C_{mm} \left(m_{NS} \right)}$$

The existence and uniqueness condition requires an aggregate replacement function slope lower than one. The following condition is therefore obtained:

$$\Psi \cdot \left[\frac{p^2}{C_{mm} \left(m_S \right)} + \frac{\left(n - p \right)}{C_{mm} \left(m_{NS} \right)} \right] < 1$$

For the M-Game we have $\frac{\partial B_M(M)}{\partial M} = \Psi^M = B_{MM}$. For the M+A-Game we have instead. $\Psi^{M+A} = \frac{\partial\left(B_M\left(M, a(M), \varepsilon(M)\right) + B_\varepsilon\left(M, a(M), \varepsilon(M)\right)\cdot \varepsilon_M(M)\right)}{\partial M}$. Solving this derivative, we obtain: $B_{MM} + B_{Ma}\cdot\frac{\partial a}{\partial M} + B_{M\varepsilon}\cdot\varepsilon_M + \left[B_{\varepsilon M} + B_{\varepsilon a}\cdot\frac{\partial a}{\partial M} + B_{\varepsilon\varepsilon}\right]\cdot\varepsilon_M + B_\varepsilon\cdot\varepsilon_{MM}$. Knowing that $\overline{B_{\varepsilon M}} = B_{\varepsilon M} + B_{\varepsilon\varepsilon}\cdot\varepsilon_M$, $\overline{B_{MM}} = B_{MM} + B_{M\varepsilon}\cdot\varepsilon_M + \varepsilon_{MM}\cdot B_\varepsilon + \overline{B_{\varepsilon M}}\cdot\varepsilon_M$ and that, as proved for Proposition 3, $\frac{\partial a}{\partial M} = \frac{B_{aM} + B_{a\varepsilon}\cdot\varepsilon_M}{D_{aa} - B_{aa}}$ replacing and rearranging we obtain $\Psi^{M+A} = \overline{B_{MM}} + \frac{\left(B_{aM} + B_{a\varepsilon}\cdot\varepsilon_M\right)^2}{D_{aa} - B_{aa}}$.

Uniqueness condition is therefore needed only for the M+A-Game, in which $\Psi^{M+A} > 0$ is a possibility. The condition can be written as:

$$\Psi^{M+A}\cdot\left[\frac{p^2}{C_{mm}(m_S)} + \frac{(n-p)}{C_{mm}(m_{NS})}\right] < 1$$

A.2. Proof of Proposition 2 and Corollary

In the M+A-Game, considering the first order condition (4), we can characterize the optimal adaptation level a^* as a function of total mitigation M. We can rewrite the FOC for adaptation as:

$$B_a\left(M, a^*(M), \varepsilon(M)\right) = D_a\left(a^*(M)\right)$$

This FOC express the reaction function in the mitigation-adaptation space $f(M)$. Differentiating it with respect to M and readjusting, we obtain the slope of this function:

$$f'(M) = \frac{\partial a^*}{\partial M} = \frac{B_{aM} + B_{a\varepsilon}\cdot\varepsilon_M}{D_{aa} - B_{aa}}$$

From the general assumptions the denominator is positive and therefore the sign of this equation depends on the numerator. The first term is negative, while the second term is positive. We have strategic substitutability (complementarity) between mitigation and adaptation if $B_{aM} + B_{a\varepsilon}\cdot\varepsilon_M < (>) 0$.

Corollary

We want to analyze the effect of introducing adaptation on total mitigation level. Looking to M+A-Game mitigation FOC (3.b) we see that adaptation has a double effect on it: first, it reduces B_M as $B_{Ma} < 0$ and, second, it increases B_ε as $B_{\varepsilon a} > 0$. Depending on which of these two effects dominates the other we can have higher or lower individual mitigation levels in the M+A-game than in the M-game. If the adaptation effect on the left-hand side of equation (3.b) is positive, then mitigation levels will increase as the cost function is convex while the benefit function is concave. Vice versa, mitigation levels will decrease. The M+A-Game mitigation levels will be higher than M-game mitigation levels if $B_{Ma} + B_{\varepsilon a} \cdot \varepsilon_M > 0$. This condition, looking to the general assumptions of the model, is equal to the numerator of $f'(M)$ which determines its sign. Therefore, we will have higher (lower) mitigation levels in the M+A-game respect to the M-Game if adaptation and mitigation are strategic complements (substitutes). Compared this with the results of Bayramoglu et al. (2018), we can see the difference effect of the introduction of adaptation under the standard assumption of substitutability. The second relation between mitigation and adaptation that we insert, reduces (or invert) the strategic substitutability of the two strategies. It follows that the introduction of adaptation in our model leads to a lower decrease of mitigation levels compared with Bayramoglu et al. (2018).

A.3. Proof of Proposition 3

For the M-Game, the conclusions on superadditivity, PEP and MCOH have already been proved by Bayramoglu et al. (2018). We focus instead on our M+A-Game version. The proofs follow the same procedure of Bayramoglu et al. (2018).

SAD

For superadditivity, we consider the general definition: $p \cdot w_S^*(p) \geq [p-1] \cdot w_S^*(p-1) + w_{NS}^*(p-1)$.
Step 1: On the right-hand side, the equilibrium levels are $M^*(p-1)$, $m_S^*(p-1)$, $m_{NS}^*(p-1)$ and $a^*(p-1)$ with $m_S^*(p-1) > m_{NS}^*(p-1)$. A quantity γ is deducted from all signatories' mitigation levels and a non-signatory's mitigation level is set at the same level so that $\overline{m_{NS}^*}(p-1) = m_S^*(p-1) - \gamma$. All others non-signatories mitigation level is kept at the original value $m_{NS}^*(p-1)$ and γ is chosen to leave the total mitigation $M^*(p-1)$ unchanged. In this way, benefits will not change but costs will drop because, from the properties of the cost function, we can conclude: $p \cdot \left[C\left(m_S^*(p-1) - \gamma\right)\right] < [p-1] \cdot \left[C\left(m_S^*(p-1)\right)\right] + C\left(m_{NS}^*(p-1)\right)$. The payoff obtained from this

marginal change is denoted as $w_S^{(1)}(p)$ and it holds that: $w_S^{(1)}(p) > [p-1] \cdot w_S^*(p-1) + w_{NS}^*(p-1)$. Step 2: if $\Psi \geq 0$, as it will be clear when we analyze mitigation levels' change in coalition size p in the next proof, for all other non-signatories it holds that $m_{NS}^*(p-1) \leq m_{NS}^*(p)$ and, because $\frac{\partial w_s}{\partial m_{NS}} > 0$, from step 2 we have $w_S^{(2)} > w_S^{(1)}$. Step 3: $\max p \cdot w_S(p) = p \cdot w_S^*(p) > p \cdot w_S^{(2)}$.Going back to the SAD condition, we see that moving from the right-hand side to the left-hand side the aggregate payoff of the enlarged coalition increases. This is because total costs decrease (step 1), all outsiders increase their mitigation level (step 2) and the players in the enlarged coalition can freely choose mitigation and adaptation (step 3).

Mitigation Cohesiveness

We want to find how individual and total mitigation levels change in the size of the coalition. Considering p as a continuous variable, what we are looking for is the sign of the derivative $\frac{\partial M}{\partial p}$.

For the M+A-Game, we can rewrite mitigation FOC (3.b) considering equilibrium mitigation levels as a function of coalition size p. Doing so, we obtain the following conditions respectively for signatories and non-signatories:

$$p \cdot \left[B_M\left(M(p), a\left(M(p)\right), \varepsilon\left(M(p)\right)\right) + B_\varepsilon\left(M(p), a\left(M(p)\right), \varepsilon\left(M(p)\right)\right) \cdot \varepsilon_M\left(M(p)\right) \right] = C_m\left(m_S(p)\right)$$

$$B_M\left(M(p), a\left(M(p)\right), \varepsilon\left(M(p)\right)\right) + B_\varepsilon\left(M(p), a\left(M(p)\right), \varepsilon\left(M(p)\right)\right) \cdot \varepsilon_M\left(M(p)\right) = C_m\left(m_{NS}(p)\right)$$

Differentiating these two conditions with respect to p and rearranging, we obtain how individual mitigation levels of signatories and non-signatories change in the coalition size. Taking in mind the difference in the term Ψ of the two games, we obtain:

$$\frac{\partial m_S}{\partial p} = \frac{\partial M}{\partial p} \cdot \frac{p \cdot \Psi}{C_{mm}(m_S)} + \frac{B_M + B_\varepsilon \cdot \varepsilon_M}{C_{mm}(m_S)}$$

$$\frac{\partial m_{NS}}{\partial p} = \frac{\partial M}{\partial p} \cdot \frac{\Psi}{C_{mm}(m_{NS})}$$

We know that $\frac{\partial M}{\partial p} = m_S + p \cdot \frac{\partial m_{NS}}{\partial p} - m_{NS} + (n-p) \cdot \frac{\partial m_{NS}}{\partial p}$. Substituting $\frac{\partial m_S}{\partial p}$ and $\frac{\partial m_{NS}}{\partial p}$, and rearranging we obtain:

$$\frac{\partial M}{\partial p}=\frac{m_S-m_{NS}+\dfrac{p\cdot\left[B_M+B_\varepsilon\cdot\varepsilon_M\right]}{C_{mm}\left(m_S\right)}}{1-\Psi\cdot\left[\dfrac{p^2}{C_{mm}\left(m_S\right)}+\dfrac{\left(n-p\right)}{C_{mm}\left(m_{NS}\right)}\right]}$$

m_S-m_{NS} is always positive as demonstrated for Lemma 1, the term $\frac{p\cdot\left[B_M+B_\varepsilon\cdot\varepsilon_M\right]}{C_{mm}\left(m_S\right)}$ is always positive from the general assumptions and the denominator is always positive because of the uniqueness and existence condition. Therefore, we can conclude the following:

$$\frac{\partial M}{\partial p}>0$$

$\frac{\partial m_s}{\partial p}><(>)0$ and $\frac{\partial m_{ns}}{\partial p}<\left(>\right)0$ if $\Psi<\left(>\right)0$.

PEP

Looking to positive externality property, we can consider coalition size p as a continuous variable and analyze the sign of $\frac{\partial w_{ns}}{\partial p}$. When $\frac{\partial w_{ns}}{\partial p}>0$, PEP holds. We can rewrite the payoff of a non-signatory as $w_{ns}=B\left(M\left(p\right),a\left(M\left(p\right)\right),\varepsilon\left(M\left(p\right)\right)\right)-C\left(m_{NS}\left(p\right)\right)-D\left(a\left(M\left(p\right)\right)\right)$.

We differentiate it with respect to p knowing from the FOCs 3.b and (4) respectively that $B_M+B_\varepsilon\cdot\varepsilon_M=C_m$ and $B_a=D_a$. Rearranging and substituting $\frac{\partial m_{NS}}{\partial p}=\frac{\partial M}{\partial p}\cdot\frac{\Psi}{C_{mm}\left(m_{NS}\right)}$ we obtain:

$$\frac{\partial \Pi_{ns}}{\partial p}=\left[B_M+B_\varepsilon\cdot\varepsilon_M\right]\cdot\left[\frac{\partial M}{\partial p}\cdot\left(1-\frac{\Psi}{C_{mm}\left(m_{NS}\right)}\right)\right]$$

$B_M+B_\varepsilon\cdot\varepsilon_M>0$ from the general assumptions of the model and $\left(1-\frac{\Psi}{C_{mm}\left(m_{NS}\right)}\right)>0$ from the existence and uniqueness condition. Since we have proved that $\frac{\partial M}{\partial p}>0$, then PEP always holds.

A.4. Preliminaries of explicit payoff functions

For the explicit payoff function of the pure mitigation game, all the assumptions of the model and the existence and uniqueness condition are always satisfied irrespective of the (positive) parameters values as in Bayramoglu et al. (2018).

Moving to the M+A-Game, we look to payoff function (6). Doing some simple partial derivatives, we find that $B_M = b - g \cdot M - f \cdot a$, $\overline{B_M} = b - g \cdot M + (e-f) \cdot a$, $B_{MM} = \overline{B_{MM}} = -g < 0$, $B_{Ma} = -f < 0$, $\overline{B_{Ma}} = -f + e >< 0$, $B_a = \beta + (e-f) \cdot M$, $B_{aa} = 0$, $B_{a\varepsilon} = 1$, $B_\varepsilon = a$, $B_{\varepsilon\varepsilon} = 0$, $\varepsilon_M = e$, $\varepsilon_{MM} = 0$, $C_m = c \cdot m_i$, $C_{mm} = c$, $D_a = d \cdot a$, $D_{aa} = d$.

From $\Psi^{M+A} = \overline{B_{MM}} + \frac{(B_{aM} + B_{a\varepsilon} \cdot \varepsilon_M)^2}{D_{aa} - B_{aa}}$ we get $\Psi = -g + \frac{(e-f)^2}{d} = \frac{(e-f)^2 - g \cdot d}{d}$. The sign of Ψ depends on the difference $(e-f)^2 - g \cdot d$. The additional assumption on existence and uniqueness condition is: $\Psi \cdot \left[\frac{p^2}{C_{mm}(m_S)} + \frac{(n-p)}{C_{mm}(m_{NS})} \right] < 1$. For our explicit payoff function, the condition is most restrictive when $p = n$ and therefore we obtain: $\frac{\Psi \cdot n^2}{C_{mm}(m_s)} < 1$ and, substituting the values and rearranging: $c \cdot d - n^2 \cdot \left(g \cdot d - (e-f)^2 \right) > 0$.

For reaction functions, we derive:

$$r_S^{'}(M_{-i}) = \frac{p \cdot \left((e-f)^2 - d \cdot g \right)}{c \cdot d - p \cdot \left((e-f)^2 - d \cdot g \right)}, \qquad R_S^{'}(M_{NS}) = \frac{p^2 \cdot \left((e-f)^2 - d \cdot g \right)}{c \cdot d - p^2 \cdot \left((e-f)^2 - d \cdot g \right)},$$

$$r_{NS}^{'}(M_{-j}) = \frac{(e-f)^2 - d \cdot g}{d \cdot g - (e-f)^2 + c \cdot d}, \quad R_{NS}^{'}(M_S) = \frac{(n-p) \cdot \left((e-f)^2 - d \cdot g \right)}{c \cdot d - (n-p) \cdot \left((e-f)^2 - d \cdot g \right)} \text{ and } f^{'}(M) = \frac{e-f}{d}$$

Reaction functions in the mitigation space will be downward (upward) sloping if $\Psi = \frac{(e-f)^2 - g \cdot d}{d} < (>) 0$. Reaction functions in the mitigation-adaptation space will be downward (upward) sloping if $e - f < (>) 0$, i.e., mitigation and adaptation will be strategic substitutes (complements) if $e - f < (>) 0$.

Looking to mitigation and adaptation levels, we obtain: $m_S = \frac{p \cdot c \cdot d \cdot \left(b \cdot d + \beta \cdot (e-f)\right)}{\left(p^2+n-p\right) \cdot \left(d \cdot g - (e-f)^2\right) + c \cdot d}$,

$m_{NS} = \frac{c \cdot d \cdot \left(b \cdot d + \beta \cdot (e-f)\right)}{\left(p^2+n-p\right) \cdot \left(d \cdot g - (e-f)^2\right) + c \cdot d}$ and $a = \frac{\left(p^2+n-p\right) \cdot \left(b \cdot (e-f) + \beta \cdot\right) g + \beta \cdot c}{\left(p^2+n-p\right) \cdot \left(d \cdot g - (e-f)^2\right) + c \cdot d}$.

The denominator of mitigation and adaptation levels is always positive because of the existence and uniqueness condition. Hence, the following conditions need to be satisfied in the parameters' choice for the M+A-Game simulations:

$$C1: b - g \cdot M + (e-f) \cdot a > 0$$

$$C2: \beta + (e-f) \cdot M > 0$$

$$C3: c \cdot d - n^2 \cdot \left(g \cdot d - (e-f)^2\right) > 0$$

$$C4: b \cdot d + \beta \cdot (e-f) > 0$$

$$C5: \left(p^2+n-p\right) \cdot \left(b \cdot (e-f) + \beta \cdot\right) g + \beta \cdot c > 0$$

Where $C1$ and $C2$ are required for the general assumptions to hold, $C3$ is the most restrictive existence and uniqueness condition, $C4$ and $C5$ are respectively the mitigation and adaptation non-negativity conditions.

Looking to $C1$ and $C2$, they take the most restrictive values in case of $f > e$ and with social optimum mitigation level (highest value) and Nash equilibrium adaptation level (highest value since with $f > e$ mitigation and adaptation are substitutes). Substituting these values, we see that these two conditions are respectively captured by $C4$ and $C5$. There will be three conditions left that we need to observe in performing simulations: $C3$, $C4$ and $C5$. In case of $f < e$, all the conditions would be less stringent.

Tables

Table 1: Mitigation game*

PARAMETERS	r'_{NS}	ICI	SAD	PEP	MCOH	WCOH	p*	$M\left(p^{*}\right)$	INI
b=10, g=1, c=1.	-0.5000	0.01	p>17	✓	✓	P>16	1	9.90	0
b=10, g=10, c=1.	-0.9091	0	p>17	✓	✓	P>16	1	0.99	0
b=10, g=100, c=1	-0.9901	0	p>17	✓	✓	P>16	1	0.09	0
b=10, g=0.001, c=1.	-0.0009	426.11	✓	✓	✓	YES	2	925.60	1.71
b=10, g=1, c=0.1.	-0.9901	0	p>17	✓	✓	P>16	1	9.99	0
b=10, g=1, c=300.	-0.0033	122.86	✓	✓	✓	✓	2	2.54	1.26
b=10, g=1, c=3000.	-0.0003	1117.86	✓	✓	✓	✓	2	0.33	1.88
b=10, g=1.9, c=3000.	-0.0006	650.91	✓	✓	✓	✓	2	0.32	1.80
b=30, g=1, c=1.	-0.5000	0.01	p>17	✓	✓	P>16	1	29.70	0

Table 2: Effect of adaptation, first set of simulations: $\Psi < 0$, $f'(M) >< 0$ *

PARAMETERS	$r_{NS}^{'}$	$f^{'}(M)$	**Ψ**	**ICI**	**SAD**	**PEP**	**MCOH**	**WCOH**	**p***	$M(p^{*})$	$a(p^{*})$	**INI**
M-Game: b=10, g=1, c=1.	-0.5000	x	x	0.01	p>17	✓	✓	P>16	1	9.90	X	0
M+A-Game: Effect of considering adaptation												
β=10, e=1, f=2, d=5	-0.4444	-0.20	-0.80	0.01	p>17	✓	✓	p>15	1	9.88	0.02	0
β=10, e=1, f=1.5, d=5	-0.4872	-0.10	-0.95	0.01	p>17	✓	✓	p>15	1	9.37	1.06	0
β=10, e=1, f=1.01, d=5	-0.4999	0	-0.99	0.01	p>17	✓	✓	p>16	1	9.88	1.98	0
β=10, e=1, f=2, d=50	-0.4949	-0.02	-0.98	0.01	p>17	✓	✓	p>15	1	9.89	0	0
β=10, e=1, f=2, d=1.000001	0	-0.99	0	0	✓	✓	✓	✓	2	0	9.99	0
β=10, e=2, f=1, d=5	-0.4444	0.20	-0.80	0.01	p>17	✓	✓	p>15	1	14.81	4.96	0
β=10, e=3, f=1, d=5	-0.1666	0.40	-0.20	0.22	p>17	✓	✓	✓	1	66.66	28.66	0
β=10, e=1.01, f=1, d=5	-0.4999	0	-0.99	0.01	p>17	✓	✓	p>16	1	9.92	2.02	0
β=10, e=2, f=1, d=50	-0.495	0.02	-0.98	0.01	p>17	✓	✓	p>15	1	10.30	0.41	0
β=10, e=1, f=2, d=1.000001	0	0.99	0	4874.37	✓	✓	✓	✓	2	0	9.99	1.98

Table 3: Effect of adaptation, third set of simulations: $\Psi >< 0$, $f\,'(M) >< 0$ *

SIMULATIONS	$r_{NS}^{'}$	$f^{'}(M)$	Ψ	ICI	SAD	PEP	MCOH	WCOH	p*	$M(p^*)$	$a(p^*)$	INI
M-Game: b=10, g=1, c=3000	-0.0003	x	x	1117.86	✓	✓	✓	✓	2	0.33	x	1.88
M+A-Game: Effect of considering adaptation												
β=30, e=1, f=2.6, d=5	-0.0002	-0.32	-0.49	0.12	✓	✓	✓	✓	2	0.01	6.00	0.0001
β=30, e=1, f=1.1, d=5	-0.0003	-0.02	-0.99	33.67	✓	✓	✓	✓	2	0.31	5.99	0.056
β=30, e=1, f=2, d=50	-0.0032	-0.02	-0.98	268.99	✓	✓	✓	✓	2	0.31	0.59	0.44
β=30, e=3.2, f=1, d=5	0	0.44	-0.03	735.26	✓	✓	✓	✓	2	0.78	6.34	0.33
β=30, e=1.1, f=1, d=5	-0.0003	0.02	-0.99	42.47	✓	✓	✓	✓	2	0.35	6.01	0.07
β=30, e=2, f=1, d=50	-0.0032	0.02	-0.98	321.37	✓	✓	✓	✓	2	0.35	0.61	0.53
β=30, e=1, f=4.2, d=10	0	-0.32	0.02	0.63	✓	✓	✓	✓	3	0.14	2.99	0.0006
β=30, e=1, f=4.3, d=10	0	-0.33	0.09	0.05	✓	✓	✓	✓	3	0.04	2.99	0.00004
β=30, e=1, f=4.5, d=12	0	-0.29	0.02	7.31	✓	✓	✓	✓	3	0.04	2.48	0.008
β=30, e=4.3, f=1, d=10	0	0.33	0.09	1182.87	✓	✓	✓	✓	3	0.69	3.22	1.31
β=30, e=4.6055, f=1, d=10	0.0001	0.36	0.29	$9.83X10^6$	✓	✓	✓	✓	100	$5.6X10^5$	$2.03X10^5$	$9.83X10^6$
β=30, e=6, f=1, d=20	0.0001	0.25	0.25	9309.66	✓	✓	✓	✓	3	0.62	1.65	1.88
β=30, e=6.099, f=1, d=20	0.0001	0.25	0.29	$4.74X10^5$	✓	✓	✓	✓	100	$1.77X10^6$	$4.52X10^5$	$4.74X10^5$

Table 4: Effect of adaptation, fourth set of simulations: $\Psi > 0$**,** $f'(M) >< 0$**)***

SIMULATIONS	r'_{NS}	$f'(M)$	**Ψ**	**ICI**	**SAD**	**PEP**	**MCOH**	**WCOH**	**p***	$M(p^*)$	$a(p^*)$	**INI**
M-Game: b=10, g=1.9, c=3000.	-0.0006	x	x	650.91	✓	✓	✓	✓	2	0.32	x	1.80
M+A-Game: Effect of considering adaptation												
β=30, e=1, f=7.5, d=20	0	-0.33	0.21	1.58	✓	✓	✓	✓	3	0.01	1.49	0
β=30, e=1, f=7.5999, d=19.8	0.0001	-0.33	0.29	0	✓	✓	✓	✓	100	2.27	0.76	0
β=30, e=7.5, f=1, d=20	0.0001	0.33	0.21	6228.60	✓	✓	✓	✓	3	0.70	1.72	2.19
β=30, e=7.5999, f=1, d=19.8	0.0001	-0.33	0.29	8.28X10^5	✓	✓	✓	✓	100	3X10^5	10^5	8.28X10^5

* For the general properties of the game (PEP, SAD, WCOH and MCOH), ✓ means that they hold for every coalition of size p. If this is not the case, p values indicated refer to intervals or specific values for which a given condition holds. For any other interval or values of p, the condition fails. If SAD holds for a given p, it means that the move from $p-1$ to p is superadditive. For $f'(M)$, Ψ, ICI, INI, $M(p^*)$ and $a(p^*)$ we round to two digits and for r'_{NS} we round to 4 digits.

CHAPTER THREE

International Environmental Agreements, Mitigation and Adaptation: an Integrated Assessment Analysis

---

Abstract

This paper focuses on the effects of the introduction of adaptation in International Environmental Agreements (IEAs) games. We empirically test the theoretical results of the literature with an application of an Integrated Assessment Model (IAM). We update a version of the Nordhaus' RICE model taking 2015 as starting year and considering 2 games: the pure mitigation game (M-Game) in which mitigation is the only policy option to reduce future climate damages, and the mitigation-adaptation game (M+A-Game) in which adaptation is also an option. We find that mitigation and adaptation are, as expected, strategic substitutes. Furthermore, as theoretical results suggest, we find that the introduction of adaptation alters the strategic relation between players mitigation (emissions) levels. In the M+A-Game, mitigation reaction functions can be upward sloping. Consequently, free riding incentives are reduced, and coalition stability is easier to achieve. Without monetary transfers, the M-Game does not lead to any stable coalition, while in the M+A-Game coalitions formed by 2 out of 6 macro regions can form. Allowing for the possibility of optimal transfers, both games lead to full cooperation. However, we note that, considering adaptation, more coalitions become stable and, in the case of social optimum, the difference between the surplus of cooperation and the free riding incentives is enlarged.

Keywords: International environmental agreements – Mitigation-adaptation game – Integrated assessment modeling

1. Introduction

The success of climate cooperation is of paramount interest for our society. Game theoretic modeling of International Environmental Agreements (IEAs) and applied Integrated Assessment models (IAMs) are the two main instruments used to explore the prospects of international negotiations for the reduction of greenhouse gas emissions.

Difficulties in reaching a large and effective agreement have been extensively investigated by game theoretic models. In a pure mitigation (or emissions) game, i.e., games without the option to adapt to combat climate change, the common result is that only small stable coalitions can form (Carraro and Siniscalco 1993, Hoel 1992 and, for a recent survey on IEAs literature, Finus and Caparros 2015). More optimistic results are found when players are asymmetric and when monetary transfers within the coalition members are possible (Finus and McGinty 2018).

The introduction of adaptation can change substantively the picture. Intuitively, when adaptation is possible, the optimal reply to a potential free rider can be more adaptation and not more mitigation. This could reduce the free riding benefit and thus foster the stability of a climate coalition. Furthermore, if the interaction between mitigation and adaptation is sufficiently large, independently upon the fact that mitigation and adaptation are strategic complements or substitutes mitigation reaction functions can be upward sloping (Ebert and Welsh 2011, 2012; Bayramoglu et al. 2018). Under this condition stable coalition size will increase compared to the pure mitigation game, and even the grand coalition can form (Bayramoglu et al. 2018 and Rubio 2018).

The purpose of this paper is to introducing data, test these theoretical results through an Integrated Assessment Model (IAM) and verify their robustness when more realistic assumptions concerning cost and benefit of climate change policies, country asymmetries, dynamic settings are accounted for.

IAMs couple a climate and an economic module in order to analyse the interactions between the socio-economic and the environmental sphere[20]. IAMs have been extensively applied to the study of climate policies (Kriegler et al. 2014 and Moore and Diaz 2015).

In a more limited number of cases they have been applied to study the problem of the formation of climate change coalitions, but with a focus on mitigation. In a pure mitigation context, IAM analysis suggests, in accordance with the theory, that stable coalition size is generally small (Finus 2008), but that participation can be increased using various monetary transfer schemes or mitigation burden sharing rules (Eyckmans and Finus 2007 and Altamirano-Cabrera et al. 2008).

Adaptation has been introduced in IAMs more recently (De Bruin et al. 2009b and Agrawala et al. 2011). However, to our knowledge, AD-IAMs have not been applied for the analysis of stability of international environmental agreements. Instead, the main fields of application have been the analysis of adaptation cost and benefits and the optimal policy mix between adaptation and mitigation (De Bruin et al. 2009a and Bosello et al. 2010)[21].

To perform our analysis, we use an updated version of the RICE-96 model (Nordhaus and Yang 1996). Model enrichments include: re-calibration of the model damage function by using the more recent insights from the literature, re-initialization of the model to year 2015; introduction of an adaptation module, benchmarking of the model on the shared social economic pathway n°2.

In particular, adaptation is introduced as an additional control “flow” variable that competes with mitigation and investment in the utility maximization process.

---

20 The first model is the aggregate DICE and it dates back to the early 90’s (Nordhaus 1993). Over the years, many regionalized versions (RICE) have been developed (Nordhaus and Young 1996, Nordhaus 2010) and other IAMs have been modeled (see for instance Tol 1995 and Bosetti et al. 2006).

21 For a survey of the main IAM studies on adaptation cost and benefits and on the optimal policy mix see Agrawala et al. 2011.

We also consider multiple coalition structures. That is, we examine how adaptation can influence size and stability of climate coalitions considering all the single coalition structures that can form starting from the 6 geo-political blocks of the model. While on the one hand considering 6 regions can be a limiting aspect of our analysis, on the other hand it suffices to verify the main properties of theoretical models keeping the integrated assessment analysis more manageable.

In the pure mitigation game, as theory suggests, we find that mitigation (emissions) reaction functions are downward sloping, i.e., when mitigation (emissions) of others increases (decrease) the best response will be to decrease mitigation (increase emissions). For what concerns stable coalitions, no coalitions form without the possibility of transfers. Considering optimal transfers large coalitions, up to the grand coalition, can become stable.

Adaptation and mitigation (emissions) strategies are substitutes, i.e., when total emissions increase so does individual adaptation and vice versa. The introduction of adaptation, as in the theoretical models, leads to changes in the strategic relationship between players' mitigation strategies. Mitigation reaction functions can become upward sloping, i.e., when other players increase mitigation now the optimal response will be to match this behaviour increasing abatement as well. In particular, in our IAM, the reaction functions of 4 out of 6 players turn to be upward sloping when also adaptation is an option. The leakage effect is attenuated (or disappears), and this is positive for coalition stability as free riding incentives are lower. Indeed, from the stability analysis we find that now some coalitions, despite small, are internally stable. While in the M-Game with no transfers cooperation was not possible, in the M+A-Game 4 coalitions formed by 2 players are internally stable. Allowing for optimal transfers, larger coalitions, up to full cooperation, are internally stable.

The model is described in section 2. Section 3 describes the game, results are presented in section 4. Section 5 concludes.

2. The model

The model we use for our analysis is an updated version of the Nordhaus' RICE model (Nordhaus and Yang 1996). Enrichments consist either in the addition of an adaptation module to the RICE model that originally considered only mitigation, or in the re-calibration of the climate change damage and adaptation functions, GDP and emission trends, based on the latest available information. The remaining equations' structure is the same as the RICE-96 model. Therefore, we are working with a Ramsey type model in which the climate change dimension is hard linked with the economic dimension. Emissions, by product of economic activity, feed back onto the economy through a reduced form climate change damage function translating temperature increases into GDP losses. Temperature increases are calculated starting from emissions through a reduced form climate module. Policy variables, mitigation or adaptation, can influence the process. The former reducing emissions, the latter reducing damages. These actions, however, cost. The intertemporal optimization process that in the model aims to maximize the discounted flow of per capita consumption, finds the optimal balance between mitigation, adaptation, investments and residual damage.

The relevant feature of RICE-like models, crucial for our analysis, is the possibility to simulate strategic behaviour across players. In this version, the six macro-regions that compose the world economy: 1) United States of America (USA), 2) Japan (JPN), 3) European Union (EU28), 4) China (CHN), 5) Former Soviet Union countries (FSU) and 6) Rest Of the World (ROW) can either play cooperatively or non-cooperatively.

In the first case there is a global central planner that maximises global welfare, thus fully internalizing the benefit of climate action. In the second case each region maximizes its own welfare disregarding effects on other. We expand also upon this feature enabling the study of intermediate coalition structures where sub groups of regions act cooperatively among themselves against non-cooperative singletons.

Addressing to the appendix A for the full model equation listing, next sections describe the main modification introduced.

2.1 Modelling adaptation

To enable the simulation of the M+A-Game, the model is expanded to include the possibility for regions to adapt. To do this the scaling factor transforming gross into net output becomes: $\Omega_i^{M+A}(t) = \dfrac{1 - b_{1_i} \cdot \mu_i(t)^{b_{2_i}} - c_{1_i} \cdot a_i(t)^{c_{2_i}}}{1 + \left[\theta_{1_i} \cdot \dfrac{T(t)}{2.5}^{\theta_{2_i}}\right] \cdot \left[1 - a_i(t)\right]}$. The adaptation cost share of GDP $c_{1_i} \cdot a_i(t)^{c_{2_i}}$ is added at the numerator. Adaptation benefits, in the form of damage reduction, are introduced at the denominator. Adaptation level $a_i(t)$ is a new decision variable that can take values between 0 (no damages are avoided) and 1 (damages are completely avoided).

We modeled adaptation as a flow variable. In each period, players have to choose the optimal adaptation level knowing that both costs and benefits will arise in that period and would not affect the following periods. An alternative modelling approach in IAMs, is to present adaptation as a stock variable: adaptation investments built adaptive capacity and protection level over time (see e.g. Bosello et al. (2010 and De Bruin (2011)). However, for our purposes the “flow” assumption is more convenient. On the one hand, both adaptation and mitigation are flow variables and hence they are more easily comparable. On the other hand, the theoretical results we want to test are all based on the concept of flow adaptation, as theoretical IEA games are mostly played over one single period.

To conclude, note that, differently from mitigation, adaptation is a private good. Individual adaptation level only reduces individual damages. There are no externalities associated with adaptation.

2.2 Initial data, trend data and parameters’ calibration

The model requires a series of initial and trend data shaping the evolution of exogenous variables, technical progress and population. The initial model year is 2015. Regional GDP and population are taken from World Bank (2019). The 2015 capital stock for the 6 regions is obtained applying to the original values (referring to 2005) used in RICE-2010 (Nordhaus 2010) economic growth data.

Initial emissions levels are derived harmonizing a mix of sources. The Seventh National Communications to the UNFCCC for the EU and Japan, (European Commission 2017, Government of Japan 2017); the Climate Action Tracker (Climate Action Tracker 2019) for the United States, China and Former Soviet Union Countries. ROW emissions are computed as a residual from global 2015 emissions. The other starting data to initialize the model are the temperature increase with respect to the pre-industrial level. which is set to nearly 1°C in 2015 (NOAA 2016), and the atmospheric concentration of GHG in billion tonnes of CO_2 equivalent, that is an updated value from RICE-2010.

Initial data for the 6 macro regions are reported in Appendix A.

Population trend replicates the socio-economic pathway SSP2 (Riahi et al. 2017), that is the "middle of the road" scenario. Technological change and emission intensity trends are also calibrated to replicate GDP and emissions trends of the SSP2.

The parameterization of the mitigation cost function is the same of the original model. We update instead the damage function calibration. The new reference point is the damage of RICE-2010 for a temperature increase of 2.5°C with respect to pre-industrial levels. We then assume that this damage encompasses residual damages and adaptation cost that we disentangled. To do this we follow the adaptation cost and benefit study from Nordhaus and Boyer (2000), also used for the calibration of AD-RICE 2012 (De Bruin 2014) and information from the latest surveys on adaptation costs and benefit, (UNFCCC 2007 and World Bank 2010a) and, for the developing world (World Bank 2006, Oxfam 2007 and UNDP 2007).

All parameters and their values are reported in Appendix A.

3 The Game

3.1 The analysis of multiple coalition structures

Starting from this set up we apply the Integrated Assessment Model to the study of climate negotiation and coalition formation. More specifically, we study all the equilibrium solutions for every possible coalition that can form starting from the 6 players of our model. This amounts to 58 possible coalition structures, from non-cooperation to full cooperation. The intertemporal maximization process gives, for each player, equilibrium values for all the simulation periods.

Then, coalition stability is analyzed. Using the "standard" open membership rule, i.e., allowing all players to freely join and leave the coalition, stability should be both internal and external. Internal stability holds when no players inside the coalition would be better off by withdrawing the agreement, while external stability holds when no players outside the coalition would be better off by joining the agreement. According to the theory, with symmetric players, the failure of external stability of a coalition of size p implies that the coalition of size $p+1$ is internally stable. In our case, with asymmetric players, this would not be the case. Thus, in this work we test only internal stability[22]. Formally:

A given coalition P is internally stable if:

$$U^{*}_{i \in P}\left(P\right) \geq U^{*}_{i \notin P}\left(P-\left\{i\right\}\right) \forall i \in P$$

Recall that, in the context of an intertemporal maximization, U_i refers to the net present value of individual utility over the n simulation periods.

In the analysis we also allow for monetary transfers within coalition members. With asymmetric players, some might gain while others might lose staying in the coalition. Allowing for transfers,

22 In our simulations, when monetary transfers are not possible, no stable coalition would form. The few internally stable coalitions are not externally stable as well.

provided that the coalition surplus is greater than the free riding incentive, the coalition can become internally stable. In our model, utility is a logarithmic function of per capita consumption. Since population is exogenous, the net present value of utility, is a homothetic transformation of the net present value of consumption (dollars). Thus, monetary transfers will enable the shift of consumption and utility from one region to another.

The specific transfer scheme considered is, as in Carraro et al. (2006), the optimal transfers scheme, that guarantees the largest possible participation.

All this implies a reformulation of the concept of internal stability which becomes "potential internal stability".

A given coalition P potentially internally stable if:

$$\sum_{i\in P} U_i\left(P\right) \geq \sum_{i\in P} U_i\left(P-\left\{i\right\}\right)$$

Coalition P is potentially internally stable if the surplus that it generates overcomes the free riding incentives. In other words, there will be enough resources to guarantee at least the free rider payoff to each signatory.

3.2 Reaction functions

One of the most important aspects of our analysis is to study the strategic relationship between mitigation and adaptation and the effects of adaptation on players mitigation strategies. To do so, we study the reaction functions in the mitigation space (in both M- and M+A-Game) and in the mitigation-adaptation space (in the M+A-Game) as done from a theoretical perspective in Ebert and Welsh (2011 and 2012) and Bayramoglu et al. (2018). The concept of reaction function is rather complicated in an intertemporal setup. Players might react differently in different periods of time. Since mitigation is expressed as a rate (a value between 0 and 1) and there is not a concept of aggregate mitigation, we focus on actual emissions that can be summed up through all periods.

Furthermore, the complexity of the model does not allow us to analytically derive the reaction function of the 6 regions. Therefore, we proceed numerically applying the following concepts.

Slope of mitigation reaction functions

Player i's individual reaction function in the mitigation, or emissions, space is downward (upward) sloping if the reaction to an increase of mitigation levels of all other players throughout all the periods, everything else equal, results in a higher (lower) aggregate (over the t periods) individual emissions level.

Strategic interaction between mitigation and adaptation

In the M+A-Game, adaptation and mitigation strategies are strategic substitutes (complements) if the optimal response to a higher (lower) total emissions level throughout all the periods is a higher (lower) individual adaptation level.

Checking for slope of reaction function we start from the Nash equilibrium. We then impose a higher mitigation level to 5 out of 6 players and test the reaction of the sixth one. In the pure mitigation game, the other decision variable, i.e., investments, is kept fixed and the effect of other players' lower emissions is isolated. In the mitigation-adaptation game, while investments are kept fixed, adaptation levels are free to adjust optimally. In this way, the effect of adaptation on mitigation reaction functions can be captured. The slope of the mitigation reaction function of the reacting country is determined comparing its total emissions level over 90 simulation years against the original non-cooperative solution.[23]

To test the strategic relationship between mitigation and adaptation, we perturb mitigation level for all players throughout all the periods. Keeping investment fixed, if the reaction of a player to higher

---

23 We are aware that, having asymmetric players, mitigation reaction functions slopes may change by considering different coalition structures. However, for objective difficulties in testing all the 58 coalition structures, we focus only on individual mitigation reaction functions in a fully non-cooperative setting.

(lower) mitigation is to decrease (increase) individual adaptation, then adaptation and mitigation strategies are strategic substitutes (complements).

4 Results

In this section we present the main results of our analysis. As said, we would like to emphasise how the introduction of adaptation can change the exit of climate change agreement especially in term of stability and size of stable coalitions. As a first preliminary step, we examine the strategic relationship between mitigation and adaptation. Then, we move to the effect of adaptation on mitigation reaction functions and finally we look into negotiation outcome to analyse the effects on coalition stability.

4.1 Result 1: Mitigation and adaptation are strategic substitutes

In the M+A-Game, when all players increase (decrease) their mitigation level over the 25 periods of the game, all other things being equal, the optimal response is to decrease (increase) adaptation levels in every period.

Figure B.1 in Appendix B, reports this result highlighting the effect of total emissions reduction (increase of total mitigation) on individual adaptation levels. With total emissions decreasing (total mitigation increasing) throughout all periods, players react by reducing their adaptation levels. This effect is more evident in later periods, when the benefits of emissions reduction manifest. Initially, the higher mitigation effort, due to climate inertia, does little to reduce damages and the need to adapt.

These results confirm that the model behaves according most of the theoretical literature showing that, increasing (total) mitigation effort reduces the marginal benefits from adaptation as there will be less damages to avoid.

This outcome can be also analytically derived from the definition of the output scaling factor of the M+A-Game:

$$\Omega_i^{M+A}(t) = \frac{1 - b_{1_i} \cdot \mu_i(t)^{b_{2_i}} - c_{1_i} \cdot a_i(t)^{c_{2_i}}}{1 + \left[\theta_{1_i} \cdot \frac{T(t)}{2.5}^{\theta_{2_i}}\right] \cdot \left[1 - a_i(t)\right]}$$

Its denominator expresses the direct interconnection between mitigation and adaptation on the benefits side, while on the cost side (numerator) the two strategies do not present direct interdependencies. Marginal benefits of adaptation are thus obtained by differentiating the denominator with respect to individual adaptation level $a_i(t)$. The marginal damage reduction due to adaptation (its marginal benefit) is $-\theta_{1_i} \cdot \frac{T(t)}{2.5}^{\theta_{2_i}}$. This term decreases as $T(t)$ decreases. Since $T(t)$ decreases in total mitigation level, then adaptation marginal benefits reduce (increase) as mitigation increases (decreases).

4.2 Result 2: Adaptation can induce upward sloping individual mitigation reaction functions.

In the pure mitigation game, the 6 macro regions exhibit the standard downward sloping mitigation (emissions) reaction function. The introduction of adaptation reverses the slope of the reaction function of 4 out of 6 regions. USA, Japan, EU and Former Soviet Union Countries present upward sloping reaction functions in the mitigation space.

Figure B.2 in Appendix B shows the reaction functions of the 6 regions in the two games. Starting from the non-cooperative Nash equilibrium, we impose a higher mitigation level to 5 players, and test the reaction of the sixth one. The table reports the percentage change in aggregate emissions (over the period 2015-2105) imposed to other players and the percentage change in individual emission (best response) of every macro-region. In the M-Game, when emissions (mitigation) of all others decrease (increases) the optimal response of each player is to increase emissions (decrease mitigation). In the M+A-Game, China and Rest of the world maintain this behaviour, while the other 4 regions react in the opposite way. Now, when emissions of all others decrease (increases) the optimal response of these 4 players is to decrease (increase) emissions. When this happens, however, the relative change in the emissions is small, meaning that reaction functions, albeit upward sloping, are rather flat.

Once again, the theoretical results are confirmed by our IAM application. According to the theory, mitigation reaction functions become upward sloping when there is a “sufficiently” strong relation between (emissions) mitigation and adaptation. This occurs when mitigation-adaptation reaction functions are steep. Interestingly, our empirical application shows exactly this (Table B1 in appendix B). China and rest of the World, the two regions maintaining downward sloping mitigation reaction functions, both in the M- and M+A-Game, are also those with the weakest interdependency between emissions and adaptation (plots 1.e and 1.g).

The intuition behind this result is that, having now adaptation as an option, players can react to higher (lower) emissions not only by reducing (increasing) their emissions but also by adapting more (less). If the strategic substitutability between the two strategies is strong, then a player reacts to higher (lower) emissions with sharp increase (decrease) in adaptation and possibly decrease (increase) mitigation.

4.3 Result 3: Adaptation enhances coalition (internal) stability and welfare surplus from cooperation.

When no transfers across coalition members are allowed, the M-Game is unable to lead to any internally stable coalition while in the M+A-Game 4 internally stable coalitions of 2 macro regions form. If monetary transfers are possible, many coalitions, including the grand coalition, are potentially internally stable in both games. However, the number of stable coalitions is higher in the M+A-game than in the M-game. Furthermore, the grand coalition in the M+A-Game leads to a higher difference between the surplus generated by the coalition and the free riding incentives.

In a situation in which monetary transfers across coalition members are not possible, adaptation enhances the possibility of cooperation. In the M+A-Game, we found the following internally stable coalitions: USA-European Union, USA-FSU Countries, European Union-FSU Countries and China-rest of the World.

Table 1 compares the performance of these 4 coalition in the pure mitigation and mitigation-adaptation game. We focus on the sign of the changes in mitigation (emissions) and adaptation levels compared to the non-cooperative case. The change in aggregate emissions level of signatories (S) and non-signatories (NS) are indicated with ΔE_S and ΔE_{NS} respectively. In the M+A-Game, the change in individual adaptation levels is indicated with Δa_i. Looking into those changes, we can get insights on the mechanism that makes those coalitions stable after the introduction of adaptation.

Table 1: Analysis of stable coalitions without transfers

Coalition	**M-Game**			**M+A-Game**			
	Internal Stability	ΔE_S	ΔE_{NS}	*Internal Stability*	ΔE_S	ΔE_{NS}	Δa_i
USA-EU28	X	-	+	✓	-	-	-
USA-FSU	X	-	+	✓	-	+	-
EU28-FSU	X	-	+	✓	-	+	-
CHN-ROW	X	-	+	✓	-	-	-

All 4 coalitions that become internally stable in the M+A-Game emit less compared to the non-cooperative case. Due to strategic substitutability between mitigation and adaptation, the individual adaptation level in all the 6 regions reduces.

Comparing M- and M+A-Game, in 2 cases, USA-EU28 and CHN-ROW, internal stability is achieved thanks to the reversion of the slope of mitigation reaction functions. In the M-Game, as coalition forms and signatories increase their mitigation level, the outsiders' response is a higher emissions level. This "standard" leakage effect implies the existence of free riding incentives that eventually make the coalitions not internally stable. On the contrary, in the M+A-Game, the mitigation reaction function of non-signatories is upward sloping. When the coalition reduces emissions, non-signatories do the same. Welfare of coalition members is positively affected by this response, and free riding incentive are low enough to make the coalition stable. In the other two cases USA-FSU and EU28-FSU, non-signatories' mitigation reaction function is downward sloping both in the M- and M+A-Game. Still, the presence of adaptation, decreases the leakage effect. When non signatories increase their emissions, the coalition members also decrease their emissions, but by less when it is possible to adapt. This attenuates the incentive to free ride respect to the M-Game. More insights on this positive effect of adaptation can be found in the next Result 4.

Qualitatively similar considerations can be derived considering the possibility of monetary transfers, that however play a very important role in the reduction of free riding incentives. Indeed, with transfer, we find that the grand coalition can form both in the M- and M+A-Game. However, some indications on the positive effect of adaptation on coalition stability can still be found. First of all, some coalitions that are not potentially internally stable in the M-Game, are potentially internally stable in the M+A-Game. In details, all the 15 possible 2-players coalitions are stable in the M+A-Game, while only 12 are stable in the M-Game. When the grand coalition, which is the final equilibrium in both games, forms, interesting insights are offered comparing the surplus of cooperation with the free riding incentive in case of full cooperation in the two games. The M-Game generates a surplus (of consumption net present value) over the free riding incentives of 1.92 USD trillion, while in the M+A-Game it increases to 1.97.

Under an optimal transfers scheme regime, the effect of adaptation is not visible in terms of equilibrium coalition's size. However, adaptation is able to enlarge the benefit of cooperation compared to free riding incentives.

4.4 Result 4: M+A-Game stable coalitions lead to higher emissions reduction

The 4 internally stable coalitions in the M+A-Game, only lead to relatively small emissions reduction compared to full cooperation. However, the M+A-Game, compared to the M-Game, achieves higher emissions reduction thanks to the lower leakage effect.

Figure B.3 in Appendix B shows the aggregate emissions reduction in the period 2015-2105 generated by the 4 internally stable coalitions[24]. The plot also considers the same coalitions, which are not stable, in the M-Game and the grand coalition (full cooperation). Each of the 4 coalitions, is able to lead to a higher emissions reduction in the M+A-Game. This happens, again, because of the positive effect that adaptation has on mitigation strategies. Looking to the performance of each of the 4 stable coalitions of the M+A-Game, we notice that the coalition between China and Rest of the World is the one achieving the largest emissions reduction. The reason is straightforward, as China is the major emitter and ROW region comprises most world's countries. Among the other three coalitions, the best performing is USA-EU28. This is also not surprising as it is a cooperation between 2 of the biggest worlds' economies. The less performing coalitions, in emissions reduction terms, are USA-FSU and EU28-FSU. However, since the 4 stable coalitions are only formed by 2 macro regions, they achieve a relatively small emissions reduction compared to the grand coalition (social optimum). Each of the possible stable coalition is excluding a large amount of global emissions from the agreement and hence the final outcome cannot be satisfactory. Monetary transfers play a fundamental role in the success of negotiation, as they allow to reach full cooperation.

---

[24] We analyse stable coalitions' performance looking only to emissions reduction and not, as in most theoretic models, welfare improvements. The welfare discounting in the maximization process makes the improvements in the net present value of utility too small to be appreciated.

5 Summary and conclusion

This paper applies an updated version of the Nordhaus' RICE model to test the effect of introducing adaptation in IEAs modeling comparing the pure mitigation (M-) and the mitigation-adaptation (M+A-) game.

Existing literature has applied integrated assessment modeling to the study of IEAs and coalition formation only in a pure mitigation context. We extend the application of IAMs to a mitigation-adaptation game. The objective of analysis reflects the one of the main theoretical models that analyse the effect of adaptation on mitigation (emissions) strategies (Ebert and Welsh 2011 and 2012), the strategic relation between mitigation and adaptation (Ingham et al 2013) and, more broadly, IEAs in the presence of adaptation (Bayramoglu et al. 2018 and Rubio 2018).

The main theoretical results of these papers are tested, and in large part confirmed, by our IAM application. Mitigation and adaptation are strategic substitutes. In presence of higher total mitigation (lower emissions and lower climate change damages) players would choose a lower adaptation level. Since IAMs are dynamic games that try to reproduce at best the realty, the benefits from emissions reduction manifest in later periods and, therefore, also the effect of higher mitigation on adaptation marginal benefits shifts. Within an intertemporal maximization, the analysis of mitigation reaction functions is also not trivial. We use as a benchmark the reaction of each player i in the non-cooperative Nash equilibrium where all other players increase their mitigation levels in every period. In the pure mitigation game, individual mitigation reaction functions are downward sloping while in the M+A-Game reaction functions can be positively sloped. Increasing the mitigation levels of all other players by 5 times, the reaction of 4 out of 6 regions (USA, Japan, EU and Former Soviet Union Countries) is now to decrease their aggregate emissions over the period 2015-2105. This result goes in the same direction of the theoretical findings of Ebert and Welsh (2011 and 2012) and Bayramoglu et al. (2018). More generally, adaptation is able to reduce the leakage effect. Consequently, coalitions in the M+A-Game lead to higher aggregate emissions reduction than in the M-Game. Indications in favour of the positive effect of adaptation on coalition stability are also found. Without the possibility of monetary transfers, no coalitions form in the M-Game. In the M+A-Game 2-players stable coalitions arise. With an optimal transfers scheme both games allow for the formation of the grand coalition. However, looking at free riding incentives, they are weaker in the M+A-Game.

Since this work is the first IAM application to consider coalition formation in a M+A-Game, there is room for several extensions and modifications. Considering a model with more than 6 macro regions, the results in terms of stable coalition size could show more evidences of the positive role played by adaptation. Furthermore, different ways of adaptation modeling are available. This paper, for a direct comparison with mitigation, considers adaptation as a flow variable. However, another option would be to consider it as a stock or, more realistically, to model both a flow and a stock component of adaptation. Finally, the relation between mitigation and adaptation could be enriched assuming an additional interdependence as done, in a theoretical work, by Ingham et al. (2013).

References:

Altamirano-Cabrera, J. C., M. Finus and R. Dellink (2008), Do Abatement Quotas Lead to More Successful Climate Coalitions? “The Manchester School”, vol. 76(1), pp. 104-129.

Agrawala, S., F. Bosello, C. Carraro, E. De Cian and E. Lanzi (2011), Adapting to Climate Change: Costs, Benefits, and Modelling Approaches. “International Review of Environmental and Resource Economics”, vol. 5, pp. 245–284.

Bayramoglu, B., M. Finus and J.-F. Jacques (2018), Climate Agreements in Mitigation-Adaptation Game. “Journal of Public Economics”, vol. 165, pp. 101-113.

Bosello F., C. Carraro and E. De Cian (2010), Climate Policy and the Optimal Balance between Mitigation, Adaptation and Unavoided Damage. “Climate Change Economics”, vol.1, pp.71-92.

Bosetti, V., C. Carraro, M. Galeotti, E. Massetti and M. Tavoni (2006), WITCH: A World Induced Technical Change Hybrid Model. "The Energy Journal", vol. 27, Special Issue: Hybrid Modeling of Energy-Environment Policies: Reconciling Bottom-up and Top-down, pp. 13-37.

Carraro, C. and D. Siniscalco (1993), Strategies for the International Protection of the Environment. "Journal of Public Economics", vol. 52 (3), pp. 309–328.

Carraro, C., J. Eyckmans and M. Finus (2006), Optimal transfers and participation decisions in international environmental agreements. "The Review of International Organizations", vol. 1(4), pp. 376-396.

Climate Action Tracker (2019), Countries, https://climateactiontracker.org/countries/, Accessed March 2019.

De Bruin, K. C., R. B. Dellink and R. S. Agrawala (2009a), Economic Aspects of Adaptation to Climate Change: Integrated Assessment Modelling of Adaptation Costs and Benefits. "OECD Environment Working Paper", No. 6.

De Bruin, K. C., R. B. Dellink and R. S. J. Tol (2009b), AD-DICE: An implementation of adaptation in the DICE model. "Climatic Change", vol. 95(1-2), pp. 63-81.

De Bruin, K. C. (2011), Distinguishing Between Proactive (Stock) and Reactive (Flow) Adaptation. "CERE WP" No. 8/2011.

De Bruin, K. C. (2014), Calibration of the AD-RICE 2012 Model. "CERE WP" No. 3/2014.

Ebert, U. and H. Welsch (2011), Optimal Response Functions in Global Pollution Problems Can be Upward-sloping: Accounting for Adaptation. "Environmental Economics and Policy Studies", vol. 13(2), pp. 129–138.

Ebert, U. and H. Welsch (2012), Adaptation and Mitigation in Global Pollution Problems: Economic Impacts of Productivity, Sensitivity, and Adaptive Capacity. "Environmental and Resource Economics", vol. 52, pp. 49–64.

European Commission (2017), REPORT FROM THE COMMISSION Seventh national communication and third biennial report from the European Union under the UN Framework Convention on Climate Change (UNFCCC) (required under the UNFCCC and the Kyoto

Protocol) And its accompanying Staff working documents: Seventh National Communication of the EU and Third Biennial Report of the EU, December 2017.

Eyckmans, J. and M. Finus (2007), Measures to Enhance the Success of Global Climate Treaties. “International Environmental Agreements: Politics, Law and Economics”, vol 7(1), pp. 73-97.

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”, vol. 2(1), pp. 29-67.

Finus, M. and A. Caparrós (2015), Game Theory and International Environmental Cooperation. The International Library of Critical Writings in Economics. Edward Elgar, Cheltenham, UK.

Finus, M. and M. McGinty (2019), The Anti-Paradox of Cooperation: Diversity May Pay. “Journal of Economic Behavior and Organization”, Vol. 157, pp. 541-559.

Government of Japan (2017), Japan’s Seventh National Communication Under the United Nations Framework Convention on Climate Change, December 2017.

Hoel, M. (1992), International Environment Conventions: The Case of Uniform Reductions of Emissions. “Environmental and Resource Economics”, vol. 2(2), pp. 141-159.

Kriegler E., J.P. Weyant, G.J. Blanford, V. Krey, L. Clarke, J. Edmonds, A. Fawcett, G. Luderer, K. Riahi, R. Richels, S.K. Rose, M. Tavoni, and D.P. van Vuuren (2014), The role of technology for achieving climate policy objectives: overview of the EMF 27 study on global technology and climate policy strategies. “Climatic Change”, vol. 123(3-4), pp. 353-367.

Nordhaus, W. D. (1993), Rolling the DICE: an Optimal Transition Path for Controlling Greenhouse Gases. “Resource and Energy Economics”, vol. 15(1), pp. 27-50.

Ingham, A., Ma, J. and Ulph, A.M. (2013), Can Adaptation and Mitigation be Complements? "Climatic Change", vol. 120(1–2), pp. 39–53.

NOAA National Centers for Environmental Information, State of the Climate: Global Climate Report for Annual 2015, published online January 2016, retrieved on August 2, 2019 from https://www.ncdc.noaa.gov/sotc/global/201513.

Nordhaus, W. D. and Z. Yang (1996), A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies. "American Economic Review", vol. 86(4), pp. 741-765.

Nordhaus, W. D. and J. G. Boyer (2000), Warming the World: Economic Models of Global Warming. Cambridge, Massachusetts- London, England: MIT Press.

Nordhaus W.D. (2010), Economic aspects of global warming in a post-Copenhagen environment. "Proceedings of the National Academy of Sciences of the U.S.A.", vol. 107(26), pp. 11721-11726.

Moore F. and D. Diaz (2015), Temperature Impacts on Economic Growth Warrant Stringent Mitigation Policy. "Nature Climate Change", vol. 5(2).

Oxfam (2007), Adapting to Climate Change: What's Needed in Poor Countries, and Who Should Pay. "Oxfam Briefing Paper", No. 104, Oxfam International Secretariat, Oxford, UK, 47 pp.

Riahi, K., P. van Vuuren, E. Kriegler, J. Edmonds, B.C. O'Neill, S. Fujimori, N. Bauer, K. Calvin, R. Dellink, O. Fricko, W. Lutz, A. Popp, J. Crespo Cuaresma, S. KC, M. Leimbach, L. Jiang, T. Kram, S. Rao, J. Emmerling, K. Ebi, T. Hasegawa, P. Havlík, F. Humpenöder, L. A. Da Silva, S. Smith, E. Stehfest, V. Bosetti, J. Eom, D. Gernaat, T. Masui, J. Rogelj, J. Strefler, L. Drouet, V. Krey, G. Luderer, M. Harmsen, K. Takahashi, L. Baumstark, J. C. Doelman, M. Kainuma, Z. Klimont, G. Marangoni, H. Lotze-Campen, M. Obersteiner, A. Tabeau and M. Tavoni (2017), The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. "Global Environmental Change", vol. 42, pp. 153-168.

Rubio, S.J. (2018), Self-Enforcing International Environmental Agreements: Adaptation and Complementarity. Working Paper, 029.2018, Fondazione Eni Enrico Mattei.

Tol, R. S. J. (1995), The Damage Costs of Climate Change: Towards More Comprehensive Calculations. "Environmental and Resource Economics", vol. 5, pp. 353-374.

UNDP (2007), Human Development Report 2007/08: Fighting Climate Change: Human Solidarity in a Divided World. United Nations Development Programme (UNDP). Palgrave McMillan, New York, NY, USA, 384 pp.

UNFCCC (2007), Investment and Financial Flows to Address Climate Change. Background paper on the analysis of existing and planned investment and financial flows relevant to the development of effective and appropriate international response to climate change. UNFCCC Secretariat, Bonn, Germany, 272 pp.

World Bank (2006), Clean Energy and Development: Towards an Investment Framework. DC2006-0002, prepared for the April 5, 2006 meeting of the Development Committee, Environmentally and Socially Sustainable Development Vice Presidency, Infrastructure Vice Presidency, The International Bank for Reconstruction and Development / The World Bank, Washington, DC, USA, 146 pp.

World Bank (2010a), Economics of Adaptation to Climate Change: Synthesis Report. The World Bank Group, Washington, DC, USA, 136 pp.

World Bank (2019), World Bank Open Data, The World Bank Group, https://data.worldbank.org/. Accessed March 2019.

Appendix

A.1 Model's variables and parameters

The following list reports the exogenous, endogenous and decision variables of the model. Variables are the same in the M- and M+A-Game except one. The mitigation-adaptation game has one additional decision variable: adaptation level.

Exogenous variables

$A_i(t)$ = level of technology
$L_i(t)$ = population (labour) level
$O(t)$ = exogenous forcing from other greenhouse gases

Endogenous variables

$K_i(t)$ = capital stock
$Y_i^G(t)$ = gross GDP
$E_i(t)$ = GHG emissions level
$M(t)$ = increase in atmospheric GHG concentration from pre-industrial levels
$F(t)$ = climate forcing from total GHG concentration
$T(t)$ = atmospheric temperature (increase) respect to pre-industrial level
$T^O(t)$ = dee temperature (increase) respect to pre-industrial level
$\Omega_i(t)$ = output scaling factor due to climate policies cost and climate damage
$Y_i^N(t)$ = net GDP (net of climate damages and policies cost)
$C_i(t)$ = consumption level
$\Phi_i(t)$ = welfare weight
$U_i(t)$ = utility function
W = social welfare

Decision variables

$I_i(t)$ = investment level
$\mu_i(t)$ = mitigation level
$a_i(t)$ = adaptation level (only in the M+A-Game)

The regional initial data of the model are the 2015 levels of GDP, population, emissions and capital stock. The values are reported in the following Table A.1.

Table A.1: Initial regional data:

	USA	**JPN**	**EU-28**	**CHN**	**FSU**	**ROW**
$Y^N(t=1)$ (billions USD)	18121	4395	16416	11065	1901	23018
$L(t=1)$ (millions)	326.65	126.10	509.72	1371.37	287.49	4625.27
$K(t=1)$ (billions USD)	27846.26	6153.29	25808.19	16795.91	4153.24	36904.84
$E(t=1)$ (CO_2eq million tonnes)	18.06	3.61	11.77	34.64	10.44	50.72

In the following Table A.2, we report the parameters of the model with their values and calibration. The values assigned are generally used for both the M- and the M+A-Game. When this is not the case (damage parameters and adaptation cost parameters), we specify the different calibration used in the two models.

Table A.2: Parameters and calibration

Regional parameters: M-Game	**Calibration**
b_1 = mitigation cost parameter	USA = 0.07 JPN = 0.05 EU-28 = 0.05 CHN = 0.15 FSU = 0.15 ROW = 0.10
b_2 = mitigation cost exponent	2.887
θ_1 = climate damage parameter	USA = 0.010365 JPN = 0.010882 EU-28 = 0.013934 CHN = 0.022530 FSU = 0.007086 ROW = 0.022982
θ_2 = climate damage exponent	3.5

Regional parameters: M+A-Game	**Calibration**
b_1 = mitigation cost parameter	USA = 0.07 JPN = 0.05 EU-28 = 0.05 CHN = 0.15 FSU = 0.15 ROW = 0.10
b_2 = mitigation cost exponent	2.887
c_1 = adaptation cost parameter	USA = 0.0363 JPN = 0.0289 EU-28 = 0.0741 CHN = 0.1759 FSU = 0.0349 ROW = 0.0159
c_2 = adaptation cost exponent	3.5
θ_1 = gross damage parameter	USA = 0.014858 JPN = 0.020765 EU-28 = 0.020810 CHN = 0.030863 FSU = 0.010737 ROW = 0.024300
θ_2 = gross damage exponent	3.5
$\sigma(t)$ = emissions/GDP ratio	Computed with initial data and with exogenous trend data
Other parameters: both games	**Values**
γ = capital elasticity of output	0.25
δ_K = capital stock depreciation rate	0.1
ρ = pure rate of social time preference	0.03
β = per decade carbon removal rate	0.64
δ_M = GHG decadence rate	0.0833
λ = climate feedback factor	1.41
τ_1 = atmospheric level temperature coefficient	0.226
τ_2 = transfer atm. to ocean coefficient	0.44
τ_3 = transfer ocean to atm. coefficient	0.02

A.2 Model's equations

The full equations list of the model is reported. The M- and M+A-Game differ only for the output scaling factor. We stress this difference by reporting the two different output scaling factors: $\Omega_i(t)$ for the M-Game and $\Omega_i^{M+A}(t)$ for the M+A-Game. All other equations are used in both games.

Equations

$$K_i(t) = [1-\delta_K]\cdot K_i(t-1) + I_i(t)$$
$$Y_i^G(t) = A_i(t)\cdot K_i(t)^{\gamma}\cdot L_i(t)^{\gamma-1}$$
$$E_i(t) = [1-\mu_i(t)]\cdot \sigma_i(t)\cdot Y_i^G(t)$$
$$M(t) = \beta\cdot\sum_{i=1}^{n} E_i(t) + [1-\delta_M]\cdot M(t-1)$$
$$F(t) = \frac{4.1\cdot\log[M(t)/M(0)]}{\log(2)} + O(t)$$
$$T(t) = T(t-1) + \tau_1\cdot\left[F(t) - \lambda\cdot T(t-1) - \tau_2\cdot\left[T(t-1) - T^O(t-1)\right]\right]$$
$$T^O(t) = T^0(t-1) + \tau_3\cdot\left[T(t-1) - T^O(t-1)\right]$$
$$\Omega_i(t) = \frac{1 - b_{1_i}\cdot\mu_i(t)^{b_{2_i}}}{1 + \theta_{1_i}\cdot\dfrac{T(t)^{\theta_{2_i}}}{2.5}}$$
$$\Omega_i^{M+A}(t) = \frac{1 - b_{1_i}\cdot\mu_i(t)^{b_{2_i}} - c_{1_i}\cdot a_i(t)^{c_{2_i}}}{1 + \left[\theta_{1_i}\cdot\dfrac{T(t)^{\theta_{2_i}}}{2.5}\right]\cdot[1 - a_i(t)]}$$
$$Y_i^N(t) = Y_i^G(t)\cdot\Omega_i^{(M+A)}(t)$$
$$C_i(t) = Y_i^N(t) - I_i(t)$$

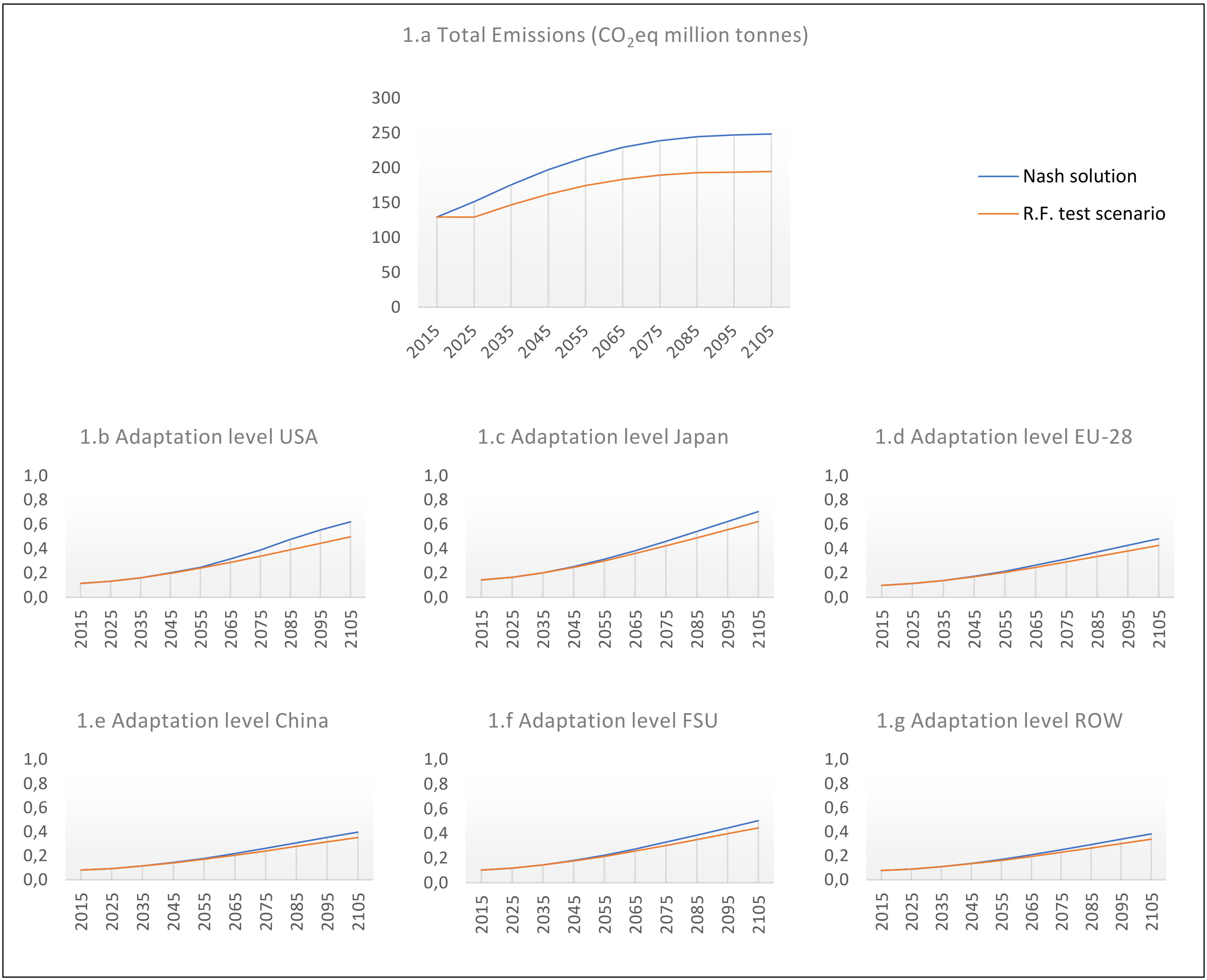
If $i \in P$, $\Phi_i(t) = p\cdot\left[\dfrac{C_i(t)/L_i(t)}{\sum_{i\in P} C_i(t)/L_i(t)}\right]$ with $p = cardinality(P)$

If $i \notin P$, $\Phi_i(t) = 1$

$$U_i(t) = L_i(t)\cdot\log\left(\frac{C_i(t)}{L_i(t)}\right)$$
$$W = \sum_t \sum_i U_i(t)$$

FIGURES

Figure B.1: Test of emissions- (mitigation-)adaptation strategic relationship*



* R.F. test scenario is obtained starting from the non-cooperative Nash solution and imposing to each player a mitigation level 5 times higher throughout all periods. Investment are kept fixed to the Nash level and the reaction in terms of adaptation levels is analyzed.

Figure B.2: Test of emissions (mitigation) individual reaction functions

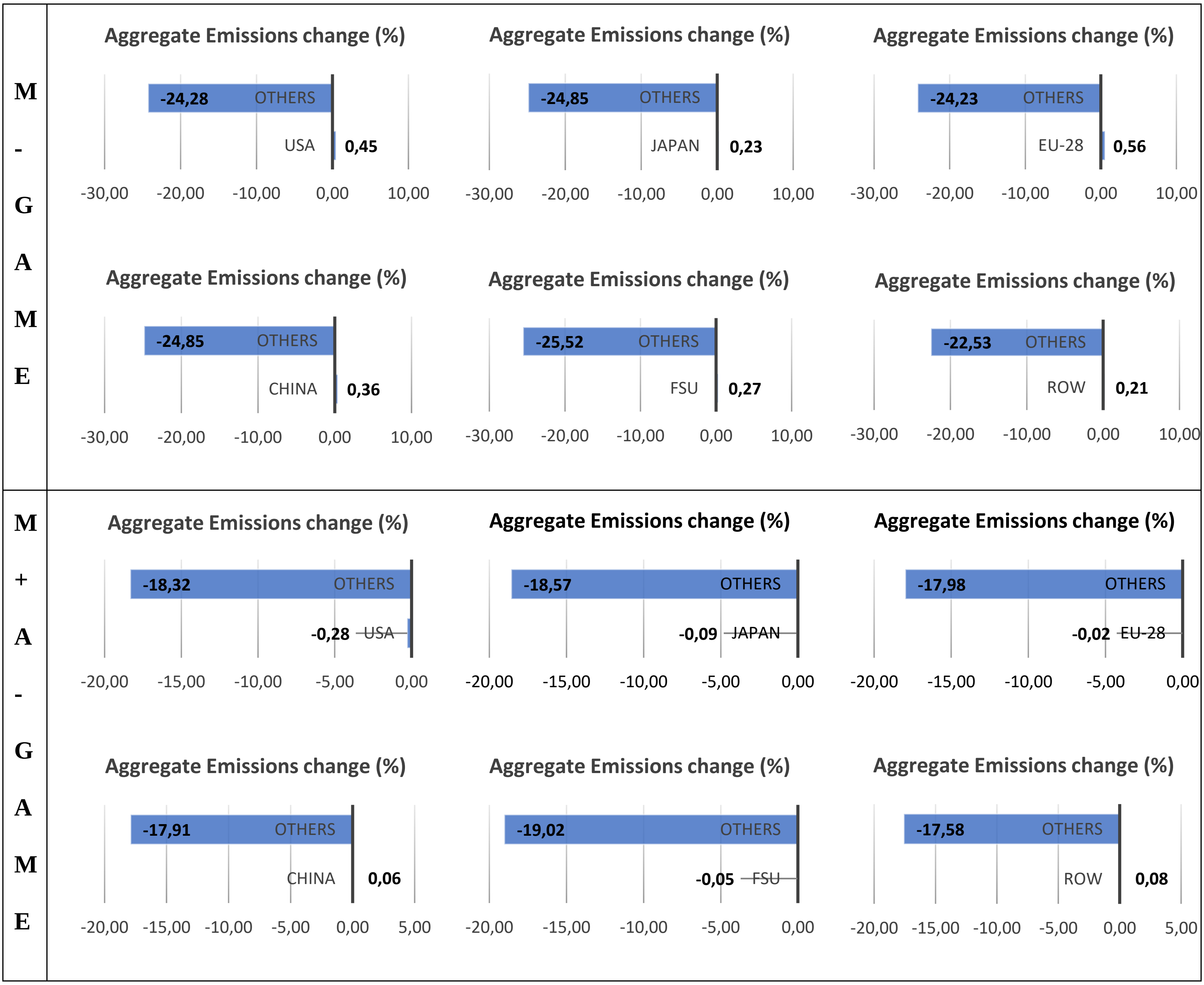
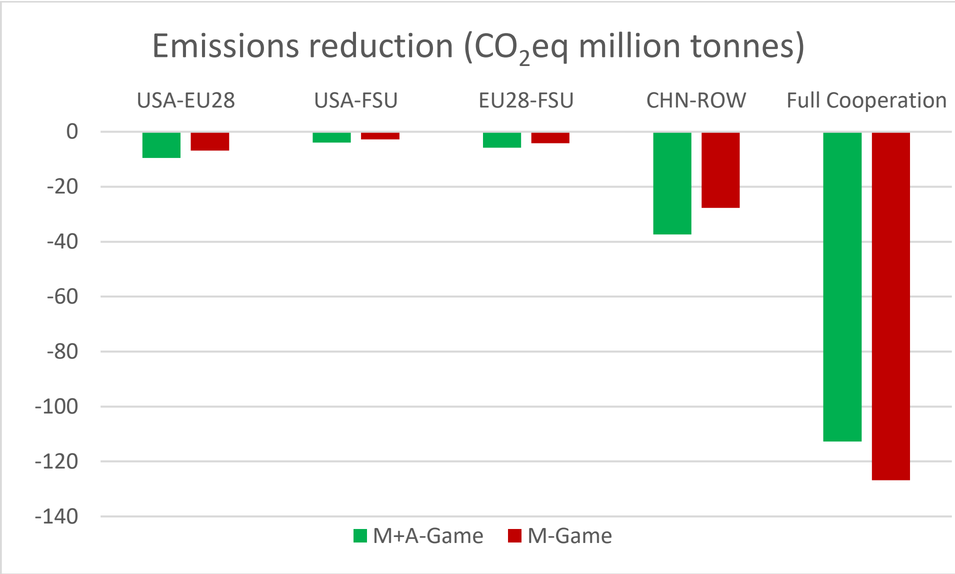


Figure B.3: Internally Stable coalitions'(in the M+A-Game) emissions reduction*



* Emissions reduction refers to aggregate emissions for the period 2015-2105 compared to the non-cooperative case. Full cooperation in the M-Game, by assumption, also implicitly includes optimal adaptation. Full cooperation results in our M+A-Game slightly differ because of calibration issues. Recall that in the social optimum aggregate welfare is maximized and there is no leakage effect.

Acknowledgements

I would like to start from the very past, from the people that believed, and helped me believe, that I could make it. They made me grow in commitment, passion and self-confidence both in studying and life. My parents and my family with their tireless support, my oldest friends and loved ones. I am lucky that many of these people are still close to me, and that others came in during my Venice time. This dissertation would not have been possible without the help and support of my Supervisor, without the chance to spend a six-months visiting period in Austria where I found unexpected and priceless guidance and support from my German mentor, and without the precious help from other colleagues and researchers. Improvements of the work were possible thanks to the useful comments and suggestions from the two external reviewers.

Finally, a fundamental role has been played by all the amazing people that surrounded me in the last three years, and that are now part of my life. PhD colleagues, flatmates and all other new friends that helped me within my work and, more importantly, outside work. The moments of leisure, fun and empathy gave me energies to manage and conclude this step of my career. With lots of memories, experiences and more expertise I now step forward.

Thank you all.

Estratto per riassunto della tesi di dottorato

L'estratto (max. 1000 battute) deve essere redatto sia in lingua italiana che in lingua inglese e nella lingua straniera eventualmente indicata dal Collegio dei docenti.

L'estratto va firmato e rilegato come ultimo foglio della tesi.

Studente: Francesco Furini____________ matricola: 956308 ________

Dottorato: Science and Management of Climate Change____________

Ciclo: 32 ________

Titolo della tesi: **Adaptation and Mitigation in the Context of International Environmental Agreements: Strategic Interactions end Effects on Negotiations' Outcome**

Abstract:

This research focuses on the analysis, from a game theoretical perspective, of International environmental agreements in the presence of adaptation. Despite its private good nature, adaptation plays a crucial role in climate agreements negotiation because of its strategic relation with mitigation. For this reason, it is very important to include both strategies in IEAs modeling.

The dissertation is a collection of three papers which expand the existing literature on IEAs in different directions: 1) the standard mitigation-adaptation game (M+A-Game) is analysed in a Stackelberg scenario; 2) the strategic relation between mitigation and adaptation and its effect on climate negotiation is analysed assuming that mitigation, attenuating climate change damages, can also affect the effectiveness of adaptation; 3) the existing theoretical results are tested through an integrated assessment model application.

The strategic relation between mitigation and adaptation, the effect of adaptation on mitigation strategies and on negotiation's outcome are analysed. Successful climate cooperation requires both large stable coalitions and high welfare improvements with respect to non-cooperation. The paradox of cooperation persists in most of the game configurations considered. Optimistic results arise only in a situation in which strategic complementarity holds both in mitigation and mitigation-adaptation space.

Abstract (Italiano):

L'obiettivo di questa ricerca è analizzare, applicando la teoria dei giochi, la formazione di accordi internazionali sul clima in presenza di adattamento. Nonostante la natura di bene privato dell'adattamento, il suo ruolo è fondamentale nell'ambito della negoziazione climatica per via della interdipendenza strategica con la mitigazione. Per questo motivo, è molto importante includere entrambe le strategie nella modellizzazione di accordi internazionali sul clima.

La tesi è composta da tre articoli che espandono lo stato dell'arte della letteratura sugli accordi internazionali sul clima in queste direzioni: 1) il gioco con mitigazione e adattamento (M+A-Game) è analizzato in un contesto di Stackelberg leadership; 2) la relazione strategica tra mitigazione e adattamento e le conseguenze sulla negoziazione climatica sono analizzate assumendo che la mitigazione, limitando gli impatti ambientali futuri, è in grado di influenzare l'efficacia dell'adattamento; 3) i risultati teorici esistenti sono testati attraverso l'applicazione di un *integrated assessment model.*

Il focus dell'analisi è sulla relazione strategica tre mitigazione e adattamento, sull'effetto dell'adattamento sulle strategie di mitigazione e sul risultato della negoziazione Il successo della cooperazione climatica richiede un'ampia partecipazione che porti a sostanziali miglioramenti nel benessere sociale rispetto ad una situazione non cooperativa. Il paradosso della cooperazione persiste nelle molteplici configurazioni analizzate, e solo in una situazione in cui sia le strategie di mitigazione sia mitigazione e adattamento sono complementi strategici le previsioni sui risultati della negoziazione sono ottimiste.

Firma dello studente

Francesco Furini