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

Michael Finus — *Department of Economics, Karl-Franzens-Universität Graz,* *e-mail: michael.finus@uni-graz.at*

Francesco Furini — *Department of Economics, Università Ca' Foscari Venezia,* *e-mail: francesco.furini@unive.it*

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-Counrot 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 mitigation 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 mitigation-adaptation game under the Stackelberg scenario. This is also true 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, Stackelberg leadership

1. Introduction

Mitigation and adaptation are two strategies to combat climate change. Mitigation directly targets at 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 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 function 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 conclusions 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 to understand the basic driving forces and incentive structure for coalition formation across the two

[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).

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]

[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

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 $B_{Ma} = B_{aM} = \partial^2 B / \partial M \partial a$.

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

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

function in the public good game is the damage function with aggregate emissions and adaptation being the arguments in this function.

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, $1 \le p \le n$. 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

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

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).

Generally, 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., $1 < p < n$.

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, $1 \le p \le n$, a coalition of size p is stable if it is internally and externally stable.

$$\text{Internal stability:} \quad w_S^*(p) \ge w_{NS}^*(p-1)$$
$$\text{External stability:} \quad w_{NS}^*(p) \ge w_S^*(p+1) \tag{2}$$

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 Bayramoglu et al. (2018), in the M+A-Game, this is equivalent to all countries choosing first their mitigation levels and then all countries choose 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]

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.

---

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

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)	$p \cdot \left[B_M(M)\left(1+R'_{NS}\right)\right] = C_m(m_S)$ (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)	$p \cdot \left[B_M(M, a_i)\left(1+R'_{NS}\right)\right] = C_m(m_S)$ (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, $R'_{NS} = \partial M_{NS} \big/ \partial M_S$ 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 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* $\Psi^{M+A} = B_{MM} + \frac{(B_{aM})^2}{D_{aa} - B_{aa}}$ *in the M+A-Game. The slopes of individual and aggregate reaction functions of signatories and non-signatories in mitigation space are given by* $r_S'(M_{-i\in P}) = \frac{p\cdot\Psi}{C_{mm}(m_S) - p\cdot\Psi}$, $R_S'(M_{NS}) = \frac{p^2\cdot\Psi}{C_{mm}(m_S) - p^2\cdot\Psi}$, $r_{NS}'(M_{-j\notin P}) = \frac{\Psi}{C_{mm}(m_{NS}) - \Psi}$ *and* $R_{NS}'(M_S) = \frac{(n-p)\cdot\Psi}{C_{mm}(m_{NS}) - (n-p)\cdot\Psi}$, *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, 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 $\frac{\left(B_{aM}\right)^2}{D_{aa}-B_{aa}} = B_{aM}\cdot\frac{\partial a_i}{\partial M}$, 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 $\frac{C_m\left(m_S\right)}{p\cdot\left(1+R_{NS}^{'}\right)} = C_m\left(m_{NS}\right)$. 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 $-1<R_{NS}^{'}<0$, 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 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 $\Psi^{M+A} = B_{MM} + \frac{\left(B_{aM}\right)^2}{D_{aa}-B_{aa}}$. *If* $\Psi^{M+A}>0$, *for any coalition size* p, *a sufficient condition for the existence of a unique second stage equilibrium is:*

---

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

$\Psi^{M+A} \cdot \left[\frac{p^2}{C_{mm}(m_S)} + \frac{(n-p)}{C_{mm}(m_{NS})} \right] < 1$ *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 externalities 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 simplicity, 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, 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* $w_{NS}^{NC}(p) > w_{NS}^{ST}(p)$ *and hence*
- $p^{ST*} \geq p^{NC*}$.

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*
  $w_S^{ST}(p) - w_S^{NC}(p) < w_{NS}^{ST}(p) - w_{NS}^{NC}(p)$ *and hence*
- $p^{ST*} \leq p^{NC*}$.

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 $W^{NC}\left(p^{NC}\right) > W^{ST}\left(p^{ST}\right)$ 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 $W^{NC}\left(p^{NC*}\right) < W^{ST}\left(p^{ST*}\right)$ 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, $M^{NC}\left(p^{NC*}\right) <,> M^{ST}\left(p^{ST*}\right)$.

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.

[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).

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 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, $2 \le p \le n$.

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-

[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 $n \ge 7$ (as assumed in our simulations), this leads to $p^{NC*} = \{3, n\}$. See Appendix A.4.

exclusiveness of the public good. Non-signatories reduce their contribution to this public good if 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. 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

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 sufficienly 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

reduce their contribution to the public good, which may not be compensated by the followers' additional mitigation effort.

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* is 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

---

[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 $INI^{NC} > INI^{ST}$.

be the only anti-paradox constellation. However, this rests on the rather unlikely assumption that 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.

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 contribution. 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 to 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 are 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). 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.

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.

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.

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.

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: $r_S^{'}(M_{-i})=-\frac{p\cdot g}{c+p\cdot g}$, $R_S^{'}(M_{NS})=-\frac{p^2\cdot g}{c+p^2\cdot g}$, $r_{NS}^{'}(M_{-j})=-\frac{g}{c+g}$ 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:

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

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

For the NC-scenario, we have:

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

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:

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

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

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

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.

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 $\Psi \cdot \left[\frac{p^2 \cdot \left(1 + R_{NS}^{'}\right)}{C_{mm}(m_S)} + \frac{(n-p)}{C_{mm}(m_{NS})} \right] < 1$. 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 $C3^{ST} = C3^{NC}$ $\left(c \cdot d - (n-p) \cdot \left(f^2 - d \cdot g\right) > 0\right.$ if $\left.c \cdot d - n^2 \cdot \left(f^2 - g \cdot d\right) > 0\right)$.

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:

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

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

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

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

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

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

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:

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

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

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

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 $M^{NC}(p) > M^{ST}(p)$ but assume the opposite: $M^{NC}(p) < M^{ST}(p)$.

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 $M^{NC}(p)<M^{ST}(p)$ so that $M^{NC}(p)>M^{ST}(p)$ must hold. Consequently, $m_{NS}^{NC}(p)<m_{NS}^{ST}(p)$ must hold from the first order conditions of non-signatories and hence for $M^{NC}(p)>M^{ST}(p)$ we must have $m_S^{NC}(p)>m_S^{ST}(p)$.

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:

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

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: $C\left(m_S\right)=\frac{c}{\varepsilon}\cdot m_S{}^{\varepsilon}$, $C\left(m_{NS}\right)=\frac{c}{\varepsilon}\cdot m_{NS}{}^{\varepsilon}$ with $\varepsilon>1$ and hence $C_m\left(m_S\right)=c\cdot m_S{}^{\varepsilon-1}$, $C_m\left(m_{NS}\right)=c\cdot m_{NS}{}^{\varepsilon-1}$. From the first order conditions with respect to mitigation in the NC-scenario we know that $\frac{C_m\left(m_S\right)}{p}=C_m\left(m_{NS}\right)$ 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 $\frac{C_m\left(m_S\right)}{p\cdot\left(1+R_{NS}^{'}\right)}=C_m\left(m_{NS}\right)$. For our polynomial cost function, we obtain $m_S^{ST\varepsilon-1}=p\cdot\left(1+R_{NS}^{'}\right)\cdot m_{NS}^{ST\varepsilon-1}$ so that $m_S^{ST}=\sqrt[\varepsilon-1]{p\cdot\left(1+R_{NS}^{'}\right)}\cdot m_{NS}^{ST}$. Basic algebraic manipulation delivers: $\Delta m^{NC}=m_S^{NC}-m_{NS}^{NC}=\left(\sqrt[\varepsilon-1]{p}-1\right)m_{NS}^{NC}$ and $\Delta m^{ST}=m_S^{ST}-m_{NS}^{ST}=\left(\sqrt[\varepsilon-1]{p\cdot\left(1+R_{NS}^{'}\right)}-1\right)m_{NS}^{ST}$. 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 $\frac{\partial M}{\partial p} > 0$ 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 Ψ 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:

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

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)

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. Any move from $p-1$ to p implies one more signatory who are leaders, and one less non-signatory who are followers, and hence by axiomatic reasoning $p \cdot w_S^*(p) \geq (>) [p-1] \cdot w_S^*(p-1) + w_{NS}^*(p-1)$ must hold. For the final move from $p-1=n-1$ to $p=n$, this must also be true 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.