FAERE

French Association
of Environmental and Resource Economists

Working papers

Climate agreements in a mitigation-adaptation game

Basak Bayramoglu - Michael Finus - Jean-François Jacques

WP 2016.17

Suggested citation:

B. Bayramoglu, M Finus, JF . Jacques *(2016).* Climate agreements in a mitigation-adaptation game. *FAERE Working Paper, 2016.17.*

ISSN number: **2274-5556**

www.faere.fr

Climate Agreements in a Mitigation-Adaptation Game*

Basak Bayramoglu[†]& Michael Finus[‡]& Jean-François Jacques[§]

April 28, 2016

Abstract

We study the strategic interaction between mitigation (public good) and adaptation (private good) strategies in a climate agreement. We show that these two strategies are strategic substitutes considering various definitions of substitutability. Moreover, adaptation may cause mitigation levels between different countries to be no longer strategic substitutes but complements. We analyze under which conditions this leads to more succesful self-enforcing agreements. We argue that our results extend to many important externality problems involving public goods.

Keywords: Climate Change, Mitigation-Adaptation Game, Public Good Agreements, Strategic Substitutes versus Complements

JEL-Classification: C71, D62, D74, H41, Q54

---

*This paper has greatly benefitted from financial support under the EU-project "EconAdapt". The views expressed do not necessarily reflect those of the European Union.

†Economie Publique, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France. E-mail: Basak.Bayramoglu@grignon.inra.fr

‡Department of Economics, University of Bath, UK. E-mail: m.finus@bath.ac.uk

§Université Paris-Est, ERUDITE (EA 437), UPEM 77454 Marne la Vallée, France and LEDa-CGEMP, Université Paris-Dauphine, Paris, France. E-mail: Jean-Francois.Jacques@u-pem.fr

1 Introduction

Climate change is probably one of the most important challenges of human mankind. The Kyoto Protocol signed in 1997 was the first global treaty with specific mitigation targets but turned out to be not sufficient to address global warming. After several years of negotiations, a successor protocol was recently signed in Paris. However, most scholars doubt that the Paris accord will be sufficient to keep the increase of the global surface temperature below 2 degrees Celsius, a widespread accepted target to avoid severe interference with the climate system.

Clearly, mitigation to address the cause of global warming is costly, participation in a climate treaty is voluntary and compliance is difficult to enforce. Due to the slow progress of curbing global warming, and the first visible impacts of climate change, in particular in developing countries, adaptation measures (like building dykes against flooding and installing air-conditioning devices against heat) have received more attention in recent years. This is reflected in the negotiations leading to the Paris accord but also in the scientific community, as for instance summarized by various recent reports by the Internal Panel on Climate Change (IPCC). In contrast to mitigation (i.e. reducing emissions), which can be viewed as a non-excludable public good, adaptation (i.e. amelioration of climate damages) is typically viewed as a private good; it only benefits the country in which adaptation measures are implemented. The key research question which we try to answer is: how does adaptation, as an additional strategy to mitigation, affect the prospects of international policy coordination to tackle climate change?

At the outset, the answer is not straightforward when considering the following points. Firstly, adding adaptation to the set of strategies will reduce the costs of addressing the impacts of global warming. This should facilitate cooperation. Secondly, anything else being equal, the optimal mix of both strategies will lead to lower mitigation levels as we show. This means that the need for policy coordination related to the pure public good "mitigation" is reduced. Put differently, the positive spillovers from cooperation are reduced. On the one hand, this may negatively impact on the global welfare gains from cooperation. On the other hand, this may also reduce the free-rider incentive. Thirdly, the benefits of each strategy are not independent of the level of the other strategy. The benefits from mitigation are high (low) if there is a low (high) level of adaptation and vice versa, the benefits from adaptation are high (low) if there is a low (high) level of mitigation. Thus, mitigation and adaptation interact strategically. Ebert and Welsch (2011 and 2012) showed in a two-player model that reaction functions in mitigation space may be upward sloping in the presence of adaptation. They conjecture that in the context of agreement formation, strategic complementarity of public good provision levels may lead to larger stable agreements than in a pure mitigation game. It is a primary concern of this paper to test their conjecture. Moreover, under those conditions when agreements are large in the presence of adaptation, we evaluate whether they are also more successful in welfare terms. This is important because membership in an agreement cannot measure success per se.

Our paper is related to four strands of literature. Firstly, there is large body of literature on the game-theoretic analysis of international environmental agreements (IEAs), which can be traced back to Barrett (1994) and Carraro and Siniscalco (1993) and of which the most influential papers have been collected in a volume by Finus and Caparros (2015), including a comprehensive overview. Our model is particularly related to those recent papers, which analyze the impact of additional strategies to mitigation on the success of coalition formation, like R&D investment to reduce mitigation costs (El-Sayed and Rubio 2014, Battaglini and Harstad, 2016 and Harstad 2012) or to generate breakthrough technologies with zero emissions (Barrett 2006, and Hoel and de Zeeuw 2010). In terms of strategic implications, there are two interesting links. Because adaptation leads to lower equilibrium mitigation levels, the positive effect on free-rider incentives and hence on participation in stable coalitions is similar to the concept of modest emission reductions as analyzed in Barrett (2002), and Finus and Maus (2008). Moreover, like the papers (e.g. Finus and Rübbelke 2013) on ancillary benefits, strategies impact not only on public but also on private benefits (i.e. impure public goods).[1] However, ancillary benefits imply one strategy (mitigation) having two independent effects (private and public), whereas in a mitigation-adaptation game there are two strategies, one having a private and one having a public impact, which are linked. The most obvious connection is of course to those recent papers which study mitigation and adaptation in a strategic context. Different from for instance Buob and Stephan (2011), Ebert and Welsch (2011, 2012), Zehaie (2009), Eisenack and Kähler (2016), we allow for more than two players and study the formation of agreements. Different from for some recent work by Barrett (2008) and Benchekroun et al. (2016) who study IEAs, we work in a much more general framework and derive most results analytically.

Secondly, there is a literature on non-convexities of negative externalities, including early contributions by Baumol and Bradford (1972), Laffont (1976) and Starrett (1972). Noticing that any public bad game can be recasted in a public good game framework, where the latter is the setting of this paper, this means non-concave payoffs. This literature does not consider agreement formation but points to the strategic interaction between public and private actions, which can result in non-convexities. We show that in our model, in the presence of amelioration, the conditions for upward-sloping reaction functions in public good provision space are exactly those related to the convexity of an agent's payoff functions with respect to other players' provision levels.

Thirdly, there is a large literature on the private provision of public goods

[1] Ancillary benefits, also called co-benefits and secondary benefits in the environmental economics literature, refers to the fact that some mitigation measures will reduce local pollutants as a by-product. In the public goods literature, this phenomenon has been referred to as joint production (e.g. Cornes and Sandler 1984).

(e.g. Bergstrom, Blume and Varian 1986, Cornes and Hartley 2007, and Fraser 1992). "Private" means non-cooperative with the possibility of co-operative agreements normally not being considered in this literature. Typically, agents maximize a utility function subject to a linear budget constraint, with utility being derived from the total level of public provision (which is the sum of individual contributions) and a private numeraire good.[2] Central conclusions which emerge are the underprovision of the public good in the non-cooperative equilibrium compared to a Pareto-optimal provision, the theorem of income neutrality, implying that a redistribution of income (within boundaries) will not affect the equilibrium total public good provision, and the fact that the difference between equilibrium provision and first best increases with the number of agents. The typical assumption is that both goods are normal goods, which gives rise to downward sloping reaction functions in public good provision levels. This assumption is convenient to prove uniqueness of the equilibrium public good provision vector. It has typically two further implications. The cross derivative of utility with respect to the public and private good is assumed to be of minor importance and the typical text book illustration assumes a Cobb-Douglas utility function which gives rise to positive cross derivatives (and downward sloping reaction functions). However, downward sloping reaction functions, usually associated with the term easy riding, is not the only possibility as pointed out by Cornes and Sandler (1986, ch. 5). Moreover, it does not seem unrealistic to consider the possibility that the public good can be a superior good which would allow for the possibility of upward sloping reaction functions. For some environmental goods there is some evidence (e.g. Bergstrom and Goodman (1973), Boercherding and Deacon (1972) and Selden and Song (1994)) of income elasticities larger than 1. Our model essentially captures this possi-

[2]This refers to the standard assumption of a pure public good with a summation technology. Alternative assumptions, like impure public goods are considered for instance in Cornes and Sandler (1994) and a departure from the summation technology, like weakest-link and best-shot technologies are analyzed for instance in Hirschleifer (1983).

bility. However, different from most of the public goods literature, we do not assume a linear budget constraint with constant prices, but, in the tradition of the IEA literature, consider (strictly) convex cost functions of private and public good provision and hence non-constant marginal costs.[3] It is important to note that in the mitigation-adaptation context it is very plausible to assume that the cross derivative is negative. However, we will show that the absolute value of the derivative is what matters and not the sign to have upward sloping reaction functions in the public good provision space. The importance of the cross derivative for public good provision extends much beyond the specific context we consider in this paper. For instance, member states of the European Community can either coordinate on policy issues like security, anti-terrorism, migration and social policy or pursue those issues nationally. That is, financial resources can either be transferred to Brussels or remain with national governments. In practice, national and international policy measures co-exist and the benefit of national (international) policy measures is often diminished by the quality of international (national) measures. Citizens can vote for improved flood protection through their local government or can invest directly into the protection of their houses. Similarly, they can vote for the improvement of the local policy force or invest in devices to secure their private homes. Money can be devoted to build and maintain a public or a private swimming pool and farmers can invest either in their own machinery and irrigation devices or to become a member of a cooperative with access to shared facilities. In each of these examples, it is likely that the benefit of the private investment impinges on the benefits of the public investment and vice versa, i.e. the cross derivative is negative. In other cases, it can be expected that the cross derivative is positive. Public spending on improved infrastructure may increase the value of houses and hence makes the private investment in flood protection and security more

[3]Unfortunately, this generalization comes at the cost that the problem can no longer be viewed in terms of income elasticities.

valuable for home owners.

Fourthly, there is quite some literature that investigates complementarities in strategic games. From the survey by Vives (2005), it appears that complementarity does not need to be the result of special assumptions but there are many interesting economic problems with this feature, though the analysis is usually more complex, requires different tools for the analysis and may suffer from multiple equilibria. For our problem, it turns out that a slight modification of standard theorems is sufficient for the analysis and simple conditions give existence and uniqueness of equilibria.

In what follows, we set out our model and its assumptions in Section 2. We present results of our two stage coalition formation model in reverse order according to backwards induction in Section 3 and 4, respectively, and summarize our main results and policy conclusions in Section 5.

2 Model

We consider n players, which are countries in our context, $i = 1, 2, \ \ldots \ , n$, with the payoff function of country i in the pure mitigation game (M-game) given by:

$$\Pi_i(Q, q_i) = B_i(Q) - C_i(q_i) \tag{1}$$

and in the mitigation-adaptation game (M+A-game) by:

$$\Pi_i(Q, q_i, x_i) = B_i(Q, x_i) - C_i(q_i) - D_i(x_i) \tag{2}$$

where it will turn out throughout the paper that the M+A-game can be viewed as a generalization of the M-game. We denote the set of players by N. In the richer M+A-game, country i can not only choose its individual mitigation level q_i but also its adaptation level x_i within its (compact and convex) strategy space $q_i \in [0, \ \bar{q_i}]$ and $x_i \in [0, \ \bar{x_i}]$ with $\bar{x_i}$ and $\bar{q_i}$ sufficiently large. Country i's payoff comprises benefits, B_i, which depend on total mitigation, $Q = \sum_{j=1}^{n} q_j$, and in the M+A-game additionally also on its individual adaptation level, x_i; the cost of mitigation is denoted by C_i, and the cost of adaptation by D_i.

If there is no misunderstanding, we drop the index i as we assume that players are ex-ante symmetric, i.e. they have the same payoff function; if we need to stress that players are ex-post asymmetric, e.g. because they chose different strategies, we will use the index. Apart from assuming that all functions, including their first and second derivatives, are continuous in their variable(s), we make the following assumptions regarding the components of the payoff functions (with the understanding that all derivatives with respect to x are only relevant in the M+A-game) where subscripts denote derivatives, e.g. $B_Q = \frac{\partial B}{\partial Q}$ and $B_{QQ} = \frac{\partial^2 B}{\partial Q^2}$.

General Assumptions

Both Games*:*

a) $B_Q > 0$, $B_{QQ} \leq 0$, $C_q > 0$, $C_{qq} > 0$.

b) $\lim_{Q \to 0} B_Q > \lim_{q \to 0} C_q > 0$.

M+A-Game:

c) $B_x > 0$, $B_{xx} \leq 0$, $D_x > 0$, $D_{xx} \geq 0$.

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

d) $B_{xQ} = B_{Qx} < 0$.

e) $\lim_{x \to 0} B_x > \lim_{x \to 0} D_x > 0$.

From a technical point of view, assumptions a and c reflect the standard assumptions of concave benefit and convex cost functions. We allow for the possibility that benefit functions can be linear such that we can revisit some simple examples, which have been considered in the literature on IEAs in the context of a pure mitigation game. We assume cost functions of mitigation to be strictly convex in order to ensure unique equilibrium mitigation levels. For adaptation, it turns out that this is not necessary. However, in assumption c, we state that if benefit functions are linear in adaptation, then adaptation cost functions must be strictly convex and vice versa. These properties of the benefit and cost functions together with assumption b and e rule out corner solutions as for instance in Kolstad (2007) in a pure mitigation game and in Barrett (2008) in a mitigation-adaptation game.

From an economic point of view, assumption a stresses that mitigation is a pure public good, i.e. the marginal benefit from mitigation depends on the sum of all (and not on individual) mitigation efforts. In contrast, assumption c stresses that adaptation is a pure private good, i.e. the marginal benefit from adaptation depends on the individual adaptation level of a country (and not on those of others). The interdependency between mitigation and adaptation is captured through assumption d. The marginal benefit from mitigation (adaptation) decreases with the level of adaptation (mitigation). For simplicity, such an interdependency is assumed away on the cost side. In order to stress this, we assume for clarity two separate cost functions.

The strategic interaction between countries is directly related to the (pure) public good nature of mitigation. Mitigation in country i generates benefits in country i but also in all other countries. Thus, mitigation levels generate positive externalities. Adaptation levels generate no direct externalities. However, they indirectly influence the strategic interaction among countries because, as will become apparent below: the higher the adaptation level in a country, the lower will be its mitigation level, irrespective whether country i acts independently or joins an agreement.

Finally note that the assumption of ex-ante symmetric players is very much in the tradition of the literature on coalition formation in general (Bloch 2003 and Yi 1997 for overviews) and on IEAs in particular (Finus and Caparros 2015 for an overview) due to the complexity of coalition formation. This does not preclude that players are ex-post asymmetric. As will become apparent below, signatories and non-signatories will typically choose different mitigation levels and hence will receive different payoffs.

We assume the General Assumptions to hold throughout the paper. If we make further assumptions, we will mention them explicitly. Our two-stage coalition formation game unfolds as follows.

Definition 1: Coalition Formation Game

Stage 1

All countries choose simultaneously whether to join coalition $P \subseteq N$ or to remain a singleton player. Countries $i \in P$ are called signatories and countries $j \notin P$ are called non-signatories.

Stage 2

All non-signatories $j \notin P$ choose their economic strategies in order to maximize their individual payoff and all signatories $i \in P$ do so in order to maximize the aggregate payoff to all coalition members. Choices of all players are simultaneous.

M-Game: Mitigation levels are chosen simultaneously.

M+A-game: Version 1: Mitigation and adaptation are chosen simultaneously. Version 2: Mitigation and adaptation are chosen sequentially; all players choose first mitigation and then adaptation.

Stage 1 is the cartel formation game, which originates from the literature in industrial organization (d'Aspremont et al., 1983) and has been widely applied in this literature (e.g. Deneckere and Davidson 1985, Donsimoni et al. 1986 and Poyago-Theotoky 1995; see Bloch 2003 and Yi 1997 for surveys) but also in the literature on IEAs (e.g. Barrett 1994, Carraro and Siniscalco 1993 and, Rubio and Ulph 2006; see Finus and Caparros 2015 for a survey). This game has also been called open membership single coalition game as membership in coalition P is open to all players and players have only the choice between joining coalition P or remaining a singleton.[4] Open membership may be defended on two grounds. In the context of the provision of a public good, it appears that one is more concerned about players leaving a coalition than joining it. Moreover, to the best of our knowledge, all international environmental treaties are of the open membership type. The assumption of a single coalition simplifies the analysis but is also in line with the historical records of IEAs with a single treaty.

Stage 2 follows the standard assumption in the literature on coalition formation (see Bloch 2003 and Yi 2003 for surveys): the coalition acts as a kind of meta player (Haeringer 2004), internalizing the externality among its members, whereas non-signatories act selfishly, maximizing their own payoff. We also follow the mainstream assumption and assume that signatories and non-signatories choose their economic strategies simultaneously.[5] In the M-game, the second stage is simple: an equilibrium mitigation vector $\boldsymbol{q}^*(P)$ is derived, given that coalition P has formed. In the M+A-game, Version 1 and 2 reflect different possible assumptions about the timing of mitigation and adaptation. As both versions lead to the same second stage equilibrium economic strategies as we show below, our results are robust.[6]

The two-stage coalition formation game is solved by backwards induction. In the second stage, given that some coalition $P \subseteq N$ has formed in the first stage, in the M+A-game, Version 1 determines simultaneously an equilib-

[4]Surveys of coalition games with other membership rules, including exclusive membership and multiple coalitions, are provided in Bloch (2003) and Yi (1997) and a systematic comparison of equilibrium coalition structures under different membership rules is conducted in Finus and Rundshagen (2009).

[5]Again, see Bloch (2003) and Yi (2003) on this. This has been called Nash-Cournot assumption in the literature on IEAs and has been contrasted with the assumption of a sequential choice, called Stackelberg assumption, where signatories act as a Stackelberg leader. The Stackelberg assumption has been considered for instance in Barrett (1994) and Rubio and Ulph (2006) in a pure mitigation game.

[6]In principle, we could also consider a Version 3 in which the timing is reversed compared to Version 2. Version 3 is considered in Zehaie (2009). However, assuming first adaptation and then mitigation is not in line with the historical development in climate change policy.

rium mitigation vector $\boldsymbol{q}^*(P)$ and an equilibrium adaptation vector $\boldsymbol{x}^*(P)$ as a Nash equilibrium between coalition P and all remaining players not in P. Version 2 may be broken down into stage 2a and 2b. In stage 2b the equilibrium adaptation vector is determined, again, as a Nash equilibrium between coalition P and the remaining singletons. Equilibrium adaptation levels in stage 2b will depend on the levels of mitigation chosen in stage 2a, which in turn depend on which coalition P has formed in stage 1. Hence, in stage 2b, we can write $\boldsymbol{x}^*(\boldsymbol{q}(P))$. Substituting this into the payoff function (1), payoffs in stage 2a are only a function of mitigation levels. This allows us to solve stage 2a for equilibrium mitigation levels, $\boldsymbol{q}^*(P)$.

It is clear that we want for technical reasons for each possible coalition P a unique equilibrium strategy vector to exist. This allows us to write $\Pi_i^*(P)$ instead of $\Pi_i(\boldsymbol{q}^*(P))$ in the M-game. and, accordingly, $\Pi_i^*(P)$ instead of $\Pi_i(\boldsymbol{q}^*(P), \boldsymbol{x}^*(P))$ in the M+A-game. Even though we provide sufficient conditions for existence and uniqueness only in the next section, we make already use of this assumption in order to save on notation and define a stable coalition P^* as follows:

$$\begin{aligned} &\text{internal stability:} \qquad \Pi_i^*(P^*) \geq \Pi_i^*(P^* \setminus \{i\}) \ \forall i \in P^* \\ &\text{external stability:} \qquad \Pi_j^*(P^*) \geq \Pi_j^*(P^* \cup \{j\}) \ \forall j \notin P^* \end{aligned}$$

It is evident that the conditions of internal and external stability de facto define a Nash equilibrium in membership strategies in the first stage. Each player i who announced to join coalition P^* should have no incentive to (unilaterally) change her strategy by leaving coalition P^* and each player j who announced not to join coalition P^* should have no incentive to (unilaterally) change his strategy and join coalition P^*, given the equilibrium announcements of all other players.

Note that by the construction of the coalition game, the equilibrium economic strategy vectors in the second stage correspond to the Nash equilibrium known from games without coalition formation if coalition P is empty or contains only one player. We also call this "no cooperation". By the same token, if coalition P comprises all players, i.e. the grand coalition forms, $P = N$, this corresponds to the "social optimum". We also call this "full cooperation". Any non-trivial coalition (i.e. a coalition of at least two players) which comprises more than one player but less than all players may be viewed as partial cooperation.

In order to evaluate the outcomes and to analyze the driving forces of coalition formation, we define some useful properties where p denotes the cardinality of P, i.e. the size of coalition P.

Definition 2: Superadditivity, Positive Externality and Cohesiveness

i) A game is (strictly) cohesive if for all $P \subset N$:

$$\sum_{k\in N} \Pi_k^*(\{N\}) \geq (>) \sum_{k\in P} \Pi_k^*(P) + \sum_{l\in\{N\setminus P\}} \Pi_l^*(P).$$

(ii) A game is (strictly) fully cohesive if for all $P \subseteq N$, $p \geq 2$ and all $i \in P$:

$$\sum_{k\in P} \Pi_k^*(P) + \sum_{l\in\{N\setminus P\}} \Pi_l^*(P) \geq (>) \sum_{k\in\{P\setminus\{i\}\}} \Pi_k^*(P\setminus\{i\}) + \sum_{l\in\{N\setminus P\cup\{i\}\}} \Pi_l^*(P\setminus\{i\}).$$

(iii) A coalition game exhibits a (strict) positive externality if for all $P \subset N$, $p \geq 2$ and for all $j \in N \setminus P$:

$$\Pi_j^*(P) \geq (>) \Pi_j^*(P \setminus \{i\}).$$

(iv) A coalition game is (strictly) superadditive if for all $P \subseteq N$, $p \geq 2$ and

all $i \in P$:

$$\sum_{k \in P} \Pi_k^*(P) \geq (>) \sum_{k \in \{P \setminus \{i\}\}} \Pi_k^*(P \setminus \{i\}) + \Pi_i^*(P \setminus \{i\}).$$

Typically, a game with externalities is strictly cohesive, with the understanding that in a game with externalities the strategy of at least one player has an impact on the payoff of at least one other player. The reason is that the grand coalition internalizes all externalities by assumption. Hence, cohesiveness motivates the choice of the social optimum as a normative benchmark, and it is the basic motivation to investigate stability and outcomes of cooperative agreements. A stronger normative motivation is related to full cohesiveness as it provides a sound justification to search for large stable coalitions even if the grand coalition is not stable due to large free-rider incentives. The fact that large coalitions, including the grand coalition, may not be stable in coalition games with positive externalities is well-known in the literature (e.g. see the overviews by Bloch 2003 and Yi 1997). Examples of positive externality games include output and price cartels and the pure mitigation game. The positive externality can be viewed as a benefit generated by the coalition, which also accrues to outsiders as these benefits are non-excludable. This property makes it attractive to stay outside the coalition. This may be true despite superadditivity, a property which makes joining a coalition attractive. In the context of the pure mitigation game, stable coalitions are typically small because with increasing coalitions, the positive externality effect dominates the superadditivity effect.[7] Whether this is also the case if adaptation is available as a second strategy is one of the key research question of this paper.

Finally note that all four properties are related to each other. For in-

[7]This is quite different in negative externality games. In Weikard (2009) it is shown that in a coalition game with negative externalities and superadditivity the grand coalition is the unique stable equilibrium.

stance, a coalition game which is superadditive and exhibits positive externalities is fully cohesive and a game which is fully cohesive is cohesive.

3 Second Stage of Coalition Formation

3.1 Equivalence of Version 1 and 2 and Symmetry

In this subsection, we establish the equivalence between Version 1 and 2 in Definition 1 and some basic implications of the ex-ante symmetry assumption regarding equilibrium mitigation and adaptation levels in the second stage. We assume the existence of a unique interior second stage equilibrium for which we establish sufficient conditions in Subsection 3.2.

Lemma 1: Equivalence of Version 1 and 2 in the M+A-Game
In the M+A-game, Version 1 and 2 are equivalent in terms of an interior second stage equilibrium.

Proof. Version 1: The first order conditions in terms of mitigation are given by

$$pB_Q(Q,x) = C_q(q) \tag{3}$$

where we may recall that p denotes the size of coalition $P \subseteq N$. For non-signatories we have $p = 1$ and for signatories $p \geq 2$ if a non-trivial coalition forms. The first order conditions for non-signatories and signatories in terms of adaptation are the same and are given by

$$B_x(Q,x) = D_x(x)\ . \tag{4}$$

Version 2: In the last stage, stage 2b, when signatories and non-signatories simultaneously choose their adaptation levels, the first order conditions of non-signatories and signatories are given by (4). These first order conditions implicitly determine adaptation x as a function of total mitigation Q. Hence, using $x(Q)$, the maximization problem, which signatories and non-signatories face in stage 2a, when choosing their mitigation levels, leads to the first order conditions $p\left[B_Q(Q, x(Q)) + B_x(Q, x(Q))\frac{\partial x}{\partial Q}\right] - C_q(q) - D_x(x(Q))\frac{\partial x}{\partial Q} = 0$, again with $p = 1$ and $p \geq 2$ for non-signatories and signatories, respectively, which, using the first order conditions (4) and rearranging terms, imply (3) above. ■

The proof above made already use of the assumption of ex-ante symmetric players for notational simplicity but holds generally, also for asymmetric players. The first order conditions (3) and (4) are instructive in several respects, with the main conclusions summarized in Lemma 2 below. Firstly, only the strict convexity of the cost function of mitigation (General Assumptions, part a) ensures that mitigation levels among signatories are unique. From (4) it is evident that this is not required for adaptation. Secondly, the first order conditions in terms of adaptation are the same for non-signatories and signatories because adaptation is a private good. However, one should therefore not mistakenly conclude that policy coordination is not required in terms of adaptation. We will show later that equilibrium adaptation level decreases in the size of the coalition and hence obtains its lowest level in the social optimum. Moreover, adaptation influences optimal mitigation levels. Thirdly, all non-signatories choose the same mitigation level $q^*_{j_{\notin P}}(p)$ and all signatories choose the same mitigation level $q^*_{i\in P}(p)$ for all p, $1 \leq p \leq n$. Moreover, $q^*_{j\notin P}(p) < q^*_{i\in P}(p)$ for all p, $1 < p < n$ and hence $\Pi^*_{j\notin P}(p) > \Pi^*_{i\in P}(p)$. Finally, in the M-game, there is only one set of first order conditions, namely (3) and what we concluded in the last point is also true.

Lemma 2: Symmetry and Equilibrium Mitigation and Adaptation

Consider an arbitrary coalition and an interior second stage equilibrium.

M-Game: *For all p, $1 < p < n$: $q^*_{j\notin P}(p) < q^*_{i\in P}(p)$ with $q^*_{j\notin P}(p) = q^*_{l\notin P}(p)$ for all $j, l \notin P$ and $q^*_{i\in P}(p) = q^*_{k\in P}(p)$ for all $i, k \in P$.*

M+A-game: $x^*_{i\in P}(p) = x^*_{j\notin P}(p)$ *for all* p*,* $1 \leq p \leq n$ *and all* $i, j \in N$. Moreover, *for all* p*,* $1 < p < n$ *:* $q^*_{j\notin P}(p) < q^*_{i\in P}(p)$ *with* $q^*_{j\notin P}(p) = q^*_{l\notin P}(p)$ *for all* $j, l \notin P$ *and* $q^*_{i\in P}(p) = q^*_{k\in P}(p)$ *for all* $i, k \in P$*.*

Both games: $\Pi^*_{j\notin P}(p) > \Pi^*_{i\in P}(p)$ *for all* p*,* $1 < p < n$*.*

Proof. Follows from the discussion above and the first order conditions (3) and (4) in the proof of Lemma 1. $q^*_{j\notin P}(p) < q^*_{i\in P}(p)$ follows from (3) and from $C_{qq} > 0$, noting that this implies $pC_q(q_{j\notin P}) = C_q(q_{i\in P})$. Finally, all players have the same benefits and costs, except signatories who have higher mitigation costs than non-signatories. ■

The importance of Lemma 2 derives from the fact that it compactly summarizes the implications of the simplification which are associated with the assumption of ex-ante symmetric players.

3.2 Existence of a Unique Interior Second Stage Equilibrium

In this subsection, we derive sufficient conditions for the existence of a unique interior second stage equilibrium for every possible coalition P of size p, $1 \leq p \leq n$. We use the concept of replacement functions, which Cornes and Hartley (2007) have shown is a convenient and elegant tool to establish existence of a unique Nash equilibrium in aggregative games. We only have to slightly modify their approach in two respects. Firstly, we view the second stage equilibrium as a Nash equilibrium between coalition P, acting de facto as a single player, and all non-signatories, who play as singletons. Secondly, in the M+A-game, and different from the M-game and the cases considered in Cornes and Hartley's paper, we need to account for the possibility of upward-sloping replacement functions as explained below. In the following, we introduce the concept of reaction and replacement functions and sketch the arguments to establish existence and uniqueness of an interior equilibrium, providing additional formal details in the proof of Proposition 1. We consider first the more comprehensive and interesting M+A-game, and briefly comment on the simpler M-game in passing. We know from the definition of the payoff function that the strategy space of each player is compact and convex and payoffs of all players are continuous and bounded in the entire strategy space. Hence, an equilibrium exists.

We first observe that the first order conditions in terms of adaptation (4) implicitly define the equilibrium adaptation levels as a function of total mitigation, $x(Q)$. Consequently, the first order conditions in terms of mitigation (3) can be written as $pB_Q(Q, x(Q)) = C_q(q)$. Now if we let $Q = q_i + Q_{-i}$, each first order condition implicitly defines q_i as a function of Q_{-i}, which is the reaction function of player i. Hence, generally, for any coalition $P \subseteq N$ we have $q_{i \in P} = r_{i \in P}(Q_{-i})$ for signatories and $q_{j \notin P} = r_{j \notin P}(Q_{-j})$ for non-signatories (setting $p = 1$ in the first order conditions of non-signatories). Clearly, reaction functions are well-known and well-suited to study the strategic interaction among players and we will use them in the next subsection for exactly this reason. However, for the purpose at hand, and given that we consider more than two players and more than one strategy, the concept of replacement functions is much simpler.

For instance, if we use the first order conditions directly and derive the individual replacement function of signatories, $q_{i \in P} = R_{i \in P}(Q)$, and of non-signatories $q_{j \notin P} = R_{j \notin P}(Q)$. The aggregate replacement function is simply derived by summing over all individual replacement functions, i.e. $Q = R(Q) = \sum_{i \in N} R_i(Q)$. The idea is illustrated in Figure 1 for the assumption of downward sloping replacement functions.[8]

"Figure 1 about here"
"Figure 2 about here"

The graphical determination of the second stage equilibrium works as follows. Firstly, the aggregate replacement function is derived as the verti-

[8]The graph assumes linear replacement functions but this does not necessarily has to be the case and is not crucial for the following arguments.

cal summation of all individual replacement functions. Notice that due to symmetry all individual replacement functions of signatories are the same, and the same applies for all non-signatories. Secondly, the intersection of the aggregate replacement function with the 45^0-degree line, point E, determines the aggregate equilibrium mitigation level because there $Q^* = R(Q^*)$ by definition. Thirdly, one draws a vertical line from point E down to Q^* on the abscissa. Finally, from the intersection point with the individual replacement functions, points e and f in the graph, one draws horizontal lines to the ordinate which gives the equilibrium individual mitigation level of signatories $q^*_{i \in P}$ and non-signatories, $q^*_{j \notin P}$.

We note that if all individual replacement functions are continuous and downward sloping over the entire strategy space, also the aggregate replacement function will have this property. If all replacement functions start at a positive value on the ordinate, all equilibrium mitigation levels will be strictly positive. Finally, the aggregate replacement function will intersect only once with the 45^0-degree line if its slope is negative over the entire domain.

The idea of upward sloping reaction functions is illustrated in Figure 2. The procedure of determining the equilibrium works exactly the same, as discussed above. However, now the absolute value of the slope of the aggregate replacement function matters. Figure 2 illustrates that the aggregate replacement function could have a slope larger than 1 everywhere, in which case it will never intersect with the 45^0-degree line. Hence, the conditions which ensure that the aggregate replacement function has a slope less than 1 are those which ensure uniqueness.

Additional Assumption

Let $A := B_{QQ} + \frac{(B_{xQ})^2}{D_{xx} - B_{xx}}$ *in the M+A-game. For all players* $i \in N$ *and* $x_i \in [0,\ \bar{x_i}]$ *and* $q_i \in [0,\ \bar{q_i}]$*:*

$$A \left[\frac{p^2}{Cqq(q_{i \in P})} + \frac{(n-p)}{Cqq(q_{j \notin P})} \right] < 1 \ .$$

The left-hand side term in the inequality listed in the Additional Assumption above is the slope of the aggregate replacement function. The sign of this slope is related to the term $A^{M+A} := B_{QQ} + \frac{(B_{xQ})^2}{D_{xx}-B_{xx}}$ in the M+A-game and $A^M := B_{QQ}$ in the M-game. If $A < 0$, replacement functions are downward sloping and no further assumptions for uniqueness are necessary. This is also true if $A = 0$ in which case individual and aggregate replacement functions are horizontal lines and hence also intersect with the 45^0-degree line only once. In the M-game $B_{QQ} \leq 0$ and hence uniqueness follows immediately. In the M+A-game, A^{M+A} can also be negative or equal to zero, but could also be positive. It is for this last possibility why we introduce the Additional Assumption as a sufficient condition which ensures that the slope is strictly smaller than 1 over the entire strategy space.

Proposition 1: Existence of a Unique Interior Equilibrium in the Second Stage

Consider an arbitrary coalition of size p, $1 \leq p \leq n$.

M-Game: *A unique interior equilibrium in the second stage always exists.*

M+A-game: *A sufficient condition for the existence of a unique interior equilibrium in the second stage, is either $A \leq 0$ or if $A > 0$, then the Additional Assumption holds.*

Proof. The left-hand side term in the Additional Assumption is the slope of the aggregate replacement function in the M+A-game. From $pB_Q(Q, x(Q)) = C_q(q_{i\in P})$ we have $q_{i\in P} = R_{i\in P}(Q) = C_q^{-1}(pB_Q(Q, x(Q)))$. $R_{i\in P}(Q)$ is continuous in Q and $q_{i\in P}$ is strictly positive if Q approaches zero because $\lim_{Q\to 0} B_Q > \lim_{q\to 0} C_q > 0$ from our General Assumptions. From the theorem of inverse functions, we have $\frac{dC_q^{-1}(q)}{dq} = \frac{1}{C_{qq}(q)}$ and hence the slope of individual replacement functions of signatories is given by $R^{'}_{i\in P}(Q) = \frac{d(C_q^{-1}(pB_Q(Q,x(Q))))}{dQ} = \frac{1}{C_{qq}(q_{i\in N})}\frac{d(pB_Q(Q,x(Q)))}{dQ} = \left(\frac{1}{C_{qq}(q_{i\in P})}pA\right)$ because $\frac{d(B_Q(Q,x(Q)))}{dQ} = A$ and of non-signatories by $R^{'}_{j\notin P}(Q) = \frac{1}{C_{qq}(q_{i\in P})}A$ respectively. In more detail, $\frac{d(B_Q(Q,x(Q)))}{dQ} = B_{QQ} + B_{Qx}\frac{dx}{dQ}$ with $\frac{dx}{dQ} = \frac{B_{xQ}}{D_{xx}-B_{xx}}$ from $B_{xx}(Q, x)dx +$

$B_{xQ}(Q,x)dQ - D_{xx}(x)dx = 0$. For the aggregate replacement function, accordingly, we derive $R^{'}(Q) = A\left[\frac{p^2}{Cqq(q_{i\in P})} + \frac{(n-p)}{Cqq(q_{j\notin P})}\right]$. Finally, if Q^* is unique, $x^*(Q^*)$ is unique because $x = f_{i\in N}(Q)$ is continuously downward sloping over the entire strategy space and $\lim_{x\to 0} B_x > \lim_{x\to 0} D_x > 0$ from our General Assumptions ensures an interior equilibrium. ■

The importance of term A will also become apparent in the next subsection.

3.3 Strategic Interaction Between Mitigation and Adaptation

In this subsection, we analyze the strategic interaction among players in terms of mitigation and the strategic relation between mitigation and adaptation for a given coalition $P \subseteq N$ of size p, $1 \leq p \leq n$. For this, we derive the slopes of the reaction functions which have been defined in the previous Subsection 3.2. For the subsequent analysis, we need to make only two additional remarks. Firstly, one can view the coalition as one player and because of symmetry all non-signatories as another player. Hence, we can define the aggregate reaction function of signatories by $Q_{i\in P} = r(Q_{j\notin P})$ and of non-signatories by $Q_{j\notin P} = r(Q_{i\in P})$, with $Q_{i\in P} = pq_{i\in P}$ and $Q_{j\notin P} = (n-p)q_{j\notin P}$ the total mitigation of signatories and non-signatories, respectively, in order to capture the strategic interaction between these two groups in a compact way. Secondly, the first order condition (4), $B_x(Q,x) = D_x(x)$, which is identical for all players, implicitly defines optimal adaptation as a function of total mitigation, $x = f_{i\in N}(Q)$, as already used in the proof of Proposition 1 above.

Proposition 2: Slopes of Reaction Functions in Mitigation and Adaptation Space

Consider an arbitrary coalition of size p, $1 \leq p \leq n$, and let primes denote the slopes of reaction functions. Further let $A := B_{QQ} + \frac{(B_{xQ})^2}{D_{xx}-B_{xx}}$ in *the M+A-game and $A := B_{QQ}$ in the M-game.*

Strategic interaction between mitigation levels in the M-game and M+A-game

The slopes of individual and aggregate reaction functions of signatories are given by $r'_{i\in P}(Q_{-i}) = \frac{pA}{C_{qq}(q_{i\in P})-pA}$ *and* $r'(Q_{j\notin P}) = \frac{p^2A}{C_{qq}(q_{i\in P})-p^2A}$*, respectively, and the slopes of non-signatories' reaction functions are given by* $r'_{j\notin P}(Q_{-j}) = \frac{A}{C_{qq}(q_{j\notin P})-A}$ *and* $r'(Q_{i\in P}) = \frac{(n-p)A}{C_{qq}(q_{j\notin P})-(n-p)A}$*.*

That is, reaction functions are always weakly downward sloping in the M-game. In the M+A-game, reaction functions are (weakly) downward sloping if $A \leq 0$ *and are (strictly) upward sloping if* $A > 0$*.*

Strategic interaction between mitigation and adaptation in the M+A-game

For each possible coalition, the slope of the individual reaction function $x = f_{i\in N}(Q)$ is given by $f'_{i\in N}(Q) = \frac{B_{Qx}}{D_{xx}-B_{xx}} < 0$.

Proof. The derivation follows the same lines as described for replacement functions in Subsection 3.2, in particular the proof of Proposition 1 and is therefore omitted. ■

The first statement sheds light whether mitigation levels are strategic substitutes or complements. In the M-game, they are always substitutes if we exclude the case $B_{QQ} = 0$ in which case reaction functions are orthogonal. In the M+A-game, this is also the case provided the term $A := B_{QQ} + \frac{(B_{xQ})^2}{D_{xx}-B_{xx}}$ is negative, again with orthogonal reaction functions for the special case if $A = 0$. However, if $A > 0$, then reaction functions are upward sloping and mitigation strategies are strategic complements.[9] Because $B_{QQ} \leq 0$, $A > 0$ if $\frac{(B_{xQ})^2}{D_{xx}-B_{xx}} > 0$ is sufficiently large, which captures the interaction between mitigation and adaptation. Intuitively, this is evident when considering the first order condition (3), $pB_Q(q_i + Q_{-i}, x(q_i + Q_{-i})) = C_q(q_i)$, using $Q = q_i + Q_{-i}$. Increasing Q_{-i} in a comparative static way (and hence Q)

[9] It is easy to show that the signs of the slopes of reaction and replacement functions are the same, they only depend on the sign of the term A.

has a direct negative effect on B_Q, namely reducing B_Q because of $B_{QQ} < 0$. Anything else being equal, this would call for a lower $C_q(q_i)$ in order for the equality to be able to hold and hence a lower q_i because $C_{qq} > 0$. However, there is also the indirect effect, which increases B_Q and hence calls for a higher q_i. Increasing Q_{-i} increases Q and calls for a lower $x(Q)$, which in turn increases B_Q because $B_{Qx} < 0$. This second indirect effect is exactly $\frac{(B_{xQ})^2}{D_{xx}-B_{xx}}$. It is important to note that even if the indirect effect dominates the direct effect, the sufficient conditions for the existence and uniqueness of second stage equilibria, as stated in the Additional Assumptions, do not need to be violated. Moreover, only the magnitude but not the the sign of the cross derivative B_{xQ} matters.

An alternative way to view this is by noticing that the second derivative of the payoff function (2) with respect to other players' mitigation levels, after inserting $x(Q)$, is exactly term A. Thus, if $A > 0$, the payoff function is not concave but convex in other players mitigation level.

Taken together, we generalize the result of Ebert and Welsch (2011, 2012) by showing that positively sloped reaction functions in mitigation space are possible for any number of players and for any degree of cooperation.

Ebert and Welsch conjecture (without proof) that upward sloping reaction functions could lead to more optimistic outcomes in a coalition formation game (i.e. larger coalitions). The intuition is that if mitigation levels are strategic substitutes, any additional increase of signatories' mitigation efforts is countervailed by a decrease of non-signatories' mitigation efforts. In the context of climate change, this has been called (carbon) leakage which makes it less attractive to join an agreement. Thus, upward sloping reaction functions may be viewed as a form of anti-leakage or matching, which may be conducive to form large stable coalitions.

The idea to relate the success of coalition formation to the slopes of reaction function is interesting. However, we have to be aware that up to now results have only been established for a given coalition P but nothing has been concluded how mitigation and adaptation changes with the degree of cooperation, which is the crucial point for the analysis of stable coalitions. We will analyze this in Section 4.

The last statement in Proposition 2 gives a clear answer to the question whether adaptation and total mitigation are substitutes or complements. They are always substitutes, irrespective of the degree of cooperation. Because the concept of substitutes and complements is not uniquely defined in the literature, Proposition 3 adds two variants to this.

Proposition 3: Alternative Views of the Strategic Interaction between Mitigation and Adaptation

Consider an arbitrary coalition of size p, $1 \leq p \leq n$ and an interior second stage equilibrium in the M+A-game.

(*1) Individual mitigation levels of non-signatories and signatories and hence also the total mitigation level are strictly lower in the M+A-game than in the M-game.*

2) Consider payoff function (1) but let the mitigation cost function be given by $\gamma C(q)$ and the adaptation cost function by $\delta D(x)$ where $\gamma > 0$ and $\delta > 0$ are strictly positive parameters. Then individual mitigation levels of signatories and non-signatories and hence also the total mitigation level decrease (increase) in γ (δ) and adaptation levels increase (decrease) in γ (δ).

Proof. The first statement can be proved by using $B_{Qx} < 0$ from the General Assumptions, noting that $x > 0$ in an interior equilibrium, and showing (using (3)) that the contradiction $Q^{M+A*}(p) \geq Q^{M^*}(p)$ is false. For individual mitigation levels one uses again (3) with $Q^{M+A*}(p) \geq Q^{M*}(p)$ and $B_{QQ} \leq 0$ and $C_{qq} > 0$ from the General Assumptions. The second statement is proved by using the first order conditions (3) and (4), the General Assumptions and $\frac{\partial x(Q)}{\partial Q} < 0$ as established in Proposition 2. ■

From the first statement we can conclude that if adaptation is available as a second strategy, less mitigation is required. Since mitigation concerns the public good part in this strategic game, one may conjecture that the incentive to leave a coalition could be less pronounced in the M+A-game than in the M-game. The driving force would be similar like in Barrett (2002) and Finus and Maus (2008) who show that modest emission reduction lead to larger stable coalitions. We test this conjecture in the next section.

The second statement relates changes of equilibrium strategies to price effects. If mitigation costs increase uniformly across players, then players will reduce their mitigation levels and increase their adaptation levels.

Thus, without doubt, considering Proposition 2 and 3 together, in our model, adaptation and mitigation are strategic substitutes. The result hinges on the certainly plausible assumption that the cross-derivative B_{xQ} is negative but would be different for $B_{xQ} > 0$.

4 First Stage of Coalition Formation

In this section, we analyze stable coalitions. In a first step, we look at the general properties of coalition formation. The purpose is to find out whether the general properties in the M+A-game are fundamentally different from those in the M-game. It will turn out that properties can be established under more general conditions in the M+A-game than in the M-game, there are differences in the two games, but they are not sufficiently pronounced to draw general conclusions about the size and the success of stable coalitions in the two games. Therefore, in a second step, we look at two specific payoff functions, which reveals interesting differences in both games.

4.1 General Properties

Proposition 4 summarizes what we know in terms of mitigation and adaptation levels when the degree of cooperation changes, i.e. the size of coalition

P (denoted by p) increases. Note that any discrete change of p (because the number of signatories must be an integer value) is captured by a continuous change and hence we can use the differential with respect to p.

Proposition 4: Equilibrium Mitigation and Adaptation and the Degree of Cooperation

Consider an arbitrary coalition of size p, $1 \leq p < n$, *and let* $A := B_{QQ} + \frac{(B_{xQ})^2}{D_{xx}-B_{xx}} ><= 0$ *in the M+A-game and* $A := B_{QQ} \leq 0$ *in the M-game. Further assume the Additional Assumption to hold in the M+A-game and let an asterisk denote equilibrium values for a given* p.

Mitigation in the M+A-game and M-game

a) Non-signatories:
i) $\frac{dq^*_{j\notin P}(p)}{dp} > 0$ *if and only if* $A > 0$;
ii) $\frac{dQ^*_{j\notin P}(p)}{dp} >< 0$ *if* $A > 0$ *and* $\frac{dQ^*_{j\notin P}(p)}{dp} < 0$ *if* $A \leq 0$;
b) Signatories:
i) $\frac{dq^*_{i\in P}(p)}{dp} > 0$ *if* $A \geq 0$ *and* $\frac{dq^*_{i\in P}(p)}{dp} >< 0$ *if* $A < 0$;
ii) $\frac{dQ^*_{i\in P}(p)}{dp} > 0$;
c) Aggregate: $\frac{dQ^*(p)}{dp} > 0$.

Adaptation in the M+A-game

d) Signatories and non-signatories: $\frac{dx^*}{dp} < 0$.

Proof. See Appendix 1. ■

Generally speaking, the change of equilibrium mitigation levels of signatories and non-signatories (statements a and b) resulting from a change of the coalition size are mostly (though not always) related to the sign of the term A and hence to the sign of the slopes of the reaction functions. Part ai confirms that non-signatories will decrease (increase) mitigation levels when the degree of cooperation increases if reaction functions are downward (upward) sloping. If a non-signatory joins the coalition, the total mitigation level of signatories, $Q_{i \in P}$, increases (Part bii), and the remaining individual non-signatories match this behavior if mitigation levels are strategic complements and undermine this effort if they are substitutes.

Clearly, moving from p to $p+1$, means one non-signatory less and hence if individual non-signatories' equilibrium provision levels $q_{j \notin P}$ drop (or remain constant) as p increases (which happens if $A \leq 0$), the total provision level of non-signatories, $Q_{j \notin P}$, will drop. However, if mitigation levels are strategic complements, then there are two opposing effects and hence overall predictions are generally not possible (Part aii).

Interestingly, despite signatories' total mitigation level always increases with the degree of cooperation (Part bii), individual mitigation levels do not necessarily have to increase (Part bi). On the one hand, one more member calls for higher individual provision levels because more players internalize the externality among them. On the other hand, before the expansion of the coalition, the new member had lower marginal mitigation costs than the old members; now when joining the coalition, the equalization of marginal mitigation costs (as a result of cost-effectiveness within the coalition) calls for a higher mitigation level of the new member but could call for lower mitigation levels of old members compared to the initial situation provided $A < 0$.

At the aggregate things are clear-cut: total mitigation level increases with the size of the coalition (Part c). As total mitigation and adaptation are strategic substitutes, it is not surprising that the opposite holds for adaptation levels (Part d). This suggests that not only in the M-game, total mitigation increases with the degree of cooperation and obtains its highest level in the social optimum, but also in the M+A-game. Because of the substitutional relation between adaptation and mitigation, for any degree of cooperation, total mitigation will be lower in the M+A-game than in the M-game as already observed in Proposition 3. Hence, the main difference between the M+A-game and the M-game relates to the fact that non-signatories may increase their mitigation levels and hence match signatories behavior if the term A is positive in the M+A-game.

We now conduct a similar analysis in terms of payoffs (see Proposition 5 below) which are ultimately relevant when it comes to evaluate the success of coalition formation (normative dimension) and the incentive to form stable coalitions (positive dimension). The normative dimension relates to cohesiveness and full cohesiveness and the positive dimension to the properties superadditivity and positive externality. Whereas (strict) cohesiveness holds trivially in an externality game, full cohesiveness is much more difficult to establish except if $A \geq 0$. In the M-game, we know this is only the case if $B_{QQ} = 0$ whereas in the M+A-game this does not constitute a special case. However, in the case of $A < 0$, things are less straightforward. The reason is that if mitigation levels are strategic substitutes, an expansion of the coalition means on the one hand higher total mitigation levels but on the other hand an increasing difference between signatories' and non-signatories' mitigation levels and hence an increasing difference in marginal mitigation costs, a source of inefficiency.

Note that $A \geq 0$ is also a sufficient condition for superadditivity to hold, which together with the positive externality property give directly full cohesiveness. Again, superadditivity could fail for some p if $A < 0$ as will become apparent from example 2 in Subsection 4.2 (see in particular footnote 11). Typically, this is the case if the absolute value of A is large and if p is small because then the leakage effect is particularly strong (i.e. reaction functions are steep and there are many non-signatories, countervailing signatories' efforts to increase mitigation). Clearly, superadditivity cannot be violated over the entire range of p as otherwise cohesiveness could not hold.

Proposition 5: Equilibrium Payoffs and the Degree of Cooperation

Let $A := B_{QQ} + \frac{(B_{xQ})^2}{D_{xx}-B_{xx}} >\!<\!= 0$ *in the M+A-game and* $A := B_{QQ} \leq 0$ *in the M-game. Further assume the Additional Assumption to hold in the*

M+A-game.

a) Both games are (strictly) cohesive.

b) In both games the positive externality property (strictly) holds.

c) In both games a sufficient condition for (strict) superadditivity is $A \geq 0$ which is also sufficient for (strict) full cohesiveness.

Proof. See Appendix 2. ■

There are three conclusions which can be derived from Proposition 5. Firstly, at a general level, the incentive structure to form large stable coalitions does not appear to be fundamentally different in the two games because both exhibit the positive externality property. Secondly, the normative motivation to search for large stable coalition can be established under sufficient conditions which are less restrictive in the M+A-game than in the M-game because $A \geq 0$ does not require linear benefit functions ($B_{QQ} = 0$) in the M+A-game. Thirdly, the same applies to superadditivity, a condition which is crucial for the stability of coalitions. In order to highlight the importance of superadditivity for stability, we provide Proposition 6.

Proposition 6: The Role of Superadditivity for Stable Coalitions

a) A non-trivial stable coalition exists in a game which is superadditive.

b) If a coalition of size $p \geq 2$ is internally stable, then the move from $p-1$ to p is superadditive.

c) If the move from $p-1$ to p is superadditive, then the payoff of signatories increases through this move .

Proof. Superadditivity implies $p\Pi^*_{i\in P}(p) \geq (p-1)\Pi^*_{i\in P}(p-1) + \Pi^*_{j\notin P}(p-1)$. a) If $p=2$, $\Pi^*_{i\in P}(p-1) = \Pi^*_{j\notin P}(p-1)$ and hence $2\Pi^*_{i\in P}(p) \geq 2\Pi^*_{j\notin P}(p-1)$ or $\Pi^*_{i\in P}(p) \geq \Pi^*_{j\notin P}(p-1)$ which is the condition for internal stability. Hence, $p=2$ is internally stable if the move from $p-1$ *to* p is superadditive. If $p=2$ is externally stable we are done. If not, then $p=3$ must be internally stable. Repeating this argument means that eventually a coalition must be externally stable, noting that the grand coalition is externally stable by definition. b) and c). We rewrite the general condition for superadditivity which gives $\Pi^*_{i\in P}(p) + (p-1)\cdot(\Pi^*_{i\in P}(p) - \Pi^*_{i\in P}(p-1)) \geq \Pi^*_{j\notin P}(p-1)$, noting that we have $\Pi^*_{j\notin P}(p-1) > \Pi^*_{i\in P}(p-1)$ from Lemma 2 and that internal stability implies $\Pi^*_{i\in P}(p) \geq \Pi^*_{j\notin P}(p-1)$. ■

Part *a* of Proposition 6 is interesting in that it establishes sufficient conditions for the existence of a non-trivial coalition. However, at this level of generality, it is not clear how large stable coalitions will be and whether they are larger in the M+A-game than in the M-game and if so on what this depends. Part *b* is similar in spirit, looking at superadditivity and internal stability in the neighborhood of a coalition of size p. The problem is that superadditivity is only a necessary condition but not a sufficient condition for internal stability. Part *c* reminds us that because non-signatories' payoffs increase with the degree of cooperation due to the positive externality property, we need for internal stability that also signatories' payoffs increases in the neighborhood of p for which superadditivity is a sufficient condition. However, even if signatories' payoffs constantly increase in p for all p, it is still difficult to predict stable coalitions. The reason is that starting from $p = 1$ in which case $\Pi^*_{i\in P}(1) = \Pi^*_{j\notin P}(1)$, gradually increasing p, we need that $\Pi^*_{i\in P}(p)$ increases faster than $\Pi^*_{j\notin P}(p-1)$ in order to have large internally stable coalitions. The central question is, however, what "faster" means. The answer is not straightforward because $\Pi^*_{i\in P}(p) < \Pi^*_{j\notin P}(p)$ for any p from Lemma 2 and hence the "fast increase" of $\Pi^*_{i\in P}(p)$ must happen within a very short interval to have internal stability at p, i.e. $\Pi^*_{i\in P}(p) \geq \Pi^*_{j\notin P}(p-1)$. Finally, to make things even worse, we cannot rule out the possibility that $\Pi^*_{i\in P}(p)$ decreases first and then increases in p and we still may have a stable coalition at the increasing part of $\Pi^*_{i\in P}(p)$. It is because of this lack of analytical tractability at the general level why all papers which analyzed stable coalitions in the M-game have considered specific payoff functions and often used simulations.

4.2 Examples

We consider two specific payoff functions which we call example 1 and 2. Both examples assume quadratic costs functions. Example 1 assumes a linear benefit function with the following payoff function

$$\Pi_{i(1)}^{M} = bQ - \frac{c}{2}q_i^2 \tag{5}$$

in the M-game and

$$\Pi_{i(1)}^{M+A} = b(1 - \gamma x_i)Q + a(1 - \lambda Q)x_i - \frac{c}{2}q_i^2 - \frac{d}{2}x_i^2 \tag{6}$$

in the M+A-game where the parameters a, b, c, d, γ and λ are assumed to be strictly positive. In example 1, $A^M = 0$ and $A^{M+A} > 0$. Example 2 assumes again a linear benefit function in terms of adaptation but a quadratic benefit function in terms of mitigation, such that we have $A^M < 0$ and $A^{M+A} ><= 0$ where the sign of A^{M+A} depends on the parameter values.

$$\Pi_{i(2)}^{M} = (aQ - \frac{b}{2}Q^2) - \frac{c}{2}q_i^2 \tag{7}$$

$$\Pi_{i(2)}^{M+A} = (aQ - \frac{b}{2}Q^2) + x_i(e - fQ) - \frac{c}{2}q_i^2 - \frac{d}{2}x_i^2 \tag{8}$$

Again, we assume all parameters a, b, c, d, e, and f to be strictly positive. For both examples we need to impose conditions such that the examples are in line with the General Assumptions and that the Additional Assumption in the M+A-game hold. This includes conditions to ensure interior second stage equilibria for every p. Those conditions as well as all subsequent results are spelled out in detail in Appendix 3. At an analytical level, the following results can be derived.

Proposition 7: Stable Coalitions in Example 1 and 2

Assume the General Assumptions as well as the Additional Assumptions to hold for example 1 and 2.

a) In example 1, $A = 0$ *and* $p^* = 2$ *and* $p^* = 3$ *in the M-game, where the second Pareto-dominates the first equilibrium. In the M+A-game,* $A > 0$ *and* $p^* \geq 3$*.*

b) In example 2, $A < 0$ *and* $p^* = 1$ *or* $p^* = 2$ *in the M-game. In the M+A-game,* $p^* \geq 3$ *if* $A \geq 0$*.*

Proof. See Appendix 3. ■

Both examples confirm the intuition that if reaction functions are upward sloping in the M+A-game, stable coalitions will be (weakly) larger in the M+A-game than in the M-game. However, in order to obtain further conclusions, we need to conduct simulations. For example 1, we would like to find out whether stable coalitions will be strictly larger in the M+A-game than in the M-game. This is simulation run 1. For example 2, we conduct three simulation runs. Simulation runs 2 and 3 assume $A > 0$ in the M+A-game, illustrating that only if the absolute value of A is large enough will stable coalitions be strictly larger than $p^* = 3$. Finally, simulation run 4 assumes $A < 0$ in the M+A-game, like in the M-game, illustrating that then stable coalitions can even be smaller in the M+A-game than in the M-game.[10,11] Apart from determing stable coalitions, the simulation runs allow us to draw interesting conclusions regarding global payoffs, in absolute but also in relative terms. We compare global payoffs in the Nash equilibrium, the social optimum and for stable coalitions. This is interesting for a particular game, but also across the two games.

[10] For simulation run 4, the term A in the M+A-game is always smaller in absolute terms than in the M-game but coalitions can be smaller. This stresses that the intuition a less negatively sloped reaction function leads to larger coalitions is wrong. It also highlights the need for simulations.

[11] Since $p^* = 1$ for some parameter values in the M+A-game in simulation run 4, superadditivity must fail when forming a two player coalition.

The main results are displayed in Table 1 to 4. The legend describes the range of parameters considered in the simulation runs. Simulation set 1 lists the total number of simulations and set 2 the number of valid runs, i.e. those simulations which observe the conditions listed in Appendix 3. If different stable coalitions emerge, set 2 is grouped according to the size of stable coalitions. For instance, in Table 1, set 2 contains 2620 simulations of which 2616 deliver a stable coalition of size 3 and 4 simulations deliver a stable coalition of size 10 in the M+A-game, which is the grand coalition in the example because $n = 10$. (All 2620 simulations deliver a coalition of size 3 in the M-game, as predicted by Proposition 7.) The average coalition size over all 2620 simulations is denoted by an upper bar in the last column. Generally, upper bars denote averages over valid simulation runs.

The interpretation of $\bar{A}$, $\bar{Q}^*(p^*)$ and $\bar{x}^*(p^*)$ are obvious where the latter two being the averages in stable coalitions of size p^*. $\bar{I}^{SO}$ ($\bar{I}^{CO}$) is the average of index $I^{SO} = \frac{\Pi^{*SO}}{\Pi^{*NE}}$ ($I^{CO} = \frac{\Pi^{*CO}}{\Pi^{*NE}}$), a relative welfare measure. Note that Π^{*SO}, Π^{*CO} and Π^{*NE} denote the total payoff in the social optimum, equilibrium coalition and the Nash equilibrium, with the superscripts SO, CO and NE, respectively, where the first two coincide if $p^* = n$ and the last two coincide if $p^* = 1$. The larger I^{SO}, the larger the difference between the social optimum and the Nash equilibrium in relative terms and hence the larger is the need for cooperation. Index I^{CO} measures the success of stable coalitions in relative terms, also relating it to the Nash equilibrium. The following comments and conclusions apply to all four simulation runs.

Firstly note that for a given p, Q^* is lower in the M+A-game than in the M-game because adaptation is available as a second strategy as we know from the previous theoretical analysis (Proposition 3). Of course, if stable coalitions are larger in the M+A-game than in the M-game, then $Q^*(p^{*M+A}) > Q^*(p^{*M})$ is possible and the same applies for averages.

Secondly, a meaningful comparison of welfare levels must be expressed in relative terms. The indexes I^{SO} and I^{CO} are lower in the M+A-game than in the M-game, and this applies not only to averages as displayed in the tables, but is also true for every simulation run. For the index I^{CO} this is even true if the grand coalition forms in the M+A-game but a much smaller coalition emerges in the M-game (simulation runs 1 and 3). The intuition why I^{SO} is larger in the M- than in the M+A-game is that cooperation is related to coordinating mitigation levels across players and mitigation generates positive externalities. In the M+A-game, even in the social optimum some adaptation is optimal and hence the gap between no cooperation and first best is smaller. The larger index I^{CO} in the M-game suggests that despite stable coalitions may be smaller, in relative terms the success is larger than in M+A-game.

The overall message is clear: under those conditions when reaction functions are upward sloping, stable coalitions may be larger in the M+A-game than in the M-game, but these are those conditions for which the relative gains from stable agreements are smaller in the M+A-game than in the M-game.

5 Conclusion

In this paper, we have analyzed how adaptation, as an additional strategy to mitigation, affects the prospects of international policy coordination to tackle climate change. More specifically, we have studied the strategic interaction between mitigation and adaptation strategies in the canonical model of international environmental agreements (IEAs). We have shown that these two strategies are strategic substitutes considering various definitions of substitutability, regardless whether countries behave non-cooperatively, partially or fully cooperative. Moreover, different from a pure mitigation game, adaptation may cause mitigation levels between different countries to be strategic complements. In those cases there is no easy-riding and the game is transformed into a matching game. We showed that this may lead to larger stable coalitions with even the grand coalition being stable. However, adaptation reduces the importance of cooperation and hence the success of stable agreements may be small in relative welfare terms. Thus, effectively, adaptation may be more important in a non-cooperative than in a partial or full cooperative setting, though it facilitates cooperation in terms of membership.

In the Introduction, we pointed out that our analysis applies not only to climate change but also to many other economic problems involving private and public goods and their strategic interaction. Examples included international policy coordination versus national policies related to anti-terrorism, migration and social policy issues. It is therefore important to recall that most of our interesting results do not depend on the sign of the cross derivative of the benefit function but only the fact that the magnitude of the cross derivative is of sufficient importance. Of course, if the cross derivative between the public and the private good were positive, both goods would no longer be strategic substitutes but complements. However, the possibility that public good provision levels are strategic complements could still emerge and our sufficient conditions for existence and uniqueness of equilibrium strategies for every possible coalition structure would carry over directly. Also the fact that the public good is underprovided as long as the grand coalition does not form is still valid. Only the welfare implications of cooperation may be expected to be higher because also the provision of the private good increases with cooperation if the cross derivative would be positive.

Our model made a couple of assumptions in order to grasp the main driving forces analytically and to render the analysis interesting. For instance, we considered one of the most widespread coalition games and stability concepts (internal and external stability in a cartel formation game) but could have considered other concepts (Bloch 1997, Finus and Rundshagen 2009 and Yi 1997). Internal and external stability implies that after a player leaves the coalition, the remaining coalition members remain in the coalition. In the context of a positive externality game, this is the weakest possible punishment after a deviation and hence implies the most pessimistic assumption about stability. This appears to be a good benchmark because we could show that with adaptation larger coalitions can be stable, including the grand coalition. What would certainly be interesting is to depart from the assumption of symmetric players in order to capture better the current discussion whether industrialized countries should support developing countries not only in their mitigation but also their adaptation efforts (Lazkano et al. (2016)). Will support in adaptation buy more mitigation? In this context, one could assume that coalition members can pool their adaptation activities as a club, deriving an additional benefit compared to non-signatories from the cost-effective production of adaptation. Essentially, this would require to model in kind-transfers apart from monetary transfers in a coalition formation model with heterogenous agents.

References

[1] d'Aspremont, C., A. Jacquemin, J.J. Gabszewicz and J.A. Weymark (1983), On the Stability of Collusive Price Leadership. "Canadian Journal of Economics", vol. 16(1), pp. 17-25.

[2] Barrett, S. (1994), Self-enforcing International Environmental Agreements. "Oxford Economic Papers", vol. 46, pp. 878-894.

[3] Barrett, S. (2002), Consensus Treaties. "Journal of Institutional and Theoretical Economics", vol. 158(4), pp. 529-547.

[4] Barrett, S. (2006), Climate Treaties and "Breakthrough" Technologies. "American Economic Review", vol. 96(2), pp. 22-25.

[5] Barrett, S. (2008), Dikes vs. Windmills: Climate Treaties and Adaptation. Johns Hopkins University, School of Advanced International Studies.

[6] Battaglini, M. and B. Harstad. (2016). Participation and Duration of Environmental Agreements. Journal of Political Economy, forthcoming.

[7] Baumol W.J. and D.F Bradford. (1972), Detrimental Externalities and Non-Convexity of the Production Set. "Economica", vol. 39, pp. 160-76.

[8] Benchekroun, H., Marrouch, W. and A.R. Chaudhuri. (2016), Adaptation Technology and Free-riding Incentives in International Environmental Agreements, in Economics of International Environmental Agreements: A Critical Approach. Routledge, UK, forthcoming.

[9] Bergstrom T. C. and R. P.Goodman (1973), Private Demands for Public Goods. "American Economic Review", vol. 63(3), pp. 280–296.

[10] Bergstrom, T., L. Blume, and H. Varian. (1986), On the Private Provision of Public Goods. "Journal of Public Economics", vol. 29.1, pp. 25-49.

[11] Bloch, F. (2003), Non-cooperative Models of Coalition Formation in Games with Spillovers. In: Carraro, C. (ed.), The Endogenous Formation of Economic Coalitions. Edward Elgar, Cheltenham, UK et al., ch. 2, pp. 35-79.

[12] Boercherding T. E. and R. T.Deacon (1972), The Demand for the Services of Non-federal Governments, "American Economic Review", vol. 62, 891–901.

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

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

[15] Cornes, R., and T. Sandler (1984). Easy Riders, Joint Production, and Public Goods. "The Economic Journal", vol. 94.375, pp. 580-598.

[16] Cornes, R. and T. Sandler. (1986), The Theory of Externalities, Public Goods and Club Goods. Cambrige University Press.

[17] Cornes, R. and T. Sandler (1994), The Comparative Static Properties of the Impure Public Good Model. "Journal of Public Economics", vol. 54(3), pp.403-421.

[18] Cornes, R.C. and R. Hartley (2007), Aggregative Public Good Games. "Journal of Public Economic Theory", vol. 9(2), pp. 201–219.

[19] Deneckere, R. and C. Davidson (1985), Incentives to Form Coalitions with Bertrand Competition. "The RAND Journal of Economics", vol. 16(4), pp. 473-486.

[20] Donsimoni, M.-P., N.S. Economides and H.M. Polemarchakis (1986), Stable Cartels. "International Economic Review", vol. 27(2), pp. 317-327.

[21] 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.

[22] 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.

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

[24] El-Sayed, A. and S. Rubio (2014), Sharing R&D Investments in Cleaner Technologies to Mitigate Climate Change, "Resource and Energy Economics", vol.38, pp. 168-180.

[25] Finus, M. and S. Maus (2008), Modesty May Pay! "Journal of Public Economic Theory", vol. 10(5), pp. 801-826.

[26] Finus, M. and B. Rundshagen (2009), Membership Rules and Stability of Coalition Structures in Positive Externality Games. "Social Choice and Welfare", vol. 32, pp. 389-406.

[27] Finus, M. and D.T.G. Rübbelke (2013), Public Good Provision and Ancillary Benefits: The Case of Climate Agreements. "Environmental and Resource Economics", vol. 56, pp. 211-222.

[28] Finus, M. and A. Caparrós (2015), Game Theory and International Environmental Cooperation: Essential Readings. Edward Elgar, Cheltenham.

[29] Fraser, C.D. (1992), The Uniqueness of Nash Equilibrium in the Private Provision of Public Goods: an Alternative Proof. "Journal of Public Economics", vol. 49, pp. 389–390.

[30] Haeringer, G. (2004), Equilibrium Binding Agreements: A Comment. "Journal of Economic Theory", vol. 117(1), pp. 140-143.

[31] Harstad, B. (2012), Climate contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations. "The Review of Economic Studies", vol. 79(4), pp. 1527-1557.

[32] Hirshleifer, J. (1983), From Weakest-Link to Best-Shot: The Voluntary Provision of Public Goods. Public Choice, vol. 41, pp. 371-386.

[33] Hoel, M. and A. de Zeeuw (2010), Can a Focus on Breakthrough Technologies Improve the Performance of International Environmental Agreements? "Environmental and Resource Economics", vol. 47, pp. 395-406.

[34] Kolstad, C.D. (2007), Systematic Uncertainty in Self-enforcing International Environmental Agreements. "Journal of Environmental Economics and Management", vol. 53(1), pp. 68-79.

[35] Laffont, J-J. (1976), Decentralization with Externalities. "European Economic Review", vol. 7, pp. 359-75.

[36] Lazkano, I., W. Marrouch, B. Nkuiya (2016), Adaptation to Climate Change: How does Heterogeneity in Adaptation Costs Affect Climate Coalitions?, "Environment and Development Economics", forthcoming.

[37] Poyago-Theotoky, J. (1995), Equilibrium and Optimal Size of A Research Joint Venture in an Oligopoly with Spillovers. "The Journal of Industrial Economics", vol. 43(2), pp 209-226.

[38] Rubio, S.J. and A. Ulph (2006), Self-enforcing International Environmental Agreements Revisited. "Oxford Economic Papers", vol. 58(2), pp. 233-263.

[39] Selden, T. and D. Song (1994), Environmental Quality and Development: is there a Kuznets Curve for Air Pollution Emissions?. "Journal of Environmental Economics and Management", vol. (27), pp. 147-162.

[40] Starrett, D.A. (1972), Fundamental Nonconvexities in the Theory of Externalities. "Journal of Economic Theory", vol. 4.2, pp. 180-199.

[41] Vives, X. 2005, Complementarities and Games: New Developments. "Journal of Economic Literature", vol. 43.2, pp. 437-479.

[42] Weikard, H-P. (2009), Cartel Stability Under an Optimal Sharing Rule. "The Manchester School", vol. 77.5, pp. 575-593.

[43] Yi, S.-S. (1997), Stable Coalition Structures with Externalities. "Games and Economic Behavior", vol. 20(2), pp. 201-237.

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

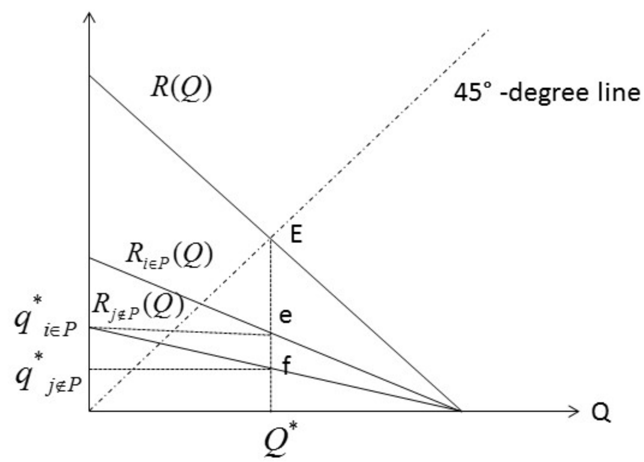

Figure 1: Downward-sloping Replacement Functions

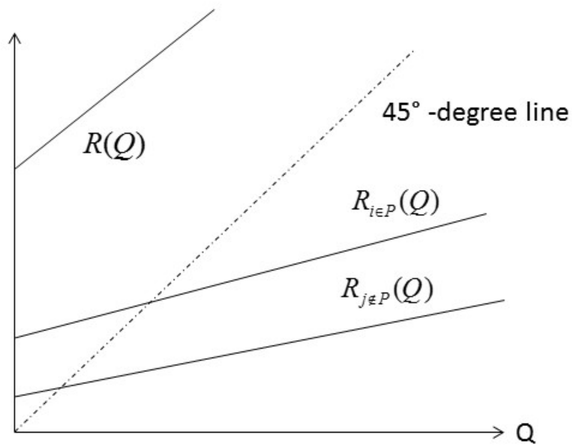

Figure 2: Upward-sloping Replacement Functions

Table 1: Example 1: Simulation run 1

		p^{*M+A} (p^{*M})	p^{*M+A} (p^{*M})	$\bar{p}^{*M+A}$ $(\bar{p}^{*M})$
		3 (3)	10 (3)	3.01 (3)
Set 2		2616	4	2620
$\bar{A}$	M+A	28.23	36.15	28.24
	M	0	0	0
$\bar{Q}^*(p^*)$	M+A	0.37	5.65	0.38
	M	1.25	1.42	1.25
$\bar{x}^*(p^*)$	M+A	0.72	0.36	0.72
$\bar{I}^{SO}$	M+A	1.25	1.01	1.25
	M	5.26	5.26	5.26
$\bar{I}^{CO}$	M+A	1.0135	1.0138	1.01
	M	1.5	1.5	1.5
$N = 10$; $a = 3000$, $\gamma = \lambda = 0.1$, parameter b moves from 300 to 500 by 10, and parameters (c, d) move from 3000 to 5000 by 100. Set 1= 9261, Set 2= 2620.				

Table 2: Example 2: Simulation run 2

		p^{*M+A} $(p^{*M}) = \bar{p}^{*M+A}$ $(\bar{p}^{*M})$
		3 (2)
Set 2		18676
$\bar{A}$	M+A	2.277
	M	-0.5651
$\bar{Q}^*(p^*)$	M+A	0.03
	M	0.04
$\bar{x}^*(p^*)$	M+A	0.43
$\bar{I}^{SO}$	M+A	1.5820
	M	4.9577
$\bar{I}^{CO}$	M+A	1.0515
	M	1.1765
$N = 10$; $a = 2.5$, $e = 1$, parameter b moves from 0.1 to 1 by 0.1, parameter c moves from 500 to 1000 by 10, parameter d moves from 2 to 2.3 by 0.1, and parameter f moves from 2 to 3 by 0.1. Set 1= 22440, Set 2= 18676.		

Table 3: Example 2: Simulation run 3

		p^{*M+A} (p^{*M})	p^{*M+A} (p^{*M})	$\bar{p}^{*M+A}$ $(\bar{p}^{*M})$
		3 (2)	10 (2)	3.1 (2)
Set 2		3961	6	3967
$\bar{A}$	M+A	4.305	5.03	4.306
	M	−0.5569	−0.35	−0.5566
$\bar{Q}^*(p^*)$	M+A	0.01	0.25	0.01
	M	0.04	0.06	0.04
$\bar{x}^*(p^*)$	M+A	1.03	0.5	1.03
$\bar{I}^{SO}$	M+A	1.0264	1.0124	1.0264
	M	4.9499	4.9827	4.9499
$\bar{I}^{CO}$	M+A	1.0014	1.0124	1.0014
	M	1.1764	1.1767	1.1764
$N = 10$; $a = 2.5$, $e = 1$, parameters (b, d) move from 0.1 to 1 by 0.1, parameter c moves from 500 to 1000 by 10, and parameter f moves from 2 to 3 by 0.1. Set 1= 56100, Set 2= 3967.				

Table 4: Example 2: Simulation run 4

		p^{*M+A} (p^{*M})	p^{*M+A} (p^{*M})	$\bar{p}^{*M+A}$ $(\bar{p}^{*M})$
		1 (2)	2 (2)	1.07 (2)
Set 2		1596	105	1701
$\bar{A}$	M+A	-511.66	-106.66	-486.66
	M	-525	-120	-500
$\bar{Q}^*(p^*)$	M+A	4.45	9.36	4.75
	M	5.65	10.83	5.97
$\bar{x}^*(p^*)$	M+A	3.04	2.71	3.02
$\bar{I}^{SO}$	M+A	1.0762	1.4540	1.0995
	M	1.169	1.7124	1.2025
$\bar{I}^{CO}$	M+A	1	1.0578	1.0036
	M	1.0328	1.0958	1.0367

$N = 10$; $c = d = 3000$, $e = 10000$, $f = 200$, parameter a moves from 3000 to 5000 by 100, and parameter b moves from 100 to 900 by 10. Set 1= 1701, Set 2= 1701.

Appendix (all or partly online)

Appendix 1: Proof of Proposition 4

In order to save on notation, we drop the asterisk and p as an argument in equilibrium mitigation and adaptation levels. Moreover, we will omit the arguments in the benefit function because they are the same for signatories and non-signatories, but keep them for the mitigation cost function because $q_{i\in P} > q_{j\notin P}$ for every p, $1 < p < n$, from Lemma 2. We use three pieces of information in order to obtain the following results.

1. Total differentiating the first order conditions (3), $pB_Q(Q,\ x(Q)) = C_q(q_i)$, using $x(Q)$ and recalling that $\frac{dB_Q(Q,x(Q))}{dQ} = A$, gives $dp \cdot B_Q + p \cdot A \cdot dQ = C_{qq}(q_{i\in P}) \cdot dq_i$ for signatories with $p \geq 2$ and $A \cdot dQ = C_{qq}(q_{j\notin P}) \cdot dq_{j\notin P}$ for non-signatories with $p = 1$.

2. Noting that the first order conditions (3) imply $p \cdot C_q(q_{j\notin P}) = C_q(q_{i\in P})$, total differentiation gives $dp \cdot C_q(q_{\ j\notin P}) + p \cdot C_{qq}(q_{j\notin P}) \cdot dq_{j\notin P} = C_{qq}(q_{i\in P}) \cdot dq_{i\in P}$.

3. $\frac{dQ_{i\in P}}{dp} = q_{i\in P} + p \cdot \frac{dq_{i\in P}}{dp}$, $\frac{dQ_{j\notin P}}{dp} = -q_{j\notin P} + (n-p) \cdot \frac{dq_{j\notin P}}{dp}$ and $\frac{dQ}{dp} = \frac{dQ_{i\in P}}{dp} + \frac{dQ_{j\notin P}}{dp}$.

Now, some basic manipulations lead to the following results. ai) $\frac{dq_{j\notin P}}{dp} = \frac{A}{C(q_{j\notin P})}\frac{dQ}{dp}$, implying because $C_{qq}(q_{j\notin P}) > 0$ and $\frac{dQ}{dp} > 0$ as we show below that the sign depends on the sign of A. aii) $\frac{dQ_{j\notin P}}{dp} = \frac{-q_{j\notin P}\left(1-\frac{Ap^2}{C_{qq}(q_{i\in P})}\right)+\frac{(n-p)A}{C_{qq}(q_{j\notin P})}\left(q_{i\in P}+p\frac{C_q(q_{j\notin P})}{C_{qq}(q_{i\in P})}\right)}{1-A\left(\frac{p^2}{C_{qq}(q_{i\in P})}+\frac{n-p}{C_{qq}(q_{j\notin P})}\right)}$ where $1 - A\left(\frac{p^2}{C_{qq}(q_{i\in P})} + \frac{n-p}{C_{qq}(q_{j\notin P})}\right) > 0$ by the Additional Assumptions and hence the denominator is always positive. We note that $\left(1 - \frac{Ap^2}{C_{qq}(q_{i\in P})}\right) > 0$ by the Additional Assumptions. Consequently, if $A \leq 0$, the nominator is negative and hence $\frac{dQ_{j\notin P}}{dp} < 0$ follows. If $A > 0$, the expression cannot be signed because an increase in p by one unit implies one non-signatory less but $q_{j\notin P}$ increases in p as shown above. bi) $\frac{dq_{i\in P}}{dp} = \frac{B_Q}{C_{qq}(q_{i\in P})} + \frac{dQ}{dp}\frac{Ap}{C(q_{i\in P})}$ where we note that the first term on the R.H.S. is positive by the General Assumptions and the sign of the second term depends on A. Hence if $A \geq 0$, then $\frac{dq_{i\in P}}{dp} > 0$, otherwise if $A < 0$ this expression cannot be signed. bii) $\frac{dQ_{i\in P}}{dp} = \frac{(q_{i\in P}-q_{j\notin P})+p\frac{C_q(q_{j\notin P})}{C_{qq}(q_{i\in P})}+q_{i\in P}\left(1-\frac{Ap^2}{C_{qq}(q_{i\in P})}\right)-\frac{(n-p)A}{C_{qq}(q_{j\notin P})}\left(q_{i\in P}+p\frac{C_q(q_{j\notin P})}{C_{qq}(q_{i\in P})}\right)}{1-A\left(\frac{p^2}{C_{qq}(q_{i\in P})}+\frac{n-p}{C_{qq}(q_{j\notin P})}\right)}$ where the denominator is positive by the Additional Assumptions, $q_{i\in P} - q_{j\notin P} > 0$ by Lemma 2, $1-\frac{Ap^2}{C_{qq}(q_{i\in P})} > 0$ by the Additional Assumptions, and $p\frac{C_q(q_{j\notin P})}{C_{qq}(q_{i\in P})} > 0$ by the General Assumptions. Consequently, if $A \leq 0$, the nominator is positive and $\frac{dQ_{i\in P}}{dp} > 0$ is evident. If $A > 0$, we use simply $\frac{dQ_{i\in P}}{dp} = q_{i\in P} + p\left[\frac{B_Q}{C_{qq}(q_{i\in P})} + \frac{Ap}{C_{qq}(q_{i\in P})}\frac{dQ}{dp}\right] > 0$. c) $\frac{dQ}{dp} = \frac{(q_{i\in P}-q_{j\notin P})+p\frac{C_q(q_{j\notin P})}{C_{qq}(q_{i\in P})}}{1-\frac{A}{C_{qq}(q_{j\notin P})}\left[p^2\frac{C_{qq}(q_{j\notin P})}{C_{qq}(q_{i\in P})}+(n-p)\right]}$ where we notice that the nominator is obviously positive and the denominator is positive by the Additional Assumptions. d) $\frac{dx}{dp} = \frac{B_{xQ}}{D_{xx}-B_{xx}}\frac{dQ}{dp}$ where the first term on the R.H.S is negative by the General Assumptions and $\frac{dQ}{dp} > 0$ as shown above. ***Q.E.D.***

Appendix 2: Proof of Proposition 5

a) Obvious and hence omitted. b) Derivations similar to those described in the proof of Proposition 4 deliver $\frac{d\Pi^*_{j\notin P}}{dp} = B_Q\left[\frac{dQ^*}{dp}\left(1-\frac{A}{C_{qq}(q_{j\notin P})}\right)\right]$ which is positive because $\frac{dQ^*}{dp} > 0$ from Proposition 4 and $\left(1-\frac{A}{C_{qq}(q_{j\notin P})}\right) > 0$ by the Additional Assumptions in the M+A-game and because of $A := B_{QQ} \leq 0$ in the M-game. c) Superadditivity implies $p\Pi^*_{i\in P}(p) \geq (p-1)\Pi^*_{i\in P}(p-1) + \Pi^*_{j\notin P}(p-1)$, $1 < p \leq n$. Consider the M+A-game. Step 1: On the right-hand side of the inequality, the equilibrium values are $Q^*(p-1)$, $q^*_{i\in P}(p-1)$, $q^*_{j\notin P}(p-1)$ and $x^*(p-1)$ with $q^*_{i\in P}(p-1) > q^*_{j\notin P}(p-1)$. Now, we deduct ε from all signatories' mitigation levels and set one non-signatory j's mitigation level to exactly the same value, i.e. $q'^*_{j\notin P}(p-1) = q^*_{i\in P}(p-1) - \epsilon$, keeping all other non-signatories mitigation level constant, choosing ε such that $Q^*(p-1)$ does not change. (Hence, $\varepsilon = \frac{\left(q^*_{i\in P}(p-1)-q^*_{j\notin P}(p-1)\right)}{p}$.) Hence, benefits do not change, but costs will drop because $pC(q^*_{i\in P}(p-1)-\epsilon) < (p-1)C(q^*_{i\in P}(p-$

$1) + C(q^*_{j\notin P}(p-1)$. We denote the payoff derived from the marginal change in step 1 for the p players by $\Pi^{\sup(1)}_{i\in P}(p)$ and hence can conclude $p\Pi^{\sup(1)}_{i\in P}(p) > (p-1)\Pi^*_{i\in P}(p-1) + \Pi^*_{j\notin P}(p-1)$. Step 2: If $A \geq 0$, for all other non-signatories $k \notin j$, $q^*_{k\notin P}(p-1) \leq q^*_{k\notin P}(p)$ and because $\frac{\partial \Pi_i}{\partial q_k} > 0$, we have from step 2, $\Pi^{\sup(2)}_{i\in P} \geq \Pi^{\sup(1)}_{i\in P}$. Step 3: $\max\limits_{q,x} p\Pi_{i\in P}(p) = p\Pi^*_{i\in P}(p) \geq p\Pi^{\sup(2)}_{i\in P}$. (Hence, moving from the right-hand side to the left-hand side of the SAD-condition the aggregate payoff of the enlarged coalition increases because total costs among the p players decrease (step 1), all outsiders increase their mitigation level (step 2) and the players in the enlarged coalition can freely choose adaptation (step 3).) A slight modification of this proof applies to the M-game for $A = 0$. ***Q.E.D.***

Appendix 3: Proof of Example 1 and 2

Example 1

For example 1, in the M+A-game, we have:

- $B_Q = b - \Lambda x_i$ with $\Lambda = b\gamma + a\lambda$;
- $B_{QQ} = 0$;
- $B_{Qx} = -\Lambda < 0$;
- $B_x = a - \Lambda Q$;
- $B_{xx} = 0$;
- $C_q = cq_i$ and $C_{qq} = c$, and $D_x = dx_i$ and $D_{xx} = d$;
- $A = B_{QQ} + \dfrac{(B_{Qx})^2}{D_{xx} - B_{xx}} = \dfrac{(-\Lambda)^2}{d} = \dfrac{\Lambda^2}{d} > 0$;
- The Additional Assumption is most restrictive if $p = n$: $\frac{An^2}{C_{qq}} < 1 \Leftrightarrow (cd - n^2\Lambda^2) > 0$;

- $q^*(p) = \frac{p(db - \Lambda a)}{cd - \Lambda^2(n-p+p^2)}$ and $x^*(p) = \frac{ca - b\Lambda(n-p+p^2)}{cd - \Lambda^2(n-p+p^2)}$ with $\frac{\partial q^*(p)}{p} > 0$ and $\frac{\partial x^*(p)}{p} < 0$ if C5 to C7 below hold, noting that $n^2 \geq n - p + p^2$ for $p \leq n$ with $n - p + p^2$ increasing in p.

We need to assume the following conditions to hold:

- C1: $1 - \gamma x_i > 0$ where x_i takes on the largest value in the Nash equilibrium;
- C2: $1 - \lambda Q > 0$ where Q takes on the largest value in the social optimum;
- C3: $B_Q = b - \Lambda x_i > 0$ where x_i takes on the largest value in the Nash equilibrium;
- C4: $B_x = a - \Lambda Q$ where Q takes on the largest value in the social optimum;
- C5: $cd - n^2\Lambda^2$;
- C6: $db - \Lambda a > 0$;
- C7: $ca - n^2 b\Lambda > 0$

where C1 and C2 are required for the payoff function to make sense, C3 and C4 are required to be in line with the General Assumptions, C5 is the sufficient condition for existence and uniqueness of a second stage equilibrium, which is most restrictive in the social optimum in the example, and C6 and C7 are required to have an interior equilibrium for every p, $1 \leq p \leq n$. Inserting the maximum values in C3 and C4, it will be apparent that these two conditions are captured by C6 and C7, respectively, and hence can be dropped.

In the M-game, we find $q^*(p) = \frac{pb}{c}$ where no conditions need to be imposed.

Internal stability in the M+A-game, $IS = \Pi_{i\in P}(p) - \Pi_{i\notin P}(p-1)$, for $p = 3$ is given by

$$IS = \frac{-4c(db - \Lambda a)\Psi}{(-cd + \Lambda^2 n + 6\Lambda^2)^2(-cd + \Lambda^2 n + 2\Lambda^2)^2} \tag{9}$$

where the denominator is clearly positive and $db - \Lambda a > 0$ from C6 above. Hence, $IS \geq 0$ if $\Psi \leq 0$ which is given by:

$$\Psi = (\Lambda^4(db - \Lambda a))n^2 + ((db - \Lambda a)\Lambda^2(3\Lambda^2 - cd))n + (\Lambda^2 cd(db - \Lambda a)). \tag{10}$$

$\Psi \leq 0$ iff $\Lambda^2 n^2 + (3\Lambda^2 - cd)n + cd \leq 0$ or $\Lambda^2 n^2 + 3n\Lambda^2 \leq cd(n-1)$. Because $cd > \Lambda^2 n^2$ from C5 above, we need $\Lambda^2 n^2 + 3n\Lambda^2 \leq \Lambda^2 n^2(n-1)$ which is true for $n \geq 3$. Inserting $n = 3$ directly into Ψ above gives also $\Psi < 0$. Hence, if $n \geq 3$, $IS > 0$ for $p = 3$ and hence $p = 2$ cannot be externally stable. If $p = 3$ is externally stable, $p^* = 3$, otherwise some $p^* > 3$ must be stable, noting that the grand coalition is externally stable by definition. In the M-game, it is easy to check that $p^* = 3$ which Pareto-dominates $p^* = 2$.

Example 2

In the M+A-game, we have:

- $B_Q = a - bQ - fx_i$;
- $B_{QQ} = -b < 0$;
- $B_{Qx} = -f < 0$;
- $B_x = e - fQ$;
- $B_{xx} = 0$;
- $C_q = cq_i$ and $C_{qq} = c$, and $D_x = dx_i$ and $D_{xx} = d$;
- $A = B_{QQ} + \dfrac{(B_{Qx})^2}{D_{xx} - B_{xx}} = (-b) + \dfrac{(-f)^2}{d} = \dfrac{f^2 - bd}{d}$ with $A < 0$ if $f^2 - bd < 0$ and $A \geq 0$ if $f^2 - bd \geq 0$;

- The Additional Assumption is most restrictive if $p = n$: $\frac{An^2}{c_{qq}} < 1 \Leftrightarrow cd - n^2(f^2 - bd) > 0$;

- $q^*(p) = \frac{p(ad-ef)}{cd-(n-p+p^2)(f^2-bd)}$ and $x^*(p) = \frac{ce-(n-p+p^2)(af-be)}{cd-(n-p+p^2)(f^2-bd)}$ with $\frac{\partial q^*(p)}{p} > 0$ and $\frac{\partial x^*(p)}{p} < 0$ if C3 to C5 below hold, noting that $n^2 \geq n - p + p^2$ for $p \leq n$ with $n - p + p^2$ increasing in p.

We need to assume the following conditions to hold:

- C1: $a - bQ - fx_i > 0$ where Q (resp. x_i) takes on the largest value in the social optimum (resp. Nash equilibrium);

- C2: $e - fQ > 0$ where Q takes on the largest value in the social optimum;

- C3: $cd - n^2(f^2 - bd) > 0$;

- C4: $ad - ef > 0$;

- C5: $ce - n^2(af - be) > 0$

where C1 and C2 are required to be in line with the General Assumptions, C3 is the sufficient condition for existence and uniqueness of an equilibrium, which is most restrictive in the social optimum in this example, C4 and C5 are required to have an interior equilibrium for every p, $1 \leq p \leq n$. Inserting the maximum values in C1, it turns out that C4 captures C1 and hence C1 can be dropped.

In the M-game, we have $q^*(p) = \frac{pa}{p^2b+c}$ where no conditions need to be imposed.

Internal stability in the M+A-game, $IS = \Pi_{i \in P}(p) - \Pi_{i \notin P}(p-1)$, for $p = 3$ is given by:

$$IS = -\frac{4c(ad-ef)^2(-f^2+bd)((-f^2+bd)n(n+3)+cd(n-1))}{(cd+6bd-6f^2-nf^2+ndb)^2(cd+2bd-2f^2-nf^2+ndb)^2} \quad (11)$$

where the denominator is clearly positive. Assume $A = 0$ and hence $IS = 0$. Assume now $A > 0$, and hence $f^2 > bd$, then $IS \geq 0$ iff $(-f^2 + bd)n(n+3) + cd(n-1) > 0$. Because $cd > n^2(f^2 - bd)$ due to C3, we have $(-f^2+bd)n(n+3)+cd(n-1) > (-f^2+bd)n(n+3)+n^2(f^2-bd)(n-1) = (f^2-bd)n(n^2-2n-3) \geq 0$ if $n \geq 3$.

Internal stability in the M-game is given by:

$$IS = -\frac{a^2c(p-1)\Omega}{2(c+bp^2+bn-bp)^2(c+bp^2-3bp+2b+bn)^2} \tag{12}$$

where the denominator is clearly positive. $IS \geq 0$ iff $\Omega < 0$ where

$$\begin{aligned} \Omega \quad = \quad & p^5b^2 - 5b^2p^4 + 2p^3cb + 2p^3nb^2 + 7b^2p^3 - 8cbp^2 - 4b^2p^2n - 3b^2p^2 \\ & + n^2b^2p - 2b^2pn + 6cbp + 2npcb + pc^2 - 2cbn - 3c^2 - 4bc + b^2n^2 \end{aligned} \tag{13}$$

Now we have:

$$\begin{aligned} \frac{\partial\Omega}{\partial p} \quad = \quad & 5b^2p^3(p-4) + cbp(6p-16) + b^2pn(6p-8) + b^2p(21p-6) \\ & + b^2n(n-2) + (6bc + 2cbn + c^2) \end{aligned} \tag{14}$$

which is clearly positive for $p \geq 4$. Inserting $p = 3$, $p = 2$ and $p = 1$ will also confirm that this term is positive, i.e. $\frac{\partial\Omega}{\partial p} > 0$ for all p, $1 \leq p \leq n$. Hence, we check Ω for $p = 3$ and find:

$$12nb^2 - 4cb + 4b^2n^2 + 4cbn > 0 \tag{15}$$

for $n \geq 3$ and hence $IS < 0$ for $p = 3$. For $p = 2$, it is easily checked that $\Omega \leq 0$ and hence $IS \geq 0$ is possible, depending on the parameter values.