# Members, joiners, free-riders, supporters

January 13, 2015

**Abstract**

We augment the standard cartel formation game from non-cooperative coalition theory, often applied in the context of international environmental agreements on climate change, with the possibility that singletons support coalition formation without becoming coalition members themselves. Rather, their support takes the form of a monetary transfer to the coalition, which increases the members' payoffs, and thereby provides an incentive for other singletons to join the coalition. We show that, under mild conditions on the costs and benefits of contributing to the public good (i.e. abatement of $CO_2$ emissions), supporters exist in equilibrium. The existence of supporters increases the size of stable coalitions, increases abatement of $CO_2$ emissions, and increases payoffs to each of three types of agents: members, free-riders, and supporters. Importantly, this result does not require *commitment*.

**Keywords**: Coalition formation; Public goods; Support; Transfers; International Environmental Agreements

**JEL** classification: C72; D02; H41; Q54

# 1 Introduction

We augment the standard cartel formation game from non-cooperative coalition theory, often applied in the context of international environmental agreements (IEAs, cf. Hoel, 1992; Carraro and Siniscalco, 1993; Barrett, 1994), with the possibility that singletons support coalition formation without becoming coalition members themselves. Rather, their support takes the form of a monetary transfer to the coalition, which increases the members' payoffs, and thereby provides an incentive for other singletons to join the coalition. In the model, a larger coalition implies an increase in provisioning of the public good, which is the prime incentive to provide transfers to the coalition. We refer to these paying singletons as *supporters*. We show that there exist equilibria (that is, stable coalition structures), with a strictly positive number of supporters. The existence of supporters increases the size of stable coalitions, increases contributions to the public good, and increase payoffs to each of three types of agents: members, free-riders, and supporters.

A standard application of the model used in the current paper is the formation of IEAs for the case of climate change mitigation (Finus, 2003). Up till now, cooperation by countries to reduce greenhouse gas emissions has taken off slowly with the Kyoto Protocol as the prime example. To the best of our knowledge, however, this protocol—as well as other IEAs—has not (yet) seen financial support by non-participating countries. One possible reason is that many IEAs, including the Kyoto Protocol, appear to be void in the sense that ratifying countries would have reached their objective also non-cooperatively (Finus and Tjøtta, 2003; Böhringer and Vogt, 2004). If this feature of IEAs is common knowledge among singletons, there is no incentive for support.

Yet, when cooperating countries would manage to agree on actions that go beyond non-cooperative behaviour, our paper demonstrates that augmenting IEAs with the option of support would significantly improve the prospects for wider cooperation. This result is substantiated in the following two examples from other fields than climate change economics. The first example is given by peace-keeping missions. Typically a small core group of countries (i.e. the coalition) joins to lead the mission in terms of sending troops and equipment (Bove and Elia, 2011), with a set of other countries supporting this coalition financially (Khanna et al., 1998; Shimizu and Sandler, 2002). In the case of UN-missions, this support is partly formalized trough a sharing rule for peace-keeping costs, although countries may opt out. In the case of non-UN missions, regional aspects and private incentives affect whether countries send troops or choose to provide financial support (Bobrow and Boyer, 1997; Gaibulloev et al., 2009).

The second example is the occurrence of private charitable giving by donors to volunteer organizations. Typically, such charities use part of their gift income to expand their activities by recruiting additional volunteers, and possibly compensating them for any costs incurred. Contrary to donating money, volunteers donate their time to work for the charity. Volunteer organizations therefore raise two types of charitable activities: a group of volunteers (i.e. the coalition) contributes their time to provide the charitable good, with a set of donors supporting the charity financially.[1]

The model that we use is a standard cartel formation game which features simultaneous membership decisions, a single coalition, and open membership. We augment this model with the option that singletons support the coalition (we discuss alternative support options in Section 6). Importantly, we separately model the process of singleton agents joining the coalition in response to support. We refer to these agents as *joiners*. Subsequently, we solve the game for its Nash equilibria in membership and support strategies, which implies that no agent has an incentive to unilaterally change its membership or its support decision. The main result shows that, under mild conditions on the costs and benefits of contributing to the public good (i.e. abatement of $CO_2$ emissions), supporters exist in equilibrium.

Our paper is inspired by a classic paper on IEA formation by Carraro and Siniscalco (1993, henceforth CS93), who analyze four types of *commitment* to expand coalitions beyond what is feasible non-cooperatively. One of these (see their Proposition 5 on *external commitment*) is the possibility that a subset of singletons makes transfers to the coalition, thereby inducing the remaining singletons to join the coalition. This resembles the setting of the current paper, but, importantly, CS93 make three assumptions that we relax (see Section 3). The first two assumptions are that CS93 fix the level of supporting transfers and they fix the number of joiners at the maximum level. We drop both assumptions by endogenizing the level of transfers as well as the number of joiners. The third assumption is that CS93 require *commitment* among the supporters. This type of commitment implies that supporters will not deviate *ex post*, even if this is profitable to them. Finus (2003) correctly observes that: *"... commitment is not compatible with the notion of self-enforcing IEAs. In fact, assuming enough commitment, any problem of cooperation can trivially be solved"*. We therefore drop the commitment assumption and, instead, impose appropriate stability conditions on the supporters.

CS93 find that commitment is required to prevent instability, without further examining the sensitivity of this result with respect to the model setup. Subsequent research has taken this result for granted, so that further analysis on external support has been largely

[1] Although the two types may partially overlap within households or individuals (Freeman, 1997), charitable giving and volunteering are generally seen as substitutes rather than complements (Feldman, 2010; Bauer et al., 2013).

neglected in the literature (as opposed to the other types of commitment analyzed by CS93, but see Ansink and Bouma, 2013). This lack of attention is somewhat surprising since the potential increase in cooperation due to support is substantial. Our paper therefore contributes to the policy discussion on the design, or architecture, of IEAs (Aldy and Stavins, 2009). Specifically, the option of support allows IEA participation by countries in differentiated ways, which is likely to broaden participation as well as the scope for agreement negotiations (Olmstead and Stavins, 2012).

In the next section we introduce the model. In Section 3 we compare our stability conditions with those by CS93. In Section 4 we present the main result on the existence of a positive number of supporters in equilibrium. In Section 5 we illustrate our existence result using a workhorse IEA model specification due to Barrett (1994). Conclusions are provided in Section 6.

# 2 An IEA model with supporters

The standard coalition (cartel) formation game is a two-stage game where agents choose to sign up as coalition members in stage 1. In stage 2, a conventional public goods game is played in which each agent chooses his contribution to the public good (i.e. abatement of $CO_2$ emissions). In this game, coalition members act as a single agent by coordinating their contributions. How contributions are coordinated and how payoffs are distributed among coalition members is not always obvious. In the domain of international environmental agreements (IEAs) the conventional assumption is that members maximize joint welfare. In a transboundary pollution setting, for example, members are assumed to equate individual marginal abatement costs to the sum of members' marginal abatement benefits (Barrett, 1994). However, other behavioural assumptions have also been discussed and applied, usually in the context of asymmetric agents. These include assumptions on permit allocation schemes and on the distribution of payoffs among coalition members (Hoel, 1992; Botteon and Carraro, 1997; Hoel and Schneider, 1997; Altamirano-Cabrera et al., 2008).

Behavioural assumptions will usually reflect the underlying motives of agents in a game or may constitute a shortcut to avoid overly complex and intractable game structures. In our paper we extend the conventional cartel formation game to allow for supporters of the coalition. Hence, agents may choose to support (i.e. subsidize) the coalition without becoming a member. This leads to a partition of the set of agents in three subsets: members, supporters and free-riders. Similar to the standard coalitional public goods game, members coordinate their contributions to the public good. Free-riders can benefit from these improvements without facing the costs. Supporters take an intermediate position.

They would offer a subsidy to the coalition in order to stimulate others' participation in the coalition (some alternatives are discussed in Section 6). The full set of behavioural assumptions concerning contributions and supporting transfers is provided in Table 1.

Table 1: Behavioural assumptions of three subsets of agents

| **Subset** | **Decision** | **Determined** | **How** |
|---|---|---|---|
| Members | Contribution | Jointly | Best response |
| Supporters | Supporting transfer | Individually | Conditional transfer |
| Supporters | Contribution | Individually | Best response |
| Free-riders | Contribution | Individually | Best response |

Table 1 shows that all decisions are made individually except for contribution decisions by coalition members. Also, all decisions are best responses except for supporting transfers, which are made conditionally. Let us explain this conditionality. Transfers will indirectly impact the level of contributions to the public good by encouraging participation and thereby increasing contributions to the public good. It is natural to assume that the transfers are conditional on (i) effectiveness and (ii) advantage. Effectiveness requires that supporting transfers are made to increase the size of the coalition by stabilizing a larger coalition than would be possible without support. Advantage requires that there will be no transfer if the amount needed to stabilize the coalition is so large that supporters would be better-off with the largest coalition that is stable without support.

Summarized, we consider a cartel formation game with members, supporters, and free-riders, with behavioural assumptions as outlined in Table 1. In this section, we introduce the model following the intuition by CS93 that supporters only make their support decision *after* the formation of a stable coalition. This implies that the model consists of two parts. Part A consists of two stages and is identical to the standard cartel formation game without support. We focus our attention on stable coalitions before turning to part B, which consists of three stages and introduces the option of supporting transfers.

## 2.1 Part A: Membership

Consider a set $N = \{1, 2, \ldots, n\}$ consisting of $n$ symmetric agents (the case of asymmetric agents is briefly discussed in Section 6). In stage 1, each agent decides to become a coalition member or not. We denote this decision by $\mu_i \in \{1, 0\}$ $\forall i \in N$. The outcome of this stage 1 is a vector $\mu = (\mu_i : i \in N)$, which partitions the set of agents $N$ in two subsets: the coalition $M = \{i : \mu_i = 1\}$ consisting of $m = |M|$ members, and its complement $N \setminus M$ consisting of all $n - m$ singletons.

In stage 2, given $M$, each agent decides how much of the public good to provide. These individual contributions are denoted $q_i$ with $q = (q_1, q_2, \ldots, q_n)$. Welfare maximization implies that members maximize the aggregate welfare to all members of the coalition, while singletons maximize their individual welfare. Welfare is based on the benefits and costs of the provisioning of public goods:

$$w_i(q) = b\left(\sum_{j\in N} q_j\right) - c(q_i) \;\forall i \in N. \tag{1}$$

The benefit function $b()$ is increasing and concave in $\sum_{j\in N} q_j$ while the cost function $c()$ is increasing and convex in $q_i$. Benefits depend on *total* contributions to the public good, while costs depend only on *individual* contributions. Given symmetric agents, contributions and payoffs can be determined using a partition function approach, based on the partition $\{M, N \setminus M\}$. Because all agents are *ex ante* symmetric, $M$ and $N \setminus M$ can be identified by their size, allowing to calculate contributions by members and singletons as follows:

$$\overline{q}(m) \equiv q_i = \arg\max_z \; m \cdot b\left(\sum_{j\neq i} q_j + z\right) - c\left(z\right) \quad \forall i \in M, \tag{2}$$

$$\underline{q}(m) \equiv q_i = \arg\max_z \; b\left(\sum_{j\neq i} q_j + z\right) - c\left(z\right) \qquad \forall i \in N \setminus M. \tag{3}$$

Let $\kappa_i(m)$ denote the welfare of agent $i$ when agent $i$ and $m-1$ other agents are in the coalition. Similarly, let $\lambda_i(m)$ denote his welfare when agent $i$ is not in the coalition formed by $m$ agents. By convexity of $c(q_i)$ we have $\underline{q}(m) \leq \overline{q}(m) \leq m\underline{q}(m)$. This notation allows us to write out $\kappa_i(m)$ and $\lambda_i(m)$:

$$\kappa_i(m) = b\big(m\overline{q}(m) + (n-m)\underline{q}(m)\big) - c\big(\overline{q}(m)\big), \tag{4}$$

$$\lambda_i(m) = b\big(m\overline{q}(m) + (n-m)\underline{q}(m)\big) - c\big(\underline{q}(m)\big). \tag{5}$$

Using this notation for welfare of members and singletons, we now turn to the stability conditions for part A. The standard approach to assess equilibria is to assess *internal and external stability* (D'Aspremont et al., 1983) of the coalition, which are derived from Nash equilibria in membership strategies. This stability concept features two conditions: (i) no member has an incentive to leave the coalition and (ii) no singleton has incentive to join the coalition. A coalition $M$ that satisfies both conditions is called *stable*:

$$\kappa_i\big(m\big) \geq \lambda_i\big(m-1\big), \tag{6a}$$

$$\lambda_i\big(m\big) \geq \kappa_i\big(m+1\big). \tag{6b}$$

## 2.2 Part B: Support

In stage 3, given an initially stable coalition $M$, each singleton $i \in N \setminus M$ decides to become a supporter $s$ or not. We denote this decision by $\sigma_i \in \{1, 0\}$ $\forall i \in N \setminus M$. The outcome of stage 3 is a vector $\sigma = (\sigma_i : i \in N \setminus M)$, which, jointly with $\mu$, partitions the set of agents $N$ in three subsets. We refer to this partition as the *coalition structure* $\{M, F, S\}$, where

$$M = \{i : \mu_i = 1\} \quad \text{is a set of } m = |M| \text{ members,} \tag{7}$$

$$F = \{i : \mu_i = 0 \wedge \sigma_i = 0\} \quad \text{is a set of } f = |F| \text{ free-riders, and} \tag{8}$$

$$S = \{i : \mu_i = 0 \wedge \sigma_i = 1\} \quad \text{is a set of } s = |S| \text{ supporters,} \tag{9}$$

and $n = m + f + s$.

In stage 4, given coalition structure $\{M, F, S\}$, each supporter chooses its supporting transfer $t(s) \geq 0$ to the coalition. The transfers are used to (i) give an (indirect) incentive to a subset $J \subseteq F$ of free-riders, to *join* the coalition (by increasing the coalition welfare), and (ii) stabilize the size $m + j(s)$ coalition, where $j(s)$ is shorthand notation for the composite function $j(t(s))$, denoting cardinality of the set of joiners $J$ as a function of supporting transfers. For now, we avoid forcing any specific structure on $j(s)$ nor on $t(s)$. Rather, the existence of equilibria with supporters depends on whether there exist functions $j(s)$ and $t(s)$ such that the stability conditions, introduced below, hold. In Section 4 (see Lemmas 1 and 2) we will see that this procedure leads to a specific form of $t(s)$ and a weak condition on $j(s)$.

Note that we do not include $J$ as a separate set in the coalition structure $\{M, F, S\}$. The reason for not doing so is that the size of $J$ may change instantaneously in response to a change in $S$. This is further discussed in the description of stability conditions, below.

Stabilization is induced by the aggregate transfer $st(s)$ being distributed among the $m + j(s)$ members/joiners through transfers $\tau_i$ with $\tau = (\tau_1, \tau_2, \ldots, \tau_n)$. A natural feature of a model with *ex ante* symmetric agents[2] is that both the transfers by supporters as well as the distribution of the sum of transfers among members is based on equal sharing (Yi, 1997). That is, $\tau_i = -t(s)$ $\forall i \in S$ and $\tau_i = \frac{st(s)}{m+j(s)}$ $\forall i \in M \cup J$. Clearly, a positive value of $\tau_i$ reflects a positive transfer and vice versa. When $t(s) = 0$, then obviously $j(s) = 0$ and in this case we will interpret supporters as free-riders. This interpretation avoids trivial equilibria with supporters that do not support.

In stage 5, given the joiners in stage 4, each agent updates its initial contribution decision from stage 2. Doing so, members/joiners now maximize their aggregate welfare

[2] Despite agents being symmetric *ex ante*, they may receive different payoffs *ex post* due to their decisions in stages 1 and 4.

while supporters and remaining free-riders still maximize their individual welfare, which extends (1) to:

$$w_i(q,\tau) = b\left(\sum_{j\in N} q_j\right) - c(q_i) + \tau_i \;\forall i \in N. \tag{10}$$

Equation (10) implies that transfers must be self-financed, consistent with CS93.

Having described all stages, we now turn to the stability conditions for part B. These are, again, based on stability conditions, derived from Nash equilibria in membership *and support* strategies. We introduce the stability conditions in their general form, without forcing any specific structure on $j(s)$ or $t(s)$.

Extending the *internal and external stability* conditions (6a) and (6b) to include supporters and joiners, we obtain a system of four stability conditions. Conditions (11a)–(11b)—based on stages 1 and 4 decisions—are the internal and external stability conditions for the set of members/joiners, stating that none has an incentive to become a free-rider and vice versa. When $s = 0$, they simplify to Conditions (6a)–(6b). Conditions (11c)–(11d)—based on stage 3 decisions—are the internal and external stability conditions for the set of supporters, stating that none has an incentive to become a free-rider and vice versa. Note that the conditionality of supporting transfers is reflected in these two stability conditions since changes in the number of supporters may instantaneously affect both $j(s)$ and $t(s)$. While the concepts of internal and external stability reflect an equilibrium state that is immune to *single* deviations, the conditionality of supporting transfers nevertheless requires that a change in transfers can *directly* affect the number of joiners. Note that this model feature is different from allowing for *multiple* deviations. As a result, the RHS of Condition (11c) allows for a change in the number of joiners when one supporter deviates to free-riding. Likewise, the RHS of Condition (11d) allows for a change in the number of joiners as well as a change in the level of transfers when one free-rider deviates to become a supporter.

Any particular coalition structure $\{M, F, S\}$ that satisfies all four conditions is called *stable*:

$$\kappa_i\big(m + j(s)\big) + \frac{st(s)}{m + j(s)} \geq \lambda_i\big(m + j(s) - 1\big), \tag{11a}$$

$$\lambda_i\big(m + j(s)\big) \geq \kappa_i\big(m + j(s) + 1\big) + \frac{st(s)}{m + j(s) + 1}, \tag{11b}$$

$$\lambda_i\big(m + j(s)\big) - t(s) \geq \lambda_i\big(m + j(s - 1)\big), \tag{11c}$$

$$\lambda_i\big(m + j(s)\big) \geq \lambda_i\big(m + j(s + 1)\big) - t(s + 1). \tag{11d}$$

Strictly speaking, our model setup does not allow for members/joiners switching to supporters or vice versa since these decisions are made in different stages of the model. This would eliminate the need to check for such switching in terms of stability. Nevertheless, since we cannot exclude such deviations to occur in practice, we still assess both conditions in the next section:

$$\lambda_i\big(m + j(s)\big) - t(s) \geq \kappa_i\big(m + j(s-1) + 1\big) + \frac{(s-1)t(s-1)}{m + j(s-1) + 1}, \tag{11e}$$

$$\kappa_i\big(m + j(s)\big) + \frac{st(s)}{m + j(s)} \geq \lambda_i\big(m + j(s+1) - 1\big) - t(s+1). \tag{11f}$$

Condition (11e) states that no supporter should have an incentive to switch and become a member, while Condition (11f) states the reverse.

# 3 Comparing stability conditions

Our stability conditions differ from those introduced by CS93, mainly because we relax three of their assumptions. First, CS93 assume commitment, which was discussed already in Section 1.

Second, CS93 assume that $t(s)$ is such that Condition (11a) holds with equality. We do not impose this assumption. We will see in Lemma 1, introduced below, that in equilibrium, $t(s)$ should be such that Conditions (11c) and (11d) hold with equality. This implies that Condition (11a) holds, but not necessarily with equality. That is, the set of members/joiners should receive transfers that exceed the amount needed to keep them from deviating. This larger transfer is needed to make sure that no free-rider has an incentive to become a supporter (and thereby having to contribute to this transfer).

Third, CS93 assume that supporters maximize the number of joiners so that $j(s) = n - m - s$. As a consequence, this assumption eliminates the existence of 'normal' free-riders: agents that do not support and are not in the coalition. We do not impose this restriction but rather endogenize the size of the set of joiners so that $j(s) \leq n - m - s$.

Driven by these assumptions, CS93 consider only two agent types, members/joiners and supporters, for which they establish three stability conditions. We now relate their conditions to ours, using our model setup and notation (we write $m' = m + j(s)$ since CS93 do not model the process of joining):

1. No member should have an incentive to deviate and free-ride, assuming that all others do not deviate: $\kappa_i\big(m'\big) + \frac{st(s)}{m'} \geq \lambda_i\big(m' - 1\big)$. This condition is equivalent to Condition (11a) and we also check the reverse deviation in Condition (11b).

2. No supporter should have an incentive to deviate and free-ride, assuming that all others would deviate too, so that the original non-cooperative coalition emerges: $\lambda_i(m') - t(s) \geq \lambda_i(m)$. This condition is implied by the stricter Condition (11c) (which drops the commitment assumption) and we also check the reverse deviation in Condition (11d).

3. No supporter should have an incentive to deviate and join the coalition as a member (supported by the remaining supporters), assuming that all others do not deviate: $\lambda_i(m') - t(s) \geq \kappa_i(m'+1) + \frac{(s-1)t(s-1)}{m'+1}$. This condition is implied (in their model setup) by the former condition. In our model it is equivalent to Condition (11e) for the case where $j(s-1) = j(s)$ and we also check the reverse deviation in Condition (11f).

# 4 Existence of equilibria with supporters

Our main result requires one restriction on the incentive of members/joiners to deviate, as a function of the number of members/joiners. Following Hoel and Schneider (1997), we use the following function to reflect (internal) stability:

$$\Phi_i(m + j(s)) = \kappa_i(m + j(s)) - \lambda_i(m + j(s) - 1). \tag{13}$$

We make the following assumption.

**Assumption 1.** $\Phi_i(m + j(s))$ is a decreasing function of $m + j(s)$. That is, an increasing number of members/joiners makes the coalition structure less stable by monotonically increasing the incentive to deviate.

By (internal and external stability) Conditions (6a)–(6b), we have $\Phi_i(m + 0) \geq 0$ and $\Phi_i(m + 1) \leq 0$. As a result, Assumption 1 implies that $\Phi_i(m + j(s))$ is negative for all $j(s) \geq 1$. Hence, by Condition (11a), transfers (from supporters) are necessary to stabilize any equilibrium with a positive number of joiners. Assumption 1 is not very restrictive (cf. Hoel and Schneider, 1997; Finus and Maus, 2008) and is a sufficient (but not a necessary) condition for the existence result in Proposition 1.

Before presenting our main result, we first present the following two lemmas that provide structure to $j(s)$ and $t(s)$, based on the system of stability conditions (11). We will use both lemmas in the proof of Proposition 1 as well as in Section 5.

**Lemma 1.** *Stability of coalition structure $\{M, F, S\}$ implies that $t(s) = \lambda_i(m + j(s)) - \lambda_i(m + j(s-1))$.*

*Proof.* Any stable coalition structure $\{M, F, S\}$ satisfies Conditions (11a)–(11d). Substitute $t(s) = \lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s-1)\big)$ into Conditions (11c)–(11d) to obtain:

$$\lambda_i\big(m + j(s)\big) - \big[\lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s-1)\big)\big] \geq \lambda_i\big(m + j(s-1)\big), \tag{14}$$

$$\lambda_i\big(m + j(s)\big) \geq \lambda_i\big(m + j(s+1)\big) - \big[\lambda_i\big(m + j(s+1)\big) - \lambda_i\big(m + j(s)\big)\big]. \tag{15}$$

Both inequalities hold with equality, which implies that any other level of $t(s)$ would violate one of them. Therefore, stability of $\{M, F, S\}$ implies that $t(s) = \lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s-1)\big)$. □

Lemma 1 states that, in equilibrium, the supporting transfers are such that supporters have no incentive to deviate and terminate their support (and vice versa). Interestingly, when a coalition structure is not stable—or, equivalently, $s$ is larger than necessary to stabilize $\{M, F, S\}$—Lemma 1 implies that $t(s) = 0$. To see why, note that for any unstable coalition structure, we have $j(s) = j(s-1)$.

A consequence of Lemma 1 is that, in searching for stable coalition structures, we can restrict ourselves to a specific functional form for $t(s)$.

**Lemma 2.** *For any stable coalition structure $\{M, F, S\}$ we have $j(s) > j(s-1)$.*

*Proof.* Since we ignore trivial equilibria with supporters that do not support, we have $t(s) > 0$. Because of this strictly positive support, and because the function $\lambda()$ in (5) is increasing in $m + j(s)$, Condition (11c)—or, alternatively, Lemma 1—dictates that $j(s) > j(s-1)$. □

Lemma 2 states that, in equilibrium, when the number of supporters decreases by one, the number of joiners decreases (strictly). Note that the related inequality $j(s+1) > j(s)$, does not necessarily hold, which we will see confirmed by an example illustrated in Figure 1.

In order to avoid confusion in the interpretation of Lemmas 1 and 2, recall that our stability conditions are based on *single* deviations, derived from Nash equilibria in membership and support strategies. These equilibria take strategies of other agents as given, so that the coalition structure that results after a deviation (e.g. one with $j(s-1)$ joiners) may not be stable itself. Stability of the 'resulting' coalition structure is irrelevant for the stability analysis of the 'current' coalition structure.[3]

We now present our main result.

**Proposition 1.** *Under Assumption 1, there exist stable coalition structures $\{M, F, S\}$ with $s > 0$.*

[3]Alternative equilibrium concepts may be employed, such as farsightedness (cf. Diamantoudi and Sartzetakis, 2015).

*Proof.* Consider coalition structure $\{M, F, S\}$. Using Lemma 1, take the supporting transfer such that supporters have no incentive to deviate and terminate their support:

$$t(s) = \lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s-1)\big). \tag{16}$$

By the proof of Lemma 1, (16) implies that Conditions (11c) and (11d) hold with equality.

We proceed to check the final two conditions of system (11). Substitute (16) in Conditions (11a)–(11b) and rearrange the resulting inequalities, using chained notation:

$$\begin{aligned} \big(m + j(s)\big)\cdot\Big(\lambda_i\big(m + j(s) - 1\big) - \kappa_i\big(m + j(s)\big)\Big) \\ \leq s\cdot\Big(\lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s-1)\big)\Big) \\ \leq \big(m + j(s) + 1\big)\cdot\Big(\lambda_i\big(m + j(s)\big) - \kappa_i\big(m + j(s) + 1\big)\Big). \end{aligned} \tag{17}$$

That is, the aggregate incentive to deviate to the $m + j(s)$ members/joiners is (weakly) smaller than the aggregate supporting transfer and this transfer is (weakly) smaller than the aggregate incentive to deviate to $m + j(s) + 1$ members/joiners.

To verify for which $s$, if any, the system of inequalities (17) holds, we first check whether the inequality holds between the first and last term. Using (13), and flipping signs, we write the inequality between the first and last term of (17) as

$$\big(m + j(s)\big)\cdot\Big(\Phi_i\big(m + j(s)\big)\Big) \geq \big(m + j(s) + 1\big)\cdot\Big(\Phi_i\big(m + j(s) + 1\big)\Big). \tag{18}$$

Recall that we focused on stable coalitions when we introduced the possibility of supporting transfers (see stage 3). Hence, using (13), the external stability condition (6b) implies that $\Phi_i(m+1) \leq 0$. Combining this inequality with Assumption 1 we have $0 \geq \Phi_i\big(m + j(s)\big) \geq \Phi_i\big(m + j(s) + 1\big)$, which implies that (18) holds with strict inequality.

Now, combining Lemmas 1 and 2—or, alternatively, since we ignore trivial equilibria with supporters that do not support—we have $t(s) = \lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s-1)\big) > 0$, and we know by transitivity that there exists a non-empty interval $\left[\underline{s}, \bar{s}\right] \ni s$ such that (17) holds.

Next, we prove by example that there exists some combination of $b\left(\sum_{j\in N} q_j\right)$ and $c(q_i)$ for which the interval $\left[\underline{s}, \bar{s}\right] \ni s$ contains at least one natural number. This example is provided in Section 5. Obviously, when the interval $\left[\underline{s}, \bar{s}\right]$ contains more than one natural number, $s$ is the smallest of these, since otherwise $j(s) = j(s-1)$, which contradicts Lemma 2. □

Note that Proposition 1 does not claim that $s$ is a natural number. This is clear from the context, though, and illustrated by an example in the next section.

As a corollary to Proposition 1, we extend our result to include Conditions (11e)–(11f). We will see that this extension requires one additional assumption.

**Assumption 2.** The function $\lambda()$ in (5) is convex in $m + j(s)$.

**Corollary 1.** *Under Assumptions 1–2, there exist stable coalition structures* $\{M, F, S\}$ *with* $s > 0$ *that satisfy Conditions* (11e)–(11f).

*Proof.* Building on the proof to Proposition 1, we proceed to check whether Conditions (11e) and (11f) hold. We start with Condition (11e) in which we substitute (16) to obtain:

$$\lambda_i\big(m + j(s-1)\big) \geq \kappa_i\big(m + j(s-1) + 1\big) + \frac{(s-1)t(s-1)}{m + j(s-1) + 1}. \tag{19}$$

Inequality (19) is similar to Condition (11b) but with one supporter less and therefore, by Lemma 2, fewer joiners.

There are two cases to consider. First, if the coalition structure $\{M, F', S'\}$, with $f' = f + 1$ and $s' = s - 1$, is stable, then (19) holds since it is the equivalent of Condition (11b) for this coalition structure.

Second, if the coalition structure $\{M, F', S'\}$ is not stable, then we know that $t(s') = 0$ (see the discussion of Lemma 1), so that the last RHS term cancels out. We are left to check whether $\lambda_i\big(m + j(s-1)\big) \geq \kappa_i\big(m + j(s-1) + 1\big)$, which, by (13), can be written as

$$\Phi(m + j(s-1) + 1) \leq 0. \tag{20}$$

Using (13), the external stability condition (6b) implies that $\Phi_i(m+1) \leq 0$ and, by Assumption 1, $\Phi_i()$ decreases further for larger numbers of joiners. Since $m + j(s-1) + 1 \geq m + 1$, we know that inequality (20), and therefore inequality (19), holds.

Because inequality (19) holds in both cases, we know that Condition (11e) is satisfied.

We now turn to Condition (11f) in which we substitute (16) to obtain:

$$\kappa_i\big(m + j(s)\big) + \frac{st(s)}{m + j(s)} \geq \lambda_i\big(m + j(s+1) - 1\big) - \lambda_i\big(m + j(s+1)\big) + \lambda_i\big(m + j(s)\big). \tag{21}$$

This inequality is implied by Condition (11a) as long as

$$\lambda_i\big(m + j(s+1)\big) - \lambda_i\big(m + j(s+1) - 1\big) \geq \lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s) - 1\big). \tag{22}$$

For any $j(s+1) \geq j(s)$, this inequality holds if $\lambda()$ is convex in $m + j(s)$, and we use Assumption 2. $\square$

Assumption 2 is restrictive. It puts a constraint on the functional forms and parameter values of the cost- and benefit functions for which the corollary holds. Nevertheless, our example in the next section with linear benefits and quadratic costs satisfies Assumption 2.

# 5 Example

In this section we analyze an example of a cartel formation game that satisfies Assumptions 1 and 2. This example both finishes the proof of our existence result in Proposition 1 and it illustrates the possibility of multiple equilibria. Our example uses a model specification with a linear benefit function and a quadratic cost function. We update (1) to:

$$w_i = \beta \left(\textstyle\sum_{j\in N} q_j\right) - \frac{1}{2}\gamma (q_i)^2 \;\; \forall i \in N, \tag{23}$$

where $\beta$ and $\gamma$ are model parameters. A well-known result is that, without the option of support, the unique stable coalition size for this model specification equals $m = 3$ whenever $n \geq 3$ (Barrett, 1994, Proposition 2).

Next, consider the possibility of supporting transfers. The model specification in (23) dictates that each agent has a dominant strategy in terms of contributions. Using (2) and (3) we have:

$$\overline{q}\big(m + j(s)\big) = \frac{\big(m + j(s)\big)\beta}{\gamma}, \tag{24}$$

$$\underline{q}\big(m + j(s)\big) = \frac{\beta}{\gamma}. \tag{25}$$

Using these contribution decisions as well as (4) and (5), we can write $\kappa_i\big(m + j(s)\big)$ and $\lambda_i\big(m + j(s)\big)$ as follows:

$$\kappa_i\big(m + j(s)\big) = \frac{\beta^2}{\gamma}\left(\frac{1}{2}\big(m + j(s)\big)^2 + \big(n - m - j(s)\big)\right), \tag{26}$$

$$\lambda_i\big(m + j(s)\big) = \frac{\beta^2}{\gamma}\left(\big(m + j(s)\big)^2 + \big(n - m - j(s)\big) - \frac{1}{2}\right). \tag{27}$$

Consider a coalition with $m = 3$ members. Take parameter values $\beta = \gamma = 1$ and $n = 50$ and consider $j(s) = 1$. Using (16) and Lemma 2 we have $t(s) = \lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s-$

$1)) = \lambda_i(3+1) - \lambda_i(3+0) = 61\frac{1}{2} - 55\frac{1}{2} = 6$. By the proof of Proposition 1, this transfer implies that stability conditions (11d) and (11c) hold with equality.

Next, we verify whether there exists a positive number of supporters $s$ such that stability conditions (11a) and (11b) hold. We do so by substituting (26) and (27) into (17), using our selected parameter values. Doing so, we obtain:

$$(3+1)\cdot\left(55\frac{1}{2}-54\right) \leq s\cdot\left(61\frac{1}{2}-55\frac{1}{2}\right) \leq (3+1+1)\cdot\left(61\frac{1}{2}-57\frac{1}{2}\right). \tag{28}$$

This chained inequality holds for the interval $s \in \left[\underline{s}=1, \bar{s}=\frac{20}{6}\right]$, containing the natural numbers 1, 2, and 3. The number of supporters in equilibrium is the smallest of these (see the last sentence of the proof to Proposition 1), so that $s = 1$. That is, a coalition structure with $3+1=4$ members/joiners is stable when there is one supporter.

Next, we illustrate stable coalition structures with $j(s) > 1$. To avoid having to check many uninteresting cases, we restrict ourselves to functions $j(s)$ such that, for any stable coalition structure $\{M, F, S\}$, we have $j(s-1) = j(s) - 1$. That is, if one supporter deviates, the remaining supporters are not able to support the original number of joiners but, rather, choose their transfers based on one joiner less. This restriction gives structure to Lemma 2 and thereby provides a one-to-one relationship between the number of supporters $s$, the number of joiners $j(s)$, and the related transfers $t(s)$, which facilitates our calculations in this section.

We can now substitute our parameter values and write out (16):

$$\begin{aligned} t(s) &= \lambda_i(3+j(s)) - \lambda_i(3+j(s-1)) \\ &= (3+j(s))^2 + (50-3-j(s)) - (3+(j(s)-1))^2 - (50-3-(j(s)-1)) \\ &= (3+j(s))^2 - (3+j(s)-1)^2 - 1. \end{aligned} \tag{29}$$

Following the same procedure as above, we calculate the equilibrium number of $s$ as a function of the number of joiners $j(s) \geq 1$ on top of the $m = 3$ coalition members.

Results of this procedure are shown in Figure 1, which displays the equilibrium number of joiners as a function of the number of supporters on top of the 3 coalition members. The horizontal bars depict the interval $[\underline{s}, \bar{s}]$ for each value of $j(s)$. Equilibrium $s$ is the smallest natural number in each of these intervals (see the last sentence of the proof to Proposition 1). Clearly, there are multiple equilibria. Figure 1 shows 11 distinct stable coalition structures that differ in the number of supporters and joiners. The equilibrium number of supporters and joiners is constrained by two factors. One factor is that, given the functional forms and parameterization used in this section, equilibrium $s$ is increasing

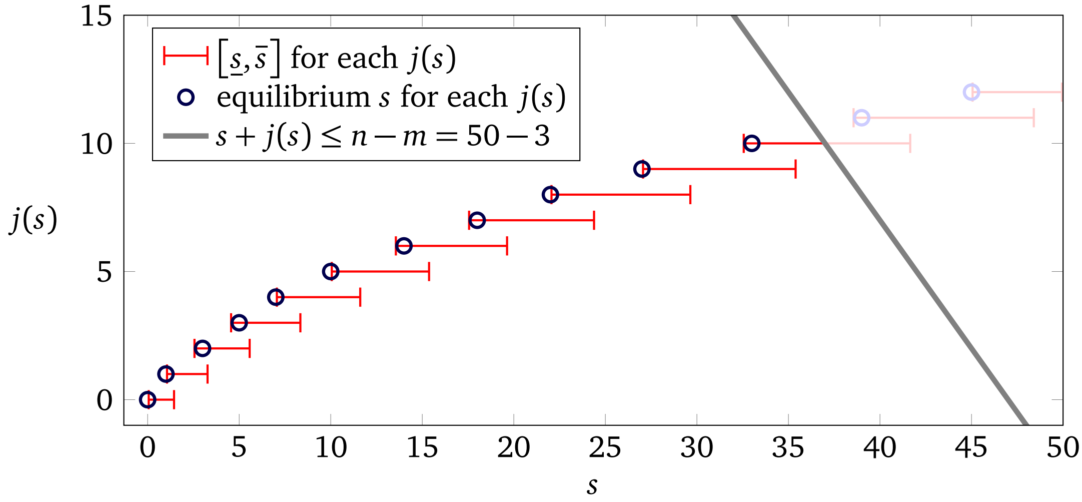


Figure 1: Equilibrium number of $j(s)$ joiners as a function of the number of $s$ supporters on top of the $m = 3$ coalition members, given the parameter values introduced in Section 5. Bars depict the interval $[\underline{s}, \bar{s}]$ for each value of $j(s)$ and the decreasing line indicates the constraint on the aggregate number of supporters and joiners, given by the total number of $n = 50$ agents minus $m = 3$ members.

and convex in the number of joiners $j(s)$ so that increasingly larger numbers of supporters are required to achieve one additional joiner. The other factor is the constraint provided by the total number of agents. Since the aggregate of joiners and supporters cannot exceed $50 - 3 = 47$, the maximum size of $m + j(s)$ is achieved with 33 supporters which allows 10 joiners.

By (10) and using the parameter values from our example we can now calculate payoffs of members/joiners, supporters, and free-riders for each equilibrium.

$$\begin{aligned} w_i &= \kappa_i\big(m + j(s)\big) + \frac{st(s)}{m + j(s)} \\ &= \frac{1}{2}\big(3 + j(s)\big)^2 + \big(47 - j(s)\big) + s\left(\frac{4 + 2j(s)}{3 + j(s)}\right) \qquad \forall i \in M \cup J, \qquad (30) \\ w_i &= \lambda_i\big(m + j(s)\big) - t(s) \\ &= \big(2 + j(s)\big)^2 + \big(47 - j(s)\big) + \frac{1}{2} \qquad \forall i \in S, \qquad (31) \\ w_i &= \lambda_i\big(m + j(s)\big) \\ &= \big(3 + j(s)\big)^2 + \big(47 - j(s)\big) - \frac{1}{2} \qquad \forall i \in F \setminus J. \qquad (32) \end{aligned}$$

In Figure 2 we plot these payoffs for each of the stable equilibria identified in Figure 1. Note that welfare of members/joiners and supporters are nearly identical, deviations only

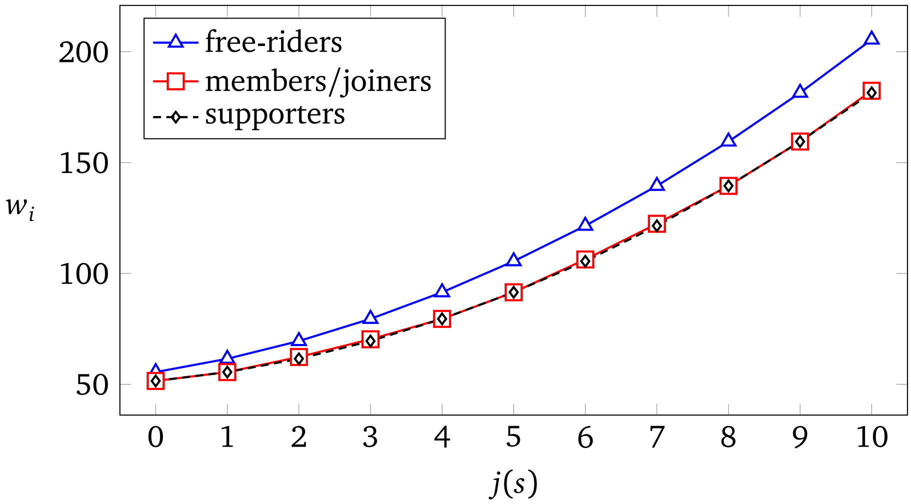


Figure 2: Individual welfare based on (10) for each of three types of agents—free-riders, members/joiners, and supporters—for each of the equilibria as identified by $j(s) \in (0, 1, \ldots, 10)$.

caused by differences between equilibrium $s$ and $\underline{s}$. Payoffs increase monotonically in $j(s)$, which demonstrates that by allowing for supporters, the cartel formation game has turned into a coordination game in which equilibria with higher numbers of supporters payoff-dominate those with fewer (cf. Barrett, 2003). Consequentially, our example illustrates that allowing for support would significantly improve the prospects for wider cooperation in cartel formation games.

Our example hosts not only multiple equilibria but—constrained by the aggregate number of supporters and joiners—it also includes each possible equilibrium number of joiners, i.e. the full range $j(s) \in (0, 1, \ldots, 10)$. This result is sensitive to our selected parameters. To illustrate this sensitivity, we check stability for the more general cost function $c(q_i) = \frac{1}{2}\gamma\,(q_i)^{\alpha}$, where the cost parameter $\alpha \geq 1$ determines convexity. We find that, compared with $\alpha = 2$, decreasing cost convexity destabilizes coalition structures with low numbers of joiners, by violating Condition (11a), Condition (11b), or both. Similarly, increasing cost convexity destabilizes coalition structures with high numbers of joiners. In our example, for $\alpha \notin [1.53, 3.61]$, no stable coalition structure with $s > 0$ exists.

# 6 Discussion and conclusion

Inspired by CS93, we augment the standard cartel formation game from non-cooperative coalition theory with the possibility of support. We show that, under mild conditions on the costs and benefits of providing the public good, supporters exist in equilibrium. In a standard IEA specification with linear benefits and quadratic costs, we demonstrate that multiple stable coalition structures exist. This multiplicity of equilibria turns the cartel formation game into a coordination game with substantial potential for additional gains by the possibility of support. Our paper therefore highlights an option for IEA design that allows countries to participate in differentiated ways, and with potential to substantially increase (i) cooperation, (ii) abatement of $CO_2$ emissions, and (iii) associated payoffs to all. This option is particularly promising for IEAs for which free-riding incentives are typically large, as is the case for climate change.

Three features of our model need some discussion here, since they could be changed or generalized in future work: (i) symmetry, (ii) the five-stage model setup, and (iii) supporters' behaviour. We do not expect that such changes or generalization would impact our main result.

First, we assume symmetric agents, which allows an analytical solution to our existence result. For *internal commitment* (i.e. some or all members, and possibly joiners too, will not deviate *ex post*), the more countries are committed, the larger are the gains from expanding the coalition (CS93). The symmetry assumption was dropped first by Botteon and Carraro (1997) who showed in a simulation that, with asymmetric agents, only little or even no *internal* commitment is necessary to reach the grand coalition (see also Petrakis and Xepapadeas, 1996; Jeppesen and Andersen, 1998). Clearly, asymmetry expands the scope for cooperation and increases the scope for larger coalitions through transfers without a need for commitment (Barrett, 2001; McGinty, 2007; Weikard, 2009; Fuentes-Albero and Rubio, 2010). In our current paper, we find, for the case of symmetric agents, that wide cooperation can be achieved *without commitment*. Hence, transferring the results on internal commitment to the current setting, we expect this will be even easier with asymmetric agents.

Second, our model has five stages, following the intuition by CS93 that supporters only make their support decision *after* the formation of a stable coalition. When this assumption is dropped, the model could potentially be simplified to a more conventional two-stage model. However, one quickly runs into problems since it is not straightforward how the number of joiners can be determined endogenously (a crucial step in our analysis) in a two-stage model.

Third, we assume that supporters offer a subsidy to the coalition in order to stimulate others' participation in the coalition. One could consider plausible alternatives, of which we mention two. One of these is that supporters target their transfer to a specific agent in order to convince him to join the coalition (i.e. the support would not be shared with the coalition). This alternative, however, would only make sense in the context of asymmetric countries, a setting that we already discussed. Another alternative is that supporters choose their support strategically. A likely strategy is choosing support to coordinate on an equilibrium that maximizes the supporters' payoff. This alternative differs from the current approach in which we only assess whether a coalition structure can be stabilized *given* some number of supporters, members, and joiners. One could envisage that this type of strategic behaviour would mimic the kind of coordination that is discussed in the context of multiple equilibria occurring in the example of Section 5.

Finally, and to close this section, recall two central observations from the paper. First, as we stated in Section 1, by including support, countries may participate in an IEA in differentiated ways, which is likely to broaden participation as well as the scope for agreement negotiations. Second, as we stated in Section 5, by including support, the cartel formation game may turn into a coordination game rather than a prisoners' dilemma-type problem. This transformation of the game is known to allow for wider and deeper cooperation (Barrett, 1998). Jointly, these observations make us optimistic about the scope of support as a possible policy tool in the design of IEAs on climate change.

## References


Aldy, J. E. and R. N. Stavins (2009). *Post-Kyoto International Climate Policy: Implementing Architectures for Agreement*. Cambridge: Cambridge University Press.

Altamirano-Cabrera, J.-C., M. Finus, and R. Dellink (2008). Do abatement quotas lead to more successful climate coalitions? *The Manchester School 76*(1), 104–129.

Ansink, E. and J. Bouma (2013). Effective support for community resource management. *Forest Policy and Economics 1*(37), 94–103.

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers 46*(S), 878–894.

Barrett, S. (1998). On the theory and diplomacy of environmental treaty making. *Environmental and Resource Economics 11*(3-4), 317–333.

Barrett, S. (2001). International cooperation for sale. *European Economic Review 45*(10), 1835–1850.

Barrett, S. (2003). *Environment and Statecraft*. Oxford: Oxford University Press.

Bauer, T. K., J. Bredtmann, and C. M. Schmidt (2013). Time vs. money: The supply of voluntary labor and charitable donations across Europe. *European Journal of Political Economy 32*(1), 80–94.

Bobrow, D. B. and M. A. Boyer (1997). Maintaining system stability contributions to peacekeeping operations. *Journal of Conflict Resolution 41*(6), 723–748.

Böhringer, C. and C. Vogt (2004). The dismantling of a breakthrough: The Kyoto Protocol as symbolic policy. *European Journal of Political Economy 20*(3), 597–617.

Botteon, M. and C. Carraro (1997). Burden-sharing and coalition stability in environmental negotiations with asymmetric countries. In C. Carraro (Ed.), *International Environmental Negotiations: Strategic Policy Issues*. Cheltenham: Edward Elgar.

Bove, V. and L. Elia (2011). Supplying peace: Participation in and troop contribution to peacekeeping missions. *Journal of Peace Research 48*(6), 699–714.

Carraro, C. and D. Siniscalco (1993). Strategies for the international protection of the environment. *Journal of Public Economics 52*(3), 309–328.

D'Aspremont, C., A. Jacquemin, J. Gabszewicz, and J. Weymark (1983). On the stability of collusive price leadership. *Canadian Journal of Economics 16*(1), 17–25.

Diamantoudi, E. and E. S. Sartzetakis (2015). International environmental agreements: Coordinated action under foresight. Forthcoming in *Economic Theory*.

Feldman, N. E. (2010). Time is money: Choosing between charitable activities. *American Economic Journal: Economic Policy 2*(1), 103–130.

Finus, M. (2003). Stability and design of international environmental agreements: The case of transboundary pollution. In H. Folmer and T. Tietenberg (Eds.), *The International Yearbook of Environmental and Resource Economics 2003/2004*, pp. 82–158. Cheltenham: Edward Elgar.

Finus, M. and S. Maus (2008). Modesty may pay! *Journal of Public Economic Theory 10*(5), 801–826.

Finus, M. and S. Tjøtta (2003). The Oslo Protocol on sulfur reduction: The great leap forward? *Journal of Public Economics 87*(9), 2031–2048.

Freeman, R. (1997). Working for nothing: The supply of volunteer labor. *Journal of Labor Economics 15*(1), S140–S166.

Fuentes-Albero, C. and S. Rubio (2010). Can international environmental cooperation be bought? *European Journal of Operational Research 202*(1), 255–264.

Gaibulloev, K., T. Sandler, and H. Shimizu (2009). Demands for UN and non-UN peacekeeping: Nonvoluntary versus voluntary contributions to a public good. *Journal of Conflict Resolution 53*(6), 827–852.

Hoel, M. (1992). International environment conventions: The case of uniform reductions of emissions. *Environmental and Resource Economics 2*(2), 141–159.

Hoel, M. and K. Schneider (1997). Incentives to participate in an international environmental agreement. *Environmental and Resource Economics 9*(2), 154–170.

Jeppesen, T. and P. Andersen (1998). Commitment and fairness in environmental games. In N. Hanley and H. Folmer (Eds.), *Game Theory and the Environment*. Cheltenham: Edward Elgar.

Khanna, J., T. Sandler, and H. Shimizu (1998). Sharing the financial burden for UN and NATO peacekeeping, 1976-1996. *Journal of Conflict Resolution 42*(2), 176–195.

McGinty, M. (2007). International environmental agreements among asymmetric nations. *Oxford Economic Papers 59*(1), 45–62.

Olmstead, S. M. and R. N. Stavins (2012). Three key elements of a post-2012 international climate policy architecture. *Review of Environmental Economics and Policy 6*(1), 65–85.

Petrakis, E. and A. Xepapadeas (1996). Environmental consciousness and moral hazard in international agreements to protect the environment. *Journal of Public Economics 60*(1), 95–110.

Shimizu, H. and T. Sandler (2002). Peacekeeping and burden-sharing, 1994-2000. *Journal of Peace Research 39*(6), 651–668.

Weikard, H.-P. (2009). Cartel stability under an optimal sharing rule. *The Manchester School 77*(5), 575–593.

Yi, S.-S. (1997). Stable coalition structures with externalities. *Games and Economic Behavior 20*(2), 201–237.