CESifo Working Papers

Members, Joiners, Free-Riders, Supporters

Erik Ansink
Cees Withagen

CESIFO WORKING PAPER NO. 5802
CATEGORY 10: ENERGY AND CLIMATE ECONOMICS
MARCH 2016

ISSN 2364-1428

CESifo *Working Paper No. 5802*

Members, Joiners, Free-Riders, Supporters

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 four types of agents: members, joiners, free-riders, and supporters.

JEL-Codes: C720, D020, H410, Q540.

Keywords: coalition formation, public goods, support, transfers, international environmental agreements.

*Erik Ansink**
Department of Economic and Social History / Utrecht University
Drift 6
The Netherlands – 3512 BS Utrecht
j.h.ansink@uu.nl

Cees Withagen
Department of Spatial Economics
VU University Amsterdam
The Netherlands - 1081 HV Amsterdam
cwithagen@feweb.vu.nl

*corresponding author

February 29, 2016
We thank Stefan Ambec, Lorenzo Cerda Planas, Gerard van der Meijden, and Hans-Peter Weikard, as well as seminar participants at CESifo, EAERE 2015, and FAERE 2015, for comments and discussion. We acknowledge financial support from FP7-IDEAS-ERC Grant No. 269788 and 339647.

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 increases payoffs to each of four types of agents: members, joiners, 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 behavior, 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) contribute their time to provide the charitable good, while a set of donors supports the charity financially.[1] The common denominator in both examples as well as in supported climate coalitions is that agents can participate in differentiated ways.

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 we take a different perspective in three respects. First, CS93 fix the number of joiners at the maximum level, implying that in equilibrium there are no free-riders; agents are either member of the original coalition, joiners to the new coalition, or supporters. In the current paper, however, we endogenize the number of joiners, allowing for the existence of free-riders in equilibrium. Second, we impose more plausible equilibrium conditions. For instance, while we check whether or not free-riders have an incentive to join or support the coalition, for the reason given above, CS93 do not check this. Moreover, CS93 claim that in their equilibrium, no supporter has an incentive to join the coalition but the condition they impose is not sufficient. We elaborate on this difference in Section 5. A third issue relates to *commitment*. 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

[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).

do not escape from making commitment assumptions. As is usual in non-cooperative game-theoretical models on IEAs it is assumed that coalition members in equilibrium choose the optimal abatement level from the perspective of the coalition. In addition, we also assume that supporters commit to certain supporting transfers, but these are different from CS93. In CS93 the transfers to which supporters commit make the joiners indifferent between joining the coalition and remaining a free rider, ignoring the incentive for supporters to actually provide such transfers. In the current paper the transfers to which supporters commit make supporters indifferent between being a supporter and becoming a free-rider, thereby explicitly accounting for the supporters' incentives. We further discuss commitment in Remark 1.

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 neglected in the literature (as opposed to the other types of commitment analyzed by CS93).[2] 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, a central tenet in climate change negotiations. Differentiated participation is usually connected with broadening participation as well as the scope for agreement negotiations (Olmstead and Stavins, 2012), a premise that we will see confirmed in this paper.

Our positive result sheds light on one alternative mechanism to increase cooperation in IEAs: 'climate clubs'. One distinctive feature of a club is that non-members can be excluded or penalized (Sandler and Tschirhart, 1997). In the case of climate clubs, this could be implemented by linking the IEA to a club good agreement, for instance a trade agreement that imposes higher tariffs to non-members. In a recent proposal for designing a climate club, Nordhaus (2015) finds that such sanctions are necessary to stabilize large coalitions with substantial abatement. In our model, supporters can be interpreted as those countries that pay the sanctions, while the members of the coalition enjoy the sharing of the supporting transfers as their club good. Self-enforcement implies that such sanctions are beneficial to supporters. Hence, an IEA with support can be interpreted as a voluntary implementation of a climate club, whilst avoiding the complexities that come with linking an IEA to a club good agreement (Finus, 2003).

[2] One exception is Carraro et al. (2006) who analyze the role of various types of transfers in stabilizing IEAs using simulations. One of the transfers they consider is a direct transfer from a free-rider to another free-rider in order to 'bribe' him to join the coalition. This transfer is comparable to the supporting transfers of our model, but we analyze such transfers analytically and in a richer setting.

In the next section we introduce the model. In Section 3 we present our main result on the existence of supporters in equilibrium. In Section 4 we illustrate this result using a workhorse IEA model specification due to Barrett (1994). In Section 5 we compare our results to those by CS93. 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 behavioral 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).

Behavioral 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 behavioral assumptions concerning contributions and supporting transfers is provided in Table 1.

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

Table 1: Behavioral 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

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 behavioral assumptions as outlined in Table 1. We proceed by introducing the model following the intuition by CS93 that supporters only make their support decision *after* the formation of a stable coalition. This implies that our 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. Our model therefore consists of five stages:

(i) All agents decide whether or not to become a coalition member;
(ii) All agents choose their initial contributions to the public good;
(iii) Singletons decide whether or not to become a supporter;
(iv) Supporters choose the level of supporting transfers, inducing a subset of free-riders to become joiners;
(v) All agents update their initial contributions to the public good.

We consider this five-stage model the proper model setup for three reasons. First, this setup is consistent with situations where a coalition is already formed and the option of support is introduced to expand cooperation. One could think of an amendment to an existing IEA, or some similar change in an agreement's institutional setting, that would allow such supporting transfers. Second, it is consistent with the setup by CS93 and therefore allows to compare results as we do in Section 5. Third, this setup allows for instantaneous switching of joiners in response to the supporting transfers offered to the coalition (this model feature is introduced in Section 2.2, below). It is not straightforward how this endogenous determination of the number of joiners can be incorporated in a more conventional two-stage model.

2.1 Part A: Membership

Consider a set $N = \{1, 2, \dots, n\}$ 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 $q = (q_1, q_2, \dots, q_n)$. 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:

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

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:

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

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

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*:

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

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 μ, partitions the set of agents N in three subsets. We refer to this partition as the *coalition structure* $\{M, F, S\}$, where

$$M = \{i : \mu_i = 1\} \quad \text{is a set of } m = |M| \text{ members,} \tag{7}$$
$$F = \{i : \mu_i = 0 \land \sigma_i = 0\} \quad \text{is a set of } f = |F| \text{ free-riders, and} \tag{8}$$
$$S = \{i : \mu_i = 0 \land \sigma_i = 1\} \quad \text{is a set of } s = |S| \text{ supporters,} \tag{9}$$

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's welfare), and (ii) stabilize the size $m + j(s)$ coalition, where $j(s)$ is shorthand notation for the composite function $j(t(s))$. This function denotes the 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 *reasonable expectations* with respect to both $j(s)$ and $t(s)$ such that the stability conditions, introduced below, hold. We expand on these expectations and their implications for $t(s)$ and $j(s)$ below and in Section 3. We will also show that such reasonable expectations are absent from CS93 in Section 5.

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 τ_i with $\tau = (\tau_1, \tau_2, \ldots, \tau_n)$. A natural feature of a model with *ex ante* symmetric agents[3] 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 τ_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 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.

Remark 1. The sequence of actions in stages 4 and 5 implies that our model gives the opportunity to joiners to take the money and run. In response to the supporters' aggregate transfer $st(s)$, a free-rider may join the coalition and receive his transfer τ_i in stage 4, while subsequently refusing to update its initial contribution decision to what is required from the coalition's perspective in stage 5. The timing of stages allows for this type of defection and it appears that, as a remedy, we need to require (a weak form of) commitment on behalf of the joiners (as well as the members who are in a similar position). Such a commitment assumption, however, happens to be standard, although implicit, in the non-cooperative game-theoretical literature on IEAs. As observed by Finus (2003) such models: *"... assume some form of commitment within coalitions*. That is, they *"... assume that countries comply with their emission reduction and transfer obligations if they form a coalition and therefore free-rider problems of real IEAs are underestimated"*. The fact that this assumption is tolerated in the literature may be due to its similarity to market exchanges in general, in which commitment is not assumed. Commitment can be avoided by making support conditional on the joiners indeed taking appropriate action and vice versa. Related to climate change one could imagine a fund for incoming support and outgoing transfers to members and joiners. This fund would be managed by a third party, conditional on meeting the emission abatement targets. For the purpose of this paper, such an approach appears to be overly tedious.

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

2.3 Equilibrium stability conditions

Having described all stages, we now turn to the equilibrium stability conditions for our model including part B. An equilibrium is defined by the coalition structure $\{M, F, S\}$ as well as the number of joiners $j(s)$ and transfers $t(s)$ by each supporter. We extend the internal and external stability conditions, introduced at the end of Section 2.1 to the current setting that includes supporters and joiners. Doing so, we need to formulate, for each agent, his expectations regarding what happens if he deviates. We will discuss these expectations for each condition that we impose on an equilibrium.

Most importantly, these expectations need to be reasonable. By employing the concepts of internal and external stability (i.e. Nash equilibria in membership strategies, cf. D'Aspremont et al., 1983) we can immediately rule out multiple deviations as unreasonable. Hence, we consider single deviations only, but that is not the end of the story. Due to the conditionality of supporting transfers (see Table 1) and as stressed in Section 2.2, the number of joiners may change *instantaneously* in response to a change in s – or, more accurately, a change in $t(s)$ – caused by a single deviation. As a result, we have to account for changes in the number of joiners (and supporting transfers) upon such single deviations. We focus our formulation of reasonable expectations on $j(s)$ and $t(s)$. For example, it could be assumed that deviations by supporters do not affect the number of joiners. Under this assumption, we have the expectation $j(s) = j$, and the internal stability condition for a supporter (introduced below in (11c)) boils down to $\lambda(m + j) - t(s) \geq \lambda(m + j)$. This condition can only hold if $t(s) = 0$ so that an equilibrium with supporters is trivial; apparently the expectation $j(s) = j$ was not so reasonable.

Let us start with considering the internal and external stability conditions for members, whose payoff is $\kappa(m + j(s)) + \frac{st(s)}{m+j(s)}$. A **member** should not want to deviate and become a **free-rider**:

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

Single deviations imply $m-1$ members and s supporters upon deviation. Our assumption is that, as a result, $t(s)$ and $j(s)$, are also unaffected. Conversely, a **free-rider**, whose payoff is $\lambda(m + j(s))$ should not want to deviate and become a **member**:

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

Single deviations imply $m + 1$ members and s supporters upon deviation. Total transfers $st(s)$ are now distributed over the new number of members/joiners. Again, our assumption is that $t(s)$ and $j(s)$ are unaffected.

Let us proceed with considering the internal and external stability conditions for supporters, whose payoff is $\lambda_i\big(m + j(s)\big) - t(s)$. A **supporter** should not want to deviate and become a **free-rider**:

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

Single deviations imply $s-1$ supporters and m members upon deviation. Given the change in s, the supporter considering to become a free-rider needs to have expectations regarding the new number of joiners. By conditionality of supporting transfers, one can reasonably assume the number of joiners to decrease by at least one. Generally, we define $j(s-1)$ as the expected number of joiners if the number of supporters decreases by one to $s-1$. Conversely, a **free rider** should not want to deviate and become a **supporter**.

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

Single deviations imply $s+1$ supporters and m members upon deviation. Both the number of joiners as well as the transfer may change as a result of the change in s. The free-rider who is considering to become a supporter needs to have expectations regarding the transfer he will have to pay, as well as the new number of joiners. With regard to joiners, by conditionality of supporting transfers, one can reasonably assume the number of joiners to increase by at least one. Then one can ask the question what transfer could provide an incentive to $j(s+1)$ free-riders to join the coalition. Generally, we define $j(s+1)$ as the expected number of joiners if the number of supporters increases by one to $s+1$. Similarly, we define $t(s+1)$ as the transfer that supporters expect to pay in this situation.

Next, we consider our two remaining conditions. A **supporter** should not want to become a **member**.

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

Single deviations imply $m+1$ members and $s-1$ supporters upon deviation. Both the number of joiners as well as the transfer may change as a result of the change in s. The supporter who is considering to become a member needs to have expectations regarding the transfer he will receive, as well as the new number of joiners. Expectations of the new number of joiners is $j(s-1)$ and was discussed for Condition (11c). With regard to transfers, one can ask the question what transfer could provide an incentive to $j(s-1)$ free-riders to join the coalition. Generally, we define $t(s-1)$ as the transfer that supporters are expected to pay if the number of supporters decreases by one to $s-1$. Conversely, a **member** should not want to deviate and become a **supporter**:

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

Single deviations imply $m-1$ members and $s+1$ supporters upon deviation. The new number of joiners is $j(s+1)$ and the new transfers are $t(s+1)$; expectations of both were already discussed for Condition (11d).

We will consider two types of stability in the next sections. A coalition structure $\{M,F,S\}$ that satisfies the first four Conditions (11a)–(11d) is called *IE-stable*, reflecting the focus of these conditions on internal and external stability of, respectively, the sets of members and supporters. A coalition structure $\{M,F,S\}$ that satisfies all six Conditions (11a)–(11f) is called *support-stable*, reflecting the additional conditions that focus on switching between members and supporters. Note that this is only a technical distinction since all IE-stable coalition structures are also support-stable under appropriate expectations on supporting transfers. This is further explained in Remark 2 at the end of Section 3.

We will now proceed to present our main results and defer comparing our stability conditions to those by CS93 until Section 5.

3 Existence of equilibria with supporters

In order to find equilibria we need to specify expectations as discussed in Section 2.3. We choose to do so by using Conditions (11c) and (11d). Suppose that $\{M,F,S\}$ joint with $j(s)$ and $t(s)$ constitutes an equilibrium. Then we make the following assumption:

Assumption 1. Stability conditions (11c) and (11d) hold with equality. That is, $t(s)=\lambda_i(m+j(s))-\lambda_i(m+j(s-1))$ and $t(s+1)=\lambda_i(m+j(s+1))-\lambda_i(m+j(s))$.

The first equality of Assumption 1 states that a supporter who considers to become a free-rider assumes that due to his action, the decrease in the number of joiners leads to a new free-rider payoff that is just equal to his original payoff as a supporter. The second equality states that a free-rider who considers to become a supporter assumes that due to his action, the increase in the number of joiners leads to a new supporter payoff that is just equal to his original payoff as a free-rider. As a result, in equilibrium, the supporting transfers are such that supporters have no incentive to deviate and terminate their support (and vice versa). When a coalition structure is not stable—or, equivalently, s is larger than necessary to stabilize $\{M, F, S\}$—Assumption 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 Assumption 1 is that, in searching for stable coalition structures, we can restrict ourselves to a specific functional form for $t(s)$. Note that in Remark 2 at the end of Section 3 we will consider an alternative expectation of $t(s+1)$.

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:

$$\Phi_i\big(m + j(s)\big) = \kappa_i\big(m + j(s)\big) - \lambda_i\big(m + j(s) - 1\big). \tag{12}$$

We make the following assumption.

Assumption 2. $\Phi_i\big(m + j(s)\big)$ 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 Conditions (6a)–(6b), we have $\Phi_i\big(m + 0\big) \geq 0$ and $\Phi_i\big(m + 1\big) \leq 0$. As a result, Assumption 2 implies that $\Phi_i\big(m + j(s)\big)$ 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 2 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 prove the following lemma that provides structure to $j(s)$, based on the system of stability conditions (11). We will use the lemma in the proof of Proposition 1 as well as in Section 4.

Lemma 1. *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, Assumption 1—dictates that $j(s) > j(s-1)$. □

Lemma 1 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 Lemma 1, 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.[4]

We now present our main results.

Proposition 1. *Under Assumptions 1–2, there exist IE-stable coalition structures* $\{M, F, S\}$ *with* $s > 0$.

Proof. Consider coalition structure $\{M, F, S\}$. By Assumption 1, supporting transfers are such that Conditions (11c) and (11d) hold with equality. Specifically, for $t(s)$ we have:

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

We proceed to check the first two conditions of system (11). Substitute (13) in Conditions (11a)–(11b) and rearrange the resulting inequalities, using chained notation:

$$\begin{aligned}\big(m + j(s)\big)\cdot\Big(\lambda_i\big(m + j(s) - 1\big) - \kappa_i\big(m + j(s)\big)\Big)& \\ \leq s\cdot\Big(\lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s-1)\big)\Big)& \\ \leq \big(m + j(s) + 1\big)\cdot\Big(\lambda_i\big(m + j(s)\big) - \kappa_i\big(m + j(s) + 1\big)\Big)&.\end{aligned} \tag{14}$$

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 (14) holds, we first check whether the inequality holds between the first and last term. Using (12), and flipping signs, we write the inequality between the first and last term of (14) as

$$\big(m + j(s)\big) \cdot \Big(\Phi_i\big(m + j(s)\big)\Big) \geq \big(m + j(s) + 1\big) \cdot \Big(\Phi_i\big(m + j(s) + 1\big)\Big). \tag{15}$$

Recall that we focused on stable coalitions when we introduced the possibility of supporting transfers (see stage 3). Hence, using (12), the external stability condition (6b) implies that $\Phi_i(m + 1) \leq 0$. Combining this inequality with Assumption 2 we have $0 \geq \Phi_i\big(m + j(s)\big) \geq \Phi_i\big(m + j(s) + 1\big)$, which implies that (15) holds with strict inequality.

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

Now, combining Assumption 1 and Lemma 1—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 (14) 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 4. 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 1. □

In our next result, we extend Proposition 1 to support-stability, which includes Conditions (11e)–(11f). This extension requires one additional assumption, but see Remark 2 below.

Assumption 3. The function $\lambda()$ in (5) is convex in $m + j(s)$.

Proposition 2. *Under Assumptions 1–3, there exist support-stable coalition structures* $\{M, F, S\}$ *with* $s > 0$.

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 (13) to obtain:

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

Inequality (16) is similar to Condition (11b) but with one supporter less and therefore, by Lemma 1, 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 (16) 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 Assumption 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 (12), can be written as

$$\Phi(m + j(s-1) + 1) \leq 0. \tag{17}$$

Using (12), the external stability condition (6b) implies that $\Phi_i(m+1) \leq 0$ and, by Assumption 2, $\Phi_i()$ decreases further for larger numbers of joiners. Since $m + j(s-1) + 1 \geq m + 1$, we know that inequality (17), and therefore inequality (16), holds.

Because inequality (16) holds in both cases, we know that Condition (11e) is satisfied.

We now turn to Condition (11f). By Assumption 1, we substitute $\lambda_i(m + j(s + 1)) - \lambda_i(m + j(s))$ for $t(s + 1)$ to obtain:

$$\kappa_i\big(m + j(s)\big) + \frac{st(s)}{m + j(s)} \geq \lambda_i\big(m + j(s + 1) - 1\big) - \lambda_i\big(m + j(s + 1)\big) + \lambda_i\big(m + j(s)\big). \quad (18)$$

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

$$\lambda_i\big(m + j(s + 1)\big) - \lambda_i\big(m + j(s + 1) - 1\big) \geq \lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s) - 1\big). \quad (19)$$

For any $j(s + 1) \geq j(s)$, this inequality holds if $\lambda()$ is convex in $m + j(s)$, and we use Assumption 3. □

Assumption 3 is restrictive. It puts a constraint on the functional forms and parameter values of the cost- and benefit functions for which the proposition holds. For example, convexity is violated in one conventional specification featuring quadratic benefit- and cost functions, as used by e.g. Barrett (1994), Diamantoudi and Sartzetakis (2006), and Rubio and Ulph (2006). Nevertheless, our example in the next section with linear benefits and quadratic costs satisfies Assumption 3.

Remark 2. Note that Proposition 2 relies on Assumption 3 only because of the specific expectations on supporting transfers that we formulated in Assumption 1. To see why, note that the assumption is required only to satisfy Condition (11f) which says that members should not have an incentive to deviate and become a supporter. We know from Condition (11a), which does not require Assumption 3, that members have no incentive to deviate and become a free-rider. Satisfying both conditions requires that the additional gain from a potential increase in the number of joiners – due to the member becoming a supporter rather than a free-rider, i.e. $\lambda(m - 1 + j(s + 1)) - \lambda(m - 1 + j(s))$ with $j(s + 1) > j(s)$ – is outweighed by the supporting transfer $t(s + 1)$. Since by Assumption 1 we have $t(s + 1) = \lambda_i(m + j(s + 1)) - \lambda_i(m + j(s))$, the reformulation of this requirement implies convexity. Had we formulated an alternative, sufficiently large, expectation of $t(s + 1)$, then it is obvious that both Conditions (11d) and (11f) would hold with inequality (to check Condition (11f), note that this condition is implied by Condition (11a) if and only if $\lambda_i(m - 1 + j(s)) \geq \lambda_i(m - 1 + j(s + 1)) - t(s + 1)$). For example, we could choose the expectation $t(s + 1) = \lambda_i(m + j(s + 1)) - \lambda_i(m + j(s)) + \lambda_i(m - 1 + j(s + 1)) - \lambda_i(m - 1 + j(s))$. Using this alternative expectation, Assumption 3 could be dropped without any implications for our other results. The question is then which expectation of $t(s + 1)$ is reasonable. In Assumption 1 we introduced an expectation of $t(s + 1)$ that is similar in form to the related expectation of $t(s)$, both of which are threshold values to stabilize the current coalition structure, based on (11c) and (11d). It is not immediately obvious why agents would expect higher supporting transfers, but such expectations may depend on the specific application of our model.

We summarize our model and main results as follows. Consider any coalition structure $\{M, F, S\}$. In addition, consider any combination of values for $j(s)$ and $t(s)$, as well as expected values for $j(s-1)$, $j(s+1)$, $t(s-1)$, and $t(s+1)$. This combination will constitute an equilibrium if and only if it satisfies Conditions (11a)–(11d) (IE-stability) or (11a)–(11f) (support-stability). While this is a general definition, we stress that the expectations should be reasonable. Our implementation of such reasonable expectations is that Conditions (11c) and (11d) hold with equality as in Assumption 1. We can further refine our equilibrium by considering, for example, that $j(s-1) = j(s) - 1$. In Section 4 we illustrate that under this refinement, there also exists a (support-stable) equilibrium.

4 Example

In this section we analyze an example of a cartel formation game that satisfies Assumptions 2 and 3. 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, as used by e.g. Botteon and Carraro (1997), Barrett (2006), and Finus and Maus (2008). We update (1) to:

$$w_i = \beta \left(\textstyle\sum_{j \in N} q_j \right) - \frac{1}{2} \gamma (q_i)^2 \ \ \forall i \in N, \tag{20}$$

where β and γ 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 (20) dictates that each agent has a dominant strategy in terms of contributions. Using (2) and (3) we have:

$$\overline{q}\big(m + j(s)\big) = \frac{\big(m + j(s)\big)\beta}{\gamma}, \tag{21}$$

$$\underline{q}\big(m + j(s)\big) = \frac{\beta}{\gamma}. \tag{22}$$

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:

$$\kappa_i\big(m + j(s)\big) = \frac{\beta^2}{\gamma}\left(\frac{1}{2}\big(m + j(s)\big)^2 + \big(n - m - j(s)\big)\right), \tag{23}$$

$$\lambda_i\big(m + j(s)\big) = \frac{\beta^2}{\gamma}\left(\big(m + j(s)\big)^2 + \big(n - m - j(s)\big) - \frac{1}{2}\right). \tag{24}$$

We will now continue by examining, first, if there exists an s such that we have a stable coalition structure with $j(s) = 1$. We will see below that such an s exists. Given its existence, we also know, by Lemma 1, that $j(s-1) = 0$. This allows us, using Assumption 1 and the stability conditions (11a)–(11d), to determine the level of s. We continue with $j(s) = 2, 3, \ldots$, constrained by $j(s) + s \leq n - m$. Note that we may not always find an equilibrium for each $j(s)$, an issue that is further discussed in the last paragraph of this section.

Consider a coalition with $m = 3$ members. Take parameter values $\beta = \gamma = 1$ and $n = 50$ and consider $j(s) = 1$. Using (13) and Lemma 1 we have $t(s) = \lambda_i\big(m + j(s)\big) - \lambda_i\big(m + j(s-1)\big) = \lambda_i\big(3 + 1\big) - \lambda_i\big(3 + 0\big) = 61\frac{1}{2} - 55\frac{1}{2} = 6$. 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 (23) and (24) into (14), using our selected parameter values. Doing so, we obtain:

$$\big(3 + 1\big) \cdot \left(55\frac{1}{2} - 54\right) \;\leq\; s \cdot \left(61\frac{1}{2} - 55\frac{1}{2}\right) \;\leq\; \big(3 + 1 + 1\big) \cdot \left(61\frac{1}{2} - 57\frac{1}{2}\right). \tag{25}$$

This chained inequality holds for the interval $s \in \big[\underline{s} = 1, \bar{s} = \frac{20}{6}\big]$, containing the natural numbers 1, 2, and 3. The number of supporters in equilibrium is the smallest of these, since otherwise $j(s) = j(s-1)$, which contradicts Lemma 1. The second and third supporter prefer to free-ride since they gain nothing by being a supporter. Hence, $s = 1$. 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 $\big\{M, F, S\big\}$, 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 1 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 (13):

$$\begin{aligned} t(s) &= \lambda_i\big(3+j(s)\big)-\lambda_i\big(3+j(s-1)\big) \\ &= \big(3+j(s)\big)^2+\big(50-3-j(s)\big)-\big(3+\big(j(s)-1\big)\big)^2-\big(50-3-\big(j(s)-1\big)\big) \\ &= \big(3+j(s)\big)^2-\big(3+j(s)-1\big)^2-1 \\ &= 2j(s)+4. \end{aligned} \tag{26}$$

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.

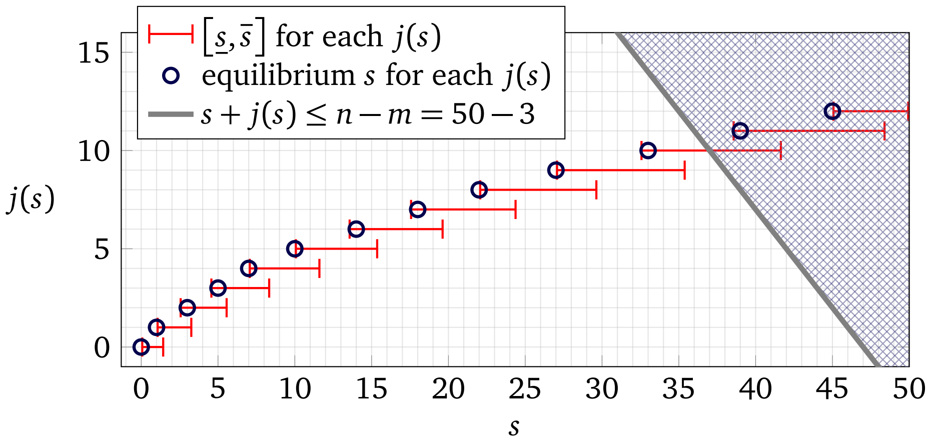

Figure 1: Equilibrium number of $j(s)$ joiners as a function of the number of s supporters on top of the $m = 3$ coalition members, given the parameter values introduced in Section 4. Bars depict the interval $[\underline{s}, \bar{s}]$ for each value of $j(s)$ and the decreasing line indicates the constraint on the aggregate number of supporters and joiners, given by the total number of $n = 50$ agents minus $m = 3$ members.

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, since otherwise $j(s) = j(s-1)$, which contradicts Lemma 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 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), (23), and (24), and using the parameter values from our example we can now calculate payoffs of members/joiners, supporters, and free-riders for each equilibrium.

$$
\begin{aligned}
w_i &= \kappa_i\big(m + j(s)\big) + \frac{st(s)}{m + j(s)} \\
&= \frac{1}{2} j(s)^2 + 2j(s) + 51.5 + 2s\left(\frac{j(s)+2}{j(s)+3}\right) \qquad \forall i \in M \cup J, && (27)\\
w_i &= \lambda_i\big(m + j(s)\big) - t(s) \\
&= j(s)^2 + 3j(s) + 51.5 \qquad \forall i \in S, && (28)\\
w_i &= \lambda_i\big(m + j(s)\big) \\
&= j(s)^2 + 5j(s) + 55.5 \qquad \forall i \in F \setminus J. && (29)
\end{aligned}
$$

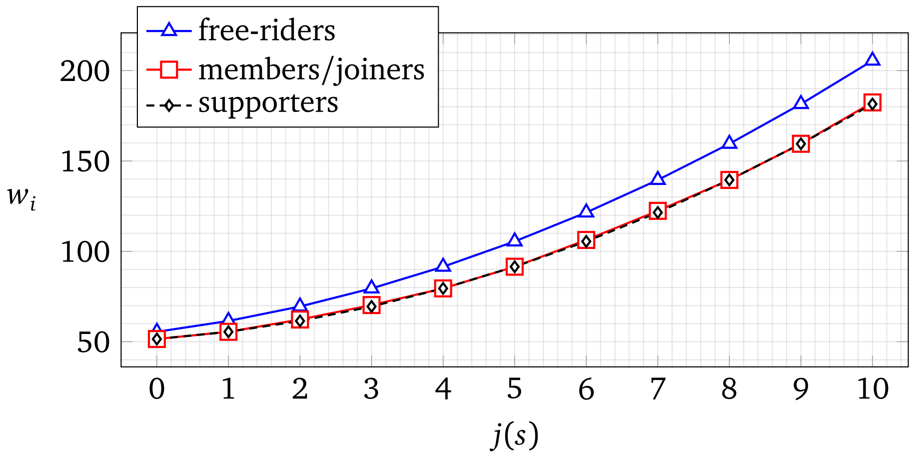

Figure 2: Individual welfare based on (10) for each of three types of agents—free-riders, members/joiners, and supporters—for each of the equilibria as identified by $j(s) \in (0, 1, \ldots, 10)$.

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 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. Table 2 illustrates the possible welfare gains in this coordination game. The table compares total welfare for four scenarios: (i) without cooperation nor support, (ii) with cooperation but no support, (iii) with cooperation and support for all levels of support, and (iv) the grand coalition. Note that both (i) and (iv) are not stable and (ii) is only stable when the option of support is not considered. Table 2 shows that, for the particular example of this section, the option of support increases total welfare considerably, up to threefold. In addition, there are clear gains from coordinating on an equilibrium with a large number of joiners. Nevertheless, there is still a large gap with total welfare in the grand coalition. Support can only bridge this gap to a small extent.

Table 2: Total welfare depending on cooperation and support.

Scenario		**Total welfare**
(i) No cooperation, no support		2475
(ii) Cooperation, no support		2763
(iii) Cooperation, support:		
	1 joiner	3045
	2 joiners	3415
	3 joiners	3870
	4 joiners	4407
	5 joiners	5023
	6 joiners	5715
	7 joiners	6480
	8 joiners	7315
	9 joiners	8217
	10 joiners	9183
(iv) Grand coalition		62500

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.

5 Comparing our results to CS93

In this section we compare our results to Proposition 5 on external commitment by CS93. The central question they address is under what conditions a set of supporters of size s is able to stabilize a coalition containing all other agents. Our stability conditions differ from those introduced by CS93, mainly because we relax two of their assumptions. First, CS93 assume a less plausible form of commitment, which was discussed already in Section 1 and Remark 1. Second, CS93 assume that supporters maximize the number of joiners so that, for given m and s, we have $j \equiv n-m-s$. As a consequence, this assumption eliminates the existence of free-riders. We do not impose this restriction but rather endogenize the size of the set of joiners so that $j(s) \leq n-m-s$.

CS93 claim that the coalition of size $m+j$ will be stable if their Condition (12) holds:

$$\left(\frac{m+j}{s}\right)\big(\lambda(m+j-1)-\kappa(m+j)\big) < \lambda(m+j)-\lambda(m). \tag{30}$$

First, they set the supporting transfer t by each supporter such that

$$\kappa(m+j)+\frac{st}{m+j} = \lambda(m+j-1). \tag{31}$$

The transfer is set in such a way that coalition members are indifferent between being in the coalition and deviating to become a free-rider. This is our Condition (11a) holding with equality, while we also assess the reverse condition in Condition (11b).

Given (31) and if (30) holds, then:

$$\lambda(m+j)-t \geq \lambda(m). \tag{32}$$

Supporters in the equilibrium with support are better off than supporters in the equilibrium without support. Note that this is a weaker condition than our Condition (11c) and we also assess the reverse condition in Condition (11d).

We also have their Condition (13):

$$\lambda(m+j)-t \geq \kappa(m+j+1). \tag{33}$$

This condition is interpreted by CS93 as: *"... no additional countries want to join the coalition"*. This interpretation is questionable. Note that the only potential 'additional countries' are supporters. Their payoff is the left hand side of this expression. The right hand side gives the payoff to members if there were one more member, however, without taking into account supporting transfers. It is unclear why a supporting agent would make the suggested comparison. Note that our Condition (11e) is the condition that is probably intended by CS93, while we also assess the reverse condition in Condition (11f).

To prove their proposition, CS93, verify three conditions on the payoffs of members, all of which follow trivially from (30).[5] In addition, they verify (33), arguing that *"... it is easy*

[5] The first is that members in the equilibrium with support have no incentive to defect: $\kappa(m+j)+\frac{st}{m+j} \geq$

to see that (12) implies ... (13)", i.e. using our numbering: (30) implies (33). This statement is not correct unless further assumptions are made, for example $\lambda(m) > \kappa(m+j+1)$, but this is not easy to justify at all.

Despite the inconsistency of their result, we reconstruct Figures 1 and 2 for Proposition 5 on external commitment by CS93. By assumption of $j \equiv n - m - s$, Figure 3 shows that j is decreasing linearly in s, in sharp contrast with Figure 1. The stability conditions employed by CS93 dictate that coalition structures with 15 supporters or less are not stable. All other coalition structures, with sufficient supporters to divide the burden of paying the supporting transfers, are stable. This allows equilibria with up to 31 joiners, notably more than the maximum of 10 joiners in our model. Yet, Figure 4 shows that this comes at the cost of low payoffs to the supporters that facilitate this high number of joiners. From 11 joiners onward, payoffs to supporters start to diverge from payoffs to members/joiners, and ultimately they decline in the number of joiners. This divergence shows that the external commitment assumption employed by CS93 can be very strong, at least in this example. Note that up to $j = 10$, the only difference with Figure 2 is that CS93 find slightly higher payoffs to supporters. This difference is due to their alternative calculation of transfers in (31).

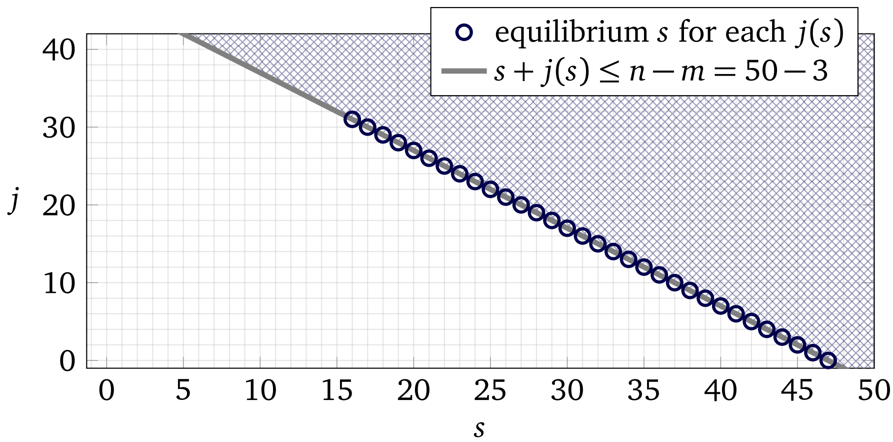

Figure 3: Identical to Figure 1 except that calculations are based on Proposition 5 on external commitment by CS93.

$\lambda(m+j-1)$. Obviously, this condition is implied by (30); it is equivalent to our Condition (11a). The second is that members in the equilibrium with support are better off than members in the equilibrium without support: $\kappa(m+j) + \frac{st}{m+j} > \kappa(m)$. This is true, since $t > 0$ and $\kappa(m+j) > \kappa(m)$. The third is that members in the equilibrium with support are better off than free-riders in the equilibrium without support: $\kappa(m+j) + \frac{st}{m+j} > \lambda(m)$. This is true since $\lambda(m+j-1) \geq \lambda(m)$ by assumption.

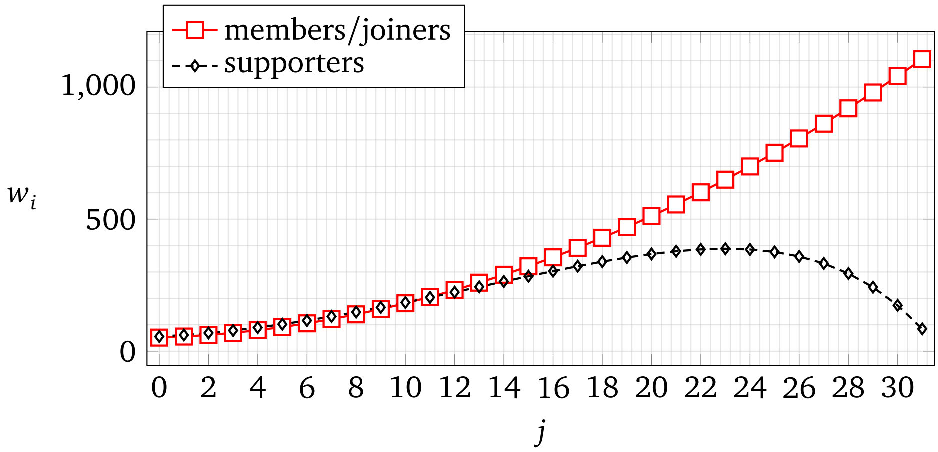

Figure 4: Identical to Figure 2 except that calculations are based on Proposition 5 on external commitment by CS93.

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' behavior. 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 agents 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; Carraro et al., 2006). 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. The crucial aspect of our model that needs to be maintained under such simplification is the joiners' behavior. Our model requires that joiners switch instantaneously in response to the supporting transfers offered to the coalition. It is not straightforward, however, how this endogenous determination of the number of joiners can be incorporated 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 agents, 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 behavior would mimic the kind of coordination that is discussed in the context of multiple equilibria occurring in the example of Section 4.

Finally, and to close the paper, 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 4, 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. and R. 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.

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.

Barrett, S. (2006). Climate treaties and "breakthrough" technologies. *American Economic Review: Papers and Proceedings 96*(2), 22–25.

Bauer, T., J. Bredtmann, and C. 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. and M. 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., J. Eyckmans, and M. Finus (2006). Optimal transfers and participation decisions in international environmental agreements. *Review of International Organizations 1*(4), 379–396.

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. Sartzetakis (2006). Stable international environmental agreements: An analytical approach. *Journal of Public Economic Theory 8*(2), 247–263.

Diamantoudi, E. and E. Sartzetakis (2015). International environmental agreements: Coordinated action under foresight. *Economic Theory 59*(3), 527–546.

Feldman, N. (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.

Nordhaus, W. (2015). Climate clubs: Overcoming free-riding in international climate policy. *American Economic Review 105*(4), 1339–1370.

Olmstead, S. and R. 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.

Rubio, S. and A. Ulph (2006). Self-enforcing international environmental agreements revisited. *Oxford Economic Papers 58*(2), 233–263.

Sandler, T. and J. Tschirhart (1997). Club theory: Thirty years later. *Public Choice 93*(3-4), 335–355.

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.