Perfectly Coalition-Proof Nash Equilibria for Overlapping Coalitions: With an Application to International Environmental Agreements

By

Emilson Caputo Delfino Silva
Department of Marketing, Business, Economics & Law, University of Alberta, Canada (emilson@ualberta.ca)

and

Xie Zhu
Department of Economics, Oakland University, USA (zhu@oakland.edu)

August 20, 2013

Abstract: The Perfectly Coalition-Proof Nash equilibrium (PCPN) concept is extended to allow for the emergence of overlapping coalitions in equilibrium. We study the efficiency and stability properties of environmental agreements to control emissions of correlated continental and global pollutants. We show that set of PCPN equilibria includes Perfectly Strong Nash (PSN) equilibria if the national damage from continental pollution is sufficiently large relative to the national damage from global pollution. We also show that: (i) continental agreements may be perfectly coalition-proof under much less restrictive circumstances; and (ii) perfect Nash equilibria for fully overlapped agreements may be superior to the Grand Coalition’s optimal allocation in the presence of coalitional operation costs.

Keywords: perfectly coalition-proof equilibrium; overlapping coalitions; climate change; correlated pollutants; international environmental agreements.

JEL Classification: C72, D62, D78, H41, H77.

Acknowledgements: We wish to thank Debraj Ray and Charles Kahn for helpful discussions that motivated us to think about how self-enforceability and coalition-proofness could be extended to settings in which overlapping coalitions can coexist. Any errors are, of course, our own.
1. Introduction

Many, if not all, social and organizational networks are environments in which overlapping (bilateral or multilateral) groups or coalitions are the norm rather than the exception. Individuals and firms are typically actively engaged in various groups (associations, clubs, alliances, etc.) at the same time. Nations are also members of various international organizations or alliances. Examples of international overlapping coalitions abound: (i) many nations participate in bilateral or multilateral free trade agreements; (ii) some nations participate in bilateral or multilateral green R&D (e.g., Carbon Capture and Storage) agreements; and (iii) various nations are simultaneously signatories of different types of international environmental agreements, such as the Protocol Montreal, the Kyoto Protocol and international agreements to control acid rain (in particular, in North America and Europe).

In this paper, we extend the coalition-proof and perfectly coalition-proof equilibrium concepts of Bernheim, Peleg and Whinston (1987) – henceforth, we shall refer to this paper as BPW – to allow for the emergence of overlapping coalitions in perfect equilibria. The extended equilibrium concepts impose an extra requirement on self-enforceability: a strategy vector is self-enforcing if: (i) the Nash equilibrium for each non-overlapping partition of the set of players is coalition proof, just as in BPW, and (ii) for each partition of the set of players involving overlapping coalitions, the Nash equilibrium for an associated partition consisting of the union of the overlapping sets and the union’s complement is also coalition proof.

To our knowledge, Chakravorti and Kahn (1994) is the only other paper that formally extends the coalition-proof concept to allow the emergence of overlapping coalitions in equilibrium. The authors propose a new equilibrium

---

1 As in BPW, we consider non-binding agreements when players can have unlimited communication. Thus, a coalition is
concept, called Universal Coalition-Proof (UCP) equilibrium, which not only permits each player to simultaneously belong to various coalitions, but also allows deviant subcoalitions to form new coalitions with players who do not belong to their original coalitions. Our proposed extension of coalition proofness is more restrictive than UCP: we do not allow a deviant subcoalition to form new coalitions with a player who is not directly or indirectly linked to any of the players who belong to the deviant subcoalition. For a particular collection of overlapping coalitions, two distinct players $i$ and $j$ who belong to the superset consisting of the union of overlapping coalitions are directly connected if both belong to (at least) one of the overlapping coalitions. These players are indirectly connected if they are not elements of the intersection between any coalition in which player $i$ is a member and any coalition in which player $j$ is a member. In our extension, indirectly linked players $i$ and $j$ (or, more generally, subcoalitions of indirectly linked players) are allowed to deviate from their coalitions and form a new coalition in which both are directly connected, but neither $i$ nor $j$ can deviate from their coalitions to form a new coalition with a player $k$ who belongs to the complement of the union of overlapping coalitions.

We provide an application of our notions of self-enforceability and coalition proofness in the context of international environmental agreements. We consider a setting in which the production of energy through the burning of fossil fuels generates two types of pollutants, carbon dioxide and sulfur dioxide. Carbon dioxide emissions produce global climate change damages. Sulfur dioxide emissions produce regional (i.e., continental) acid rain damages. As all nations

---

2 Recent scientific evidence shows that mitigating greenhouse gases and controlling air pollutants are two closely related tasks (see, e.g., IPCC (2001), EEA (2004), Bollen et al. (2009) and Defra (2010). Enhanced fuel efficiency can reduce both types of pollutants. The linkage between the two types of pollutants is also observed in abatement spillovers, in the sense that technical measures of abatement aiming at reducing one type of pollutant may reduce or increase the other type of pollutant. For instance, selective catalytic reduction (SCR) on gas boilers reduces both methane – a greenhouse gas – and nitrogen oxides (EEA, 2004).
face both types of damages, they may wish to join different, potentially overlapping, international environmental agreements.

International environmental agreements (IEAs) have been formed around the globe to mitigate transboundary air pollutions. For example, the Convention on Long Range Transboundary Air Pollution (CLRTAP), which entered into force in 1983, together with eight subsequent protocols to the Convention, aims to reduce sulfur dioxide emissions, nitrogen oxides emissions and other transboundary air pollutants in countries in the United Nations Economic Commission for Europe (UNECE) region. In North America, Canada and the United States signed in 1991 the Canada-United States Air Quality Agreement to deal with transboundary acid rain problems. The two countries signed the Ozone Annex to the Agreement in 2000 to reduce ground-level ozone.

Conventional wisdom suggests that a Pareto efficient climate change mechanism requires participation and cooperation of all countries emitting greenhouse gases. But, accumulated evidence of failures in producing a fully participatory global agreement to reduce greenhouse gas emissions motivated Silva and Zhu (2011) to consider the efficiency of alternative IEAs. They examined the efficiency properties of IEAs designed to control emissions of both greenhouse gases and transboundary air pollutants, without requiring the formation of a fully participatory protocol for climate change. The authors considered the efficiency of perfect equilibria for fully overlapped coalitional structures in which all nations belong to at least one IEA, there is no IEA containing all nations and there is complete interconnection among distinct IEAs – i.e., IEAs have at least one member nation in common. The authors demonstrate

---

3 See, e.g., Caplan et al. (2003), for a discussion on the efficiency of an international carbon dioxide emissions permit market without global participation.

4 See, e.g., Caplan and Silva (2005) and Silva and Zhu (2009), for efficient international environmental mechanisms that control both greenhouse gases and air pollution.
that non-cooperative national environmental policy making can lead to a Pareto-efficient subgame perfect equilibrium provided that the forward looking policy makers anticipate proportionally equitable international income transfers within all coalitions. They also show that there is a large set of efficient fully overlapped coalitional structures that produce national payoffs identical to those produced by the efficient Grand Coalition.

Due to the lack of a global authority with coercive power, the success of efficient fully overlapped IEAs in controlling both climate change and air pollution hinges on the voluntary participation and compliance of sovereign nations. In other words, the IEAs must be self-enforcing. In providing the global public good of mitigating climate change, free riding incentives may prevail and lead countries to defect from an efficient coalitional structure. The task of the current paper is to go beyond the efficiency issue to also investigate the stability of perfect equilibria for fully and partially overlapped coalitional structures. We also identify stable perfect equilibria for all other relevant coalitional structures.

In what follows, if the subgame perfect equilibrium for a partition of the set of players is Perfectly Coalition-Proof Nash (PCPN) equilibrium, then the coalitional structure produced by the partition is stable. As usual, if a PCPN equilibrium is Pareto efficient, it is also a Perfectly Strong Nash (PSN) equilibrium. We demonstrate that the subgame perfect equilibrium for any fully overlapped coalitional structure is a PSN equilibrium if the national damage from air pollution is at least twice as large as the national damage produced by greenhouse gas emissions. In addition, we also show that continental agreements may be self-enforcing and hence emerge in PCPN equilibria under several circumstances. Finally, we demonstrate that in the presence of coalition operation costs, the subgame perfect equilibrium for some fully overlapped coalitional structures may Pareto dominate the subgame perfect equilibrium for a fully participatory, global,
coalitional structure – i.e., the coalitional structure produced by the Grand Coalition.

The literature on the formation of IEAs can be divided into two branches according to whether countries are assumed to be myopic or farsighted. Of particular relevance to this paper are the papers that utilize the concept developed by d’Aspremont et al. (1983) for cartel formation in a non-cooperative game framework (e.g., Carraro and Siniscalco (1993), Barrett (1994)) and the papers that have recently introduced the idea of farsightedness into the analysis of IEAs. Osmani and Tol (2009), for example, demonstrate that farsighted stable coalitions can achieve higher welfare and environmental quality than the stable coalitions that are produced by myopic countries who rely on the stability concept developed by d’Aspremont et al. (1983).

The most important difference between our framework and those reviewed above is that our framework does not adopt concepts originating from cooperative game theory; rather, since coalition-proofness is a refinement of Nash equilibrium, its foundation is deeply rooted into non-cooperative game theory. Coalition formation in our case rises due to the players’ equilibrium selection arising in unrestricted pre-game communication. An agreement made by the players to select a particular Nash equilibrium is non-binding.

By focusing on the coalition-proof equilibrium concept, we also consider farsighted players. When contemplating a deviation, any coalition member considers the potential subsequent deviations that the first deviation may trigger. This is in keeping with the logic of coalition-proofness in BPW. However, unlike in BPW, each player also foresees the effects of a first deviation on coalitions in which he is not a member provided that these other coalitions overlap with the coalition in which he is a member. Our framework, therefore, considers the

---

effects of farsighted behavior for the formation of non-binding IEAs. Another
novelty of our approach is that we consider stability of IEAs in a setting with two
types of correlated environmental externalities.

The paper is organized as follows. Section 2 shows how BPW’s coalition-
proof and perfectly coalition-proof equilibrium concepts can be extended to allow
for the emergence of overlapping coalitions in the equilibrium refinement. Section
3 builds the basic model of a world featuring four nations that may form IEAs to
combat two types of correlated environmental evils. Section 4 examines the
subgame perfect equilibrium for each relevant coalitional structure. Section 5
adopts the adapted coalition-proof refinement in order to select PCPN equilibria –
i.e., it applies the adapted equilibrium concepts in order to identify the set of
stable coalitional structures. Section 6 considers the effects of coalition operation
costs, which increase with membership size at an increasing rate. Section 7
concludes the paper.

2. Extension of Coalition-Proof Nash Equilibrium

In our framework we apply two types of Nash equilibrium refinements – the
notions of Perfectly Strong Nash equilibrium and Perfectly Coalition-Proof Nash
equilibrium – to select stable perfect Nash equilibria in dynamic games with
overlapping coalitions.

Consider a game with \( n \) players. The \( n \)-player game can be described as
follows: \( \Gamma = \left[ \left\{ u_i \right\}_{i=1}^{n}, \left\{ S_i \right\}_{i=1}^{n} \right] \), where \( S_i \) represents the strategy set for player \( i \)
and \( u_i : \prod_{k=1}^{n} S_k \to R \) denotes player \( i \)’s payoff function. The notion of Strong
Nash (SN) equilibrium (see Aumann (1959) and BPW) is defined as follows:
Definition 1. $s^* \in \prod_{i=1}^n S_i$ is an SN equilibrium if and only if for all subsets $X$ of $\{1,...,n\}$ and for all $s_X \in \prod_{i \in X} S_i$ there exists an agent $j \in X$ such that $u_j(s^*) \geq u_j(s_X, s_{-X}^*)$, where $s_{-X}^* = \{s^*_i\}_{i \not\in X}$.

The following quote taken from BPW is very useful for our purposes:

“…while the Nash concept defines the equilibrium only in terms of unilateral deviations, Strong Nash equilibrium allows for deviations by every conceivable coalition. We believe, however, that the Strong Nash concept is actually “too strong.” In particular, coalitions are allowed too much freedom (in fact, complete freedom) in choosing their joint deviations while the whole set of players must originally be concerned with arriving at an agreement that is immune to deviations by any coalition, no deviating group of players (including the coalition of the whole) faces a similar restriction. In environments with unlimited private communication, however, any meaningful agreement to deviate must also be self-enforcing (i.e., immune to deviations by subcoalitions). This inconsistency in the Strong Nash concept most clearly manifests itself in the stringent requirement that a Strong Nash equilibrium must be Pareto efficient (within the entire feasible payoff space of the game). As a result of this requirement, Strong Nash equilibria almost never exist.” (BPW, page 3.)

We follow BPW’s reasoning and extend its Coalition-Proof Nash (CPN) and Perfectly Coalition-Proof Nash (PCPN) equilibrium concepts in order to allow for the coexistence of overlapping coalitions in equilibrium. As pointed out by BPW, it is important to note that if a CPN equilibrium is Pareto efficient, it is also a SN equilibrium. Similarly, if a PCPN equilibrium is Pareto efficient, it is also a Perfectly Strong Nash (PSN) equilibrium.
Before we provide our extension to BPW’s self-enforceability and coalition-proofness, it is useful to present the original concepts. For the $n$-player game $\Gamma = \left( \{u_i\}_{i=1}^n, \{S_j\}_{j=1}^n \right)$, let $J$ be the set of subsets of $\{1,\ldots,n\}$. For $h \geq 1$, $m \geq 1$ and $n \geq h + m$, let $J_{\{h,\ldots,h+m\}} \equiv \{h, h+1, \ldots, h+m\} \in J$. For $h = 1$ and $m = n-1$, we have $J_{\{1,\ldots,n\}} \equiv \{1,\ldots,n\}$. Hence, $J_{\{h,\ldots,h+m\}}$ is a consecutively ordered subset of $J_{\{1,\ldots,n\}}$.

Let $J_{(h,m)}$ be the set of subsets of $J_{\{h,\ldots,h+m\}}$. An element of $J_{(h,m)}$ is denoted $J_{(h,m)}$. Let $S_{J_{(h,m)}} \equiv \prod_{i \in J_{(h,m)}} S_i$. For the case $\{1,\ldots,n\}$, we will write $S$. Let $-J_{(h,m)}$ be the complement of $J_{(h,m)}$ in $\{1,\ldots,n\}$. For every $s_J \in S_{J_{(h,m)}}$, let

$$\Gamma / s_J \equiv \left[ \left\{ u_i \right\}_{i \in J_{(h,m)}}, \left\{ S_j \right\}_{j \in J_{(h,m)}} \right]$$

be the game induced on coalition $J_{(h,m)}$ by the actions $s_J$ for coalition $-J_{(h,m)}$, where $\bar{u}_i : S_{J_{(h,m)}} \to R$ is given by

$$\bar{u}_i \left( s_{J_{(h,m)}} \right) \equiv u_i \left( s_{J_{(h,m)}}, s_{-J_{(h,m)}} \right)$$

for every $i \in J_{(h,m)}$ and $s_{J_{(h,m)}} \in S_{J_{(h,m)}}$.

BPW (p.6) defines self-enforceability and coalition-proofness recursively as follows:

**Definition 2.** (i) In a single player game $\Gamma$, $s^* \in S$ is a **Coalition-Proof Nash equilibrium** if and only if $s^*$ maximizes $u_i(s)$.

(ii) Let $n > 1$ and assume that Coalition-Proof Nash equilibrium has been defined for games with fewer than $n$ players. Then,

(a) For any game $\Gamma$ with $n$ players, $s^* \in S$ is **self-enforcing** if, for all $J_{(h,m)} \in J$, $s^*_{J_{(h,m)}}$ is a Coalition-Proof Nash equilibrium in the game

$$\Gamma / s^*_{-J_{(h,m)}}.$$
(b) For any game $\Gamma$ with $n$ players, $s^* \in S$ is a **Coalition-Proof Nash equilibrium** if it is self-enforcing and if there does not exist another self-enforcing strategy vector $s \in S$ such that $u_i(s) > u_i(s^*)$ for all $i = 1, \ldots, n$.

According to BPW, “...an agreement is coalition-proof if it is efficient within the class of self-enforcing agreements, where self-enforceability requires that no coalition can benefit by deviating in a self-enforcing way” (BPW, p.6).

We now turn to our extension of BPW’s self-enforceability and coalition-proofness concepts. Let $o \geq 2$ denote the number of coalitions in a particular collection of overlapping coalitions, $J_{(h_i,m_i)}$, $J_{(h_2,m_2)}$, $\ldots$, $J_{(h_o,m_o)}$, where $h_i \geq 1$, $m_i \geq 1$, $n \geq h_i + m_i$, $l = 1, \ldots, o$. Without loss of generality, assume that $J_{(h_l,m_l)} \cap J_{(h_{l+1},m_{l+1})} \neq \emptyset$, $k = 1, \ldots, o - 1$.

Let $\bigcup_{(h,m)}^o = \bigcup_{l=1}^o J_{(h_l,m_l)} = J_{(h_1,m_1)} \cup J_{(h_2,m_2)} \cup \ldots \cup J_{(h_o,m_o)}$ and $-\bigcup_{(h,m)}^o$ be the complement of $\bigcup_{(h,m)}^o$ in $\{1, \ldots, n\}$. Since $S_{(h,m)} = \prod_{i \in J_{(h,m)}} S$, let $S_{\bigcup_{(h,m)}^o} = \prod_{l=1}^o S_{J_{(h_l,m_l)}}$. For every $s_{-\bigcup_{(h,m)}^o} \in S_{-\bigcup_{(h,m)}^o}$ let

\[
\Gamma / s_{-\bigcup_{(h,m)}^o} = \left[ \{ \overline{u}_l \}_{l \in \bigcup_{(h,m)}^o}, \{ S_l \}_{l \in \bigcup_{(h,m)}^o} \right]
\]

be the game induced on coalition $\bigcup_{(h,m)}^o$ by the actions $s_{-\bigcup_{(h,m)}^o}$ for coalition $-\bigcup_{(h,m)}^o$, where $\overline{u}_l: S_{\bigcup_{(h,m)}^o} \rightarrow R$ is given by $\overline{u}_l \left( s_{\bigcup_{(h,m)}^o} \right) = u_i \left( s_{\bigcup_{(h,m)}^o \setminus \bigcup_{(h,m)}^o}, s_{\bigcup_{(h,m)}^o \setminus \bigcup_{(h,m)}^o} \right)$ for every $i \in \bigcup_{(h,m)}^o$ and $s_{\bigcup_{(h,m)}^o} \in S_{\bigcup_{(h,m)}^o}$. Our extension of **Definition 2** is limited to imposing an additional restriction on the concept of self-enforceability (i.e., replacing (ii.a) with (ii.a')):
(a’) For any game $\Gamma$ with $n$ players, $s^* \in S$ is *self-enforcing* if, for all *non-overlapping* \( J_{\{h,m\}} \in J \), \( s^*_{J_{\{h,m\}}} \) is a Coalition-Proof Nash equilibrium in the game $\Gamma / s^*_{J_{\{h,m\}}}$, and for any collection of overlapping coalitions, \( J_{\{h_l,m_l\}}, l = 1,\ldots,o \), \( s^*_{\bigcup_{\{h,m\}}} \) is a Coalition-Proof Nash equilibrium in the game $\Gamma / s^*_{\bigcup_{\{h,m\}}}$. In words, for a strategy vector to be self-enforcing it needs to “pass” two tests: (i) the Nash equilibrium for any partition of non-overlapping coalitions must be coalition proof, just like in BPW; and (ii) for any partition containing overlapping coalitions, the Nash equilibrium for a partition consisting of two relevant sets – namely, a coalition produced by the union of overlapping coalitions and the complement of such a coalition – must be coalition proof. When coalitions overlap, the players who belong to a particular set of overlapping coalitions make strategic choices taking the strategic choices of all players who do not belong to this set of overlapping coalitions (i.e., players who belong to the complement of the union of the overlapping coalitions) as given, and vice-versa. The equilibrium strategies are determined according to two relevant coalitions, the superset formed by the union of the overlapping coalitions and its complement. Given such sets, the self-enforceability concept of BPW is readily applicable.

As it is stated in the first quote of BPW shown in page 9, in settings with unlimited private communication, “…any meaningful agreement to deviate must also be self-enforcing (i.e., immune to deviations by subcoalitions).” In our view, when coalitions overlap, any agreement to deviate is self-enforcing if and only if it is immune to deviations by all subcoalitions formed by players who belong to at least one of the overlapping coalitions. To see this, consider again the coalitions \( J_{\{h_l,m_l\}}, l = 1,\ldots,o \), described above. The set of all subcoalitions include all subcoalitions that can be formed by players who belong to any particular coalition
$J_{(h,m)}$, $x \in \{1,\ldots,o\}$, only and all subcoalitions that can be formed by players who belong to any pair of coalitions $J_{(h,m)}$ and $J_{(h,m)}$, $y, z \in \{1,\ldots,o\}$, $y \neq z$, but not to both; that is, for all $i \in J_{(h,m)} \cup J_{(h,m)}$ such that $i \notin J_{(h,m)} \cap J_{(h,m)}$. In sum, BPW’s equilibrium concepts for single shot games can be extended to allow for the emergence of overlapping coalitions provided that $n \geq 4$ and subsets of $\{1,\ldots,n\}$ of cardinality greater or equal to 3 can be understood as being supersets produced by the union of overlapping coalitions.

To illustrate, consider an example in which $n = 4$. The proper subsets of $\{1,2,3,4\}$ are

(i) singleton coalitions: $\{1\}, \{2\}, \{3\}, \{4\}$;

(ii) two-member coalitions: $\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}$; and

(iii) three-member coalitions: $\{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}$.

To exemplify our notation, consider the set $\{1,2,3\}$. This corresponds to $J_{\{1,2,3\}}$. The set of subsets of $J_{\{1,2,3\}}$ is denoted $J_{\{1,2\}}$, since $h = 1$ and $m = 2$. When we refer to $J_{\{1,2\}} \in J_{\{1,2\}}$ we mean that $J_{\{1,2\}}$ is any element of $\{\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$.

Suppose that pre-play communication among the players yields the following CPN equilibrium structure: $\{\{1,2\}, \{2,3\}, \{4\}\}$; i.e., coalitions $\{1,2\}, \{2,3\}$ and $\{4\}$ coexist in the CPN equilibrium. Coalitions $\{1,2\}$ and $\{2,3\}$ are pairwise overlapped; player 2 belongs to both coalitions. Players 1 and 2 are directly connected in $\{1,2\}$. Players 2 and 3 are directly connected in $\{2,3\}$. Players 1 and
3, however, are indirectly connected through their direct connection to the pivotal player 2. In this structure, there are two disjoint sets; namely, the set of overlapped coalitions, \(\{\{1,2\},\{2,3\}\}\), and the singleton, \(\{4\}\). Note that the elements of a set of overlapped coalitions are coalitions and the union of such coalitions generates a set whose elements are all players who are directly or indirectly connected in the set of overlapped coalitions: the set \(\{1,2,3\}\), since \(\{1,2\}\cup\{2,3\}=\{1,2,3\}\).

The crucial difference between our framework and BPW’s is that in our framework a set such as \(\{1,2,3\}\) does not need to “materialize” (i.e., to form) in the equilibrium refinement in order to be treated as such. Provided the two sets of pairwise overlapped coalitions, \(\{1,2\}\) and \(\{1,3\}\), are selected in the equilibrium refinement, their union produces the set \(\{1,2,3\}\), and it is the latter and its complement in \(\{1,...,4\}\) which are the relevant coalitions for the notions of self-enforceability and coalition-proofness. In other words, for a game \(\Gamma\) with 4 players, let \(J_{\{1,2\}\cup\{2,3\}}=\{1,2\}\cup\{2,3\}=\{1,2,3\}\) and \(-J_{\{1,2\}\cup\{2,3\}}=\{4\}\). Given this arrangement, the notions of self-enforceability and coalition proofness presented in the modified Definition 2 (i.e., using (ii.a’) rather than (ii.a)) are readily applicable.

This example also allows us to demonstrate that our extension expands the sets of circumstances under which BPW’s self-enforceability and coalition-proofness concepts are applicable. First note that BPW’s CPN equilibrium concept is not readily applicable in this case. Rather, it is applicable to either a coalitional structure \(\{(1,2),\{3,4\}\}\) in which \(\{3,4\}\) is the complement of \(\{1,2\}\) or a coalitional structure \(\{(2,3),\{1,4\}\}\) in which \(\{1,4\}\) is the complement of \(\{2,3\}\).
In BPW, the coalitions cannot overlap. The trick that allows us to expand BPW’s equilibrium concept to the coalitional structure \(\{\{1,2\},\{2,3\},\{4\}\}\) is the statement that self-enforceability must hold for the coalition formed by the union of \(\{1,2\}\) and \(\{2,3\}\) (whenever these coalitions coexist); namely, the set \(\{1,2,3\}\). Underlying this trick is the requirement that all players that belong to a union of overlapped coalitions observe all deviations that occur within all overlapped coalitions. In BPW, all members of a coalition observe all deviations that occur within the coalition, but outsiders do not. In our framework, we also assume that a player who does not belong to a set of overlapped coalitions cannot observe deviations that occur within any of the overlapped coalitions. But, in our framework, a player \(i\) who is not a member of a coalition \(\{j,k\}\) observes (and reacts) to a deviation that occurs within this coalition provided that at least one of the members of this coalition - say, \(k\) - is a member of a coalition with \(i\), \(\{i,k\}\). Given this assumption, that deviations that occur within any overlapped coalition are common knowledge for all members of the union of overlapped coalitions, we can extend BPW’s CPN equilibrium concept to the case \(\{\{1,2\},\{2,3\},\{4\}\}\) - it becomes formally identical to the case \(\{\{1,2,3\},\{4\}\}\).

It is important to note that the notion of self-enforceability must hold for the set produced by the union of overlapping coalitions if deviations within any overlapped coalition are common knowledge for all members of the set of overlapped coalitions. Since a deviation within an overlapped coalition produces external benefits or costs for players who are indirectly connected, these players will generally find it desirable to react, rather than hold their actions fixed, when a deviation occurs. This implies that all players who belong to the set of overlapped coalitions can “…contemplate deviations from the deviation” (BPW, p.7). For
example, in the coalitional structure \(\{\{1,2\},\{2,3\},\{4\}\}\), suppose that player 1 deviates from \(\{1,2\}\) to form the singleton coalition \(\{1\}\). Having observed this deviation, player 3 may also contemplate deviating from \(\{2,3\}\) to form the singleton coalition \(\{3\}\). Knowing this, player 1 must also account for this possibility when it decides to deviate from \(\{1,2\}\).

For the same coalitional structure, consider the following alternative deviations. Player 2 decides to deviate from both \(\{1,2\}\) and \(\{2,3\}\) to form the singleton coalition \(\{2\}\). The deviations have the immediate effects of producing the singleton coalitions \(\{1\}\) and \(\{3\}\), with players 1 and 3 facing external benefits or costs from player 2’s actions. Having observed the actions of player 2, suppose that players 1 and 3 decide to form the two-member coalition \(\{1,3\}\), which did not exist in the original set of overlapped coalitions. Our framework captures this possibility: when the notion of self-enforceability holds for the union \(\{1,2,3\}\), all players consider the pros and cons of all possible deviations from \(\{1,2,3\}\), which include the formation of subgroups \(\{2\}\) and \(\{1,3\}\).

Since in what follows we consider perfect equilibria for multistage games, we must also extend BPW’s definitions for perfectly self-enforcing strategies and perfectly coalition-proof Nash equilibrium in order to allow for the coexistence of overlapping coalitions in dynamic games. As in BPW, let \(t\) denote the number of stages in an extensive form game. Then, the equilibrium concept is defined inductively as demonstrated below (see BPW, p. 10):

**Definition 3.** (i) In a single player, single stage game \(\Gamma\), \(s^* \in S\) is a *Perfectly Coalition-Proof Nash* equilibrium if and only if \(s^*\) maximizes \(u_i(s)\).
(ii) Let \((n, t) \neq (1, 1)\). Assume that Perfectly Coalition-Proof Nash equilibrium has been defined for all games with \(v\) players and \(\tau\) stages, where \((v, \tau) \leq (n, t)\), and \((v, \tau) \neq (n, t)\).

(a) For any game with \(n\) players and \(t\) stages, \(s^* \in S\) is perfectly self-enforcing if: (1) for all non-overlapping \(J_{(h, m)} \in J\), \(s^*_{J_{(h, m)}}\) is a Perfectly Coalition-Proof Nash equilibrium in the game \(\Gamma / s^*_{J_{(h, m)}}\); (2) for any collection of overlapping coalitions, \(J_{(l, m)}\), \(l = 1, \ldots, o\), \(s^*_{\bigcup_{l} (h, m)}\) is a Coalition-Proof Nash equilibrium in the game \(\Gamma / s^*_{\bigcup_{l} (h, m)}\); and (3) the restriction of \(s^*\) to any proper subgame forms a Perfectly Coalition-Proof Nash equilibrium in that subgame.

(b) For any game with \(n\) players and \(t\) stages, \(s^*\) is a Perfectly Coalition-Proof Nash equilibrium if it is perfectly self-enforcing, and if there does not exist another perfectly self-enforcing strategy vector \(s\) such that \(u_i(s) > u_i(s^*)\) for all \(i = 1, \ldots, n\).

3. **International Environmental Agreements for Correlated Pollutants**

Consider a world consisting of four identical nations and two regions, North America and Europe. Nations 1 and 2 are located in North America and nations 3 and 4 are located in Europe. Normalizing the population in each country to unity, the representative consumer’s utility in nation \(i\) is \(u_i = x_i + v(e_i) - a_i^2 - \sigma g^2\), where \(x_i\), \(e_i\), \(a_i\) and \(g\) are respectively quantities consumed of a numeraire good, energy, acidic deposition and a greenhouse gas. The consumer’s benefit from energy consumption is denoted \(v(e_i) \equiv be_i - ce_i^2\), where \(b > 1\) and \(c \geq 2.3028\).
These restrictions guarantee interior solutions in all maximization problems examined below. For computational ease, we assume that \( c = 3 \) henceforth.

We assume that one unit of energy consumption leads to one unit of sulfur dioxide emissions and one unit of emissions of the greenhouse gas. The sulfur emissions flow in both directions across the borders of the USA and Canada and of the two European nations. Let the parameter \( \alpha \) denote the fraction of nation \( j \)'s sulfur emissions that deposit in nation \( j \), \( j = 1, 2 \). The fraction of nation \( j \)'s sulfur emissions that travel to nation \( -j \) in the same region is hence \( 1 - \alpha \). Total sulfur depositions received by nation \( j \) are represented by the sum of its own sulfur depositions and sulfur spillovers from nation \( -j \) in the same region, i.e.,
\[
a_j = \alpha e_j + (1 - \alpha) e_{-j}, \quad j = 1, 2, \quad -j = 1 \text{ if } j = 2 \text{ and vice-versa.}
\]
Similarly, total sulfur depositions received by nation \( k \) is
\[
a_k = \alpha e_k + (1 - \alpha) e_{-k}, \quad k = 3, 4, \quad -k = 3 \text{ if } k = 4 \text{ and vice-versa.}
\]
For symmetry, we assume that \( \alpha = 1/2 \) and hence
\[
a_1 = a_2 = (e_1 + e_2)/2 \text{ and } a_3 = a_4 = (e_3 + e_4)/2.
\]
The global quantity of greenhouse gas emissions is
\[
g = \sum_{i=1}^{4} e_i.
\]

The parameter \( \sigma \in [0, 1] \) is a sensitivity index, which measures how sensitive each nation is to the damage caused by climate change. Given the quantities of the emissions generated, the parameter \( \sigma \) also measures the relative magnitude of acid rain and climate change damages. For example, in any symmetric equilibrium, \( e_i = a_i \equiv e \), \( i = 1, \ldots, 4 \), and \( g = 4e \). Hence, the damage caused by acid rain to nation \( i \) is \( a_i^2 = e^2 \). The damage caused by climate change to nation \( i \) is \( \sigma g^2 = 16\sigma e^2 \). Thus, if \( \sigma = 1/16 \), the damage caused by acid rain is equal to the damage caused by climate change to each nation. If \( \sigma = 1/32 \), the acid rain
damage is twice as harmful as the climate change damage.\textsuperscript{6} We can therefore also refer to $\sigma$ as the ‘damage-relativity index’. We shall show below that the value of $\sigma$ is very important for the efficiency properties of PCPN equilibria; in particular, we show that PSN equilibria exist for a sufficiently low $\sigma$ value.

Throughout, we consider IEAs in which the parties into an IEA agreement delegate authority to an international arbitrator to split the surplus of the coalition according to the Nash bargaining formula. The arbitrator implements intra-coalition transfers after coalition members make non-cooperative and simultaneous energy consumption choices. Thus, all the sequential games considered here fit squarely within the realm of non-cooperative games. Within any non-singleton coalition, each nation is free to choose its most desirable energy consumption – i.e., non-singleton coalition members are not restricted to commit to energy consumption levels that are most desirable from the collective perspective of the coalitions in which they belong. This non-cooperative approach – minimalist with respect to collective commitments – is both pragmatic and useful in light of ubiquitous failures of non-singleton coalition members to make credible collective commitments in IEAs.

We consider various games. First, we consider the non-cooperative game in which the nations behave as separate and individual entities. After this, we consider all the possible variations in which at least one bilateral coalition is in place. In such cases, the game involves two stages, one in which decisions concerning correlated emissions are made and another in which income transfer

\textsuperscript{6} Keeping uncertainties of monetizing the impacts of climate change and air pollution in mind, let us take the US and EU as examples to gain some idea of the relative magnitude of the damages. An analysis by Chestnut and Mills (2005) of the US Acid Rain Program established by Title IV of the 1990 Clean Air Act Amendments estimates annual benefits of the program in 2010 at $122 billion in 2000 US dollars. According to the Second Perspective Report by the US EPA (2011), the benefits of the 1990 Clean Air Act Amendments of reducing nitrogen oxides, sulfur dioxide, volatile organic compounds and fine particles were estimated to reach almost $2 trillion in 2006 US dollars for the year 2020. In Nordhaus and Boyer (2000), the annual damage of a 2.5°C warming to the US was estimated to be $28 billion in 1990 dollars. For the EU, the health damages in 2020 due to baseline (2009) pollution of nitrogen oxides, sulfur dioxide, ammonia, volatile organic compounds and fine particles were estimated by AEA Technology plc (2011) to have a median value of £474.112 billion using the value of a life year (VOLY) approach based on PPP-adjusted UNECE average values in 2005 prices. Nordhaus and Boyer (2000) estimated the annual damage of a 2.5°C warming to OECD Europe to be 2.83% of GDP.
decisions are made for those nations that belong to non-singleton coalitions. Within a coalition, we assume that the coalition members agree on utilizing the Nash bargaining formula to split the coalition’s surplus and on hiring an international arbitrator to implement intra-coalition income transfer.  

4. Relevant International Environmental Agreements (IEAs)  

We compute the subgame-perfect equilibrium for each of nine distinct types of coalitional structures which may be selected by the players during pre-game communication: structure 0, (\{1\},\{2\},\{3\},\{4\}); structure 1, (\{1,2\},\{3,4\}); structure 2, (\{1,2\},\{3\},\{4\}); structure 3, (\{1,3\},\{2\},\{4\}); structure 4, (\{1,4\},\{2\},\{3\}); structure 5, (\{1,3\},\{2,4\}); structure 6, (\{1,2,3\},\{4\}); structure 7, (\{1,2,3,4\}); and structure 8, (\{1,2\},\{2,3\},\{3,4\}). We will show that the subgame perfect equilibria for coalitional structures 7 and 8 are Pareto efficient while the subgame perfect equilibria for coalitional structures 0 – 6 are inefficient. Note that the union of the overlapping coalitions in coalitional structure 8 produces \{1,2,3,4\}; hence, the self-enforceability requirements for this coalitional structure are identical to those that hold for the Grand Coalition, in structure 7. Structures 0 – 6 represent all possible types of inefficient policy schemes; the inefficient policy schemes that are not included are “redundant” in the sense that each has a subgame perfect equilibrium that is identical to a subgame perfect equilibrium for one of the structures 0 – 6. Consider, for example, the partially overlapped coalitional structure (\{1,2\},\{2,3\},\{4\}). We will demonstrate in subsection 4.7 that the subgame perfect equilibrium for this structure is identical to the subgame perfect equilibrium for structure 6. Note also that the union of the overlapping coalitions in the structure (\{1,2\},\{2,3\},\{4\}) produces a list of interconnected

---

7 We assume that there is no agency problem in delegation. Coalition members delegate authority to an international arbitrator and can fully monitor this agent. To simplify matters, in most of what follows, we also assume that all coalition formation transactions – i.e., first and second stage actions, “writing” the agreement, hiring and monitoring the arbitrator, and implementing international transfers – are costless. We consider coalition operation costs for non-singleton coalitions in section 6.
members which corresponds to the list of directly connected members in the three-member coalition in structure 6; namely, $\{1,2,3\}$. This, in turn, implies that both types of coalitions are subject to the same self-enforceability requirements in the coalition formation stage.

In addition, as we will show in subsection 4.9, the fully overlapped coalitional structure ($\{1,2\},\{1,3\},\{1,4\}$) is redundant because it yields the same subgame perfect equilibrium as structures 7 and 8. It should also be noted that the union of the overlapping coalitions in ($\{1,2\},\{1,3\},\{1,4\}$) generates $\{1,2,3,4\}$, implying that this coalitional structure is subject to the same self-enforceability requirements as structures 7 and 8 in the coalition formation stage.

### 4.1. Structure 0: No IEA Formed

Let us start by examining the singleton coalitional structure. In this case, the game ends in the first stage. The budget constraint of the representative consumer in nation $i$ is $x_i + e_i = w_i$, $i = 1,\ldots,4$, where $w_i$ denotes nation $i$’s income. We assume that $w_i = w$ for all $i$. We can hence write nation $i$’s numeraire consumption as $x_i = w - e_i$, $i = 1,\ldots,4$.

Taking energy consumption choices of all other countries as given, nation $j$, $j = 1,2$, chooses $e_j \geq 0$ to maximize

$$w + (b-1)e_j - 3e_j^2 - \frac{(e_j + e_j)^2}{4} - \sigma(e_j + \sum_{i \neq j} e_i)$$

where $-j = 2$ if $j = 1$ and vice-versa. Taking energy consumption choices of all other countries as given, nation $k$, $k = 3,4$, chooses $e_k \geq 0$ to maximize

$$w + (b-1)e_k - 3e_k^2 - \frac{(e_k + e_k)^2}{4} - \sigma(e_k + \sum_{i \neq k} e_i)$$

where $-k = 4$ if $k = 3$ and vice versa. The first order conditions for an interior solution to the maximization of (1a) and (1b) respectively are

\[\sigma'(e_j + \sum_{i \neq j} e_i) + 2e_j = 0\]

\[\sigma'(e_k + \sum_{i \neq k} e_i) + 2e_k = 0\]
\[ b - 6e_j = 1 + \frac{(e_j + e_{-j})}{2} + 2\sigma \left( \sum_{i=1}^{4} e_i \right), \quad j = 1, 2, \quad (1c) \]

\[ b - 6e_k = 1 + \frac{(e_k + e_{-k})}{2} + 2\sigma \left( \sum_{i=1}^{4} e_i \right), \quad k = 3, 4. \quad (1d) \]

Equations (1c) and (1d) inform us that each nation equalizes the marginal benefit and marginal cost of energy consumption when deciding on the amount of energy consumption in the nation. Without participating in any IEA, a nation’s marginal cost of energy consumption includes the amount of numeraire consumption the nation gives up and the marginal damages the nation suffers from acid rain and climate change. The solution to equations (1c) and (1d) is

\[ e_{i}^{0*} = e_{i}^{\sigma*} = \frac{b - 1}{7 + 8\sigma}, \quad i = 1, \ldots, 4. \]

where we use the superscript \( z^{*} \) to denote the value of a variable realized in the non-cooperative equilibrium for structure \( z \), \( z = 0, \ldots, 8 \). This equilibrium yields the following payoffs:

\[ u_{i}^{0*} = u_{i}^{\sigma*} = w + \frac{(3 - 8\sigma)(b - 1)^2}{(7 + 8\sigma)^2}, \quad i = 1, \ldots, 4. \quad (1e) \]

### 4.2. Structure 1: Continental Agreements

Consider now the coalitional structure consisting of two international coalitions, one in each continent. In this case, the game has two stages. In the second stage, the international arbitrator within each IEA promotes international income transfers according to the Nash bargaining formula, utilizing \( u^{0} \) as the status quo utility. Let \( t_j \) denote the monetary transfer paid (if positive) or received (if negative) by nation \( j \) in North America, \( j = 1, 2 \). Similarly, let \( t_k \) denote the monetary transfer paid (if positive) or received (if negative) by nation \( k \) in Europe, \( k = 3, 4 \). We assume that \( t_j + t_{-j} = 0 \) and \( t_k + t_{-k} = 0 \). The budget
constraint of the representative consumer in nation \( j \) is \( x_j + e_j = w - t_j \). Similarly, the budget constraint of the representative consumer in nation \( k \) is \( x_k + e_k = w - t_k \). Hence, the utilities of the representative consumers in nations \( j \) and \( k \) are

\[
\hat{u}_j = w + (b-1)e_j - 3e_j^2 - \frac{(e_j + e_{-j})^2}{4} - \sigma \left(e_j + \sum_{i \neq j} e_i\right)^2 - t_j,
\]

(2a)

\[
\hat{u}_k = w + (b-1)e_k - 3e_k^2 - \frac{(e_k + e_{-k})^2}{4} - \sigma \left(e_k + \sum_{i \neq k} e_i\right)^2 - t_k.
\]

(2b)

The payoff for the international arbitrator in North America is \( \pi_{12}(\hat{u}_j, \hat{u}_k) \equiv (\hat{u}_j - u^0)(\hat{u}_k - u^0) \) while the payoff for the international arbitrator in Europe is \( \pi_{34}(\hat{u}_j, \hat{u}_k) \equiv (\hat{u}_j - u^0)(\hat{u}_k - u^0) \). Having observed the nations’ choices of energy consumption in the second stage of the game, the international arbitrator in North America chooses \( \{t_1, t_2\} \) to maximize \( \pi_{12}(\hat{u}_j, \hat{u}_k) \) subject to the income redistribution constraint \( t_1 + t_2 = 0 \). The first order conditions are

\[
(\hat{u}_j - u^0) - (\hat{u}_k - u^0) = 0 \quad \text{and} \quad t_1 + t_2 = 0.
\]

Thus, \( \hat{u}_j = \hat{u}_k \) and

\[
t_1(e_1, e_2) = \frac{e_1(b-1-3e_1) - e_2(b-1-3e_2)}{2} = -t_2(e_1, e_2).
\]

(2c)

Plugging equations (2c) into the national payoff functions (2a) implies that each nation in North America earns an average payoff equal to:

\[
w + \frac{1}{2} \left[ e_j (b-1-3e_j) + e_{-j} (b-1-3e_{-j}) - \frac{(e_j + e_{-j})^2}{2} - 2\sigma \left(e_j + \sum_{i \neq j} e_i\right)^2 \right].
\]

(2d)

In the first stage of the game, nation \( j \) chooses \( e_j \geq 0 \) to maximize (2d), taking every other nation’s choice of energy consumption as given. The first order condition for an interior solution is
\begin{equation}
\begin{split}
b - 6e_j = 1 + e_j + e_{-j} + 4\sigma \left( \sum_{i=1}^{4} e_i \right).
\end{split}
\tag{2e}
\end{equation}

Applying similar reasoning, the first order condition for an interior solution that governs the behavior of nation \( k \) in the first stage is
\begin{equation}
\begin{split}
b - 6e_k = 1 + e_k + e_{-k} + 4\sigma \left( \sum_{i=1}^{4} e_i \right).
\end{split}
\tag{2f}
\end{equation}

Equations (2e) and (2f) tell us that in this structure a nation chooses energy consumption so that the marginal benefit of energy consumption equals the marginal cost of energy consumption, which includes the amount of numeraire consumption the nation gives up and the marginal damages the two nations in the continental agreement suffer from acid rain and climate change.

The solution to equations (2e) and (2f) is
\begin{equation}
\begin{split}
e_i^* = \frac{b - 1}{8(1 + 2\sigma)}, \quad i = 1, \ldots, 4,
\end{split}
\end{equation}

The equilibrium payoffs for structure 1 are
\begin{equation}
\begin{split}
u_i^* = u^* = w + \frac{(b - 1)^2}{16(1 + 2\sigma)^2}, \quad i = 1, \ldots, 4.
\end{split}
\tag{2g}
\end{equation}

### 4.3. Structure 2: Continental Agreement in North America Only

In this coalitional structure, there is one continental environmental agreement in North America and no international agreement in Europe. In this structure, the behavior of the international arbitrator in North America is the same as in structure 1. Thus, we obtain equations (2c) and the payoff function (2d).

In the first stage of the game, nation \( j \), \( j = 1, 2 \), chooses \( e_j \geq 0 \) to maximize (2d), and nation \( k \), \( k = 3, 4 \), chooses \( e_k \geq 0 \) to maximize (1b), taking energy consumption choices of other countries as given. The first order conditions yield equations (1d) and (2e). We obtain the following equilibrium outcomes in structure 2:
\begin{align}
    e_j^{2*} &= \frac{(7 - 4\sigma)(b - 1)}{8(7 + 11\sigma)}, \quad j = 1, 2, \quad (3a) \\
    e_k^{2*} &= \frac{(2 + \sigma)(b - 1)}{2(7 + 11\sigma)}, \quad k = 3, 4, \quad (3b) \\
    u_j^{2*} &= w + \frac{(49 - 71\sigma - 104\sigma^2)(b - 1)^2}{16(7 + 11\sigma)^2}, \quad j = 1, 2, \quad (3c) \\
    u_k^{2*} &= w + \frac{3[16 + \sigma(24\sigma - 19)](b - 1)^2}{16(7 + 11\sigma)^2}, \quad k = 3, 4. \quad (3d)
\end{align}

4.4. Structure 3: Inter-Continental Agreement Between Nations 1 and 3

In this coalitional structure, there is only one bilateral intercontinental agreement formed by nations 1 and 3. In this case, the budget constraints of the representative consumers in nations 1 and 3 can respectively be written as

\[ x_1 + e_1 = w - t_1 \quad \text{and} \quad x_3 + e_3 = w - t_3, \]

which allow us to write numeraire consumption in these two nations as \( x_1 = w - e_1 - t_1 \) and \( x_3 = w - e_3 - t_3 \). The payoff for the international arbitrator in this agreement is

\[ \pi_{13}(\hat{u}_1^3, \hat{u}_3^3) = (\hat{u}_1^3 - u_0^0)(\hat{u}_3^3 - u_0^0). \]

Thus, in the second stage, the arbitrator implements transfers that satisfy \( \hat{u}_1^3 = \hat{u}_3^3 \) and \( t_1 + t_3 = 0 \). These conditions imply that

\[ t_1(e_1, e_2, e_3, e_4) = \frac{1}{2} \left[ e_1(b - 1 - 3e_1) - \frac{(e_1 + e_2)^2}{4} \right] - \left[ e_3(b - 1 - 3e_3) - \frac{(e_3 + e_4)^2}{4} \right] \]

and \( t_3(e_1, e_2, e_3, e_4) = -t_1(e_1, e_2, e_3, e_4) \). Plugging these transfer functions into the national payoff functions yield the following average payoff functions

\[ w + \frac{1}{2} \left[ e_1(b - 1 - 3e_1) + e_3(b - 1 - 3e_3) - \frac{(e_1 + e_2)^2 + (e_3 + e_4)^2}{4} - 2\sigma \left( \sum_{i=1}^{4} e_i \right)^2 \right]. \quad (4a) \]
In the first stage, nations 1 and 3 non-cooperatively choose energy consumption levels to maximize (4a). Nation 2 chooses energy consumption in order to maximize (1a) (for \( j = 2 \)) and nation 4 chooses energy consumption in order to maximize (1b) (for \( k = 4 \)). All choices are simultaneous and each nation takes the choices of all other nations as given. The first order conditions for the problems solved by nations 2 and 4 are identical to conditions (1c) – for \( j = 2 \) – and (1d) – for \( k = 4 \) – respectively. The first order conditions for the problems solved by nations 1 and 3 are

\[
b - 6e_1 = 1 + \frac{e_1 + e_2}{2} + 4\sigma \left( \sum_{i=1}^{4} e_i \right), \quad (4b)
\]

\[
b - 6e_3 = 1 + \frac{e_3 + e_4}{2} + 4\sigma \left( \sum_{i=1}^{4} e_i \right). \quad (4c)
\]

The solution to the system of equations (4b)-(4c) is

\[
e^*_{3h} = \frac{(3-2\sigma)(b-1)}{3(7+12\sigma)}, \quad h = 1, 3; \quad \text{and} \quad e^*_{4l} = \frac{(3+2\sigma)(b-1)}{3(7+12\sigma)}, \quad l = 2, 4.
\]

The subgame perfect equilibrium for this structure yields the following payoffs:

\[
u^*_{3h} = w + \frac{(9 - 14\sigma - 28\sigma^2)(b-1)^2}{3(7+12\sigma)^2}, \quad h = 1, 3, \quad (4d)
\]

\[
u^*_{4l} = w + \frac{(9 - 10\sigma + 20\sigma^2)(b-1)^2}{3(7+12\sigma)^2}, \quad l = 2, 4. \quad (4e)
\]

### 4.5. Structure 4: Inter-Continental Agreement Between Nations 1 and 4

This coalitional structure is identical in all respects to coalitional structure 3 except that the inter-continental agreement is for nations 1 and 4 rather than for nations 1 and 3. In the second stage, the international arbitrator implements transfers that satisfy \( \hat{u}^4_1 = \hat{u}^4_4 \) and \( t_1 + t_4 = 0 \). These conditions imply that
and \( t_4(e_1,e_2,e_3,e_4) = -t_1(e_1,e_2,e_3,e_4) \). Plugging these transfer functions into the national payoff functions yield the following average payoff functions

\[
w + \frac{1}{2} e_1(b - 1 - 3e_1) + e_4(b - 1 - 3e_4) - \frac{(e_1 + e_2)^2 + (e_3 + e_4)^2}{4} - 2\sigma \left( \sum_{i=1}^{4} e_i \right)^2 \]  

(5a)

In the first stage, nations 1 and 4 non-cooperatively choose energy consumption levels to maximize (5a). Nation 2 chooses energy consumption in order to maximize (1a) (for \( j = 2 \)) and nation 3 chooses energy consumption in order to maximize (1b) (for \( k = 3 \)). All choices are simultaneous and each nation takes the choices of all other nations as given. The first order conditions for the problems solved by nations 2 and 3 are identical to conditions (1c) – for \( j = 2 \) – and (1d) – for \( k = 3 \) – respectively. The first order conditions for the problems solved by nations 1 and 4 are (4b) and

\[
b - 6e_4 = 1 + \frac{e_3 + e_4}{2} + 4\sigma \left( \sum_{i=1}^{4} e_i \right) \]  

(5b)

The solution to the system of equations (1c), (1d), (4b) and (5b) is

\[
e^{*}_h = \frac{(3 - 2\sigma)(b-1)}{3(7 + 12\sigma)}, \quad h = 1, 4, \]  

\[
e^{*}_l = \frac{(3 + 2\sigma)(b-1)}{3(7 + 12\sigma)}, \quad l = 2, 3. \]  

The subgame perfect equilibrium for this structure has the following payoffs:

\[
u^{*}_h = w + \frac{(9 - 14\sigma - 28\sigma^2)(b-1)^2}{3(7 + 12\sigma)^2}, \quad h = 1, 4, \]  

(5c)

\[
u^{*}_l = w + \frac{(9 - 10\sigma + 20\sigma^2)(b-1)^2}{3(7 + 12\sigma)^2}, \quad l = 2, 3. \]  

(5d)
4.6. Structure 5: Two Bilateral Inter-Continental Agreements.

In this coalitional structure, the behaviors of nations 1 and 3 are identical to the ones we examined in structure 3. Thus, we obtain equations (4b) and (4c). The transfers promoted within coalition \{2,4\} by its international arbitrator imply that nations 2 and 4 make energy consumption choices to maximize (6a):

\[
w + \frac{1}{2} \left[ e_2 (b - 1 - 3e_2) + e_4 (b - 1 - 3e_4) - \frac{(e_1 + e_2)^2 + (e_3 + e_4)^2}{4} - 2\sigma \left( \sum_{i=1}^{4} e_i \right)^2 \right]. \tag{6a}
\]

The first order conditions are

\[
b - 6e_2 = 1 + \frac{e_1 + e_2}{2} + 4\sigma \left( \sum_{i=1}^{4} e_i \right), \tag{6b}
\]

\[
b - 6e_4 = 1 + \frac{e_3 + e_4}{2} + 4\sigma \left( \sum_{i=1}^{4} e_i \right). \tag{6c}
\]

The solution to the system of equations (6b)-(6c) is

\[
e_i^{5^*} = e_i^{5^*} = \frac{b - 1}{7 + 16\sigma}, \quad i = 1,\ldots, 4.
\]

The subgame perfect equilibrium payoffs in this structure are

\[
u_i^{5^*} = u_i^{5^*} = w + \frac{3(b - 1)^2}{(7 + 16\sigma)^2}, \quad i = 1,\ldots, 4. \tag{6d}
\]

4.7. Structure 6: Trilateral Inter-Continental Agreement

Consider now the coalitional structure in which nation 4 is a singleton and nations 1, 2 and 3 belong to a trilateral intercontinental agreement. Let \(t_l\) denote the monetary transfer paid (if positive) or received (if negative) by nation \(l\), \(l = 1,2,3\). Since transfers are redistributive, \(t_1 + t_2 + t_3 = 0\). The budget constraint of the representative consumer in nation \(l\) yields \(x_i = w - e_i - t_i\). The payoff for the international arbitrator is \(\pi_{123}(\hat{u}_1^6, \hat{u}_2^6, \hat{u}_3^6) = \Pi_{i=1}^3 (\hat{u}_i^6 - u_i^0)\), where
\[ \hat{u}_l^* = w + e_l \left( b - 1 - 3e_l \right) - a_l^2 - \sigma g^2 - t_l \quad \text{and} \quad a_l = \frac{e_l + e_2}{2} \quad \text{if} \quad l = 1, 2 \quad \text{and} \quad a_l = e_3 + e_4 \quad \text{if} \quad l = 3. \]

In the second stage of the game, the international arbitrator chooses \( \{t_1, t_2, t_3\} \) to maximize \( \pi_{123} (.) \) subject to \( t_1 + t_2 + t_3 = 0 \). The solution is given by \( u_1^* = \hat{u}_2^* = \hat{u}_3^* \) and \( t_1 + t_2 + t_3 = 0 \).

The second stage equations imply that each nation \( l \) in the first stage chooses \( e_l \geq 0 \) to maximize the following average payoff function:

\[
\hat{u}_l^* = w + \frac{1}{3} \left[ (b - 1) \left( \sum_{i=1}^{3} e_i \right) - 3 \left( \sum_{i=1}^{3} e_i^2 \right) - \frac{(e_1 + e_2)^2}{2} - \frac{(e_3 + e_4)^2}{4} - 3\sigma \left( \sum_{i=1}^{4} e_i \right)^2 \right]. \tag{7a}
\]

Nation 4 chooses \( e_4 \geq 0 \) to maximize (1b) (for \( k = 4 \)). All choices are non-cooperative and simultaneous. Each nation makes its choice taking the choices of all other nations as given. The first order conditions are (1d) (for \( k = 4 \)) and

\[
b - 6e_j = 1 + e_1 + e_2 + 6\sigma \left( \sum_{i=1}^{4} e_i \right), \quad j = 1, 2, \tag{7b}
\]

\[
b - 6e_3 = 1 + \frac{e_3 + e_4}{2} + 6\sigma \left( \sum_{i=1}^{4} e_i \right). \tag{7c}
\]

The solution to the system of equations (1d), (7b) and (7c) is

\[
e_j^{6*} = \frac{7 - 4\sigma - (b - 1)}{4(14 + 37\sigma)}, \quad j = 1, 2; \quad e_3^{6*} = \frac{4 - 3\sigma - (b - 1)}{2(14 + 37\sigma)}; \quad e_4^{6*} = \frac{4 + 7\sigma - (b - 1)}{2(14 + 37\sigma)}.
\]

The equilibrium payoffs are as follows:

\[
u_l^{6*} = w + \frac{(146 + 111\sigma - 581\sigma^2)(b - 1)^2}{12(14 + 37\sigma)^2}, \quad l = 1, 2, 3, \tag{7d}
\]

\[
u_4^{6*} = w + \frac{(48 + 83\sigma + 367\sigma^2)(b - 1)^2}{4(14 + 37\sigma)^2}. \tag{7e}
\]

Now consider the partially overlapped coalitional structure (\{1,2\}, \{2,3\}, \{4\}). In this case, there is a continental agreement formed by nations 1 and 2 and a
trans-continental agreement formed by nations 2 and 3. Our previous analysis of intra-coalition income transfers enables us to say that the income transfers promoted by the arbitrator in the continental agreement will lead to $\hat{u}_1 = \hat{u}_2$ and the income transfers promoted by the arbitrator in the trans-continental agreement will lead to $\hat{u}_2 = \hat{u}_3$. Thus, knowing that $\hat{u}_1 = \hat{u}_2 = \hat{u}_3$, nations 1, 2 and 3 will motivated to maximize the average payoff function of the three nations described by expression (7a). Since the behavior of nation 4 will be the same as in the coalitional structure ({1,2,3},{4}), the subgame perfect equilibrium for the partially overlapped coalitional structure ({1,2},{2,3},{4}) will be identical to the subgame perfect equilibrium for the non-overlapped coalitional structure ({1,2,3},{4}).

4.8. Structure 7: the Grand Coalition

If the Grand Coalition (GC) is formed, the budget constraint of the representative consumer in nation $i$ is $x_i + e_i = w - t_i$, $i = 1,...,4$, where $t_i$ denotes the monetary transfer paid (if positive) or received (if negative) by nation $i$ under the GC and $\sum_{i=1}^{4} t_i = 0$. We write numeraire consumption in nation $i$ as

$$x_i = w - t_i - e_i, \ i = 1,...,4.$$ 

The payoff for the international arbitrator is $\pi_{1234}(\hat{u}_1^7, \hat{u}_2^7, \hat{u}_3^7, \hat{u}_4^7) = \prod_{i=1}^{4} (\hat{u}_i^7 - u^0)$

where $\hat{u}_j^7 = w + e_j (b - 1 - 3e_j) - \frac{(e_1 + e_2)^2}{4} - \sigma g^2 - t_j, \quad j = 1,2, \quad$ and $\hat{u}_k^7 = w + e_k (b - 1 - 3e_k) - \frac{(e_3 + e_4)^2}{4} - \sigma g^2 - t_k, \quad k = 3,4$. The arbitrator’s choices yield $\sum_{i=1}^{4} t_i = 0$ and $\hat{u}_j^7 = \hat{u}_k^7$, for $j = 1,2$ and $k = 3,4$. 

29
The second stage optimal equations imply that each nation’s payoff function in the first stage corresponds to the following average payoff function:

\[
\begin{equation}
w + \frac{1}{4} \left[ (b - 1) \left( \sum_{i=1}^{4} e_i \right) - 3 \left( \sum_{i=1}^{4} e_i^2 \right) \right] - \frac{(e_1 + e_2)^2 + (e_3 + e_4)^2}{2} - 4\sigma \left( \sum_{i=1}^{4} e_i^2 \right)^2. \tag{8a}
\end{equation}
\]

In the first stage, nation \( h, h = 1, \ldots, 4 \), chooses \( e_h \geq 0 \) to maximize (8a), taking the choices of all other nations as fixed. The first order conditions are

\[
\begin{align*}
& b - 6e_j = 1 + e_1 + e_2 + 8\sigma \left( \sum_{i=1}^{4} e_i \right), \quad j = 1, 2, \tag{8b} \\
& b - 6e_k = 1 + e_1 + e_2 + 8\sigma \left( \sum_{i=1}^{4} e_i \right), \quad k = 3, 4. \tag{8c}
\end{align*}
\]

Equations (8b) and (8c) inform us that all nations find it desirable to internalize all continental (i.e., acid rain) and global (i.e., climate change) externalities. The solution to the system of equations (8b) – (8c) is

\[
e_i^{7*} = e_i^{7*} = \frac{b - 1}{8(1 + 4\sigma)} \quad i = 1, \ldots, 4.
\]

The equilibrium payoffs are

\[
u_i^{7*} = u_i^{7*} = w + \frac{(b - 1)^2}{16(1 + 4\sigma)}. \tag{8d}
\]

4.9. **Structure 8: Fully Overlapped Environmental Agreements**

Consider now the following fully overlapped coalitional structure: \( \{1,2\}, \{2,3\}, \{3,4\} \). There are: (i) two continental agreements, one in North America (i.e., \( \{1,2\} \)) and the other in Europe (i.e., \( \{3,4\} \)); and (ii) one inter-continental agreement (i.e., \( \{2,3\} \)). The payoff of the North American arbitrator is

\[
\pi_{12}(\hat{u}_1^8, \hat{u}_2^8) = (\hat{u}_1^8 - u^0)(\hat{u}_2^8 - u^0),
\]

the payoff of the European arbitrator is

\[
\pi_{34}(\hat{u}_3^8, \hat{u}_4^8) = (\hat{u}_3^8 - u^0)(\hat{u}_4^8 - u^0)
\]

and the payoff of the inter-continental arbitrator is

\[
\pi_{23}(\hat{u}_2^8, \hat{u}_3^8) = (\hat{u}_2^8 - u^0)(\hat{u}_3^8 - u^0).
\]
As in structure 1, \( t_j \) denotes the monetary transfer paid (if positive) or received (if negative) by nation \( j \) in North America, \( j = 1, 2 \), and \( t_k \) denotes the monetary transfer paid (if positive) or received (if negative) by nation \( k \) in Europe, \( k = 3, 4 \). We also let \( t_{23} \) denote the transfer paid (if positive) or received (if negative) by nation 2 in the inter-continental agreement. The corresponding transfer for nation 3 is \( t_{32} \). The amounts of numeraire good consumed are

\[
x_h = w - e_h - t_h, \quad h = 1, 4, \quad x_2 = w - t_2 - t_{23} \quad \text{and} \quad x_3 = w - t_3 - t_{32}.
\]

In the second stage, the three arbitrators play a non-cooperative and simultaneous game. Each arbitrator chooses intra-coalition transfers taking the choices of the other arbitrators as given. Hence, the optimal solutions in the second stage satisfy \( \hat{u}_1^8 = \hat{u}_2^8 = \hat{u}_3^8 = \hat{u}_4^8 \) and \( t_1 + t_2 = t_3 + t_4 = t_{23} + t_{32} = 0 \). These equations, in turn, imply that each nation’s payoff function in the first stage corresponds to the average payoff function (8a). Hence, the first order conditions for the first stage yield equations (8b) and (8c). In sum, the subgame perfect equilibrium for the fully overlapped coalitional structure \( \{1,2\}, \{2,3\}, \{3,4\} \) corresponds to the subgame perfect equilibrium for the coalitional structure consisting of the coalition of the whole.

Proposition 1 summarizes our first set of important results:

**Proposition 1.** The subgame-perfect equilibrium for the GC is Pareto efficient. The subgame perfect equilibrium for the fully overlapped coalitional structure \( \{1,2\}, \{2,3\}, \{3,4\} \) is also Pareto efficient. Hence, these two subgame perfect equilibria are identical.

Suppose now that during pre-play communication the players select the fully overlapped coalitional structure \( \{1,2\}, \{1,3\}, \{1,4\} \). In the second stage of the induced game, the three international arbitrators make intra-coalition transfers that equate coalition payoffs for the nations. In the first stage, each nation chooses
energy consumption anticipating that national payoffs are equalized; hence, they make choices that internalize intra- and inter-coalition environmental externalities, yielding an allocation identical to the allocation produced by the Grand Coalition. Since the same argument holds for any self-enforcing fully overlapped coalitional structure, we may write the following proposition:

**Proposition 2.** The subgame perfect equilibrium for any fully overlapped coalitional structure is Pareto efficient. Hence, the subgame perfect equilibrium for any fully overlapped coalitional structure is identical to the subgame perfect equilibrium for the GC.

Silva and Zhu (2011) also obtain the equivalence result described in Proposition 2. As in this earlier paper, the reason for the equivalence result is the fact that the nations anticipate that the income transfers implemented within the overlapped coalitions will perfectly align their objectives. Thus, it is individually rational to make the energy consumption choice that internalizes intra- and inter-coalition externalities. Unlike Silva and Zhu (2011), we examine coalition formation below and select the stable coalitional structures utilizing our adaptation to BPW's equilibrium concepts defined in section 2.

5. **Perfectly Coalition-Proof Nash Equilibrium**

All players fully anticipate all outcomes described above and can select one of the (perfect) Nash equilibrium during pre-game communication. Suppose initially that $\sigma = 0$. In this case, the nations do not experience damages from climate change; they only face acid rain damages. Since acid rain is a continental phenomenon, a coalitional structure consisting of continental agreements is a natural candidate for a self-enforcing coalitional structure. This is indeed one of the messages of Proposition 3 – it also states that the GC and all fully overlapped coalitional structures are self-enforcing. Since the perfect equilibria for all these self-enforcing coalitional structures are Pareto efficient, we can write:
**Proposition 3.** If $\sigma = 0$, all PCPN equilibria are PSN equilibria. The set of PSN equilibria contains the subgame perfect equilibria for: (i) $\{1,2\}, \{3,4\}$; (ii) the GC; and (iii) all fully overlapped coalitional structures.

**Proof.** Suppose that $\sigma = 0$. Close inspection of the conditions that characterize the subgame perfect equilibrium for $\{1,2\}, \{3,4\}$ reveals that it corresponds to the subgame perfect equilibrium for the GC. In addition, the combination of results (1e), (2g), (3c), (3d), (4d), (4e), (5c), (6d), (7d), (7e) and (8d) yields: (i) $u^* = u^* > \max \{u^0, u^1, u^2, u^3, u^4, u^5, u^6\}$, $\forall i$; and (ii) $u^v = u^v = u^v > u^v = u^v$.

These conditions imply that the GC and $\{1,2\}, \{3,4\}$ are self-enforcing coalitional structures. Utilizing Proposition 1, we conclude that the subgame perfect equilibrium for the GC is a PSN equilibrium. This, in turn, implies that the subgame perfect equilibrium for $\{1,2\}, \{3,4\}$ is also a PSN equilibrium. Since the self-enforceability requirements for any fully overlapped coalitional structure are identical to the self-enforceability requirements for the GC, all fully overlapped coalitional structures are self-enforcing. From Proposition 2, it follows that the subgame perfect equilibrium for any fully overlapped coalitional structure is a PSN equilibrium.

Note that if $\sigma = 0$ the set of PCPN equilibria coincides with the set of PSN equilibria because the subgame perfect equilibria for all other coalitional structures are Pareto inefficient and thus cannot be perfectly coalition-proof. For $\sigma \in (0,1]$, the combination of results (1e), (2g), (3c), (3d), (4d), (4e), (5c), (5d), (6d), (7d), (7e) and (8d) yields:

$$u^* > \max \left\{u^0, u^1, u^2, u^3, u^4, u^5, u^6\right\}, \quad (9a)$$

$$u^* > \begin{cases} u^* = u^* = u^* \quad & \text{if } \sigma < (\geq) 0.691778, \\ u^* > (\leq) u^* \quad & \text{if } \sigma < (\geq) 0.0336877. \end{cases} \quad (9b, 9c)$$
Considering conditions (9a) – (9c) together, we conclude that the subgame perfect equilibrium for the GC is perfectly coalition-proof for $\sigma \in (0,0.0336877]$. This, in turn, implies that the subgame perfect equilibria for all fully overlapped coalitional structures are also perfectly coalition-proof for $\sigma \in (0,0.0336877]$. From Proposition 3, we also know that the subgame perfect equilibria for the GC and all fully overlapped coalitional structures are perfectly coalition-proof if $\sigma = 0$. Since these PCPN are Pareto efficient, they are PSN equilibria for $\sigma \in [0,0.0336877]$. These findings are summarized in the proposition below.

**Proposition 4.** If $\sigma \in [0,0.0336877]$, the subgame perfect equilibria for the GC and all fully overlapped coalitional structures are PSN equilibria.

For $\sigma > 0.0336877$, the subgame perfect equilibria for the GC and all fully overlapped coalitional structures are not perfectly coalition-proof because they are not self-enforcing: one nation – say, nation 4 – has an incentive to defect from the GC or from any fully overlapped coalitional structure in order to form a singleton coalition, holding the actions of all other nations fixed. However, as demonstrated by the following comparison of equilibria payoffs, the subgame perfect equilibrium for the coalitional structure $\{1,2,3\}\cup\{4\}$ is not perfectly coalition-proof because it is not self-enforcing: holding the actions of $\{4\}$ fixed, nations 1 and 2 find it mutually beneficial to defect from $\{1,2,3\}$ in order to form the North American continental coalition $\{1,2\}$:

$$u_1^{2*} = u_2^{2*} > u_1^{6*} = u_2^{6*} = u_3^{6*} \quad \text{for} \quad \sigma \in [0,1], \quad (10a)$$

$$u_1^{2*} = u_2^{2*} > (\leq)u^0 \quad \text{if} \quad \sigma < (\geq) 0.8273166. \quad (10b)$$

Condition (10a) follows from comparing (3c) and (7d). Conditions (10b) follow from comparing (1e) and (3c). Together, conditions (10a) and (10b) inform us that nations 1 and 2 have incentives to deviate from $\{1,2,3\}$ holding the actions of $\{4\}$ fixed: (i) for $\sigma \leq 0.8273166$, nations 1 and 2 can foresee that it is feasible and
mutually rational to produce the coalitional structure \(\{1,2\},\{3\},\{4\}\) and that the sub-coalition \(\{1,2\}\) is self-enforcing; (ii) for \(\sigma > 0.8273166\), nations 1 and 2 find it advantageous to deviate from \(\{1,2,3\}\) in order to produce the singletons \(\{1\}\), \(\{2\}\) and \(\{3\}\). Conditions (10a) and (10b) also imply that subgame perfect equilibria for partially overlapped coalitional structure are not self-enforcing.

Let us now consider the self-enforceability and coalition-proofness properties of all other subgame perfect equilibria for non-overlapped coalitional structures for \(\sigma\) in \((0.0336877,1]\). Results (1e), (2g), (3c), (3d), (4d), (4e), (5c), (5d) and (6d) yield

\[u^* > \max\{u^0, u^1, u^3, u^4, u^5\}\text{, for } \sigma \in [0,1], \quad (10c)\]

\[u^* > u^2 = u^3\text{ for } \sigma \in (0,1], \quad (10d)\]

\[u^* > (\leq) u^2 = u^3\text{ if } \sigma < (\geq) 0.55603598, \quad (10e)\]

\[u^* > (\leq) u^2 = u^3 = u^4 = u^*\text{ if } \sigma < (\geq) 0.380138355, \quad (10f)\]

\[u^2 = u^3 = u^4 = u^* > u^5\text{ for } \sigma \in (0,1], \quad (10g)\]

\[u^1 = u^3 = u^4 = u^* > u^0\text{ if } 0.0405 < \sigma < 0.294098, \quad (10h)\]

\[u^1 = u^3 = u^4 = u^* = u^0\text{ if } \sigma = 0.0405 \text{ or } \sigma = 0.294098, \quad (10i)\]

\[u^1 = u^3 = u^4 = u^* < u^0\text{ if } \sigma < 0.0405 \text{ or } \sigma > 0.294098. \quad (10j)\]

We start by claiming that the subgame perfect equilibrium for the continental structure, \((1,2),\{3,4\}\) is the PCPN equilibrium for \(\sigma \in (0.0336877,0.55603598]\). First, note that conditions (10d) and (10e) reveal that the continental coalitional structure is self enforcing for \(\sigma \leq 0.55603598\). In other words, coalitional structure 2 does not emerge. It remains to show that there does not exist another self-enforcing coalitional structure that Pareto dominates it for \(\sigma \in (0.0336877,0.55603598]\). Clearly, the Nash equilibrium for coalitional structure 0, \((\{1\},\{2\},\{3\},\{4\})\), and subgame perfect equilibrium for coalition
structure 5, \( \{1,3\}, \{2,4\} \), are Pareto dominated by the subgame-perfect equilibrium for the continental structure – see (10c). Coalitional structures 3 and 4, \( \{1,3\}, \{2\}, \{4\} \) and \( \{1,4\}, \{2\}, \{3\} \), respectively, are self-enforcing for \( \sigma \in [0.0405, 0.294098] \) – see (10h) – (10j). However, the subgame perfect equilibria for these coalitional structures cannot be perfectly coalition-proof for \( \sigma \in [0.0405, 0.294098] \) because they are Pareto dominated by the subgame perfect equilibrium for the continental coalitional structure for \( \sigma < 0.380138355 \) – see (10c) and (10f). Hence, the subgame perfect equilibrium for the continental coalitional structure is the PCPN equilibrium for \( \sigma \in (0.0336877, 0.55603598) \).

Suppose now that \( \sigma > 0.55603598 \). Conditions (10b) and (10e) reveal that coalitional structure 2 is self-enforcing for \( \sigma \in (0.55603598, 0.8273166) \). Since there is no other self-enforcing coalitional structure for \( \sigma \) values in such an interval, the subgame perfect equilibrium for coalitional structure 2 is the PCPN equilibrium for \( \sigma \in (0.55603598, 0.8273166) \).

Finally, suppose that \( \sigma > 0.8273166 \). For \( \sigma \in (0.8273166, 1] \), coalitional structure 2 is not self-enforcing, since nations 1 and 2 benefit from deviating from \( \{1,2\} \), holding the actions of nations 3 and 4 fixed. The deviation produces the singleton coalitional structure. Note that even though the subgame perfect equilibria for coalitional structures 1, 5, 7 and 8 Pareto dominate the Nash equilibrium for the singleton coalitional structure, they are not self-enforceable. In addition, the subgame perfect equilibria for coalitional structures 3 and 4 are not self-enforceable because they do not Pareto dominate the Nash equilibrium for the singleton coalitional structure. Therefore, the Nash equilibrium for the singleton coalitional structure is coalition-proof for \( \sigma \in (0.8273166, 1] \).
Combining all results of this section, we can write the following proposition:

**Proposition 5.** If $\sigma \in [0, 0.0336877]$, the subgame perfect equilibria for the GC and all fully overlapped coalitional structures are PSN equilibria. If $\sigma \in (0.0336877, 0.55603598]$, the subgame perfect equilibrium for the continental coalitional structure is the PCPN equilibrium. If $\sigma \in (0.55603598, 0.8273166]$, the subgame perfect equilibrium for coalitional structure 2 is the PCPN equilibrium. Finally, if $\sigma \in (0.8273166, 1]$, the Nash equilibrium for the singleton coalitional structure is coalition-proof.

If we postulate that all nations use PSN as their selection criterion, we see that only the GC and fully overlapped coalitional structures would be selected for sufficiently small $\sigma$. It must be noted that if the national damage caused by acid rain is at least twice as large as the national damage caused by climate change, PSN equilibria do exist. The problem of using PSN as the Nash refinement occurs for $\sigma > 0.0336877$, since in such circumstances there does not exist a PSN equilibrium. But, as demonstrated by Proposition 5, PCPN equilibria feature at least one continental agreement for $\sigma \leq 0.8273166$, which is quite reassuring. It is particularly important to note that for $\sigma \in (0.0336877, 0.55603598]$ the subgame perfect equilibrium for the continental coalitional structure, (\{1,2\},\{3,4\}), is perfectly coalition-proof. Since regional (e.g., continental) pollution damages appear to be quite important relative to climate change damages, a key message from our analysis is that policy makers should pay closer attention to the overall benefits produced by regional or continental environmental agreements. They may be stable second-best alternatives to fully participatory global environmental agreements.

### 6. Coalition Operation Costs: Dominant Fully Overlapped Structures

We have demonstrated that the subgame perfect equilibrium for any fully overlapped coalitional structure is a PSN equilibrium for a sufficiently low
relativity-damage index. However, we also showed that any of these equilibria is isomorphic to the subgame perfect equilibrium for the GC. In this section, we demonstrate that the subgame perfect equilibrium for a fully overlapped coalitional structure may be superior to the subgame perfect equilibrium for the GC in the presence of coalition operation costs.

We consider a setting in which the operation cost faced by any non-singleton coalition rises with the coalition’s membership size at an increasing rate.\(^8\) This is a common assumption used in club theory.\(^9\) As in club theory, the operation cost of a non-singleton international environmental coalition may be increasing and strictly convex in membership size due to the various types of activities that such an organization carries out; e.g., the scheduling and organization of periodic meetings, policy agenda coordination, development of synergistic R&D, income and technological transfers, etc. For a realistic appraisal of the gamma of activities that such an organization may carry out, see the activities that the Global Environment Facility execute in its mandate as the body responsible for the implementation of the financial mechanisms for the United Nations Framework Convention on Climate Change (UNFCCC) and the United Nations Convention to Combat Desertification (UNCCD), among others.\(^10\)

For concreteness, suppose that the administrative cost incurred by the international arbitrator of a coalition of size \(m\), \(m \in \{2, 3, 4\}\), is \(\chi m^2\), where \(\chi > 0\) is a fixed technological parameter. Allocating this cost equally among coalition members implies that the administrative cost paid by each member is \(\chi m\).

---

\(^8\) For simplicity, we assume that the operation cost of a singleton coalition is zero.

\(^9\) See, e.g., Cornes and Sandler (1996) for a comprehensive description of club theory and review of the literature.

It seems logical and realistic to postulate that a non-singleton coalition’s administrative cost is small relative to the benefits that a coalition produces in terms of (partial or total) internalization of environmental externalities. We formalize this statement with the requirement that, for some \( \sigma \in (0, 0.0336877) \), administrative costs are not sufficiently large to change the rankings (9a) – (9c).

Given \( \sigma \), the payoff for each nation in the subgame perfect equilibrium for the GC is now \( u^*(\sigma) - 4\chi \). We wish to restrict the value that \( \chi \) can take, given \( \sigma \in (0, 0.0336877) \), so that \( u^*(\sigma) - 4\chi \geq u_i^*(\sigma) \).

Let \( \chi \equiv \frac{[u^*(\sigma) - u_i^*(\sigma)]}{4} \). To make things clear, we shall fix \( \sigma \).

Letting \( \sigma = 1/32 = 0.03125 \), we have \( \chi = 0.000025832(b-1)^2 \). It can now be easily verified that the following inequalities hold for \( \sigma \in (0, 0.03125) \):

\[
\begin{align*}
& u^* - 4\chi > \max\{u_1^*, u_2^*, u_3^*, u_4^*, u_5^*, u_6^*, u_7^*\} - 2\chi, \quad (11a) \\
& u^* - 4\chi > u_l^*(\sigma) - 3\chi, \quad l = 1, 2, 3, \quad (11b) \\
& u^* - 4\chi > \max\{u_0^*, u_5^*, u_4^*, u_2^*, u_3^*, u_4^*, u_5^*\}. \quad (11c)
\end{align*}
\]

Hence, the rankings (9a) – (9c) hold for \( \chi = 0.000025832(b-1)^2 \) and \( \sigma \in (0, 0.03125) \).

Letting \( u^* \) denote the payoff that each nation gets in the subgame perfect equilibrium for coalitional structure \( \{\{1,2\}, \{2,3\}, \{3,4\}\} \) in the absence of administrative cost, it is straightforward to check that the payoff each nation gets in the subgame perfect equilibrium for the same structure in the presence of administrative cost is \( u^* - 3\chi \). Since \( u^* = u^* \), the subgame perfect equilibrium for \( \{\{1,2\}, \{2,3\}, \{3,4\}\} \) Pareto dominates the subgame perfect equilibrium for the GC. This, in turn, implies that the subgame perfect equilibrium for the GC is not
perfectly coalition-proof, since the deviation \((\{1,2\},\{2,3\},\{3,4\})\) is beneficial for all nations. These important results are summarized in the following proposition.

**Proposition 6.** Suppose that the administrative cost paid by each member of a non-singleton coalition of size \(m\), \(m \in \{2,3,4\}\), is \(\chi m\), where \(\chi > 0\). Then, for \(\sigma \in (0,0.03125)\) and \(\chi \in (0,\bar{\chi}]\), \(\bar{\chi} = 0.000025832(b-1)^2\), the subgame perfect equilibrium for \((\{1,2\},\{2,3\},\{3,4\})\) is a PSN equilibrium and represents a strict Pareto improvement with respect to the subgame perfect equilibrium for the GC. Therefore, the latter is not perfectly coalition-proof.

Clearly, the subgame perfect equilibrium for \((\{1,2\},\{2,3\},\{3,4\})\) is not the only PSN equilibrium in the presence of administrative cost. For example, the subgame equilibrium for \((\{1,2\},\{1,3\},\{1,4\})\) is also a PSN equilibrium, since it is identical to the subgame perfect equilibrium for \((\{1,2\},\{2,3\},\{3,4\})\) – each nation faces an administrative cost equal to \(3\bar{\chi}\). This logic holds for every fully overlapped coalitional structure in which there are three overlapping coalitions and each coalition contains two members. Let us refer to these fully overlapped coalitional structures as “simple fully overlapped bilateral coalitional structures.”

**Proposition 7.** Suppose that the administrative cost paid by each member of a non-singleton coalition of size \(m\), \(m \in \{2,3,4\}\), is \(\chi m\), where \(\chi > 0\). Then, for \(\sigma \in (0,0.03125)\) and \(\chi \in (0,\bar{\chi}]\), \(\bar{\chi} = 0.000025832(b-1)^2\), the subgame perfect equilibrium for any simple fully overlapped bilateral coalitional structure is a PSN equilibrium.

Propositions 6 and 7 are remarkable. When the administrative cost paid by each member of a non-singleton coalition is increasing on membership size, the set of PSN equilibria includes subgame perfect equilibria for simple fully overlapped bilateral coalitional structures only. In addition, the subgame perfect equilibrium for the GC is not even a PCPN equilibrium! Policy makers who are
concerned about climate change damages should consider the benefits of forming overlapped bilateral international environmental agreements since such benefits may exceed the benefits of forming a fully participatory global agreement to reduce emissions of greenhouse gases.

7. Conclusion

This paper contributes to the literature on endogenous formation of international environmental agreements in that we consider the potential formation of non-binding overlapping environmental agreements to combat two types of correlated environmental evils, one that causes continental pollution damages and another that causes global pollution damages. We extend the notions of coalition-proof and perfectly coalition-proof equilibria in order to consider the stability of coalitional structures in which overlapping coalitions may coexist.

We identify the entire set of coalitional structures that are perfectly coalition-proof and demonstrate that there are Pareto efficient perfectly coalition-proof equilibria if the national damage caused by the continental pollutant is at least twice as high as the national damage caused by greenhouse gas emissions. We also show that continental agreements may be perfectly coalition-proof even when the national damage caused by greenhouse gas emissions exceeds the national damage caused by the continental pollutant. Therefore, the coexistence of continental and global pollutants may produce a large set of circumstances under which perfectly coalition-proof equilibria Pareto dominate the Nash equilibrium for the singleton coalitional structure. Finally, we also show that the perfectly coalition-proof equilibrium for a simple fully overlapped bilateral coalitional structure may be superior than the allocation implied by a fully participatory global agreement in the presence of coalition operation costs. In such a case, the subgame perfect equilibrium for the Grand Coalition is not perfectly coalition-proof!
References


