An Inquiry into the Impact of Regime Change on International Climate Coalition Stability

Zach Pence

April 16, 2018

Abstract

In this paper we analyze the impact of regime change on international climate change negotiation. Using a cooperative game theoretic approach, we introduce a parameter to represent the climate policy ideology of each nation’s leadership. We show how nations form coalitions and choose their optimal abatement levels under their current regime, and analyze how changes in leadership impact the stability of coalitions induced by treaties. Ultimately, we find that the stability conditions are difficult to meet, and require assistance from treaties to increase the chance of being met. To conclude, we recommend methods to structure treaties to minimize the risk of regime changes altering coalition structure, and discuss these recommendations in the context of the Kyoto Protocol and the Paris Climate Accord.
Contents

1 Introduction ........................................... 3

2 Review of Literature .................................. 5

3 The Model ................................................. 9

4 International Coalitions ............................. 11
   4.1 The Core .............................................. 12
   4.2 The Impact of Regime Change .................... 13
   4.3 Stability Conditions ................................. 16

5 Policy Recommendation ............................. 18
   5.1 Enforcing Treaties ................................. 18
   5.2 Kyoto Protocol and Paris Agreement ............ 19

6 Conclusion ................................................ 21

References ............................................... 22
1 Introduction

Climate change is one of the most pressing international issues facing the world, threatening every nation to a certain degree. Unfortunately, many of the industrial processes which have large economic benefits to countries also contribute substantially to global warming. As a result, nations are incentivized to continue to produce emissions, and rely on other countries to enact climate policy which benefits them. This situation is often modelled in game theory as a generalized prisoners’ dilemma.

The prisoners’ dilemma, perhaps one of the most widely known results in game theory, is used to highlight a disparity between individual optima and social optima.\(^1\) This is typically the game structure observed in climate negotiations. However, this does not mean that nations will never cooperate; there can often be higher payoffs through cooperation, so we may expect to see nations negotiate treaties in an attempt to cooperate. This behavior can be seen in several examples, such as the Montreal Protocol on Substances the Deplete the Ozone Layer in 1987, the Kyoto Protocol to the United Nations Framework Convention on Climate Change in 1998, and most recently the Paris Agreement in 2015. These treaties highlight an important aspect of international cooperation on environmental issues, in showing that we expect some nations to agree to cooperate with each other to reduce their emissions.

To capture this element of cooperation in models, we typically use a subfield of game

\(^1\)As an example of the prisoner’ dilemma, suppose two people, \(P_1\) and \(P_2\), have just been arrested for committing a crime, and are being interrogated separately. Suppose that each person has two options: they can confess to committing the crime, or they can lie and claim that they did not commit it. If both players lie, then neither will be charged with the crime; however, they will both receive 5 years in prison for lesser infractions related to the main crime. If they both confess, they will both get 10 years in prison for the crime. However, if one lies and the other confesses, the liar will get 15 years for not cooperating, while the person who confesses will get 1 year as a reward for their cooperation.

Under this setup, both players are incentivized to confess, as it is an individually optimal strategy; for example, consider \(P_1\) without loss of generality. If \(P_1\) lies, they get 5 years if \(P_2\) lies, and 15 years if \(P_2\) confesses. However, if \(P_1\) confesses, they get 1 year if \(P_2\) lies, and 10 years if \(P_2\) confesses. Since 1 < 5 and 10 < 15, each player is best off by confessing regardless of the opponent’s choice, resulting in them both getting 10 years. Note that if they had both lied, they would have both gotten 5 years, which is the socially optimal outcome. However, they are incentivized to deviate from this result since they could get 1 year by confessing if the other person lies. Thus everyone using their individually optimal strategies results in a non-optimal outcome.
theory called cooperative game theory. In cooperative game theory, we call groups of nations who agree to cooperate *coalitions*. It is easy to see that a set of coalitions must be a partition of the set of every nation, so we can say that coalitions are formed when a treaty induces a partition of nations. However, this approach also suggests that once an equilibrium partition is found, we may expect this partition to last. In other words, over time nations will continue to cooperate once a treaty has been agreed upon. Although this assumption holds in theory, there are many observed examples where this assumption has failed. We can consider, for example, Canada’s withdrawal from the Kyoto Protocol in 2012, or the United State’s declaration of intent to withdraw from the Paris Climate Accord in 2017.

Given this disconnect between the theory and observed phenomena, the goal of this paper is to create a model which can account for these sorts of changes. We note that often, as in the previous two examples, this break in equilibrium behavior occurs after some sort of change in political power. Recall that Canada withdrew from the Kyoto Protocol shortly after Peter Kent became the Minister of the Environment, and the United States announced its planned withdrawal from the Paris Climate Accord shortly after Donald Trump took office as the President. Therefore, in this paper we propose examining the affect of regime change on coalition stability.

In Section 2 of this paper, we will review the current literature on environmental game theory. In doing so, we will introduce the existing models and their applications, including literature on treaty negotiation. In Section 3, we will construct a model for choosing abatement. This model will account for a regime’s climate policy, and demonstrate how nations participating and not participating in an agreement will decide upon their abatement levels. To do so we introduce a parameter which will influence every nation’s benefit function based on the ideology of the current regime.

In Section 4, we will examine how a change in this parameter will influence coalition stability. We will introduce the core as a solution concept for cooperative games, and
show how changes in our regime parameter influence equilibrium coalition structure. In Section 5, we will use these conditions to provide policy recommendations on how to best mitigate the risk of regime change in treaty structure. Primarily, we recommend that treaties include better enforced penalties. These penalties must be constructed such that the shift in a nation’s benefit of emission function, induced by a change in regime, is met with a shift in the nation’s cost of emission function, induced by the enforced penalties. Furthermore, this shift in the cost function must be such that the value of the benefit function minus the cost function at the current level of emission is the same before and after the regime change. To conclude, we look at the Kyoto Protocol and Paris Agreement as examples, and demonstrate why they did not meet these conditions.

2 Review of Literature

Much of the literature concerning climate negotiation uses game theoretical models. Of this literature, the majority focuses on examining these negotiations as a static decision process, in which abatement decisions are made once at some point in time. Generalizations of the prisoners’ dilemma are the most widely used in this scenario. While the basic model is concerned with a discrete strategy space \{Abate, Pollute\}, much of the academic research is more concerned with a continuous strategy space in which a nation’s total pollution with zero abatement is given, and the decision is the proportion of that pollution to abate. In Finus (2003), a model is presented in which each nation has a benefit function for their own level of pollution, and a cost function for the total amount of pollution among \(n\) nations. Using the simple economic concept that profit is total benefit minus total cost, traditional economic optimization techniques can be used in this model to determine best response functions. Importantly, Finus (2003) suggests that players may not always choose to act non-cooperatively, given that the costs of pollution are shared among nations, and each nation’s individual action contributes to the collective cost. Thus, he
points out that cooperation is to be expected among some nations.

It follows that while the prisoners’ dilemma game has been used frequently to analyze international climate policy, there are potential limitations. The first of these limitations, as hinted at and addressed by Finus (2003), is that it does not account for subsets of nations cooperating with each other. The studies which look to address this problem are in the subfield of game theory known as cooperative game theory. We define two terms which are important for cooperative techniques.

**Definition.** Let $N$ be the set of all agents participating in a game, and let $\mathcal{P}(N)$ be the power set of $N$. Then we call a set $X \in \mathcal{P}(N)$ a coalition.

In the context of international climate games, this definition is implies that any set of nations who chose to cooperate form a coalition. The proceeding definition deals with the case where every nation acts cooperatively with each other.

**Definition.** The grand coalition is the trivial coalition $N \in \mathcal{P}(N)$.

Intuitively, it follows that the grand coalition is the set of all nations in the climate negotiation context.

The cooperative game theoretic approach to international environmental decision making focuses on one large coalition of countries, not necessarily the grand coalition, participating in an environmental agreement, with every other country acting as an individual agent (Chander, 2007). The use of cooperative game theory differs from strictly non-cooperative approaches in that this core coalition is able to work together to reach an outcome that is more efficient for every member, and reduce the total cost of pollution to every nation by increasing abatement. It is important to note that when nations form coalitions, the coalitions still behave non cooperatively when interacting with other coalitions; however, the member nations of any given coalition behave cooperatively with other member nations.

As hinted above, a natural extension of this concept is the idea that there may be mul-
multiple coalitions of nations, each acting in the best interest of their respective coalitions. For example, Maskin (2003) showed that cooperative game theoretic techniques could still be used in international issues without the existence of a grand coalition. Given that members of each coalition act in that coalition’s best interest, Maskin (2003) introduces the concept of “coalition externalities”, which give us a set of conditions for when coalitions may remain the same, merge, or separate. As such, each coalition must account for the possibility of coalition externalities from other coalitions’ decisions. It is worth noting that in his paper “The Gamma-core and Coalition Formation”, Chander (2007) address the conflicting models of the core and of multiple partitions. He accepts that multiple coalitions may contribute to the original $\gamma$-core grand coalition model. In doing so, Chander (2007) suggests that eventually, if a grand coalition exists, over time the other coalitions may still collapse, converging at the initial model of only one nontrivial coalition.

Chander’s concept of the $\gamma$-core has actually become a popular tool in cooperative environmental games. Before Chander introduced the concept, the primary core models used in environmental cooperative game theory were the $\alpha$ and $\beta$ core models. Although the details of these models are technical, the primary differences are that each one uses a different characteristic function to assess the value of a coalition. Thus, the primary reason that the $\gamma$-core is more widely used is that many researchers believe that it assigns values in a manner in which makes the model more realistic (Lardon, 2011). Thus, the $\gamma$-core may produce a more powerful model.

Although in general core models assume a grand coalition, it is possible to treat smaller coalitions that have already formed as a grand coalition acting with itself to solve the game, and then for other players to best respond to the outcome of this game. This is suggested by Carraro & Moriconi (1997), who examined the process of a nation deciding whether to sign a climate treaty with a coalition allowed (and predicted), but not required to form. The key piece presented here, hinted at in the other papers discussed thus far, is that each country’s payoff function increases monotonically as the cardinality
of the participating coalition increases. Thus, there is a sense that increased cooperation is beneficial for every country, although there is still incentive to not participate in this cooperation (Carraro & Moriconi 1997). Notably, due to the underlying prisoner’s dilemma, the coalition is expected to be made up of many fewer countries than exist in the system, despite the benefit to all of its existence (Barrett, 2005). The results of these works allow us to assume that we will always have a nontrivial partition, ie. the partition will neither consist of entirely singletons nor will the partition be the grand coalition.

Note that these partitions are formed by the treaty game. There are several different variations of the treaty game, typically involving a sequential step process during a negotiation and ratification period. This process usually takes the form of negotiate, decide whether to ratify, and then decide the level of abatement. As far as the negotiation process goes, Attanasi et al. (2015) suggest a bargaining game best represents the process. The general idea is that one country proposes a treaty, and the other countries accept or reject it. If it is rejected, then another country proposes a treaty, and the process continues until a treaty is agreed upon. Ideally, as the process goes on new treaties proposed will incorporate elements from previous ones which were well received, so that the final treaty contains the best elements from the proposals.

In considering treaty formation, there are also several other factors which must be considered. For example, Baer et al. (2000) suggest that equality of costs can become an issue even with the presence of asymmetric nations; for example, a large nation may over time feel as if they are being treated unfairly if they have the same emissions cap as other nations, since they may believe they should fairly be allowed more emissions. Alternatively, a large nation who produced a lot of emission might feel that they were being treated unfairly if they had a higher level of abatement to maintain, even though other nations may view it as fair. Another potential consideration is the role of behavioral theory. Ostrom (2009) suggests some leaders may be more inclined to cooperate, and that perceptions of different leaders integrity can be an important factor as well. Overall,
there are many different methods which have been used to model treaty negotiation, some including or excluding different factors. Wood (2011) demonstrates that there is not necessarily one agreed upon model or set of criteria to include, yet points out that there is some overlap of common concepts, such as bargaining.

In the existing literature, once a treaty has been negotiated, coalitions typically maintain each nation’s payoffs over time. Thus, there is no accounting for the phenomena where different decision makers may alter payoff functions. This paper argues that this is not necessarily the case, and that these payoffs as perceived by decision makers change when the decision makers themselves change. We plan to address this discrepancy by introducing a parameter for each nation which represents the decision maker’s general climate policies. This parameter will influence the nation’s benefit function for polluting, with the assumption that a leader who prioritized the environment more will cause a downward shift in the nation’s pollution benefit curve. Since this regime parameter influences the payoff function, the current regime during the treaty negotiation process will influence the nations decision on whether or not to participate. Additionally, since coalition stability is dependent upon each member nation’s payoff function, a change of regime introduces a risk of nations leaving coalitions.

3 The Model

Before we can analyze the impact of regime change, we need to establish a basic model to use. Thus, we begin by constructing a model which will allow us to analyze regime change and coalition stability.

Let \( N = \{1, 2, ..., n\} \) be the set of all nations, let \( E_i \) be the interval of possible levels of emissions that nation \( i \) can produce, and define \( \beta_i : E_i \to \mathbb{R} \) as nation \( i \)'s benefit function from emitting at the emission level \( e_i \in E_i \). This function has the properties that \( \beta_i' > 0, \beta_i'' \leq 0 \) for all \( i \in N \). The condition on the first derivative comes from the assumption
that countries can increase their benefits by polluting more (recall that this function does not account for the cost of pollution). The condition on the second derivative simply represents diminishing marginal returns, i.e. the amount of benefit a nation gets from polluting decreases as their pollution increases.

Similarly, we need to define an emissions cost mapping $\varphi_i : \prod_{j=1}^{n} E_j \to \mathbb{R}$. The domain of this function is the Cartesian product of every nation’s possible emission levels because the cost of pollution is incurred by every nation, not just the producers. This function has the properties that $\varphi'_i > 0$, $\varphi''_i \geq 0$. The condition on the first derivative comes from the assumption that the cost increases as the total amount of pollution increases, while the condition on the second derivative represents the increasing marginal cost of emissions, i.e. that the cost of adding more emissions to the system increases as the amount of existing pollution increases.

Both of these functions will be used in constructing payoff functions, yet first we introduce a “regime policy” variable, $\rho_i \in [0, 1]$. We can think of this as a continuum of potential policies a nation’s regime may have on the true benefits of producing pollution. On the extremes, a regime who wanted to eliminate polluting industries entirely would have $\rho_i = 0$, and a nation with a regime that wanted to leave the market to reach its own optimal pollution level would have $\rho_i = 1$. Using this variable we can construct each nation’s payoff function as

$$\pi_i = \rho_i \beta_i(e_i) - \varphi_i \left( \sum_{j=1}^{n} e_j \right).$$

(1)

Using Equation 1, we can easily construct the objective functions for treaty participants and for non-signees. We know that the participant’s goal is to maximize the collective payoff, which can be expressed as the sum of their individual payoffs. Suppose $k$ of the $n$ countries signed the treaty, and let $\Pi_k = \sum_{i=1}^{k} \pi_i$ be the payoff function for the coalition. Recall that in Equation 1, the image of both $\beta_i$ and $\varphi_i$ is in $\mathbb{R}$ for all $i$. Thus,
using the field axioms\(^2\), it is clear to see that
\[
\Pi_k = \sum_{i=1}^{k} \rho_i \beta_i(e_i) - \sum_{i=1}^{k} \phi_i \left( \sum_{j=1}^{n} e_j \right). 
\]
(2)

Thus, the participating nations must choose an emissions vector,
\[
\vec{e}^* = \begin{pmatrix}
e^*_1 \\
\vdots \\
e^*_k
\end{pmatrix},
\]
which maximizes Equation 2 subject to constraints imposed by the treaty.

The next stage, which is when the non-signees decide their abatement, follows directly from Equation 1. Since each non-participating nation only needs to worry about their own objective function, they only need to find their own optimal emission level \(e^*_i\), using Equation 1 as their objective function. Note that this stage occurs after the participants decide their abatement, as this impacts \(\phi_i \left( \sum_{j=1}^{n} e_j \right)\). Following this process, we can view the participating nations as a coalition, and the non-participants as coalitions of singletons.

4 International Coalitions

When international treaties are agreed upon, they partition the set of countries into coalitions. That is, they create some \(m \leq n\) subsets of \(N, C_1, ..., C_m\), such that \(C_i \cap C_j = \emptyset\) for all \(i, j\), and \(\bigcup_{i=1}^{m} C_i = N\). This partition will take the form of one nontrivial coalition \(S \subset N\), with every nation \(i \notin S\) forming a trivial coalition (Chander, 2007). Thus, we will use cooperative techniques on the one nontrivial coalition, and the other countries will best respond to this coalition’s emission level in a non-cooperative setting.

\(^2\)While the field axioms are generally well known, Rudin (1976) is a reputable source for reference.
4.1 The Core

Suppose that following the treaty negotiations, the nations partition into one coalition of treaty participants, and singleton coalitions of non-participants. Call the coalition $S$, and note $N/S = \{n \in N : n \notin S\}$. Recall from Equation 2 that the payoff function for $S$ is the sum of payoff functions for the participants, and the objective is to find $\vec{e}^*$ in order to maximize this payoff. Call this objective function $\Pi_S$. Since these participating nations are no longer acting in a non-cooperative setting, we cannot use traditional non-cooperative methods to optimize this objective function. Instead, we will analyze the core.

In order to analyze this core, we need a value function or characteristic function $\nu : \mathcal{P}(S) \to \mathbb{R}$, where $\mathcal{P}(S)$ is the power set of $S$. $\nu$ must also satisfy the following properties:

(i) $\nu(\emptyset) = 0$, and

(ii) $\nu(A \cup B) \geq \nu(A) + \nu(B)$ for any disjoint subsets $A, B \in \mathcal{P}(S)$.

In essence, this function gives the value that the member nations get in the current partition (Chandler, 2007). The reasoning for condition (i) is obvious, as there is no value when no countries are in the coalition. The reasoning for condition (ii) is to ensure that $S$ is a stable coalition, so no country is better off by leaving.

Let $s = |S|$, and using the set of participating nations $S$ and the value function $\nu$, we can formally define the core. The definition we provide is adopted from Owen (1995).

**Definition.** The **core** of a game is the set of all $s$-vectors such that the following conditions are satisfied:

(i) $\sum_{i \in S} \pi_i = \nu(S)$, and

(ii) For any $T \subseteq S$, $\sum_{i \in T} \pi_i \geq \nu(T)$.

We denote the core as $C(S, \nu)$.

It is important to note that the core is a set of optimal vectors, not a single value. This is based upon the assumption of transferable utility, allowing nations in the core to
distribute payoffs among themselves freely. Therefore, the actual emissions vector chosen by the participating coalition does not necessarily matter; all that matters is that it is in the core, and it will be a stable solution. However, it is an important fact that \( \sum_{i \in S} \pi_i \) remains the same; our transferable utility assumption only allows for redistribution of payoffs, not for changing the value \( \nu(S) \). Thus, we can deduce by equation 2 that \( \sum_{i \in S} e_i \) is the same, regardless of how emissions levels are distributed within the participating coalition. By equation 1, it follows that each nation \( k \in N \setminus S \) maximizes the objective function

\[
\pi_k = \rho_k \beta_k(e_k) - \varphi_k \left( \sum_{i \in S} e_i \right) + \varphi_k \left( \sum_{j \in N \setminus S} e_j \right)
\]

\[
= \rho_k \beta_k(e_k) - \varphi_k \left( \sum_{j=1}^{n} e_j \right).
\]

Thus, we can see that each nonmembers objective function is unchanged, and they respond non-cooperatively with the participating coalition as if it were a single player.

### 4.2 The Impact of Regime Change

Without the contribution of \( \rho_i \) to each \( \pi_i \), finding the core would essentially be the end of the story as far as coalition stability; the participating coalition would form, choose some \( \vec{e} \in C(S, \nu) \), and this partition would be stable. Of course, over time the member nations may decide to change their emission and redistribute their payoffs, but the game structure will remain stable. However, under a change in regime, there is a risk that some \( \rho_i \) may change dramatically, which would in turn affect \( \pi_i \) and thus \( \Pi_S \). Hence, there is a risk that regime change can threaten coalition stability. Therefore, we need to establish a set of stability conditions based upon \( \rho_i \). First, however, we must analyze how a change in \( \rho_i \) can affect \( \Pi_S \). To do so we consider two cases: the case where \( \rho_i \) increases, and the case where \( \rho_i \) decreases.
Lemma 1. Let $e_{-i}$ be invariant\(^3\), then for some $i \in N$,

(i) $e_i$ increases if, and only if, $\varphi_k \left( \sum_{j=1}^n e_j \right)$ increases for all $k \in N$.

(ii) Similarly, $e_i$ decreases if, and only if, $\varphi_k \left( \sum_{j=1}^n e_j \right)$ decreases for all $k \in N$.

Proof. The proofs of (i) and (ii) are similar, so we only show (i).

(⇒) Let $\varepsilon > 0$, and suppose $e_i$ increases to $e_i + \varepsilon$. Then, it follows that the new amount of total emission is given by $\left( \sum_{j=1}^n e_j \right) + \varepsilon$. Recall that $\varphi_k$ is strictly monotonic for all $k \in N$ by our first derivative condition, so it follows that $\varphi_k \left( \left[ \sum_{j=1}^n e_j \right] + \varepsilon \right) > \varphi_k \left( \sum_{j=1}^n e_j \right)$.

(⇐) Let $\delta_k > 0$ for all $k \in N$, and suppose $\varphi_k \left( \sum_{j=1}^n e_j \right)$ increases to $\varphi_k \left( \sum_{j=1}^n e_j \right) + \delta_k$ for all $k \in N$. Since $\varphi_k$ is strictly monotonic for all $k \in N$, it follows that there exists an $\varepsilon > 0$ such that $\sum_{j=1}^n e_j + \varepsilon \mapsto \varphi_k \left( \sum_{j=1}^n e_j \right) + \delta_k$ for all $k \in N$. Since $e_{-i}$ is invariant, it follows that $\sum_{j=1}^n e_j + \varepsilon = \left( \sum_{j,j \neq i}^n e_j \right) + (e_i + \varepsilon)$. Clearly, $(e_i + \varepsilon) > e_i$, completing the proof. \(\square\)

Note that in reality, there is an element of timing involved with these changes; the change in regime will always precede the change in emissions level. However, proving both directions of this lemma allows us to see how closely related regime policy and emissions level really are, and is mathematically more precise. Using this result, we can proceed to the following proposition, which shows us how a change in regime for one nations affects every other nations’ well being.

Proposition 1. Let $e_{-i}$ be invariant, then

(i) If $\rho_i$ increases for some $i \in N$, then $\varphi_k \left( \sum_{j=1}^n e_j \right)$ increases for all $k \in N$.

(ii) Similarly, If $\rho_i$ decreases for some $i \in N$, then $\varphi_k \left( \sum_{j=1}^n e_j \right)$ decreases for all $k \in N$.

Proof. The proofs for (i) and (ii) are similar, so we only show (i).

Let $\rho_i^a$ increase to $\rho_i^b$. Then, clearly $\rho_i^b \beta_i(e_i) < \rho_i^a \beta_i(e_i)$ and $\rho_i^a \beta'_i(e_i) < \rho_i^b \beta'_i(e_i)$ for all $e_i \in E_i$.

We know by our first order condition on $\pi_i$ that nation $i$ will emit at the level where

---

\(^3\)Note that throughout this paper, we use the notation $e_{-i}$ to represent $\{e_j : j \in N, j \neq i\}$. Additionally, we use the term invariant to mean “unchanging”, i.e. does not change in our analysis.
\( \rho_i \beta_i'(e_i) = \varphi_i' \left( \sum_{j=1}^{n} e_j \right) \), so initially nation \( i \) emits at \( e_i^a \) such that \( \rho_i \beta_i'(e_i^a) = \varphi_i' \left( \sum_{j=1}^{n} e_j \right) \).

However, observe that \( \rho_i \beta_i'(e_i^a) > \varphi_i' \left( \sum_{j=1}^{n} e_j \right) \), so nation \( i \) will change their emission level from \( e_i^a \) to \( e_i^b \). Recall that \( \rho_i \beta_i''(e_i) \leq 0 \) and \( \varphi_i'' \left( \sum_{j=1}^{n} e_j \right) \geq 0 \) for all \( i \in N \), so clearly we need \( e_i^b > e_i^a \) to satisfy \( \rho_i \beta_i'(e_i^b) = \varphi_i' \left( \sum_{j=1}^{n} e_j \right) \). Thus, by Lemma 1 it follows that \( \varphi_k \left( \sum_{j=1}^{n} e_j \right) \) increases for all \( k \in N \).

This result is significant in our analysis of coalition stability, as along with Lemma 1, it demonstrates that a change of regime in any nation can change the optimal level of emission and costs of emission for every nation. It also leads to a corollary which will be important in our conditions for coalition stability.

**Corollary 1.** Let \( e_{-i} \) be invariant, then

(i) If \( \rho_i \) increases for some \( i \in N \), then \( \pi_k \) decreases for all \( k \in N \setminus \{i\} \).

(ii) Similarly, if \( \rho_i \) decreases for some \( i \in N \), then \( \pi_k \) increases for all \( k \in N \setminus \{i\} \).

**Proof.** The proofs for (i) and (ii) are similar, so we only show (i).

Let \( \varepsilon > 0 \), and suppose \( \rho_i \) increases to \( \rho_i + \varepsilon \). Then, by Proposition 1, we know that for all \( k \in N \) there exists some \( \delta_k > 0 \) such that each nations cost increases to \( \varphi_k \left( \sum_{j=1}^{n} e_j \right) + \delta_k \) (note that this hold even when \( k = i \)). Recall that \( e_{-i} \) is invariant, which implies that \( \beta_{-i}(e_{-i}) \) is invariant under this increase, limiting our analysis to changes in \( \pi_{-i} \). Observe that under this transformation, \( \beta_k(e_k) - \varphi_k \left( \sum_{j=1}^{n} e_j \right) \mapsto \beta_k(e_k) - \left( \varphi_k \left( \sum_{j=1}^{n} e_j \right) + \delta_k \right) \).

Clearly, \( \beta_k(e_k) - \varphi_k \left( \sum_{j=1}^{n} e_j \right) \geq \beta_k(e_k) - \varphi_k \left( \sum_{j=1}^{n} e_j \right) - \delta_k \). The result follows by Equation 1.

Based upon equation 2, it is clear to see why this corollary is so important, as the entire objective function of the participating coalition is constructed by summing these payoff values. Thus, it follows that a change in regime threatens to alter the entire objective function of the participating coalition. If the sum of payoffs is dramatically changed, then the entire core of the initial game could cease to satisfy the core conditions under the new game.
4.3 Stability Conditions

The results obtained in Section 4.2 provide us with enough information to determine our stability conditions. Before formally introducing these conditions, we will work through the intuition to understand how they follow from our previous results. Recall that for any vector in the core, the sum the elements in the vector must be equal to the value of the coalition. Also note that the value function itself does not change, so we can deduce that if some nation’s payoff changes, at least one other nation’s payoff must change in the opposite direction. We also know from Corollary 1 that a change in one nation’s regime induces a change in every other nations payoff. Intuitively, it follows that the nations which experienced the regime change must experience a change in payoff which balances out all the other nations’ payoff changes to maintain stability.

We also know that for any vector in the core, the sum of any combinations of elements of the vector must be greater than the value which that combination of nations would achieve if they formed their own partition. Thus, when a nation’s payoff decreases, we can see that the change must be small enough that for every possible subset containing this nation, it still cannot do better by leaving the participating coalition. Therefore, we can intuitively deduce from Corollary 1 that when some nation experiences a regime change, every nation must still be best off by remaining in the participating coalition compared to every possible subset of participating nations.

The conditions we have just described are the two stability conditions we need, and it is clear how the intuition behind them follows from the definition of the core and the results from section 4.2. However, this intuition alone is not sufficient to prove these conditions, so we will state them formally and provide a proof.

**Proposition 2.** Let $e_{-i}$ be invariant, and suppose without loss of generality that some $\rho_k$ increases. Then, the participating coalition $S$ is stable if the following conditions are met:

(i) $\sum_{i \in S} \Delta \pi_i = 0$, and
(ii) For any \( T \subseteq S \), \( \sum_{i \in T} \Delta \pi_i \geq \nu(T) - \sum_{i \in T} \pi_i \),

where \( \pi_i \) represents nation \( i \)'s payoff before the change in \( \rho_k \).

**Proof.** (i) Recall that for any optimal vector \( \vec{\pi} = (\pi_1, \ldots, \pi_s) \in C(S, \nu) \), where \( s = |S| \), we need

\[
\sum_{i \in S} \pi_i = \nu(S). \tag{3}
\]

By Corollary 1, we know that the change in \( \rho_k \) induces a change in \( \pi_i \) for all \( i \in S \). We can define this change as \( \Delta \pi_i \), and clearly the total change in \( \Pi_S \) is given by \( \sum_{i \in S} \Delta \pi_i \).

This means that the new sum of payoffs is given by \( \sum_{i \in S} \pi_i + \sum_{i \in S} \Delta \pi_i \). By equation (3), which is from the definition of the core, it follows that our new stability condition is \( \sum_{i \in S} \pi_i + \sum_{i \in S} \Delta \pi_i = \nu(S) \), which implies \( \sum_{i \in S} \Delta \pi_i = 0 \).

(ii) Recall that for any optimal vector \( \vec{\pi} \in C(S, \nu) \), we need

\[
\sum_{i \in T} \pi_i \geq \nu(T) \text{ for every } T \subseteq S. \tag{4}
\]

So, let \( T \in \mathcal{P}(S) \) be arbitrary, and define \( \Delta \pi_i \) as in (i). Then, it follows that the new sum of payoffs for \( T \) is given by \( \sum_{i \in T} \pi_i + \sum_{i \in T} \Delta \pi_i \). By equation (4), which is from the definition of the core, it follows that our new stability condition is \( \sum_{i \in T} \pi_i + \sum_{i \in T} \Delta \pi_i \geq \nu(T) \), which implies \( \sum_{i \in T} \Delta \pi_i \geq \nu(T) - \sum_{i \in T} \pi_i \). Since \( T \) was arbitrary, it follows that this condition must be met for all \( T \in \mathcal{P}(S) \).

**Corollary 2.** Let \( e_{-i} \) be invariant, and suppose without loss of generality that some \( \rho_k \) increases. Then

\[
\Delta \pi_k = - \sum_{i \in S \setminus \{k\}} \Delta \pi_i \tag{5}
\]

is a necessary condition for coalition stability.

Corollary 2 follows directly from Proposition 2, so we omit the proof.
We can see from Proposition 2 that ensuring coalition stability appears extremely difficult under even the slightest change in some $\rho_k$; especially given equation (5), it can appear that any minor regime change threatens the stability of an international environmental agreement. However, we know from real-world observation that this is not necessarily the case; the structure of coalitions only seems threatened when the change in regime is associated with a fairly large ideological shift with respect to climate change. However, we argue that this prediction is actually fairly accurate. Consider that up to this point, we have been viewing treaties as simply a means of inducing a partition. However, if this were the only purpose of treaties, we would expect a treaty to simply take the form of a list of countries in the participating coalition, which is clearly not the case. Thus, we can really think of our stability conditions as the incentives from the treaty being sufficient to induce the conditions from Proposition 2.

5 Policy Recommendation

Based upon the findings in section 4, we will examine what sorts of measures should be included in treaties to maximize the likelihood that a climate coalition remains stable after a regime change. Generally, we observe that treaties tend to work pretty well in the cases of small ideological changes in leaders, so we will focus on the case where there is a large change in regime climate policy.

5.1 Enforcing Treaties

The primary takeaway from the findings of section 4 is that treaties need to have better enforcement mechanisms. Barrett (2005) notes that while treaties need enforcement mechanisms to be successful, most do not have them. We propose two potential solutions based upon our model: the first is that nations could be held to lower emissions levels if some member nation’ regime changes to a more conservative climate policy. This would
allow some nations to lower the total costs, lowering the change in total utility, and working towards the stability conditions. However, an obvious issue with this approach is that it would punish some nations for another nation’s regime change. Thus, we may expect some nations to be hesitant to sign a treaty with such a provision. Nevertheless, it could work in the case where a smaller nation experienced the regime change, as it would take substantially less effort from other nations to lower their emissions appropriately.

An alternative, and likely more agreeable solution may be include a provision aimed to penalize nations for leaving, or significantly altering the established emissions levels. This method allows for the possibility that member nations do not have to alter their actions, so long as the punishment works. The idea here is based on the fact that an increase in some $\rho_i$ leads to a shift in $\rho_i\beta_i$. Using the fact that the payoff is given by benefit minus cost, the goal would be for the punishment of increasing emission to shift nation $i$’s cost function. Since $\pi_i$ is made of the difference between the benefit and functions, the right punishment could shift the cost such that even with the benefit shift, $\pi_i$ remained the same. In this case, no other nations would have to change their emission, and the core would remain stable.

5.2 Kyoto Protocol and Paris Agreement

The Kyoto Protocol and the Paris Agreement are two of the most prominent examples of treaties which lost member nations following major regime changes. Not surprisingly, they are also two examples of treaties which lacked strong enforcement provisions. We can begin by examining the Kyoto Protocol. The first mention of a nation failing to emit at an agreed upon level occurs in Article 18, in which it says that “Parties to this Protocol shall... approve appropriate and effective procedures... to address cases of non compliance” (United Nations, 1998, pp. 15). However, it fails to specify what sorts of consequences may actually be used, instead leaving it to be established over time. From a strategic perspective with regards to negotiation, the is a good move; without openly
including harsh punishments, nations may be more likely to sign. However, in the case of a dramatic ideological shift resulting from a regime change, the loose threat may not be substantial to act as a credible threat. This is especially true when we consider that the goal of the threat is not to actually apply the sanction, but rather to convince the nation that they would not be better off by increasing their emission. Thus, if an adequate consequence were decided after a nation had dramatically changed their emission, there is a risk that it would be too late to prevent the core from changing.

Additionally, the only mention in the treaty of nations leaving the protocol is in Article 27. There are two main points worth bringing up here: the first being that the protocol states that “At any time after three years from the date on which this Protocol has entered into force for a Party, that Party may withdraw from this Protocol” (United Nations, 1998, pp.18). While this gives nations a way out, also specifying that a withdraw will not be granted until one year after request, there is no mention of what happen should a nation unofficially withdraw (ie. ignoring the agreement). While this would likely fall under the clause in Article 18, the lack of any exact answer may cause a new regime to act under the assumption that the cost of this will not be substantial.

The Paris Agreement follows a similar blueprint to the Kyoto Protocol with respect to sanctions and enforcement. In Article 15, the agreement states that members who fail to comply will be treated in a manner which is “non-adversarial and non-punitive” (United Nations, 2015, pp. 19). As in the Kyoto Protocol, this certainly seems appropriate for a cooperative agreement, yet it may not be convincing enough given a dramatic regime change. Similarly, Article 28 in the Agreement appears to be copied almost verbatim from Article 27 of the Kyoto Protocol, replacing the word Protocol with Agreement (United Nations, 2015). However, in this case the effectiveness of the three year requirement was seen, as the United States cannot withdraw at the time of this paper, despite Donald Trump announcing that he planned to do so. While this is a positive, it also suggests that the implied shift in cost was not great enough to deter the Trump Administra-
tion. Although nothing rational is certain in this situation, perhaps an Article suggesting stronger consequences for withdrawing may have been sufficient to deter him. It is also worth noting that only requiring nations to remain for three years may have been insufficient. Given the size of the United States, their eventual withdrawal may raise emissions to the point where the costs of member nations are too high, and a stable core cannot be maintained.

6 Conclusion

By creating a cooperative model, we were able to find a set of stability conditions under regime change. Our approach focused on using a treaty to partition nations into one nontrivial coalition. The nations in this coalition behaved cooperatively, and nations outside the coalition behaved non-cooperatively, treating the coalition as another player. We defined nations payoff as the difference between their benefit from emission and their cost of emission, and introduced a regime policy factor which acted as a shift on the benefit curve. After proving several interesting properties of the model, we found conditions for which the coalition would remain stable under regime changes.

Our stability conditions ultimately appeared very difficult to achieve, supporting the use of treaties to achieve stability. However, we noted the given a significant enough ideological shift in some nation’s climate policy as the result of a regime change, treaties may need more consequences and greater enforcement. Since the stability conditions in our model showed a strong relationship between emission level, regime climate policy, cost of emission, and payoff, we recommended that treaties focus on consequences designed to shift out the cost function, with the goal of leaving payoffs unchanged.

In our analysis, we focused on the case where only one nation faced a regime change. The goal of this analysis was to keep the model simple enough to use use and understand, and to show how the different pieces of the model interacted. However, an interesting
extension to this work would be to relax the assumption that only one nation experiences a regime change. This would complicate the model, but might provide some valuable insight.

References


