Climate Change and Game Theory:

a Mathematical Survey

Peter John Wood
Crawford School
The Australian National University

CCEP working paper 2.10, October 2010

Abstract
This paper examines the problem of achieving global cooperation to reduce greenhouse gas emissions. Contributions to this problem are reviewed from noncooperative game theory, cooperative game theory, and implementation theory. We examine the solutions to games where players have a continuous choice about how much to pollute, and games where players make decisions about treaty participation. The implications of linking cooperation on climate change with cooperation on other issues, such as trade, is also examined. Cooperative and non-cooperative approaches to coalition formation are investigated in order to examine the behaviour of coalitions cooperating on climate change. One way to achieve cooperation is to design a game, known as a mechanism, whose equilibrium corresponds to an optimal outcome. This paper examines some mechanisms that are based on conditional commitments, and their policy implications. These mechanisms could make cooperation on climate change mitigation more likely.

Centre for Climate Economics & Policy
Crawford School of Economics and Government
The Australian National University
ccep.anu.edu.au

The **Centre for Climate Economics & Policy** (ccep.anu.edu.au) is an organized research unit at the Crawford School of Economics and Government, The Australian National University. The working paper series is intended to facilitate academic and policy discussion, and the views expressed in working papers are those of the authors. Contact for the Centre: Dr Frank Jotzo, frank.jotzo@anu.edu.au.

Citation for this report:

Wood, P.J. (2010), *Climate Change and Game Theory: a Mathematical Survey*, CCEP working paper 2.10, Centre for Climate Economics & Policy, Crawford School of Economics and Government, The Australian National University, Canberra.

Climate Change and Game Theory: a Mathematical Survey

Peter John Wood *

October 20, 2010

Abstract

This paper examines the problem of achieving global cooperation to reduce greenhouse gas emissions. Contributions to this problem are reviewed from non-cooperative game theory, cooperative game theory, and implementation theory.

We examine the solutions to games where players have a continuous choice about how much to pollute, and games where players make decisions about treaty participation. The implications of linking cooperation on climate change with cooperation on other issues, such as trade, is also examined. Cooperative and non-cooperative approaches to coalition formation are investigated in order to examine the behaviour of coalitions cooperating on climate change.

One way to achieve cooperation is to design a game, known as a mechanism, whose equilibrium corresponds to an optimal outcome. This paper examines some mechanisms that are based on conditional commitments, and their policy implications. These mechanisms could make cooperation on climate change mitigation more likely.

Key Words and Phrases. Climate change negotiations; game theory; implementation theory; coalition formation; subgame perfect equilibrium.

*Dr Peter John Wood, Resource Management in Asia-Pacific Program, The Crawford School of Economics and Government, The Australian National University, Canberra, ACT 0200, Australia, Email: Peter.J.Wood@anu.edu.au

Contents

1 Introduction **3**

2 Non-cooperative Games and Climate Change **6**

2.1 Normal Form Games and the Nash Equilibrium 6

2.2 Extensive Form Games and the Subgame Perfect Equilibrium 10

3 Coalitions **19**

3.1 Cooperative Game Theory and the Core 20

3.2 Coalition Formation and Externalities 22

4 Implementation Theory **26**

5 Conclusion **32**

References **35**

1 Introduction

Game theory can help us understand how strategic behaviour interacts with one of the most important social and environmental challenges of our time, climate change. A key reason why achieving international cooperation to address climate change is difficult is that there are strong free-rider incentives. These incentives arise because climate change mitigation is a global public good everyone benefits from there being less global warming, and everyone has an incentive for someone else to take on the burden of emission reductions. This is compounded by the fact that because of sovereignty issues, international institutions are weak compared to national ones. Game theory, which analyses the mathematics of strategic behaviour, can help us obtain a better understanding of how the incentive to free-ride works, identify the potential barriers to cooperation, and find approaches to facilitate a cooperative outcome. This paper surveys the game theoretic literature that relates to climate change, with an emphasis on approaches that try to find ways to facilitate cooperation.

Other surveys on the application of game theory and climate change mitigation include Finus (2001, 2003); Barrett (2003); Finus (2008). What is different about this survey is that it also includes a detailed discussion on implementation theory and its policy implications.

Game theory is often applied by assuming that the game is given, and used to predict the behaviour of participants. But an area of game theory known as implementation theory treats the desired outcome as given, and asks how to design a process that leads to this outcome (Jackson, 2001). An example of such as process could be the negotiations for an international environmental agreement. This approach may help us design processes that are more likely to lead to cooperative outcomes.

Addressing the free-rider incentives associated with climate change mitigation requires that we find mechanisms to facilitate cooperation between states. One such approach is international treaty-making.

The difficulties with finding cooperation were illustrated in the 2009 Copenhagen climate negotiations, which resulted in a political accord, but where after years of negotiations there remained too much disagreement between nations to arrive at a binding international treaty. There has been an ongoing political debate about the role of the United Nations, and whether more could be achieved in negotiations involving smaller groups of countries. The Copenhagen negotiations may have made the latter more likely. The lead US climate change negotiator, Todd Stern, stated: “You cant negotiate in a

group of 192 countries. Its ridiculous to think that you could" (Little, 2010). Nicholas Stern has offered a different perspective, stating that "The fact of Copenhagen and the setting of the deadline two years previously at Bali did concentrate minds, and it did lead... to quite specific plans from countries that hadn't set them out before", and that it was vital to stick with the UN process, whatever its frustrations (Black, 2010).

Game theory can provide useful insights when considering debates such as these. In fact, there has been a parallel debate in the game theory literature (see Section 3) on whether cooperation is more likely to arise from a grand coalition of all countries, or from smaller coalitions that do not include every country. A grand coalition, if it existed, would lead to more cooperation. But it is may not be the case that such a coalition would be stable. It is also not clear how such a coalition would form in practice. Game theory provides insight both into the stability of coalitions, and the implications of different processes for forming coalitions.

When using a model to help understand a problem, it is important to be aware of the limitations of the model. Many applications of game theory require that decision makers are rational. That is, they have clear preferences, form expectations about unknowns, and make decisions that are consistent with these preferences and expectations. These assumptions may not be consistent with experimental psychology. Ostrom (2009) has considered the the role that human behaviour considerations relate to cooperation problems, and applied this to climate change. She found that a 'surprisingly large number of individuals facing collective action problems do cooperate'. She also found that cooperation is more likely if people gain reputations for being trustworthy reciprocators; reliable information is available about costs and benefits of action; individuals have a long-term time horizon; and are not in a highly competitive environment.

In some of the situations that we describe here, countries are assumed to be the players in the game. That is, they are assumed to have clear preferences, usually based on the aggregate welfare of the countries citizens. In reality, different citizens have greatly different preferences, and the decision making is based on a political process. Game theoretic models can also be used to investigate the political process of decision making in a country, although this could lead to models becoming more complicated. Game theory can provide insights into the process of treaty ratification, by treating it as an extensive form game.

Despite the above limitations of game theoretic methodologies, many of the mechanisms described here are important because their game theoretic solutions are cooperative.

In other words, these mechanisms have game theoretic solutions that maximise the total sum of the utilities of the players. If humans are more cooperative than assumed in our models, the models could work as a lower benchmark, and at least as much cooperation as predicted by the models could be observed. Mechanisms that are expected to lead to cooperation can also be further tested using behavioural experiments.

Whether such mechanisms lead to full cooperation in practice depends on how well they can be implemented. Many of the mechanisms described here require that players will avoid backtracking on conditional commitments. This has both national policy implications and policy implications for the design of international climate agreements.

In Section 2, we introduce games where players make decisions independently. We discuss both the normal form representation of a game and the extensive form representation of a game. We investigate the role of solution concepts including the Nash equilibrium and the subgame perfect equilibrium. We study examples such as the prisoners dilemma; repeated prisoners dilemmas; games of treaty participation; games of treaty ratification; and the ultimatum game, which relates to issues such as fairness and reciprocity. We investigate a basic framework for studying what happens when countries have a continuous choice about how much they reduce their emissions.

In Section 3, we examine situations where players can cooperate with each other and form coalitions, which may then behave non-cooperatively when interacting with other coalitions. We discuss an interesting result from cooperative game theory, due to Chander & Tulkens (1997), that if a grand coalition for reducing emissions was to dissolve into singletons when any coalition breaks away, then full cooperation is possible. We also discuss non-cooperative mechanisms for coalition formation, and apply this to the question of whether cooperation is more likely among a grand coalition, or among several smaller coalitions. One mechanism for coalition formation relates to the issue of international carbon market linkage.

In Section 4, we look at applications of implementation theory to climate change. We examine mechanisms for getting players to agree to a socially optimal outcome. We also look at some mechanisms for providing public goods when there may not be strong institutions, what this says about the role of conditionality in international negotiations, and how that relates to emissions reductions in an international context. We make some comments about what individual countries can do, and how international agreements could be designed, to facilitate mechanisms such as the ones discussed.

Section 5 concludes, and includes a discussion of the subgame perfect equilibrium.

2 Non-cooperative Games and Climate Change

In non-cooperative games, players make decisions independently. We define some of the relevant ways of representing non-cooperative games and their solution concepts. We illustrate these definitions with a number of examples that are relevant to climate change.

2.1 Normal Form Games and the Nash Equilibrium

Definition 2.1 The *normal form representation* of a game specifies

1. the set of *players* in the game (in the context of climate change these will often be *countries*), N;
2. a set S of *strategy combinations*, each strategy combination assigns a strategy to each player;
3. and the set of payoffs $\Pi = \{\pi_i : i \in N\}$ received by each player for each possible strategy combination. Each payoff π_i assigns a real number (the utility[1]) to a strategy combination.

The normal form representation of a game is sometimes also known as the strategic form of a game.

When we consider a player i and strategy combination s, we will often write s_{-i} to denote the strategies of players other than i, and write $s = (s_i, s_{-i})$.

Definition 2.2 A *Nash equilibrium* for a normal form representation of a game is a strategy combination $s^* = (s_i^*, s_{-i}^*)$ where for all players $i \in N$, we have that

$$\pi_i(s_i^*, s_{-i}^*) \geq \pi_i(s_i, s_{-i}^*). \tag{1}$$

[1]It is possible to define strategic games more generally in terms of a preference relation for each player on the set of strategy combinations (Osborne, 2003, Chapter 2). It follows from ordinal utility theory that if a preference relation satisfies certain axioms, then it is representable by a utility function (Berger, 1980, Chapter 2), (Ok, 2007, Section B.4).

Assessing the impact on utility of climate change is complicated by several factors (Garnaut, 2008a), (Stern, 2006): the damages are uncertain, so players are interested in impact on *expected utility*; damages can include impacts on non-market goods such as ecosystems; many impacts occur in the future and affect future generations, so their valuation depends on a discount rate that is likely to take into account the pure rate of time preference, the marginal elasticity of consumption, and the expected rate of economic growth; and depends on the risk aversion of the player. Because unmitigated climate change presents potentially catastrophic risks, the possible impact of highly damaging outcomes can dominate the expected damage function (Weitzman, 2009).

In other words, in a Nash equilibrium every strategy is the best response to the Nash equilibrium strategies of the other players.

An important variation of the concept of a normal form game allows players to play *mixed strategies*. Instead of choosing a particular strategy, each player assigns a probability to each strategy (Osborne & Rubinstein, 1994, Chapter 3).

Example 2.1 (The Prisoner's Dilemma). The problem of achieving cooperation to reduce greenhouse gas emissions is related to a normal form game known as the *prisoner's dilemma*. All countries are collectively better off if they reduce their emissions, but each country is individually better off if they continue to pollute. We shall now describe a two player prisoner's dilemma. Each player has two possible strategies $\{Pollute,\ Abate\}$. Each player prefers the situation where she plays *Pollute* and the other plays *Abate* to the situation where both play *Abate*; and prefers that to the situation where they both play *Pollute*; and prefers that to the situation where she plays *Abate* and the other plays *Pollute*. An example of a prisoner's dilemma is illustrated below, using the *payoff matrix* notation. The two rows correspond to the two possible actions of the first player; the two columns correspond to the possible actions of the second player; the numbers in each box correspond to the payoffs for each player, with the payoff for the first player listed first.

$$\begin{array}{cccc} & & \text{Player 2} & \\ & & \textit{Abate} & \textit{Pollute} \\ \text{Player 1} & \textit{Abate} & (10,10) & (0,11) \\ & \textit{Pollute} & (11,0) & (1,1) \end{array} . \tag{2}$$

The strategy pair $(Pollute, Pollute)$ is a Nash equilibrium because given that the second player chooses *Pollute*, the first player is better off choosing *Pollute* than choosing *Abate*, and vice-versa. None of the other strategy combinations are Nash equilibira because in each case at least one player can improve their payoff by changing their strategy. The strategy pair $(Abate, Abate)$ is known as the *social optimum*, because the collective payoff (the sum of each player's payoff) is maximised. For this example the Nash equilibrium has a much lower collective payoff than the social optimum.

Brams & Kilgour (2009) describe how one way of resolving a prisoners dilemma (and other non-cooperative games) is to introduce a voting procedure, where all players are committed to choose a common strategy that is determined by a collective voting procedure. Voting transforms the game into a game with a cooperative outcome, by reducing

the size of the set of strategy combinations. But there are barriers to implementing this at an international level, because countries are highly reluctant to surrender their sovereignty.

Climate change is similar to a prisoner's dilemma, but countries don't just make a decision about whether to pollute or not, they make a decision about how much to reduce their emissions. We now describe a game, based on (Finus, 2001, Chapter 9), that models this situation.

Example 2.2 (The Global Emissions Game with Continuous Strategy Space). This game has a continuous strategy space in that each player chooses how much pollution to emit, rather than whether to pollute or not. This game describes a global pollutant, in that the the damages from the pollutant on each player depend on the total amount of pollution emitted by all of the players. This game could apply to greenhouse pollution, and also to pollutants that affect the ozone layer. This game does not examine the dynamic aspects of pollution.

Players can be thought of as countries. We assume that the set of players, N, has size n. Let e_i be the emissions from country i. The utility π_i of country i is given by

$$\pi_i = \beta_i(e_i) - \phi_i\Big(\sum_{j\in N} e_j\Big) \tag{3}$$

where β_i are the emissions benefit functions and have the property that the derivative is strictly positive ($\beta_i' > 0$) and the second derivative is not positive ($\beta_i'' \leq 0$); ϕ_i are the emissions damage functions and we assume that their derivatives is strictly positive ($\phi_i' > 0$) and the second derivative is non-negative ($\phi_i'' \geq 0$). In other words, the marginal benefits from emissions decrease with emissions, but the marginal damages from emissions increase.

To calculate the Nash equilibrium, we first work out what the best response for country i is if the emissions for all of the other countries are given. This is done by differentiating (3) with respect to e_i. The first order conditions

$$\frac{\partial \pi_i}{\partial e_i} = 0 \tag{4}$$

imply that

$$\beta_i'(e_i) = \phi_i'\Big(\sum_{j\in N} e_j\Big). \tag{5}$$

By taking the total derivative of (5) and applying the implicit function theorem, it is possible to show (see (Finus, 2001, p. 126) or (Finus, 2003, Appendix 2)) that e_i can be

expressed as a function of the emissions of the other countries. We call this function the *best reply function*, and we write $e_i = r_i(e_{-i})$ where e_{-i} is the emissions from countries other than i. It also follows that for $j \neq i$,

$$\frac{dr_i}{de_j} = \frac{\phi_i''}{\beta_i'' - \phi_i''}. \tag{6}$$

It is interesting to note that because ϕ_i'' is non-negative and β_i'' is not positive, it follows that (6) implies that if some country j reduces its emissions compared to the Nash equilibrium, then country i's best reply is to increase its emissions. This is because if j reduces their emissions, the total damages are lower, so the marginal damage function ϕ' is not as steep.

The Nash equilibrium can be obtained by substituting the best reply functions into each other and solving for the remaining variable. Suppose that the emission benefit functions are given by

$$\beta_i(e_i) = b(de_i - \frac{1}{2}e_i^2), \tag{7}$$

and the emission damage functions are given by

$$\phi_i(e_i) = \frac{c}{2}\Big(\sum_{j\in N} e_j\Big)^2. \tag{8}$$

Then the Nash equilibrium emissions are given by

$$e_i^* = \frac{bd}{b+2c}. \tag{9}$$

If there were no damages from emissions (so that $c = 0$) then the Nash equilibrium would be $e_i^* = d$. The social optimum is given by $e_i = bd/(b+4c)$. So the Nash equilibrium does involve some emission reductions, but less than optimal emission reductions.

Situations where the non-cooperative outcome is sub-optimal are known as *social dilemmas*. The above situation assumes that all participants have complete information about the payoffs for each other; assumes that decisions are made independently; does not take into account communication between the participants; and does not consider how a central authority could enforce agreements among participants about their choices. If these assumptions are not true, then it is much less certain that a suboptimal non-cooperative outcome will occur (Ostrom, 2009). When there is communication, decisions are not made independently, and participants can make enforceable agreements, cooperation may be more likely. But if participants do not have complete information, cooperation may become more difficult, because players could have an incentive to misrepresent their preferences.

2.2 Extensive Form Games and the Subgame Perfect Equilibrium

The normal form representation of a game hides the sequential nature of strategy and decision making. By contrast, extensive form games study the sequential nature of games explicitly. An extensive form game represents the game as a tree. At each node of the tree (which is sometimes referred to a 'stage' of the game), except for terminal nodes, one of the players makes a decision that determines which node is reached next. Terminal nodes determine the payoffs of the game.

Definition 2.3 An *extensive form game with perfect information* (Osborne, 2003, Chapter 5) specifies

1. the players N in the game;
2. a set of sequences of nodes in the game (*terminal histories*) with the property that no terminal history is a proper subsequence of any other terminal history;
3. a function (known as the *player function*) that assigns a player to any sequence h that is a proper subsequence of a terminal history – the player function can be thought of as specifying the player whose turn it is after h;
4. the payoffs for each player at each possible end node.

Given a history h, the set of all *actions* available to the player who moves after h is

$$A(h) = \{a \ : \ (h, a) \text{ is a history}\}.$$

A *strategy* of a player i in an extensive game with perfect information is a function that assigns an action in $A(h)$ to each history h after which it is a player i's turn to move. A strategy combination s determines a terminal history $O(s)$, known as the *outcome* of s. Associated with an extensive form game is a normal form representation that we will call the *strategic form* of the extensive form game. The strategic form has the same players and strategy combinations as the extensive form game, and the payoffs are given by the payoffs at the end nodes of each outcome of the extensive form game. If the longest terminal history of a game is finite, then we say that it has *finite horizon*.

The strategy combination s^* in an extensive game with perfect information is a *Nash equilibrium* if for every player $i \in N$ and strategy s_i,

$$\pi_i(O(s^*)) \geq \pi_i(O(s_i, s^*_{-i})). \tag{10}$$

The Nash equilibrium of an extensive form game is the Nash equilibrium of its strategic form.

Let Γ be an extensive form game with perfect information and player function P. For any non-terminal history h of Γ, the *subgame* $\Gamma(h)$ following the history h is the following extensive game:

1. the players are the same as those for Γ;
2. the terminal histories are sequences h' such that (h, h') is a terminal history of Γ;
3. the player $P(h, h')$ is assigned to the proper subhistory h' of the terminal history (h, h');
4. the payoff in $\Gamma(h)$ associated with h' is equal to the payoff in Γ associated with (h, h').

Definition 2.4 A *subgame perfect equilibrium* is a strategy combination constituting a Nash equilibrium in every subgame of the entire game. Equivalently, for every player $i \in N$, every history h after which it is player i's turn to move, and strategy s_i,

$$\pi_i(O_h(s^*)) \geq \pi_i(O_h(s_i, s^*_{-i})) \tag{11}$$

where $O_h(s)$ is the terminal history consisting of h followed by the actions generated by playing strategy s after h.

We will make extensive use of the subgame perfect equilibrium. For games with finite horizon, where there is not an infinite or indefinite amount of nodes, it is possible to calculate the subgame perfect equilibrium using a process known as *backwards induction*. The subgame perfect equilibria for the 'last' subgames are calculated first. Then taking these actions as given, we calculate the equilibria for preceding subgames and so on.

Game theoretic solution concepts such as the subgame perfect equilibrium are useful for understanding strategic behaviour, but have limitations for understanding human behaviour. For example, in a particular game known as the ultimatum game, humans behave quite differently to what has been predicted by the subgame perfect equilibrium (Güth *et al.* , 1982). It has been argued by Fehr & Gächter (2000) that the ultimatum game provides evidence that economic agents don't just base their decisions on pure self interest, and reciprocal considerations play an important role in people's actions. It has also been argued (Barrett, 2003, pp. 299–301) that the ultimatum game also provides

evidence that an international environmental agreement is more likely to be stable if it is perceived by its parties to be fair.

Example 2.3 (The Ultimatum Game). In the ultimatum game, there are two players and a sum of money. The first player proposes how to divide up the sum of money, and the second player chooses whether to accept or reject the proposal. If the second player rejects the proposal, neither player receives anything.

Assume that there is a smallest division of the sum of money available (1 cent say), that we denote by ε. Assume that the total amount of money available is equal to 1 ($1 say) and that 1 is an integer multiple of ε. The ultimatum game can be represented by an extensive form game with two stages. In the first stage the first player chooses an amount of money $x \in [0, 1]$ which is also an integer multiple of ε. In the second stage the second player chooses whether to accept the offer or not. If the second player accepts, the payoffs are $(1 - x, x)$; if not, the payoffs are $(0, 0)$.

Because the ultimatum game has finite horizon, it is possible to find the subgame perfect equilibrium using backwards induction. We first consider the subgames where the second player either accepts or rejects an offer from the first player. For any offer $x > 0$, the second player's optimal response is to accept the offer. In the subgame when the offer is $x = 0$, the second player is indifferent about whether to accept or not. There are therefore two equilibrium strategies for the second player. Either to all accept all payoffs (including $x = 0$), or to accept all payoffs except for $x = 0$.

Let us now consider the subgame perfect equilibrium strategy for the first player. There are two possibilities:

- If the second player accepts all offers, the first player's optimal strategy is to make the offer $x = 0$, and then recieve the payoff 1.
- If the second player accepts all offers except $x = 0$, the first player's optimal strategy is to make the offer $x = \varepsilon$, and recieve the payoff $1 - \varepsilon$.[2]

Both of the above possibilities are subgame perfect equilibria, but unless the first player is certain that the second player will accept all offers including $x = 0$, they are better off making the offer $x = \varepsilon$.

Let us now characterise the Nash equilibria of the ultimatum game. The first player chooses an amount x in the unit interval $[0, 1]$ that is a multiple of ε. The second player

[2]If there was no smallest division of the sum of money, then no offer $x > 0$ would be optimal, because $x/2$ would be better. In this case the only subgame perfect equilibrium corresponds to the offer $x = 0$.

chooses a function

$$f : [0, 1] \mapsto \{Accept, Reject\}.$$

A strategy combination (x, f) is a Nash equilibrium if $f(x) = Accept$ and there is no $y < x$ such that $f(y) = Accept$. The first player would not want to decrease their offer because the second would reject it; the second would not want to reject the offer because then they would get nothing. Another Nash equilibrium is the combination $x = 0$, and $f(x) = Reject$ for all x. So any possible off could be a Nash equilibrium. In this sense the Nash equilibrium is a weaker concept than the subgame perfect equilibrium.

Experiments where people have played the ultimatum game have consistently found that the first player will usually offer significantly more money to the other player than the subgame perfect equilibrium, and the second player will be unlikely to accept the offer if they are offered less than 30 per cent of the total amount (Güth *et al.* , 1982).

Equity considerations play an important role in many proposals for how greenhouse gas emissions could be allocated in a post-Kyoto agreement. They are one of the reasons why many proposals (such as Baer *et al.* (2000); Garnaut (2008b); Meyer (2000)) suggest that all countries should eventually be allocated the same amount of per-capita emissions. A shorter transition to equal per-capita emissions would be fairer than a longer transition because a longer transition rewards high per-capita emitters for having high per-capita emissions. But even a very short transition to equal per-capita emissions could be considered to be unfair because different countries have different historical emissions. Stern (2009) states (p. 153) that

> To suggest that we should all be entitled to emit roughly equal amounts by 2050 is to say that, at the end of the drinking spree, we should be using glasses of the same size. It is difficult to see this as a particularly equitable division of the entitlements to the reservoir, since this type of equality takes no account of all the 'drinking' that has gone on over the previous two hundred years.

One alternative approach is for a global total emissions budget that takes into account historical emissions (Pan *et al.* , 2000; Project Team of the Development Research Centre of the State Council, 2009). But it would mean that many countries (such as the United States) would have already used up significantly more than their emissions budget, and would have to purchase their allowances off other countries. It would be extremely unlikely that the United States Senate would ratify such an arrangement.

There are several ways of characterising outcomes that could be considered to be fair. If n different players are dividing up a good (such as a cake), then an outcome is *proportional* if each player perceives that they get a portion that is at least $1/n$ of the good. An outcome is *envy-free* if no player prefers the outcome for another player to their own outcome. A procedure (such as an extensive form game) is proportional or envy-free if it will lead to outcomes that are proportional or envy-free. An example of a two player game that is envy-free and proportional is 'divide and choose one player cuts a cake in half, and the other chooses a piece. Brams & Taylor (1996) have surveyed envy-free and proportional mechanisms for multiple players in detail.

Fleurbaey (1994) relates axiomatic work on fairness to the situation where each player has both non-transferable 'personal resources' and transferable 'external resources'. He examines envy-free allocations that take into account both type of resources and shows that envy-free allocations satisfy various axioms that are consistent with the idea of 'equal treatment of equals'. Both axiomatic and algorithmic approaches to fairness could provide useful insights to equity issues in climate negotiations. Because an agreement is more likely to be stable if it is perceived to be fair, this could have implications for the stability of climate agreements.

It is possible to modify Definition 2.3 so that players can make simultaneous moves. Instead of having the player function assign a player to a subhistory, it assigns a set of players. The game also needs to be consistent – the actions corresponding to a subhistory is the same as the actions of the players assigned by the player function to that subhistory. The reader is referred to (Osborne, 2003, p. 206) or (Osborne & Rubinstein, 1994, p. 102) for the formal definition of an extensive form game with perfect information and simultaneous moves.

Example 2.4 (The Treaty Participation Game). This example is based on Chapter 7 of Barrett (2003). This and related games are sometimes known as conjectural variation models (Finus, 2001, Section 13.2), cartel formation games, or open membership single coalition games (Finus & Rundshagen, 2003). We consider the situation that there are two players and the final payoffs are the same as for the prisoner's dilemma (Example 2.1). This game can be divided into three stages.

Stage 1 All players simultaneously choose whether to be a signatory or a non-signatory.

Stage 2 Signatories choose whether to play *Abate* or *Pollute*, with the objective of maximising their collective payoff.

Stage 3 Non-signatories choose simultaneously whether to play *Abate* or *Pollute*.

The subgame perfect equilibrium can be determined by backwards induction, so consider Stage 3 first. The Nash equilibrium of the prisoner's dilemma is for players to play *Pollute*, so non-signatories will play *Pollute*.

We now consider the Stage 2 subgame. If there is one signatory, they will anticipate that the non-signatory will play *Pollute* in Stage 3, and so will also play *Pollute*. If both countries are signatories, they will collectively choose to play *Abate*, because that will maximise their collective payoff.

In the Stage 1 game, if country Y decides not to become a signatory, then country X is indifferent about becoming a signatory. If country Y decides to become a signatory, country X is strictly better off if it becomes a signatory. Country X is therefore not worse off by becoming a signatory regardless of the other players strategy. The subgame perfect equilibrium therefore has all countries becoming signatories. When countries can make a continuous choice about their abatement, they will still choose the optimal abatement level (Barrett, 2003, p. 207).

The extension of the treaty participation game to more than two players has been investigated in (Barrett, 1994) and (Barrett, 2003, Chapter 7). Using a framework similar to that of Example 2.2, Barrett considers an agreement where signatories maximise their collective benefits, while non-signatories maximise their individual benefits. Each player is assumed to have the same emissions cost and benefit functions, we assume that they satisfy the properties described in Example 2.2. Suppose that there are n players, and α is the proportion of players that sign an international environmental agreement, so that it has $n\alpha$ signatories. Let $\pi_n(\alpha)$ be the payoff for a non-signatory, and let $\pi_s(\alpha)$ be the payoff for a signatory. An international environmental agreement is said to be *self-enforcing* if

$$\pi_n(\alpha - 1/n) \leq \pi_s(\alpha) \quad \text{and} \quad \pi_n(\alpha) \geq \pi_s(\alpha + 1/n). \tag{12}$$

In other words, an agreement is self-enforcing if no signatory can benefit from dropping out of the agreement and no non-signatory can benefit from joining the agreement. Barrett found that self-enforcing agreements would be likely to have significantly less that full participation. A similar result has been obtained by Carraro & Siniscalco (1993).

This illustrates a serious barrier to full international cooperation – even when there is an international agreement, countries can have an incentive to not comply with the agreement, or to not participate in the agreement, possibly by dropping out of the agreement.

Measures that may encourage compliance and participation include reciprocal measures, side payments, issue linkage, and trade restrictions (Barrett & Stavins, 2003)[3]. One possible reciprocal measure is for countries to reduce their emissions by a lower amount if there is less participation (Barrett, 2003, Chapter 11). Another possible method is to threaten to dissolve the treaty altogether (see Chander & Tulkens (1997) or Chapter 10 of Barrett (2003)). The problem with these punishments in the context of greenhouse gas emissions is that they hurt signatories as much as non-signatories. Threats to substantially increase greenhouse gas emissions are unlikely to be credible and involves impacts that are experienced decades into the future. An alternative way to punish non-cooperation is to link cooperation with another issue, such as trade. Another issue that can be linked to cooperation on reducing emissions is cooperation on research and development. It may however be difficult to prevent the benefits from research and development cooperation from spilling over to other countries (Barrett, 2003, p. 310).

Cooperation on global warming is automatically linked to trade through a phenomenon known as *carbon leakage*. If a country unilaterally reduces emissions, it could lead to reduced production of some internationally traded emissions intensive goods. This can in turn increase the price of the good. The increased price could then drive increased production of the good in an overseas country that has not reduced its emissions, leading to economic benefits and an increase in emissions for the non-cooperating country.

There are several ways that trade can be linked with cooperation. One way is through trade restrictions. There is a precedent for this – trade restrictions were incorporated into the Montreal Protocol on Ozone Depleting Substances. It has been suggested that the trade restrictions "were indispensable to the protocol's effectiveness" and also helped to drive the ratification process (Benedick, 1991).

The issues of carbon leakage and free riding can also be addressed through border tax adjustments. When a country has a price on carbon, a border tax adjustment consists of either: (i) the imposition of a carbon price on imported products that corresponds to a similar tax borne by domestic products; and/or (ii) an exemption from paying a carbon price for the production of exported products. It is likely that border tax adjustments would be allowed under World Trade Organisation rules (Tamiotti *et al.* , 2009). Under the Montreal Protocol, countries accounted for their production of ozone depleting substances, subtracted their exports, and added their imports. Countries were effectively accounting for their consumption of ozone depleting substances. If a country applies border tax

[3]Trade restrictions can also be thought of as a form of issue linkage

adjustments on both exports and imports when it imposes a carbon price, it is effectively putting a price on the *consumption* of emission intensive goods rather than the *production* of emissions.

Measures such as border tax adjustments could discourage free-riding, but there are risks if they are implemented in a way that is not considered to be fair. For example, suppose that the United States imposed border tax adjustments on steel imports from India, whose per-capita emissions are over ten times lower than the United States. This would be widely perceived to be unfair, would increase tension between developed and developing countries, and undermine cooperation.

Barrett (1997) examined the role of trade sanctions by analysing a game structure involving both countries and polluting firms. Barrett found than for some choices of parameters, when there were trade sanctions there would be two equilibria. One with no signatories and one with all countries being signatories. The equilibrium with everyone being signatories is preferable and this one can be realised by introducing a minimum participation level into the treaty. The treaty only becomes effective if at least a minimum amount of countries have become signatories. A similar result was obtained by Lessmann *et al.* (2009), who used an integrated assessment model and found that the imposition of tariffs would increase the level of participation of a treaty. The role of minimum participation has been examined in more detail by Carraro *et al.* (2009), who investigated extensive form games that have an initial stage where countries decide on a minimum participation level.

It is also possible to link trade with cooperation by applying a tax to fossil fuels that are exported to a non-cooperating country. Hoel (1994) has suggested that policies that affect both the supply and demand of fossil fuels are superior to policies that affect only the supply or only the demand of fossil fuels. A cartel that exports fossil fuels will capture less rents if other countries reduce their consumption due to an international climate agreement. It would then be in the interests of the cartel to apply a tax on the exported fossil fuel (Bråten & Golombek, 1998). A final way that trade is linked to cooperation is in international negotiations through implicit or explicit threats to directly link trade to cooperation.

If an in an international climate agreement is self-enforcing, for reasons to do with issue linkage or otherwise, will the agreed targets be more likely to be close to socially optimal, or less likely? A related question is whether binding or non-binding targets are more likely to be strong targets. Game theory suggests that when an agreement is self-

enforcing, players will act under the assumption that other players will comply with the agreement; when an agreement is not, players are likely to assume that other player will not comply. If an agreement had strong penalties for non-participation, countries may be willing to accept targets than they would otherwise accept in order to participate. This may suggest that binding targets are more likely to be close to socially optimal targets than non-binding ones.

However, when countries agree to binding targets, the risks associated with these targets being costly is greater. There is less risk associated with a country agreeing to a non-binding target, because if a non-binding target is difficult to comply with, little is lost by not complying. Victor (2007) asserts that with international cooperation on the North Sea, the Baltic Sea, and acid rain in Europe, nonbinding commitments backed by senior politicians were more effective than binding commitments. For the European acid rain regime, ambitious non-binding commitments to reduce nitric oxide and nitrogen dioxide pollutants were adopted by a smaller number of countries alongside a less ambitious binding convention to address the same pollutant. A domestic mechanism for implementing such an approach is described in Section 4 of Wood & Jotzo (2009).

One way to increase the likelihood of a cooperative outcome in a game is to repeat it. Repeated prisoner's dilemmas are discussed in (Osborne & Rubinstein, 1994, Chapter 8) and (Finus, 2001, Chapter 5). For repeated prisoner's dilemmas with finite horizon, the only Nash equilibrium consists of players not cooperating in each turn of the game. When games have infinite horizon the 'folk theorems' of game theory tell us that these games have a huge amount of different subgame perfect equilibria. These results suggest that cooperative behaviour is more likely if players have a long term perspective, and have a strategy for punishing players who do not cooperate. Because repeated games often have a large amount of subgame perfect equilibira, a stronger concept, known as the 'renegotiation proof equilibrium' has been developed (Farrell & Maskin, 1989).

Axelrod (1984) uses an experimental approach to study repeated versions of a prisoner's dilemma that used the following payoff matrix:

$$\begin{array}{cccc} & & \text{Player 2} & \\ & & \textit{Cooperate} & \textit{Defect} \\ \text{Player 1} & \textit{Cooperate} & (3,3) & (0,5) \\ & \textit{Defect} & (5,0) & (1,1) \end{array} . \tag{13}$$

Axelrod organised two computer tournaments where players would submit algorithms that determine whether to play a cooperative or noncooperative choice on each move, taking

into account the history of the game so far. The first tournament received 14 entries and each game would consist of 200 moves. The second tournament received 62 entries, this time each game would have a 0.00346 chance of ending after each move (so the game would not have finite horizon).

In both tournaments an algorithm called *Tit for Tat* won. *Tit for Tat* starts by cooperating, then in subsequent moves it plays the preceding move played by its opponent. Axelrod analysed the highest scoring strategies and found that they would have several properties in common: they were *nice*, in that they would not defect before their opponent does; they were *forgiving*, they would fall back to cooperating if their opponent does not continue to defect; but they would also be *retaliatory* in that they would immediately defect after an "uncalled for" defection from the other player.

Because greenhouse gas emissions are an ongoing process, and climate negotiations are a repeated process, the problem of climate change mitigation is in many ways like a repeated game. For this reason, cooperation is more likely than it is for a prisoners dilemma. But there are important ways in which climate change is not a repeated game. The damages from greenhouse gas emissions depend largely on cumulative emissions, they are also largely experienced in the future, to the extent that they also affect future generations.

The games that so far have been described in this section treat countries as the players in the game. But the subgame perfect equilibrium can also tell us about the interplay between international politics and domestic politics in climate negotiations. After an international treaty is negotiated, it then has to be ratified by its participants. This can be modelled as a two stage game. In Stage 1, the players negotiate the treaty; in Stage 2, each country decides whether to ratify the treaty. For some countries, for example the United States, ratification can be difficult. The United States requires 67 out of 100 Senate votes in order to ratify a treaty. By backwards induction, for negotiators in Stage 1 to play the subgame perfect equilibrium, they will take into account that a treaty will have to be sufficiently aligned with the domestic interests of the United States, in order for it to be ratified by the United States (Barrett, 2003, p. 148).

3 Coalitions

There have been debates in the game theory literature on whether a cooperative outcome is more likely to arise from a 'grand coalition' of all countries, or from smaller coalitions.

Game theory analyses coalitional behaviour from a variety of perspectives. One such perspective is a cooperative game theory approach, which we examine in Section 3.1. Another perspective is described in Section 3.2 where we examine non-cooperative approaches to coalition formation, and the role of externalities.

3.1 Cooperative Game Theory and the Core

Cooperative game theory investigates situations where groups of players may form coalitions that enforce cooperative behaviour. For cooperative games, the outcomes of interest consist of a partition of the players into coalitions, and actions for each coalition. Players in a coalition behave cooperatively with each other, and non-cooperatively with respect to other players and coalitions. The core is a concept that can be used to analyse the stability of a grand coalition of all players.

Definition 3.1 Let N be a set of n players. A *coalition* is a subset S of N. A *payoff vector* (also known as an imputation) for N is an n-dimensional real vector $\pi = (\pi_1, \ldots, \pi_n)$, and we write $\pi(S) = \sum_{i \in S} \pi_i$ for any coalition $S \subseteq N$. A *characteristic function* υ (also known as a coalitional function) is a function which assigns a real number to each coalition.

An n-player *game in coalitional form with transferrable utility* (also called a TU-game) is defined by a set of players N, and characteristic function υ, and denoted (N, υ). The *core* of (N, υ) is defined by

$$C(N, \upsilon) = \{\ \pi : \pi(N) = \upsilon(N) \text{ and } \pi(S) \geq \upsilon(S) \text{ for all } S \subseteq N\ \}. \tag{14}$$

The core is the set of possible outcomes in which no coalition can break away from a grand coalition in such a way that all of its members are better off. The core, being a set, always exists, but can be empty.

Example 3.1 (The γ-Core of Chander & Tulkens (1997)). This example is based on (Chander & Tulkens, 1997), which is also discussed in Chapter 13 of (Finus, 2003) and Chander & Tulkens (2008). We use the same basic framework as in Example 2.2. Let $\pi_i(e^S, e^{N \setminus S})$ be the payoff for a country i in a coalition S which has e^S emissions, and with the other countries emitting $e^{N \setminus S}$ emissions. Assume that each of the countries in $N \setminus S$ maximise their individual benefits, while countries in S maximise their collective benefits. The γ-characteristic function of a coalition S is the sum of the utilities of each member os S, assuming that members of $N \setminus S$ behave non-cooperatively. It is given by

$$\upsilon_\gamma(S) = \sum_{i \in S} \pi_i(e^S, e^{N \setminus S}). \tag{15}$$

The core of the associated TU-game can be thought of as the set of possible payoff vectors for the countries in a grand coalition where no coalition will benefit if the grand coalition dissolves into singletons when any coalition breaks away from it. The payoffs depend both on a country's emissions and a transfer t_i of payments received by country i that satisfies $\sum_{i\in N} t_i = 0$. The total payoff for country i is given by

$$\pi_i = \beta_i(e_i) - \phi_i\Big(\sum_{j\in N} e_j\Big) + t_i. \tag{16}$$

Chander and Tulkens show that the γ-core is non-empty by constructing a payoff vector that is contained in it. Let $\bar{e}_i$ be country i's Nash equilibrium emissions and let e_i^* be country i's social optimum emissions. The values for t_i chosen are

$$t_i = \Big(\beta_i(\bar{e}_i) - \beta_i(e_i^*)\Big) - \frac{\phi_i'(\sum_{j\in N} e_j^*)}{\sum_{k\in N} \phi_k'(\sum_{j\in N} e_j^*)}\Big(\sum_{k\in N} \beta_k(\bar{e}_k) - \beta_k(e_k^*)\Big). \tag{17}$$

This choice of t_i corresponds to an element of the γ-core if any of the following conditions hold:

1. damage functions are linear;
2. for all $S \subset N$ with $|N\backslash S| \geq 2$, and for all $i \in S$, $\sum_{k\in N\backslash S} \phi_k'(e^*) \geq \phi_i'(\bar{e})$; or
3. countries are symmetric.

The result of Example 3.1 suggests that socially optimal emission reductions could be possible, but it has been questioned whether this outcome is feasible. The threat that each countries will break into singletons if one or more countries leave the grand coalition may not be credible. Finus pointed out (Finus, 2001, Section 13.3.3) that the cost sharing rule provides countries with an obvious incentive to misrepresent their environmental preferences and abatement costs.

Although the cost sharing rule (17) may not be practical or feasible, it is still important because it demonstrates that the core can be non-empty. This is significant because it has been shown (Serrano, 1995), (Okada & Winter, 2002) that it is possible to design extensive form games (which can be thought of as a bargaining game) whose subgame perfect equilibria are elements of the core. This relates to the 'Nash program' (Nash, 1953; Serrano, 1997) to link cooperative and non-cooperative game theory by finding non-cooperative procedures that yield cooperative outcomes as their solution concepts.

We note that the core for a global warming game that does not assume that countries in $N\backslash S$ dissolve into singletons has been studied by Uzawa (2003). In this case the core may be empty. Uzawa also investigated the situation where utility is non-transferrable.

3.2 Coalition Formation and Externalities

The fully cooperative result from Chander and Tulkens described in Example 3.1 contrasts with the less cooperative results from Barrett (1994) and Carraro & Siniscalco (1993) that we discussed in Section 2.2. This has lead to a debate in the game theory literature about whether cooperation on climate change is best achieved among all countries working together, or among smaller coalitions. The debate has been surveyed by Tulkens (1998) and ten years later by Chander & Tulkens (2008). Tulkens (1998) described the results of Barrett (1994) and Carraro & Siniscalco (1993) as the small stable coalitions (SSC) thesis, and the results of Chander & Tulkens (1997) as the grand stable coalition (GSC) thesis. The role of coalitions in the different approaches is different – under the SSC approach, the 'bad guys' who do not cooperate are singletons, outside of any coalition; under the GSC approach, the 'bad guys' who do not cooperate form a coalition.

When there are coalitional externalities, assumptions about the coalitions that do not contain a particular player change the value of the characteristic function for that player. This in important when analysing issues such as the core, and the stability of a grand coalition. An alternative to using characteristic functions is a 'partition function' that also takes as its input a partition of the other players into coalitions.

The approach of Barrett (1994) and Carraro & Siniscalco (1993) has the property that the number of non-trivial coalitions is restricted to one; the use of a partition function facilitates going beyond this assumption. The following definition of a partition function is from Maskin (2003).

Definition 3.2 Let N be a set of n players and let $\mathscr{C}$ be a partition of N into disjoint coalitions. For each partition $\mathscr{C}$ and coalition $C \in \mathscr{C}$, the *partition function* $v(\cdot,\cdot)$ assigns a number $v(C,\mathscr{C})$, which is interpreted as the payoff for coalition C given the partition $\mathscr{C}$.

Finus & Rundshagen (2003) have applied partition functions to climate change coalitions. They consider a two-stage game, each stage can also be analysed as a game: in the first stage countries choose their coalitions; and in the second stage, coalitions choose their optimal strategy. They consider a large variety of different approaches to how countries choose their coalitions, including the approach of Barrett (1994). These approaches model the process of coalition formation as an extensive form non-cooperative game. The size and nature of the coalitions that form depend very much on this process. Some of these processes (such as the Barrett (1994) approach) have very small coalitions, but in

some cases a grand coalition was possible. Buchner & Carraro (2006) have also used this two-stage process, and incorporated it with a six-region economic model *FEEM-RICE*. How coalition formation can be treated as a non-cooperative game has been discussed in more general context by Bloch (1996), Ray & Vohra (1997), Yi (1997), and Maskin (2003). Yi (1997) also found that different rules of coalition formation lead to different predictions about stable coalition structures.

For some games coalition formation imposes a positive or negative externality on other players (Maskin, 2003), (Yi, 1997), (de Clippel & Serrano, 2008). With the basic framework that we use to analyse climate change (Example 2.2), coalition formation imposes a positive externality – when a group of countries form a coalition, their emissions will be lower than when they act individually in a non-cooperative way.

Definition 3.3 Let $\mathscr{C}$ be a partition of a set N of n players into disjoint coalitions. Let $C_1, C_2 \in \mathscr{C}$ be two of these coalitions, and let $\mathscr{C}_{12}$ be the partition that forms when C_1 and C_2 merge. For some other coalition $C \in \mathscr{C}$, we say that C_1 and C_2 impose *no coalition externality* on C if merging has no effect, i.e.

$$\upsilon(C, \mathscr{C}) = \upsilon(C, \mathscr{C}_{12});$$

the coalitions C_1 and C_2 impose a *positive coalition externality* on C if merging increases C's payoff

$$\upsilon(C, \mathscr{C}) < \upsilon(C, \mathscr{C}_{12});$$

the coalitions C_1 and C_2 impose a *negative coalition externality* on C if merging decreases C's payoff

$$\upsilon(C, \mathscr{C}) < \upsilon(C, \mathscr{C}_{12}).$$

A large variety of non-cooperative processes for coalition formation – the first stage of the games studied by Finus & Rundshagen (2003) – have been investigated. Some of them involve players making simultaneous moves, some involve sequential moves. WIth the exception of the treaty participation game, these games can lead to partitions with more than one non-trivial coalition. We list some of these games below:

- The *treaty participation game* that was described in Section 2.2 has been studied by Carraro & Siniscalco (1993), Barrett (1994), Finus & Rundshagen (2003) and others. A variant of this game is where players also consider the impact of other defections that could arise if a player defects from a coalition (Carraro & Moriconi, 1997; Finus & Rundshagen, 2003), and leads to a more cooperative outcome.

- The *equilibrium binding agreement game* was introduced by Ray & Vohra (1997) and has also been studied by Finus & Rundshagen (2003). The starting point is a grand coalition C. Then a smaller coalition, c, may split away from the grand coalition if it improves its collective payoff by doing so. In the following step, members of either c or $C\backslash c$ may propose further deviations. This process continues until no group of players want to split up into a finer partition. Such a partition in known as an equilibrium binding agreement.

- *Open membership games* have been studied by Yi (1997) and Finus & Rundshagen (2003). These games model an environment where players can freely join coalitions and no outsider is excluded from a coalition. In this game, players simultaneously announce an 'address', and players that announce the same address are in the same coalition.

- *Exclusive membership games* have been studied by Finus & Rundshagen (2003), Hart & Kurz (1983), Yi (1997), and Yi & Shin (2000). Players first simultaneously list the players who they are willing to join a coalition with. In one type of exclusive membership game, known as the Δ-Game, two players are in the same coalition if and only if they are on each others list. In another exclusive membership game, the Γ-Game, a group of players are in the same coalition if and only if their lists are all identical. In their model with symmetric countries, Finus and Rundshagen found that larger coalitions were sustained by the Γ-Game than by the open membership game or the treaty participation game.

- Bloch (1996) and Finus & Rundshagen (2003) have examined the *sequential move unanimity game*. We start with an exogenous ordering of players. The first player proposes a coalition to which they would like to belong; each prospective member then is asked (according to the same ordering) whether they accept the proposal; if all proposed members agree, then a coalition is formed and the remaining players may form a coalition according to the same process; if a proposed member disagrees, they can then propose their own coalition.

- Maskin (2003) introduced a sequential process that is also based on an exogenous ordering, and proved that when externalities were negative, a grand coalition forms (for up to three players). A counter-example was provided by de Clippel & Serrano (2008) to this statement when there was more than three players. Maskin also

provided examples of positive externality games where a grand coalition would not form.

- Aghion *et al.* (2007) compared two specific bargaining processes, in order to understand whether multilateral approaches are more likely to lead to cooperation on trade or bilateral processes were. They only modelled three players, and found that for the processes that they investigated, a grand coalition would form, even when coalitional externalities were positive.

Table 6.2 of Finus & Rundshagen (2003) compared the equilibrium coalition structures for some of the processes above. The coalition structures for the treaty participation game were the least cooperative compared to the others.

The processes described above is a non-cooperative approach to coalition formation. A significant question in game theory is which non-cooperative processes can implement concepts in cooperative game theory. We will discuss how to design non-cooperative games with cooperative solutions in the next section.

The fact that more cooperation is likely to occur with exclusive membership games than with open membership games and the the treaty participation game could have implications for how to get the most cooperation from a coalition formation process. In some ways, the exclusive membership games are similar to what arises when countries with emissions trading schemes are considering the possibility of linking their carbon markets. Countries that establish an emissions trading scheme may want to it it with those of other countries for efficiency reasons. But countries would be reluctant to link with a country whose emissions trading rules are significantly different (Jotzo & Betz, 2009), or whose mitigation commitment is much less ambitious. This suggests that carbon market linkage has important strategic implications.

For example, suppose that a country A was considering linking its emissions trading scheme to country B, and that B already has its emissions trading scheme linked to that of another country C. Suppose that A does not want to allow a certain type of offset to be used as a compliance mechanism, but C allowed the use of this offset. A would then be highly reluctant to link its emissions trading scheme to that of B, because B could import permits from C, whose carbon price could be influenced by these particular offsets. In other words, each country has a 'list' of who they are willing to link with, and link their carbon markets if and only if they are on each others list. This is exactly the situation described by the Δ-Game above.

4 Implementation Theory

Implementation theory addresses the key game-theoretic question that needs to be answered in order to address a social dilemma. How can non-cooperative games be designed so that their solution (often a Nash equilibrium or subgame perfect equilibirum) corresponds to a socially optimal outcome? After a brief formal treatment of the concepts from implementation theory, we will examine some mechanisms that relate to public good provision or pollution reduction, and discuss their policy relevance for climate change mitigation. The reader is referred to Jackson (2001) for a more detailed summary of the main concepts of implementation theory.

Let N be a set of n *players*, and let A be a set of possible *outcomes*. Let a player i have a preference relation R_i on A; if player i prefers an outcome a to another outcome b, or is indifferent, we say that aR_ib. An example of a preference relation is when each player i assigns a utility u_i to each outcome, in which case, aR_ib if and only if $u_i(a) \geq u_i(b)$.

A *social choice correspondence* F maps profiles of preferences $R = \{R_1, \ldots, R_n\}$ into the set of outcomes, i.e. $F(R) \subset A$. When $F(R)$ is a singleton, F is called a *social choice function*. A social correspondence tells us what outcomes are desirable, given a preference profile. We have made extensive use of the *social optimum*, a social choice function that maximises the sum of the utilities of each player. Other examples of social choice correspondences include the properties of *proportionality* and being *envy-free*, that were discussed in Section 2.

A *mechanism* or *game form* is a pair (M, g) consisting of a product of 'message spaces' or 'strategies' $M = M_1 \times \ldots \times M_n$, and an *outcome function* $g : M \to A$. The main difference between a mechanism an non-cooperative game is that the result of the mechanism is given by an outcome, rather than a payoff. A *solution concept* S specifies the behaviour of players who have preferences R, given a mechanism (M, g). Given (M, g, R), S specifies a subset of M. The outcome function will then lead to an *outcome correspondence* that is given by

$$O_S(M, g, R) = \{a \in A \ : \ \text{there exists } m \in S(M, g, R) \text{ such that } g(m) = a\}. \qquad (18)$$

Important examples of solution concepts include the Nash equilibrium and the subgame perfect equilibrium.

A social choice correspondence is *implemented by the mechanism* (M, g) via a solution concept S if the outcome correspondence coincides with the image of the social choice

correspondence. In other words,

$$O_S(M, g, R) = F(R). \tag{19}$$

A field that is closely related to implementation theory is *mechanism design*. The mechanism design problem involves finding mechanisms where the outcome correspondence contains the the social choice correspondence, but where there could be other solutions as well, i.e. $O_S(M, g, R) \supset F(R)$.

The use of subgame perfect equilibrium as a solution concept is particularly important, because there exist situations where a choice function cannot be implemented in a single stage via Nash equilibrium, but can be implemented in several stages via subgame perfect equilibrium (Moore & Repullo, 1988).

Example 4.1 This illustrative example is based on Moore & Repullo (1988). Suppose that there is a club with a set N of members that are designing a constitution – a mechanism (M, g) for making decisions. This mechanism could, for example, be a voting procedure, or a consensus based decision procedure. A social choice function F, together with the member's preferences R, determine the decision $F(R)$ that would be preferred. The members preferences R, may be known to each other, but unknown to outsiders, such as a court. For this reason, instead of directly using $F(R)$ to make a decision, the constitution specifies an outcome function g, based on messages M, both of which can be verified.

An interesting property of this mechanism is that there is no social planner, such as a government, that implements the mechanism. An example of such a club could be the UNFCCC, where the decision making body (the 'conference of parties') mostly makes decisions using consensus.

Because mitigation of climate change is a global public good, it is useful for us to consider non-cooperative games whose solutions implement a public good. We shall now consider some more examples of games that do this.

Example 4.2 (Provision Point Mechanisms). Bagnoli & Lipman (1989) describe a relatively simple game for providing public goods by using voluntary contributions. Each player voluntarily commits any amount of their choice towards the cost of the public good. The public good is considered to be discrete – the example of a single streetlight or multiple streetlights is described. Players 'pledge' to make contributions

towards completion of the project. If the total amount of contributions is enough to provide the public good, then players must pay and the good is provided. If the total amount of contributions is not enough, each player's contribution is refunded and the public good is not provided. Bagnoli and Lipman model this process with a normal form game. They show that this game has a solution that satisfies a solution concept known as 'undominated perfect equilibrium'. This solution provides the public good and implements the core of the economy.

An extensive form version of this game is described by Admati & Perry (1991). They call this game the *subscription game*. For simplicity, assume that there are two players. Players alternate in pledging contributions to complete the project. The game ends if and when the total amount of contributions exceeds the cost of the good. Let c_i be the total contribution from player i, let k be the cost of the public good, let v be the benefit of the public good for each player and let T be the first time such that $c_1 + c_2 > k$. We assume that the payoffs for each player are given by

$$\pi_i(T, c_1, c_2) = \begin{cases} \delta^T(v - c_i) & \text{if } c_1 + c_2 \geq k \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

Admati and Perry prove that the subgame perfect equilibrium of this game is a cooperative outcome. Admati and Perry also consider the game where players are not refunded their commitments if the good is not provided. In this case, in equilibrium the good will not be provided unless the value of the good to each player is greater than the cost of the good.

Marks & Croson (1998) have performed experiments which suggest that the provision point mechanism can be successful. Advantages of the subscription game are that it is reasonably simple, and that it does not require a strong sanctioning institution such as a government that can enforce the desired contributions to public goods.[4] Bagnoli and Lipman state

> "Even the analysis of mechanisms which are put forth as 'plausibly useful', such as Groves-Clarke taxes, is focused on mechanisms that a government might actually wish to impose and rarely on mechanisms which private individuals might jointly use. Perhaps for this reason, the literature on private provision of public goods has basically ignored the implementation literature,

[4]Some sort of institution may still be required to ensure players do not renege on their commitments, and to provide the public good once it has been paid for.

> hypothesized particular games, and demonstrated, among other things, that these games do not have efficient outcomes." (Bagnoli & Lipman, 1989, p. 596)

Example 4.3 (Using Bargaining to Resolve a Non-cooperative Game). Attanasi *et al.* (2010) develop a bargaining process that they call a *confirmed proposal* mechanism that can lead to cooperative outcomes. They describe two mechanisms: a game form with 'confirmed conditional proposals', and a game form with 'confirmed unconditional proposals'. We describe the game form with confirmed conditional proposals below. There is an underlying game (such as a prisoner's dilemma) that determines the payoffs for each player. There are two players, with 'strategy spaces' S_1, S_2 for the underlying game. Pairs of strategies can be thought of as outcomes, and the underlying game treats these outcomes as strategies that determine utility functions (which in turn determine the players' preference profiles). The mechanism proceeds as follows:

Stage 1.1 Player 1 communicates to Player 2 their intention to follow a strategy $s_1^1 \in S_1$, if the bargaining process arrives at an agreement.

Stage 1.2 Player 2 responds to Player 1's proposal by communicating their intention to follow strategy $s_2^1 \in S_2$ if Player 1 is willing to confirm their strategy.

Stage 1.3 Player 1 has a choice about whether to confirm their strategy or not. If so, then the two players choose strategies (s_1^1, s_2^1); if not, the players proceed to Stage 2.

Stage 2.1 The reply of Player 2 in Stage 1.2 becomes Player 2's new proposal, i.e. $s_2^2 = s_2^1$.

Stage 2.2 Player 1 announces their intention to follow strategy $s_1^2 \in S_1$, which must be different from their proposal from Stage 1.1 (i.e. $s_1^2 \neq s_1^1$).

Stage 2.3 Player 2 must choose whether or not to confirm the strategy profile (s_1^2, s_2^2). If so, the bargaining process ends; if not, they return to Stage 1, but with the proposal of Player 1 in Stage 1.1 being the same as their proposal from Stage 2.2, and the proposal of Player 2 in Stage 1.2 being different from their proposal in Stage 2.1.

Attanasi *et al.* (2010) show that when the underlying game is a prisoner's dilemma (so that the players' preference profiles lead to a prisoner's dilemma), the game has a subgame perfect equilibrium that induces the cooperative outcome in the first bargaining stage.

In other words, the cooperative outcome is implemented by the confirmed conditional proposal mechanism, when the player's preferences lead to a prisoner's dilemma. They also found that experiments using this mechanism with human subjects sustained high amounts of cooperation.

The above bargaining mechanism has some similarities to scenarios that can occur in international climate negotiations. Sometimes a country will state what they are prepared to so as part of a 'comprehensive international agreement' or something similar. They may later confirm whether their proposal is an actual commitment or not, depending on the proposals by other parties.

Another form of conditionality that takes place during climate negotiations is when countries state that they will make an unconditional commitment, but are willing to increase their emission reductions based on the commitments of others. For example, at Copenhagen the EU had an unconditional commitment to reduce its emissions by 20 percent compared to 1990 levels by 2020, but would be willing to reduce its emissions by 30 percent compared to 1990 levels if there was a sufficient commitment from other countries. Australia made an unconditional commitment to reduce its emission by 5 percent compared to 1990 levels by 2020, and a commitment to increase that by up to 15 percent if certain commitments were met, and to 25 percent if certain other conditions were met. These kinds of approaches are in many ways similar to the provision point mechanism of Example 4.2, but are also similar to the 'matching abatement commitment' approach described below.

Example 4.4 (Matching Abatement Commitments). A game where players commit to reducing their emissions by a multiple of other player's targets on top of their unconditional commitments is considered by Boadway *et al.* (2009). They apply a mechanism that is originally due to Guttman (1978) to climate change mitigation. Each country chooses a *matching rate* and its level of *direct abatement*. The game proceeds as follows:

Stage 1 Each country i simultaneously chooses matching rates m_{ij} that correspond to country j's direct abatement.

Stage 2 Each country i simultaneously chooses their direct abatement levels a_i. After Stage 2, the total abatement commitment of country i is

$$A_i = a_i + \sum_{j \neq i} m_{ij} a_j. \tag{21}$$

Stage 3 Countries engage in trading of their emission quotas to equalise the marginal benefits of emissions across all countries.

Boadway *et al.* (2009) show that when the preferences of the players are as described in Example 2.2, the subgame perfect equilibrium of this process achieves the efficient level of pollution abatement. They extend their model to a situation with two time periods, and treat the pollutant as a stock pollutant (so that it can build up in the atmosphere). They show that the above process also efficiently allocates emissions across periods.

The previous three mechanisms all have cooperative outcomes, and all are based on some sort of conditionality. The fact that their solutions are cooperative suggests that a cooperative approach to climate mitigation is possible. This contrasts with less optimistic views, such as from Brennan (2009), who states that the grounds for hope are "decidedly thin". The mechanism from Boadway et al. (2009) is particularly promising for two reasons. Firstly, it has been shown to work in a situation that models climate change pollution (Example 2.2); secondly, unlike the provision point mechanism of Example 4.2, there is not a risk that a provision point wont quite be reached, which would prevent the public good being provided. It is significant that these mechanisms are based on conditionality because conditionality plays a role in international negotiations on targets.

These mechanisms require that countries can make a commitment that they cannot backtrack from at each stage of the mechanism. This suggests that if an international legal architecture is devised for cooperation on climate change, a mechanism that makes 'legally binding' conditional commitments possible would be desirable. One way to make backtracking difficult could be to repeat the game. A novel approach to get countries to make commitments that they will not backtrack from is described below.

Example 4.5 (A Deposit Based Compliance Mechanism). A two-stage mechanism to provide public goods when there are not strong institutions has been described by Gerber & Wichardt (2009). We assume that there is an underlying public goods game such as the game in Example 2.2. In Stage 1, each player is required to pay a deposit. In Stage 2, there are two possible outcomes. If not all players paid the deposit in Stage 1, then the deposits are refunded and the underlying public goods game is played. If all players paid the required deposit, then in Stage 2 players are required to make a pre-specified contribution to the public good. If a player makes the contribution, they get their deposit back. If their contribution is less than what was specified, they do not.

Gerber and Wichardt show that provided the deposits satisfy a certain inequality, and the payoffs for each player are greater when all players contribute the specified contribution than when nobody does, then it is a subgame perfect equilibrium for each player to contribute the specified amount to the public good. The mechanism discourages players from reneging from their commitments because by making a deposit prior to the contribution stage, they make it a dominant strategy to comply with the agreement. The action of paying the deposit can be thought of as a way for players to execute their own punishment, rather than have to punish anyone else.

An institution is required to collect deposits, monitor players' contributions, and refund deposits. The institution does not have to implement the provision of the public good itself, or enforce punishments of free-riders.

In many of the situations described here, such as Example 2.2, players' preferences are known. An important issue in implementation theory is how to find mechanisms that induce players to reveal their preferences. A significant problem with achieving international cooperation is that players often have strong incentives to misrepresent their abatement costs and environmental preferences. In the climate negotiations, countries have an incentive to exaggerate their abatement costs in order to negotiate a weaker target for themselves or reduce the likelihood of being committed to a target. The issue of private information in international environmental agreements is discussed by Batabyal (1996). An auction mechanism that induces players to reveal their true abatement costs[5] has been described by Montero (2008), and applications to global warming have are described in (Montero, 2007).

5 Conclusion

In its simplest form, climate change mitigation is a prisoner's dilemma. The prisoner's dilemma has a Nash equilibrium that involves players acting non-cooperatively in a manner that is socially sub-optimal. When countries have a continuous choice about how much to pollute, the Nash equilibrium involves much more pollution than is optimal. This is why climate change is sometimes known as a social dilemma.

[5]Mechanisms that induce players to bid truthfully and pay the cost of the externality that they impose are known as Vickrey-Clarke-Groves mechanisms. Another mechanism with these properties is described by Dasgupta *et al.* (1980).

Normal form games such as this help us to understand the free-rider problem, but do not tell is about the sequential nature of strategic behaviour. Being able to do this helps us to address the social dilemma.

Extensive form games that have more than one stage, such as the treaty participation game, can have solutions that are more cooperative as their subgame perfect equilibrium. For two players, the treaty participation game implements a cooperative outcome. But for more than two players, there is only partial cooperation. Ways to address this may include the use of punishments; and issue linkage, possibly involving trade. Trade measures are promising but there are risks if this is done in a way that is not perceived to be fair.

When game theory is used to help us understand coalitions, outcomes that are more cooperative than the treaty participation game are possible. A socially optimal outcome has been predicted by Chander & Tulkens (1997), using cooperative game theory and the concept of the γ-core. However, this outcome is based on a threat that might not be credible, so may not be realistic. But many non-cooperative models of coalition formation have subgame perfect equilibria that are more cooperative than predicted by the treaty participation game, including several that were studied by Finus & Rundshagen (2003). Because one of the coalition formation processes, the exclusive membership game, has as a significantly more cooperative solution that some of the others, it may be the case that carbon market linkage can help facilitate a cooperative outcome. When countries that have emissions trading schemes make a decision about whether to link their carbon markets, the possibility that this could facilitate cooperation and coalition formation should be a consideration.

There are several strong results about mechanisms that implement a cooperative outcome via subgame perfect equilibrium when there is a social dilemma. These include provision point mechanisms (Example 4.2), bargaining based on confirmed proposals (Example 4.3), and approaches where countries 'match' each others pollution abatement commitments (Example 4.4). All of these approaches make use of conditionality. This suggests that when countries are willing to increase their emission reduction commitment if others do the same, cooperation is more likely. It also suggests that cooperation would be more likely if an international mechanism were to exist that would allow countries to make a binding conditional commitment. Approaches that discourage 'backtracking' are more likely to be successful.

Game theoretic approaches inform our understanding of participation and compliance in international agreements, the role of coalitions, and the role of conditionality when

bargaining over emission reductions. This can help us understand the social dilemma associated with climate change and provide insights that may help us address it.

Acknowledgements

The author would like to thank Stephen Howes, Frank Jotzo, and David Pannell for comments on earlier versions of the text. This work was supported by the Australian Government Department of the Environment, Water, Heritage and the Arts through the Environmental Economics Research Hub.

References

Admati, A.R, & Perry, M. 1991. Joint Projects without Commitment. *Review of Economic Studies*, **58**(2), 259–76.

Aghion, P., Antras, P., & Helpman, E. 2007. Negotiating free trade. *Journal of International Economics*, **73**(1), 1 – 30.

Attanasi, G., Gallego, A. G., Georgantzis, N., & Montesano, A. 2010. *Non-cooperative games with confirmed proposals.* Working Paper. LERNA Travaux No. 10.02.308.

Axelrod, R. M. 1984. *The evolution of cooperation.* Basic Books, New York.

Baer, P., Harte, J., Haya, B., Herzog, A. V., Holdren, J., Hultman, N. E., Kammen, D. M., Norgaard, R. B., & Raymond, L. 2000. Climate Change: Equity and Greenhouse Gas Resposibility. *Science*, **289**(5488), 2287.

Bagnoli, M., & Lipman, B. L. 1989. Provision of Public Goods: Fully Implementing the Core through Private Contributions. *Review of Economic Studies*, **56**(4), 583–601.

Barrett, S. 1994. Self-Enforcing International Environmental Agreements. *Oxford Economic Papers*, **46**, 878–894.

Barrett, S. 1997. The strategy of trade sanctions in international environmental agreements. *Resource and Energy Economics*, **19**, 345–361.

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

Barrett, S., & Stavins, R. 2003. Increasing Participation and Compliance in International Climate Change Agreements. *International Environmental Agreements: Politics, Law and Economics*, **3**, 349–376.

Batabyal, A. 1996. An Agenda for Design and Study of International Environmental Agreements. *Ecological Economics*, **19**, 3–9.

Benedick, R. E. 1991, 1998. *Ozone Diplomacy: New Directions in Safeguarding the Planet.* Enlarged edn. Cambridge, Massachusetts: Harvard University Press.

Berger, J. O. 1980. *Statistical decision theory, foundations, concepts, and methods.* Springer-Verlag, New York.

Black, R. 2010. Copenhagen climate summit undone by 'arrogance'. *BBC News Online*, March.

Bloch, F. 1996. Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division. *Games and Economic Behavior*, **14**, 90–123.

Boadway, R., Song, Z., & Tremblay, J.-F. 2009 (May). *The Efficiency of Voluntary Pollution Abatement when Countries can Commit.* Working Papers 1205. Queen's University, Department of Economics.

Brams, S. J., & Kilgour, D. M. 2009. How Democracy Resolves Conflict in Difficult Games. *In:* Levin, S. A. (ed), *Games, Groups, and the Global Good.* Springer.

Brams, S. J., & Taylor, A. D. 1996. *Fair Division: From cake-cutting to dispute resolution.* Cambridge University Press, Cambridge.

Bråten, J., & Golombek, R. 1998. OPEC's Response to International Climate Agreements. *Environmental and Resource Economics*, **12**, 425–442.

Brennan, G. 2009. Climate change: a rational choice politics view. *The Australian Journal of Agricultural and Resource Economics*, **53**(3), 309–326.

Buchner, B., & Carraro, C. 2006. Parallel Climate Blocs: Incentives to Cooperation in International Climate Negotiations. *In:* Guesnerie, R., & Tulkens, H. (eds), *The Design of Climate Policy Conference Volume of the 6th CESIfo Venice Summer Institute.* MIT Press, Cambridge, Massachusetts.

Carraro, C., & Moriconi, F. 1997. International Games on Climate Change Control. *Fondazione Eni Enrico Mattei.*

Carraro, C., & Siniscalco, D. 1993. Strategies for the international protection of the environment. *Journal of Public Economics*, 309–328.

Carraro, C., Marchiori, C., & Oreffice, S. 2009. Endogenous Minimum Participation in International Environmental Treaties. *Environmental and Resource Economics*, **42**, 411–425.

Chander, P., & Tulkens, H. 1997. The Core of an Economy with Multilateral Environmental Externalities. *International Journal of Game Theory*, **26**(3), 379–401.

Chander, P., & Tulkens, H. 2008. Cooperation, Stability, and Self-enforcement in International Environmental Agreements. *Pages 165–186 of:* Guesnerie, R., & Tulkens, H. (eds), *The Design of Climate Policy.* MIT Press, Cambridge, Mass.

Dasgupta, P., Hammond, P., & Maskin, E. 1980. On Imperfect Information and Optimal Pollution Control. *Review of Economic Studies*, **47**, 857–860.

de Clippel, G, & Serrano, R. 2008. *Bargaining, Coalitions and Externalities: A Comment on Maskin.* Working Paper. Brown University.

Farrell, J., & Maskin, E. 1989. Renegotiation in Repeated Games. *Games and Economic Behaviour*, **1**, 327–160.

Fehr, E., & Gächter, S. 2000. Fairness and Retaliation: The Economics of Reciprocity. *The Journal of Economic Perspectives*, **14**(3), 159–181.

Finus, M. 2001. *Game theory and international environmental cooperation.* Edward Elgar, Cheltenman, UK ; Northampton, MA.

Finus, M. 2003. Stability and design of international environmental agreements: the case of transboundary pollution. *Pages 82–158 of:* Folmer, H., & Tietenberg, T. (eds), *International yearbook of environmental and resource economics 2003/4*. Edward Elgar, Cheltenman, UK.

Finus, M. 2008. Game Theoretic Research on the Design of International Environmental Agreements: Insights, Critical Remarks and Future Challenges. *International Review of Environmental and Resource Economics*, **2**, 29–67.

Finus, M., & Rundshagen, B. 2003. Endogenous Coalition Formation in Global Pollution Control: A Partition Function Approach. *Pages 199–244 of:* Carraro, C. (ed), *The Endogenous Formation of Economic Coalitions*. Edward Elgar, Cheltenman, UK.

Fleurbaey, M. 1994. On Fair Compensation. *Theory and Decision*, **36**, 277–307.

Garnaut, R. 2008a. *The Garnaut climate change review : final report*. Cambridge University Press, Port Melbourne, Vic. :.

Garnaut, R. 2008b. *Targets and Trajectories*. Supplementary Draft Report. Commonwealth of Australia.

Gerber, A., & Wichardt, P. C. 2009. Providing public goods in the absence of strong institutions. *Journal of Public Economics*, **93**, 429–439.

Güth, W., Schmittberger, R., & Schwarze, B. 1982. An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behavior and Organization*, **3**, 367–388.

Guttman, J. M. 1978. Understanding Collective Action: Matching Behaviour. *The American Economic Review*, **68**(2), 251–255.

Hart, S., & Kurz, M. 1983. Endogenous Formation of Coalitions. *Econometrica*, **51**(4), 1047–1064.

Hoel, M. 1994. Efficient Climate Policy in the Presence of Free Riders. *Journal of Environmental Economics and Management*, **27**, 259–274.

Jackson, M. O. 2001. A crash course in implementation theory. *Social Choice and Welfare*, **18**, 655–708.

Jotzo, F., & Betz, R. 2009. Australias emissions trading scheme: opportunities and obstacles for linking. *Climate Policy*, **9**, 402–414.

Lessmann, K., Marschinski, R., & Edenhoffer, O. 2009. The effects of Tariffs on Coalition Formation in a Dynamic Global Warming Game. *Economic Modelling*, **26**(3), 641–649.

Little, A. 2010. Copenhagen Accord is the priority, says U.S. climate envoy. But what about a binding treaty? *Grist Magazine*, January.

Marks, M., & Croson, R. 1998. Alternative rebate rules in the provision of a threshold public good: An experimental investigation. *Journal of Public Economics*, **67**, 195–220.

Maskin, E. 2003. *Bargaining, Coalitions, and Externalities*. Presidential Address to the Econometric Society. Institute for Advanced Study, Princeton.

Meyer, A. 2000. *Contraction and Convergence: The Global Solution to Climate Change.* Green Books for the Schumacher Society, Totnes, U.K.

Montero, J.-P. 2007. An auction mechanism in a climate policy architecture. *Pages 327–339 of:* Aldy, Joseph E., & Stavins, Robert N. (eds), *Architectures for Agreement: Addressing Global Climate Change in the Post-Kyoto World.* Cambridge University Press, New York.

Montero, J.-P. 2008. A Simple Auction Mechanism for the Optimal Allocation of the Commons. *American Economic Review*, **98**(1), 496–518.

Moore, J., & Repullo, R. 1988. Subgame Perfect Implementation. *Econometrica*, **56**(5), 1191–1220.

Nash, J. 1953. Two-person Cooperative Games. *Econometrica*, **21**(1), 128–140.

Ok, E. 2007. *Real Analysis with Economic Applications.* Princeton: Princeton University Press.

Okada, A., & Winter, E. 2002. A Non-cooperative Axiomatization of the Core. *Theory and Decision*, **53**(1), 1–28.

Osborne, M. J. 2003. *An Introduction to Game Theory.* Oxford University Press, USA.

Osborne, M. J., & Rubinstein, A. 1994. *A Course in Game Theory.* The MIT Press.

Ostrom, E. 2009. *A Polycentric Approach for Coping with Climate Change.* Policy Research Working Paper 5095. World Bank.

Pan, J., Chen, Y., Wang, W., & Li, C. 2000. *Carbon Budget Proposal: Global Emissions under Carbon Budget Constraint on an Individual Basis for an Equitable and Sustainable Post-2012 International Climate Regime.* Working Paper. Research Centre for Sustainable Development, Chinese Academy of Social Sciences.

Project Team of the Development Research Centre of the State Council, People's Republic of China. 2009. Greenhouse gas emission reduction: A theoretical framework and global solution. *Pages 389–408 of:* Garnaut, R., Song, L., & Woo, W. T. (eds), *China's New Place in a World in Crisis.* Canberra, Australia: ANU E Press.

Ray, D., & Vohra, R. 1997. Equilibrium Binding Agreements. *Journal of Economic Theory*, **73**, 30–78.

Serrano, R. 1995. A Market to Implement the Core. *Journal of Economic Theory*, **67**(1), 285–294.

Serrano, R. 1997. A comment on the Nash program and the theory of implementation. *Economics Letters*, **55**(2), 203–208.

Stern, N. 2006. *The economics of climate change : the Stern review.* Cambridge, UK ; New York : Cambridge University Press.

Stern, N. 2009. *The Global Deal : Climate Change and the Creation of a New Era of Progress and Prosperity.* PublicAffairs ; New York.

Tamiotti, L., Olhoff, A., Teh, R., Simmons, B., Kulaçoğlu, V., & Abaza, H. 2009. *Trade and Climate Change*. Tech. rept. World Trade Organisation and United Nations Environment Programme.

Tulkens, H. 1998. Cooperation versus free-riding in international environmental affairs: two approaches. *Pages 30–44 of:* Hanley, N., & Folmer, H. (eds), *Game Theory and the Environment*. Edward Elgar, Cheltenham, England.

Uzawa, H. 2003. Global Warming as a Cooperative Game. *Pages 193–239 of:* Uzawa, Hirofumi (ed), *Economic Theory and Global Warming*. Cambridge University Press, New York.

Victor, D. G. 2007. Fragmented carbon markets and reluctant nations: implications for the design of effective architectures. *Pages 133–160 of:* Aldy, Joseph E., & Stavins, Robert N. (eds), *Architectures for Agreement: Addressing Global Climate Change in the Post-Kyoto World*. Cambridge University Press, New York.

Weitzman, M. 2009. On Modeling and Interpreting the Economics of Catastrophic Climate Change. *The Review of Economics and Statistics*, **91**(1).

Wood, P. J., & Jotzo, F. 2009. *Price Floors for Emissions Trading*. Research Report 36. Environmental Economics Research Hub, The Australian National University.

Yi, S. S. 1997. Stable Coalition Structures with Externalities. *Games and Economic Behaviour*, **20**, 201–237.

Yi, S.-S., & Shin, H. 2000. Endogenous formation of research coalitions with spillovers. *International Journal of Industrial Organization*, **18**(2), 229 – 256.