International Environmental Agreements

Aart de Zeeuw

Tilburg Sustainability Center, Department of Economics, Tilburg University, 5000 LE Tilburg, the Netherlands; email: A.J.deZeeuw@uvt.nl

Annu. Rev. Resour. Econ. 2015. 7:151–68

First published online as a Review in Advance on April 2, 2015

The *Annual Review of Resource Economics* is online at resource.annualreviews.org

This article's doi:
10.1146/annurev-resource-100814-124943

JEL codes: C70, D62, F42, H77, Q20

Keywords

global pollution, games, noncooperative, cooperative, dynamics

Abstract

The regulation of environmental externalities at the global level requires international agreements between sovereign states. Game theory provides an appropriate theoretical tool for analysis. However, game theory can result in a wide range of outcomes, and therefore it is important to discuss the assumptions and mechanisms of the different approaches and to relate these with what is observed in practice. The basic picture is not optimistic: If there are large gains of cooperation, the stable coalition is small. This grim picture challenges the perspective and design of international agreements. This article discusses and compares the different approaches: noncooperative, cooperative, dynamic, and evolutionary. Asymmetries and the options for side payments are considered. At the end, some more optimistic ways forward are presented.

1. INTRODUCTION

Global environmental problems such as ozone depletion and climate change require voluntary cooperation by sovereign states to internalize the negative externalities from the emissions in each state. One of the wicked problems of our time is that this issue can be viewed as a so-called prisoners' dilemma game, which predicts that free-rider incentives induce a noncooperative equilibrium in which countries consider only their own costs and benefits and therefore emit too much. In practice, countries try to negotiate an international agreement on some global or cross-boundary environmental problem. The best known agreements are probably the series of Conferences of the Parties (COPs) under the Framework Convention on Climate Change, which was established by the United Nations in its Agenda 21 meeting in Rio de Janeiro in 1992. The Kyoto Protocol of 1997, in particular, looked like a major step forward, but in retrospect it has not achieved much, and the current COPs are struggling to construct some new agreement. However, there have also been success stories. For example, a series of international agreements in Europe on reduction of emissions of sulfur dioxide, which causes acid rain, has led to substantial reductions. Probably the most successful agreement was the Montreal Protocol in 1987 and subsequent amendments, with the purpose of phasing out emissions of CFCs, which deplete the ozone layer. This agreement has now been signed and ratified by almost all countries. The intriguing questions are why some agreements are successful and others are not and why cooperation exists when basic theory predicts that it will fall apart.

Many arguments have been put forward to explain the success of the Montreal Protocol. Technological development produced CFC substitutes that were not so expensive, and the danger of ozone layer depletion and the increased risk of skin cancer was well understood and generally accepted. Moreover, the Montreal Protocol introduced trade sanctions on countries that did not take part in the agreement, which gave these countries an incentive to join. This first argument shows that the difference between costs and benefits plays an important role. The second argument shows that the static prisoners' dilemma game as a theoretical framework is far from sufficient for predicting potential cooperation. It is well known in game theory that in a repeated prisoners' dilemma, credible strategies, including some form of punishment, can keep the players on the cooperative path (the folk theorem).

Progress is usually made in the interaction between the development of theory and the observations in practice. The ultimate purpose is to gain more insight that can be used in the design of international environmental agreements (IEAs) that work and thus increase welfare. A nice book studying many treaties in detail and connecting them to theory is Barrett (2003). Another recommended survey on this topic is Finus (2003). This article focuses on a part of the development of the theory for IEAs. It does not give a full survey but focuses on the part that has dominated the discussion in this area. Incentives to free ride may exist, but incentives to increase cooperation may exist as well. The implication is that an equilibrium may exist when leaving or joining some coalition of cooperators does not pay off. This stability concept originates from industrial organization, in which this concept was used to investigate collusive price leadership (d'Aspremont et al. 1983), and was introduced into the literature on IEAs by Barrett (1994). The model shows that, for a finite number of players, a stable coalition (or partial cooperation) exists but that this coalition is small. Moreover, the larger the possible gains of cooperation are, the smaller the stable coalition is. This is bad news for IEAs.

The model above was challenged from different directions. In the realm of noncooperative game theory, it was argued that the free-rider incentive gets smaller if players realize that a deviation will trigger more deviations, and these possible consequences may yield larger stable coalitions. This concept of farsighted stability was developed by Chwe (1994) and was introduced into the literature on IEAs by Diamantoudi & Sartzetakis (2002). The other main challenge came from a cooperative game theory approach. The grand coalition may very well be in the core of the game and in that sense stable. Chander & Tulkens (1995) formulated the γ-core of the game, which means that any deviation is met by noncooperative behavior of the other players. Such a threat may keep the players in the grand coalition. This approach is very similar to trigger strategies in noncooperative repeated games, referred to above: The threat of a punishment in the form of noncooperative behavior may keep the players on the cooperative path. The conclusion is that the whole spectrum—from a small stable coalition to a grand stable coalition—fits in some theoretical framework. Anything goes: One can find support for either a pessimistic view or an optimistic view of the problem. This spectrum of possible outcomes is driven mainly by the assumptions of the players' strategies and, more particularly, of the threats that the players formulate. These threats have to be credible, of course, but another issue is whether this type of strategy is realistic in the international policy arena. The first purpose of this article is to describe the different models that are used and to discuss the assumptions.

Many papers have extended and modified the basic models above, but reviewing the entire literature is not possible within the scope of this article. We discuss three more aspects. One is asymmetries. The countries involved are not the same. The first question is what the best-performing stable coalition looks like in this case. The second question is how countries that lose in the cooperative outcome can stay on board. Side payments are the standard theoretical answer, but what do side payments look like in the international policy arena? A second aspect is whether the underlying prisoners' dilemma game can be twisted a bit or shifted into a coordination game by considering a broader cost-benefit structure. This approach may suppress the free-rider incentives and result in better outcomes. Finally, a third aspect is whether the design of the agreement should fully change. For example, we show that an IEA on investment in clean technology, with the purpose of reducing the costs of adoption of these technologies, may yield much better results while still respecting the stability conditions of the agreement.

The success of IEAs depends on many things, and not only on the aspects described with the models in this paper. Leadership and environmental stewardship are also important, but free-rider incentives must be overcome. Economists predicted the failure of the Kyoto Protocol, despite the heroic attempts of the European Union to keep it going. At the same time, economists faced the challenge of a different design or a different perspective on these issues that may be more successful. Some of the global environmental problems are pressing, and resolving these problems is therefore important. This article sheds some light on parts of the theory that have been developed.

Section 2 presents three versions of the basic noncooperative model, with the grim picture of small stable coalitions. Section 3 discusses the challenges from other game-theoretic models and their assumptions. Section 4 develops extensions to asymmetries. Section 5 presents different perspectives and designs, with the purpose of achieving a higher level of success with IEAs, and Section 6 concludes.

2. CARTEL STABILITY

The literature on the theory of IEAs (Hoel 1992, Carraro & Siniscalco 1993, Barrett 1994) started with a concept that originated in industrial organization. d'Aspremont et al. (1983) investigate the stability of collusive price leadership. With a continuum of firms, there is always an incentive for a cartel member to join the competitive fringe so that this market structure is unstable. With a finite number of firms, however, finding a stable cartel is always possible. This news may be bad in that literature but is good for IEAs. However, d'Aspremont et al. show that the stable cartel is small, which is bad news for IEAs.

2.1. The Basic Model

The basic idea in these papers is to modify the underlying prisoners' dilemma game: If a player switches to noncooperative behavior, the remaining cooperators change their behavior as well. That is, they no longer take into account the interests of the deviating player. Thus, the deviating player has free-rider benefits but also loses some cooperative benefits. At some coalition size, leaving the coalition no longer makes sense. This equilibrium is not a Nash equilibrium for the static game, but it is a Nash equilibrium for the two-stage game, in which the players first decide whether to be a member of the coalition and then decide on their actions as a coalition or as an individual outsider. This equilibrium is usually characterized by the requirements of internal and external stability. Internal stability means that a member of the coalition does not have an incentive to leave the coalition, and external stability means that an outsider does not have an incentive to join the coalition. Formally, if the number of countries is n and the size of the coalition is k, internal stability and external stability require, respectively, that

$$\begin{aligned}&\pi^{\mathrm{o}}(k-1)\leq\pi^{\mathrm{m}}(k), k\geq 1,\\&\pi^{\mathrm{m}}(k+1)\leq\pi^{\mathrm{o}}(k), k\leq n-1.\end{aligned} \tag{2.1}$$

Here π denotes the net benefits, which depend on the size of the coalition; m denotes coalition member; and o denotes coalition outsider. With a very simple model, we can show that the size of the stable coalition is small. Suppose that the countries can choose levels of abatement or emission reduction a_i with costs $0.5{a_i}^2$ and with total abatement as the benefits for each country. Formally, the net benefits are

$$\pi_i=\sum_{j=1}^{n}a_j-0.5a_i^2, i=1,2,\ldots,n. \tag{2.2}$$

An individual country chooses $a_i=1$, but a coalition of k countries maximizes

$$\max_{a_1,\ldots,a_k}\left(k\sum_{j=1}^{n}a_j-0.5\sum_{i=1}^{k}a_i^2\right). \tag{2.3}$$

It follows that a coalition member chooses $a_i=k$ and that net benefits of the coalition members and the outsiders become

$$\begin{aligned}&\pi^{\mathrm{m}}(k)=k^2+n-k-0.5k^2,\\&\pi^{\mathrm{o}}(k)=k^2+n-k-0.5.\end{aligned} \tag{2.4}$$

Applying the internal and external stability conditions (Condition 2.1) yields $k=2$ or $k=3$ as possible sizes of the stable coalition, regardless of the number of countries n. This is the grim result to which the literature on IEAs usually refers. This result does not change if we add a parameter to Equation 2.2, reflecting the relative weights of the costs and benefits. The model is a bit specific in the sense that outsiders always do the same, regardless of the coalition size. We can easily change that outcome and draw a more precise conclusion.

2.2. An Extended Model

Suppose that the benefits of total abatement are specified as lower (quadratic) costs of the business-as-usual emission level e, and suppose that the parameter p denotes the relative weights of the emission costs and the abatement costs. Each country minimizes total costs c_i, given by

$$c_i = 0.5p\left(e - \sum_{j=1}^{n} a_j\right)^2 + 0.5a_i^2, i = 1, 2, \ldots, n. \quad (2.5)$$

After straightforward calculations, the optimal abatement levels and the total costs of coalition members and outsiders become, respectively,

$$a^{\mathrm{m}}(k) = \frac{kpe}{1 + (k^2 + n - k)p}, \; a^{\mathrm{o}}(k) = \frac{pe}{1 + (k^2 + n - k)p}, \quad (2.6)$$

$$c^{\mathrm{m}}(k) = 0.5\frac{p(1 + k^2 p)e^2}{\left[1 + \left(k^2 + n - k\right)p\right]^2}, \; c^{\mathrm{o}}(k) = 0.5\frac{p(1 + p)e^2}{\left[1 + \left(k^2 + n - k\right)p\right]^2}. \quad (2.7)$$

First note that the optimal abatement level of the outsiders now depends on the size of the coalition k: Abatement becomes smaller when the size k becomes larger. This result is intuitively clear because a larger coalition yields more abatement so that outsiders can relax their abatement (leakage effect). It is straightforward to show that the size k of the coalition, which is both internally and externally stable, is either 1 or 2. It is 1 if the parameter p is large, i.e., $p \geq 1/[n - 4 + 2\sqrt{(n^2 - 3n + 3)}]$, and it is 2 otherwise. A large p also means that the externalities are large and thus that the possible gains of cooperation are large. Therefore, the size of the stable coalition is always small but is even smaller when the possible gains of cooperation are large. This is really bad news for IEAs. These results are robust for all specifications essentially describing the same type of problem (e.g., Finus 2003).

2.3. A Third Model

We conclude this section with another canonical model, also in the noncooperative realm, but in which the abatement decisions are yes or no, for example, describing the adoption of clean technology (Barrett 2006). Suppose that the costs of adoption are c and the individual benefits are b, with $b < c$ and $nb > c$, so that we again have a prisoners' dilemma game. It is easy to see that the smallest k with $kb > c$ is the size of the stable coalition. The coalition is internally stable because $0 < kb - c$, and it is externally stable because $(k + 1)b - c < kb$ (see Condition 2.1 above). Thus, for a given cost of adoption c, we get a small stable coalition k if the benefits of adoption b are large. Again we see that this concept of cartel stability has the property that large gains of cooperation are not realized because, in such a case, the stable coalition is small. We need to change the design, strategies, or perception of the game to induce a better outcome.

3. ALTERNATIVE MODELS

3.1. Cooperative and Repeated Game Theory

The main challenge to the grim results of the models in Section 2 came from the realm of cooperative game theory using the core concept. A feasible allocation is in the core of the game if it cannot be blocked or improved upon by a subset of players. Chander & Tulkens (1995) argue that the grand coalition may well be stable in the sense of the core by showing that no subcoalition can improve by deviating. It is essential to make an assumption on what happens in case of a deviation. Chander & Tulkens introduce the γ-core, which simply assumes that the remaining countries switch to noncooperative behavior. Other core concepts, such as the α-core, assume stronger punishments to deviations, but such an assumption is not needed here to get the desired result. In the symmetric case, the result is trivial. For example, take the second model in Section 2 with costs given by Equation 2.7. If a subcoalition of k countries deviates from the grand coalition of n countries and the other countries switch to noncooperative behavior, the costs are given by Equation 2.7. The costs of the members of the subcoalition increase because $c^{\mathrm{m}}(k) > c^{\mathrm{m}}(n)$, so the grand coalition cannot be blocked. In the asymmetric case, transfers are needed; we discuss this issue in Section 4 below.

This idea is similar to the idea of trigger strategies in noncooperative repeated games (Friedman 1986). Players start with cooperative behavior but switch to noncooperative behavior if some deviation is detected, and this threat deters deviations if the discount factor is sufficiently high. Formally, suppose that the second game in Section 2 is repeated. The optimal abatement level and the total costs of a country that deviates from full cooperation in a stage game are given by

$$a^{\mathrm{d}}(n) = \frac{p(1+np)e}{(1+p)(1+n^2p)},\ c^{\mathrm{d}}(n) = 0.5\frac{p(1+np)^2e^2}{(1+p)(1+n^2p)^2}. \tag{3.1}$$

This case is different from the case of an outsider to a coalition of $n-1$ countries because, in the latter case, the coalition adjusts its behavior. It is easy to check that $c^{\mathrm{d}}(n) < c^{\mathrm{m}}(n)$ (given in Equation 2.7) so that the deviator has lower cost in that stage game but will trigger noncooperative behavior of all countries in the future, yielding higher costs $c^{\mathrm{m}}(n) < c^{\mathrm{o}}(0)$ (given in Equation 2.7) in all future stages of the game. Deviation is deterred if the total discounted costs increase. Therefore, deviation is deterred if the discount factor δ, $0 < \delta < 1$, is sufficiently high; that is,

$$\frac{c^{\mathrm{m}}(n)}{1-\delta} < c^{\mathrm{d}}(n) + \frac{\delta c^{\mathrm{o}}(0)}{1-\delta} \Rightarrow \frac{c^{\mathrm{m}}(n) - c^{\mathrm{d}}(n)}{c^{\mathrm{o}}(0) - c^{\mathrm{d}}(n)} < \delta < 1. \tag{3.2}$$

Countries can keep each other in the cooperative mode by threatening to switch to noncooperative behavior if one country or one group of countries deviates. This reaction of the other countries in the coalition converts what seems to be a free-rider benefit into a cost. The essential difference with the basic model in Section 2 is that, in that model, the reaction of these other countries is more moderate: If a country leaves the coalition, the remaining coalition does not fall apart but continues cooperating so that only some cooperative benefits are lost. This threat is not strong enough to sustain the grand coalition.

The policy conclusion is simple: Countries should formulate strategies with strong threats, and everything will be fine. In practice, however, things may not be so simple. Strategies have to be credible. Sophisticated game theory has formulated strategies that are credible and that induce cooperation, but we have not seen these strategies resonate in the practice of policy. The Montreal Protocol introduced trade sanctions on countries that did not take part in the agreement; such sanctions can be seen as a threat for not joining the grand coalition. However, sanctions have not been implemented, largely because technological development (e.g., in the production of sprays and refrigerators) overcame the problem of CFCs and ozone depletion. The COPs following the Kyoto Protocol discussed fines on noncompliance, but again such fines have not been implemented. The question is whether we can come up with a different IEA design that is easier to implement. Before we discuss this question, we continue with the theoretical structure of IEAs because the two extremes discussed so far have intermediates that are interesting to consider.

3.2. Farsightedness

Chwe (1994) developed the concept of farsighted stability, which Diamantoudi & Sartzetakis (2002) introduced into the literature on IEAs. The basic idea is that countries realize that a deviation may trigger further deviations of others, but they do not assume a full breakdown of the coalition. In this perspective, internal stability is too weak because this condition assumes that no other deviations take place, but the γ-core approach is too strong because it assumes that the coalition falls apart completely. Furthermore, a sequence of deviations comes to an end when a new stable situation is reached. Deviations are deterred if an outsider in such a new situation is worse off than a member of the coalition in the initial situation. In this way, a set of farsighted stable coalitions can be constructed. For example, take again the second model in Section 2, with costs given by Equation 2.7. Recall that a coalition of size 2 is stable for a sufficiently small p. Thus, a coalition of size 3 cannot be stable, but a coalition of size 4 may be farsighted stable. Indeed, it is easy to show that an outsider to a coalition of size 2 is worse off than a member of a coalition of size 4. In this way, a set of farsighted stable coalitions can be constructed, and the largest one is usually close in size to the grand coalition. This analysis is good news for IEAs. It implies that countries can coordinate on a large stable coalition, without strong threats in the case of deviations, but only with some forward-looking behavior. However, there is a drawback when this model is put in a dynamic setting in which detecting deviations takes time.

We obtain a dynamic version of the second model in Section 2 by changing Equation 2.5 or the total costs c_i to

$$c_i = \sum_{t=0}^{\infty} \delta^t \left(0.5pe^2(t) + 0.5a_i^2\right), i = 1, 2, \ldots, n, \tag{3.3}$$

where δ denotes the discount factor, subject to

$$e(t+1) = e(t) - \sum_{i=1}^{n} a_i(t), t = 0, 1, \ldots, e(0) = e_0. \tag{3.4}$$

Here e_0 are the excess emissions that have to be brought down to 0. When the countries cooperate, this target is reached faster and with lower costs than in the absence of cooperation. The difference game (Equations 3.3 and 3.4), with a coalition of size k and individual outsiders, can be solved by dynamic programming (de Zeeuw 2008). It has a state e (the level of excess emissions), but it is not a stock-pollutant problem. We could formulate it as a stock-pollutant problem, but the analysis would become much more complicated. Moreover, IEAs are usually stated in emission levels. If the value functions are denoted by $V^{\mathrm{m}}(e) = 0.5c^{\mathrm{m}}e^2$ and $V^{\mathrm{o}}(e) = 0.5c^{\mathrm{o}}e^2$ for coalition members and outsiders, respectively, the optimal abatement levels become (omitting the dependence on k)

$$a^{\mathrm{m}} = \frac{\delta k c^{\mathrm{m}} e}{1 + \delta\left(k^2 c^{\mathrm{m}} + (n-k)c^{\mathrm{o}}\right)}, a^{\mathrm{o}} = \frac{\delta c^{\mathrm{o}} e}{1 + \delta\left(k^2 c^{\mathrm{m}} + (n-k)c^{\mathrm{o}}\right)}. \tag{3.5}$$

The cost parameters c^{m} and c^{o} of the value functions have to satisfy

$$c^{\mathrm{m}} = p + \frac{\delta c^{\mathrm{m}}\left(1 + \delta k^2 c^{\mathrm{m}}\right)}{\left[1 + \delta\left(k^2 c^{\mathrm{m}} + (n-k)c^{\mathrm{o}}\right)\right]^2}, c^{\mathrm{o}} = p + \frac{\delta c^{\mathrm{o}}\left(1 + \delta c^{\mathrm{o}}\right)}{\left[1 + \delta\left(k^2 c^{\mathrm{m}} + (n-k)c^{\mathrm{o}}\right)\right]^2}. \tag{3.6}$$

With these cost parameters, the same analysis can be done as with the costs in Equation 2.7. One can easily show that the coalition with internal and external stability is small but that farsighted stability yields a sequence of stable coalitions approaching the grand coalition. However, in this dynamic setting, one should pay attention to the facts that the detection of deviations takes time and that only in the next period can a reaction take place. This is not a repeated game, because the game changes over time with a changing emission level $e(t)$ (the state of the system).

We now introduce the concept of dynamic farsighted stability. The idea is that, for a different level of emissions, the deviator becomes an outsider to the smaller stable coalition in the next period. This mechanism implies that the deviator has free-rider benefits for one period but has to face the consequences afterward. Only if these deviations are deterred is the coalition dynamic farsighted stable. To analyze this situation, we need the value function of a deviator with parameter c^{d}. The state-space analog of Equations 3.1 and 3.2 can be derived in a tedious, but straightforward, manner:

$$c^{\mathrm{d}} = p + \frac{\delta c^{\mathrm{o}+}}{1 + \delta c^{\mathrm{o}+}} \frac{(1 + \delta k c^{\mathrm{m}})^2}{\left[1 + \delta\left(k^2 c^{\mathrm{m}} + (n-k)c^{\mathrm{o}}\right)\right]^2}, \tag{3.7}$$

where $c^{\mathrm{o}+}$ is the parameter of the value function of an outsider to the smaller stable coalition in the next period. Deviations are deterred if $c^{\mathrm{m}} < c^{\mathrm{d}}$, but the analysis is not easy. The system is complicated because we have to simultaneously solve for the cost parameters and the stable sets. Therefore, we have to resort to numerical analysis. However, in this case countries can coordinate on a large stable coalition only if the weighing parameter p is very small. The intuition is clear: If p is large, emissions are quickly reduced. Because deviations can be detected only in the next period, the emissions will have been substantially reduced at that point, and the threat of triggering a smaller stable coalition may not be sufficient to deter deviations. This breakdown happens if p is sufficiently large. Again, if the externalities are large, sustaining a large stable coalition is difficult.

In a way, the circle is closed, and we are back to the grim outcome. The static model in Section 2 with internal and external stability predicts that stable coalitions are small and that they are even smaller when the emission costs are weighed more heavily than the abatement costs. This grim picture disappears when strategies with strong threats are considered, but these strategies may not be realistic in practice. The grim picture almost disappears when farsightedness is considered, but the pessimistic scenario is reestablished in a dynamic context. The conclusion is still that we need to change the design or the perception of the game to induce a better outcome, but before we discuss this topic, we consider what happens if we introduce asymmetries into the models above.

4. ASYMMETRIES

Most theories on IEAs assume that countries are the same, but in fact countries are very different. Countries differ in costs of abatement, in vulnerability to environmental damage, and in size. This heterogeneity raises two fundamental questions. One is the effect on the stable coalitions' size and composition (Fuentes-Albero & Rubio 2010). In this respect, it is important to distinguish whether countries in the coalition have the option to share the costs. This issue may have an effect on stability but also arises in comparing cooperation with noncooperation: Cooperation is collectively rational but is not necessarily individually rational in the case of asymmetries. This discussion raises the second fundamental question: How can countries be compensated by other countries if the former happen to be worse off or if such compensation can keep them in the coalition?

4.1. The Basic Model Again

We start by returning to the first model in Section 2 and introduce parameters β_i and γ_i to reflect the differences in benefits of abatement and in costs of abatement:

$$\pi_i = \beta_i \sum_{j=1}^{n} a_j - 0.5\gamma_i a_i^2, i = 1, 2, \ldots, n. \tag{4.1}$$

Suppose that a coalition S of size k forms and plays a noncooperative game with $n - k$ individual outsiders. The optimal abatement levels of the coalition members and the outsiders become

$$a_i = \frac{\sum_{j \in S} \beta_j}{\gamma_i}, i \in S, a_i = \frac{\beta_i}{\gamma_i}, i \notin S. \tag{4.2}$$

High levels of β (high benefits) and low levels of γ (low costs) lead to high levels of abatement. However, the optimal abatement level of a coalition member depends on the benefits of the other coalition members. Therefore, the success of an IEA depends not only on the size of the stable coalition but also on which countries join the IEA. What are the requirements for internal and external stability in this case? It is sufficient to focus on internal stability because if external stability does not hold for country i outside coalition S, internal stability must hold for country i in coalition S plus country i. If country i leaves coalition S, the abatement of country i and the total abatement reduce by, respectively,

$$\sum_{j \in S \setminus \{i\}} \frac{\beta_j}{\gamma_i}, \sum_{j \in S \setminus \{i\}} \left(\frac{\beta_j}{\gamma_i} + \frac{\beta_i}{\gamma_j} \right), \tag{4.3}$$

because country i no longer takes into account the other countries in S and vice versa. It is straightforward to show that internal stability requires that, for all countries i in coalition S,

$$2\gamma_i \beta_i^2 \sum_{j \in S \setminus \{i\}} \frac{1}{\gamma_j} \geq \left(\sum_{j \in S \setminus \{i\}} \beta_j \right)^2, \ i \in S. \tag{4.4}$$

One can easily show that, if either all the benefit parameters β_i are the same or all the cost parameters γ_i are the same, the stable coalition cannot be larger than 3 (Pavlova & de Zeeuw 2013). Fuentes-Albero & Rubio (2010) derive this result, which is again the usual grim result in this type of model. However, if abatement benefits and abatement costs are asymmetrical, the stable coalition can be larger.

To make the analysis tractable, we assume that we have two types of countries with parameters (β_1, γ_1) and (β_2, γ_2), and we write $\beta = \beta_1/\beta_2$ and $\gamma = \gamma_1/\gamma_2$. Suppose that the coalition S consists of k_1 countries of the first type and k_2 countries of the second type. The internal stability condition (Condition 4.4) for each type of country becomes

$$2(k_1 + k_2\gamma - 1) \geq \left(k_1 + \frac{k_2}{\beta} - 1 \right)^2, 2\left(\frac{k_1}{\gamma} + k_2 - 1 \right) \geq (k_1\beta + k_2 - 1)^2. \tag{4.5}$$

If both β and γ are larger than 1 or smaller than 1, the size of the stable coalition $k_1 + k_2$ cannot be larger than 3. Without loss of generality, we can take $\beta > 1$ or $\beta_1 > \beta_2$, which means that countries of type 1 are more vulnerable to environmental damage and have more benefits of abatement than do countries of type 2. Furthermore, we take $\gamma < 1$ or $\gamma_1 < \gamma_2$, which means that countries of type 1 have lower costs of abatement than do countries of type 2. We can then prove the following result (Pavlova & de Zeeuw 2013): If β is sufficiently large and γ is sufficiently small, the largest stable coalition consists of two countries of type 1 and all countries of type 2. This sounds like good news for IEAs. However, in the case of asymmetries, a large stable coalition is not necessarily a good thing. The coalition of three countries of type 1 is also stable, and we can prove that total abatement is actually higher in this case. The intuition is clear: Countries of type 1 have higher benefits of abatement and lower costs, and therefore having one additional country of type 1 on board is preferable to having all the countries of type 2. There is again a trade-off between the size of the stable coalition and success in terms of emission reductions. We can get a large stable coalition, but only if countries are highly asymmetric, and in this case, a small stable coalition with countries with high abatement levels is preferred. Empirical analyses with global models and asymmetries between key regions in the world (e.g., Botteon & Carraro 1997) also conclude that the size of the successful stable coalition is small.

Satisfying the internal stability conditions may be easier if transfers between the coalition members are possible: Keeping certain countries on board by sharing the surplus of cooperation and lowering incentives to leave the coalition may be more successful (McGinty 2007). Before discussing below how realistic this scenario is, we first investigate the consequences. In the case of transfers, we have to consider potential internal stability (e.g., Pintissalgo et al. 2010), which means that, within the coalition, a proper cost allocation makes the coalition internally stable. That the transfers add up to zero simply implies that the potential internal stability condition is the sum of the internal stability conditions (Equation 4.5), which leads to

$$k_1\left[2(k_1-1+k_2\gamma)\beta^2-\left((k_1-1)\beta+k_2\right)^2\right]+ \\ k_2\left[2\left(k_1+(k_2-1)\gamma\right)-\gamma\left(k_1\beta+k_2-1\right)^2\right]\geq 0. \tag{4.6}$$

The largest stable coalition consists again of two countries of type 1 and all countries of type 2, but the conditions on β and γ for this scenario to hold can be somewhat relaxed. However, a large β, i.e., a large asymmetry in benefits of abatement, is necessary. Moreover, if the condition on γ is relaxed, i.e., there is a smaller asymmetry in costs of abatement, the asymmetry in β has to be even larger. In the case of transfers, total abatement with this large stable coalition may be higher than total abatement with the small stable coalition consisting of three countries of type 1. However, this result holds only in a small area of the parameter space (β, γ). Because transfers add an extra complication to the formation of IEAs, we can conclude that, in the case of asymmetries, theory predicts again that the stable coalition will be small.

4.2. Cooperative Game Theory Reconsidered

As we see in Section 3.2, the γ-core approach from cooperative game theory is more optimistic and leads to the conclusion that the grand coalition is stable. This result is obvious when the countries are the same but becomes more interesting in the case of asymmetries. In such a situation, a subcoalition may block the grand coalition, even if the remaining coalition falls apart. For example, although the total costs under full cooperation are lower than the total cost in the noncooperative equilibrium, some of the countries may very well be worse off under full cooperation and can thus block the grand coalition. Transfers, however, provide a solution in this case. Chander & Tulkens (1995) formulate a game in which each country minimizes total costs, which are split into abatement costs c and damage costs d, both functions of the emission level e. With a slight twist of the earlier notation, the objectives are given by

$$c_i(e_i)+d_i(e_i), i=1,2,\ldots,n. \tag{4.7}$$

Chander & Tulkens show, for linear damages d, that the transfers

$$t_i=\mu_i\left[\sum_{j=1}^{n}c_j^{\mathrm{c}}-\sum_{j=1}^{n}c_j^{\mathrm{nc}}\right]-\left[c_i^{\mathrm{c}}-c_i^{\mathrm{nc}}\right], \mu_i=\frac{d_i'}{\sum_{j=1}^{n}d_j'}, i=1,2,\ldots,n, \tag{4.8}$$

make sure that the grand coalition is in the γ-core of the game. The superscripts c and nc denote cooperation and noncooperation, respectively. Noncooperation can refer to the noncooperative equilibrium or to partial cooperation when a subcoalition deviates and encounters noncooperative behavior of the other countries. The sum of the transfers is 0, $t_i < 0$ if a transfer is received, and $t_i > 0$ if a transfer is paid. The idea of the transfer rule (Equation 4.8) is that each country pays a certain share of the increase in total abatement costs under full cooperation minus its own increase in abatement costs. The relative marginal damages determine the countries' shares. For example, these shares in the model (see Equation 4.1) in the previous section simply become $\beta_i/\Sigma\beta_j$. Linear damages simplify the proof because outsiders always choose the same emission level, as we see from the model in Section 4.1, but the result can be generalized under reasonable conditions (Chander & Tulkens 1997).

Germain et al. (2003) extend the result to a dynamic model, with discrete time and a finite time horizon, in which each country minimizes the discounted sum of the abatement costs $c_i(e_i)$ and damages d_i, as functions of the stock of pollutants s, subject to the accumulation of this stock of pollutants that is given by

$$s(t) = (1-\alpha)s(t-1) + \sum_{i=1}^{n} e_i(t), t = 1, 2, .., T, s(0) = s_0, \tag{4.9}$$

where α denotes the natural rate of degradation. In a dynamic model, noncooperative equilibria depend on the available information and on the level of commitment (e.g., van der Ploeg & de Zeeuw 1992): The Markov-perfect equilibrium results in the dynamic programming framework. The question in this context is whether the grand coalition can be sustained as an element of the γ-core of the game. Backward induction provides the answer. For a given level of the stock of pollutants $s(T-1)$, the game in the last period T with costs of emissions $c_i[e_i(T)]$ and damages of the stock of pollutants $d_i[s(T)]$ is a static game, and therefore there are transfers that yield the result. In the penultimate period $T-1$, the discounted full-cooperative costs δv_i for the last period, as a function of the stock of pollutants $s(T-1)$, are added to the damages $d_i[s(T-1)]$, and the same story applies. With backward induction, v_i simply becomes the cooperative value function of dynamic programming. A transfer rule of the form in Equation 4.8 at each point in time t with

$$\mu_i = \frac{F_i'\left(s^c(t)\right)}{\sum_{j=1}^{n} F_j'\left(s^c(t)\right)}, F_i\left(s^c(t)\right) = d_i\left(s^c(t)\right) + \delta v_i\left(s^c(t)\right), i = 1, 2, \ldots, n, \tag{4.10}$$

where s^c is the full-cooperative stock of pollutants, yields the result under reasonable conditions. This is good news for IEAs, but the result relies on two assumptions. First, it relies on the feasibility of the strong threats as a characteristic of the γ-core concept, as we discuss above. Second, it relies on the feasibility of transfers between countries in the policy arena. We discuss the latter issue in the next subsection.

4.3. Transfers or Side Payments

We discuss above the instability property of the prisoners' dilemma game: Incentives to free-ride on cooperation induce a noncooperative equilibrium with lower welfare. Another problem is that an asymmetric prisoners' dilemma game may have the property that, even if the countries manage to cooperate and to achieve the highest joint welfare, some of the countries may be worse off than they would be in a noncooperative equilibrium. Transfers or side payments may be needed to keep countries cooperating.

A nice example is the so-called acid rain game in Europe, analyzed by Mäler (1989). He formulates a game with abatement costs c, as functions of emissions e of sulfur dioxide and nitrogen oxides, and damage costs d, as functions of dry and wet depositions q of these substances (causing, for example, acidification of soils). Winds take these substances across borders, which turns the problem into a game between the countries in Europe. The objectives are given by

$$c_i(e_i) + d_i(q_i), i = 1, 2, \ldots, n, \underline{q} = A\underline{e}, \tag{4.11}$$

where the matrix A is the transportation matrix that connects the vector of depositions q with the vector of emissions e. Data on this transportation matrix are available from a detailed monitoring program, and reasonable estimates on the abatement costs c are available as well. However, data on the damage costs of acidification are not very reliable. To get results, Mäler assumes that the damages are linear and that the observed emissions e are the emissions in the noncooperative equilibrium. In this way, he derives the marginal damages and the emissions in the full-cooperative outcome. The purpose of the analysis is to show the uneven distribution of abatement requirements and cost reductions in Europe. Some countries, such as the United Kingdom (UK), are worse off under cooperation compared with the noncooperative equilibrium. The reason is that strong northeastern winds take UK-emitted substances to Scandinavia, whereas the UK receives very little from elsewhere because there are no countries to the southwest of the UK. To keep the UK cooperating, compensation is needed. Moreover, some countries, especially those in Central Europe, are required to abate more than other countries, especially those on the border of Europe. These former countries may also require some compensation for their extra effort in achieving the lowest joint costs, even though the cooperation is individually rational for them.

Thus, transfers or side payments may be needed to achieve individual rationality under cooperation. Furthermore, Section 4.1 shows that transfers relax the internal stability conditions, and Section 4.2 shows that transfers are required to make sure that the grand coalition is in the γ-core of the game. This result raises the question of how realistic transfers or side payments in the international policy arena are. In the history of IEAs, monetary side payments between countries are rare. Exceptions are the 1911 North Pacific Fur Seal Treaty, under which the United States and the Soviet Union were required to pay Canada and Japan 15% of their annual harvest of pelts (e.g., Barrett 2003), and the 1972 agreement between France and the Netherlands under which the latter agreed to pay France 35% of the costs to reduce the salinity of the Rhine River. This discussion does not mean that countries do not employ the idea of side payments; rather, such payments occur in another form. For example, in the 1990 London Amendment to the Montreal Protocol on ozone depletion, the phaseout of CFC emissions was coupled with technology transfers to developing countries. Issues are more often linked in the negotiations: A country is willing to give in on one issue if another country is willing to give in on another issue. This perspective also helps in understanding why things happen. For example, in the 1973 treaty with Mexico regarding the Colorado River, the United States agreed to pay the costs of mitigating the salinity of the river just before it flows into Mexico. This behavior of the United States looks irrational in isolation but becomes clearer when it is linked with other issues. Kneese (1988) argues that the United States expected something else in return, such as a preferential position for the import of oil from Mexico.

The topic of issue linkage is widely discussed in various parts of the literature. In his seminal book on negotiations, Raiffa (1982) stresses the importance of linkage in multi-issue bargaining to achieve joint gains. The political science literature pays a lot of attention to this topic as well. Sebenius (1983) argues that linkage of issues can create a zone of possible agreements between parties but also points to the risks of increasing complexity and inducing unintended consequences. For example, granting a preferential position in trade may jeopardize a free trade agreement: Issue linkage may create too many hostages. Keohane (1984) argues that institutions are needed to facilitate issue linkage and that one should look for roughly offsetting issues. Cesar &

de Zeeuw (1996) show that linking issues, which are mirror images of each other, make the full-cooperative outcome of the linked game individually rational so that it can be sustained as a noncooperative equilibrium of the repeated game.

The conclusion is that transfers or side payments are feasible but complicated. In practice, monetary side payments are not often observed, but issue linkage may be an interesting alternative. However, this alternative is not without complications because of both the higher costs of negotiations and the unintended consequences of such an agreement. In fact, there is a trade-off between increasing the options for agreement and increasing the costs. In Section 5, we discuss other negotiation setups.

5. ALTERNATIVE APPROACHES

In Section 4.3, issue linkage is discussed as a form of a side payment, with the purpose of erasing or lowering the incentives to deviate from cooperation. In general, the grim outlook of the basic models above is expected to improve if cooperation becomes more beneficial. For example, if cooperation also means that countries can share the R&D costs for the development of technology, this additional positive externality increases the incentives to cooperate. A small additional benefit of cooperation in the basic model of Section 2 gives larger stable coalitions (e.g., Bargiacchi 2006). Breton et al. (2010) assume that each coalition member punishes outsiders for the irresponsible behavior and that the costs of punishing are lower than the costs of being punished. McGinty (2010) assumes that the marginal benefit of abatement differs between coalition members and outsiders. If the marginal benefit is higher for coalition members so that the returns to abatement increase as a function of coalition size k, the prisoners' dilemma game can turn into a coordination game. A nice example is the MARPOL treaty, intended to limit oil leakage from tankers (Barrett 2003): Only after the decision to focus the treaty on a technology standard (segregated ballast tanks) did it become a success. The reason is that adopting this technology standard has network externalities because members of the agreement have an incentive to ban noncomplying tankers from their ports. A coordination game has two noncooperative equilibria, in which the countries choose either high or low abatement. Therefore, countries have only to coordinate their behavior; they do not have to be afraid of deviations (this type of game also turns up in the context with tipping points, discussed below). Breton et al. and McGinty position IEAs in a dynamic, evolutionary framework in which the successful strategies spread in the population of countries. Stability is now defined as the equilibrium of an evolutionary game. These researchers show that the size of an evolutionary stable coalition can be much larger than the size of a static stable coalition (depending on the parameters). This evolutionary game approach may be relevant in the light of recent changes in the COPs. We consider this approach in more detail below in Section 5.1.

5.1. Evolutionary Game Approach

The failure of the Kyoto Protocol can be seen as a failure to reach a unanimous top-down agreement on emission reductions. At the COP in Copenhagen in 2009, a group of countries took the initiative to announce emission reductions from the bottom up, hoping that other countries would follow. The European Union (EU) had the strategy of stepping up to a 30% emission reduction if the other developed countries matched the contribution of the EU and the developing countries took appropriate measures. In preparation for the 2015 COP in Paris, countries are invited to pledge their contributions. The idea is that in this way countries may stimulate each other to move to higher emission reductions. At the same time, they may threaten to refrain from higher emission reductions if other countries do not deliver. This scenario can be modeled by the conditional strategies discussed in Section 3.1: Countries start to cooperate and continue to cooperate as long as the other countries cooperate. This strategy works if the discount factor is sufficiently high. However, not all countries can be expected to be willing to initially use this reciprocal strategy. The question then is whether these countries are willing to switch to the reciprocal strategies later so that at the end full conditional cooperation may result. This result will occur if the reciprocal strategies are successful. We can analyze this situation as an evolutionary game (e.g., Nowak & Sigmund 2006).

Suppose that the countries can either abate or not so that we have a model of the type described in Section 2.3. We interpret the benefits b as a contribution to total abatement. We again consider an asymmetry with two types of countries. Countries of type 1 contribute b_1 to total abatement and have costs c_1, whereas countries of type 2 contribute b_2 to total abatement and have costs c_2. We assume that $b_1 > b_2$ and $c_1 > c_2$ so that countries of type 1 can be seen as developed countries with high emissions and high costs of emission reductions and countries of type 2 can be seen as less developed countries. Each country has one of two strategies: no abatement (or defect) or a reciprocal (or trigger) strategy. The trigger strategies in Section 3.1 are extended to the following strategy: Cooperate if, in the previous period, at least a certain number of countries have cooperated (Taylor 1976). This trigger is further specified as a certain number of countries of type 1 and a certain number of countries of type 2.

The question is whether a form of evolution in which countries may switch from one strategy to the other can drive out the no-abatement strategies so that full conditional cooperation results. This scenario can be modeled with so-called replicator dynamics

$$\begin{aligned} \dot{\rho}_i(t) &= \rho_i(t)\Big(\pi_i^{\mathrm{r}} - \big[\rho_i(t)\pi_i^{\mathrm{r}} + \big(1-\rho_i(t)\big)\pi_i^{\mathrm{d}}\big]\Big) \Rightarrow \\ \dot{\rho}_i(t) &= \rho_i(t)\big(1-\rho_i(t)\big)\big(\pi_i^{\mathrm{r}} - \pi_i^{\mathrm{d}}\big), \rho_i(0) = \rho_{i0}, i = 1, 2, \end{aligned} \tag{5.1}$$

where ρ_i denote the fractions of reciprocators of type i, π^{r} denotes the net benefits of a reciprocator, and π^{d} denotes the net benefits of a defector. If π^{r} is higher than the average net benefits, the fraction of reciprocators will increase. π^{r} and π^{d} are performance measures that arise in the interaction of the strategies in a repeated prisoners' dilemma game. These performance measures depend on the current fractions of reciprocators so that the two differential equations in Equation 5.1 are linked. In biology, this performance is interpreted as fitness, and success is passed on in reproduction. Here success is observed and spreads to more countries. Boyd & Richerson (1988) use the same model, with a biological interpretation, and determine the fitness measures in a symmetric game. Ochea & de Zeeuw (2015) determine π^{r} and π^{d} in the asymmetric case. They show numerically, using the replicator dynamics in Equation 5.1, that full conditional cooperation can indeed occur, provided that the initial fractions of reciprocators are sufficiently high, that the countries interact for a sufficiently long period of time (in not-too-large groups), and that the thresholds in the trigger strategies are sufficiently high.

The policy conclusion is that a bottom-up approach, with reciprocal strategies, may work but that the reciprocators have to be tough: Countries have to threaten that they will refrain from emission reductions if an insufficient number of other countries continue to cooperate. This is bad news again for IEAs because this tough behavior has not yet been observed in practice. But the good news is that it may work in situations with a group of countries that are initially not willing to cooperate, not even conditionally. Evolution in the sense that this group adopts reciprocal strategies, because these strategies prove to be successful, may take care of this issue. An important condition, however, is that the initial willing-to-cooperate group is sufficiently large. In the next subsection, we turn to the effect of tipping points on IEAs.

5.2. Tipping Points

A significant portion of environmental damages, such as climate change, is expected to occur in the form of a nonmarginal shock or a catastrophe (Lenton et al. 2008). At a so-called tipping point, the climate system will move into another domain of attraction, with a sudden drop in welfare. The current assumption is that this tipping point can be avoided by a sufficiently high abatement level. For example, if we keep the level of atmospheric CO_2 concentration below 350 parts per million by volume (or keep the rise in global mean temperature to below 2°C above the preindustrial level), we can avoid these catastrophic damages. Catastrophe is a threat posed by nature, and the question is what the effect is on IEAs. Barrett (2013) shows that the underlying prisoners' dilemma game can turn into a coordination game. We can easily see this possibility by introducing catastrophic damage X into the basic game (Equation 2.2):

$$\begin{aligned}\pi_i &= A - X - 0.5a_i^2, A = \sum_{j=1}^{n} a_j, A < \overline{A},\\ \pi_i &= A - 0.5a_i^2, A \geq \overline{A}, i = 1, 2, \ldots, n,\end{aligned} \tag{5.2}$$

where $\overline{A}$ denotes the abatement level needed to avoid the catastrophe. In the absence of catastrophes, total abatement under full cooperation is equal to n^2, with net benefits equal to $0.5n^2$ for each country (see Equation 2.4). If $\overline{A} > n^2$ so that more abatement is needed to avoid the catastrophe, the first question is whether catastrophic damage X is high enough such that catastrophe avoidance is collectively rational. The condition for catastrophic damage X is high enough if (assuming symmetry in contributions to abatement)

$$\overline{A} - 0.5\left(\frac{\overline{A}}{n}\right)^2 > 0.5n^2 - X. \tag{5.3}$$

The second question is whether catastrophe avoidance, if the other countries do the same, is individually rational. The optimal alternative for each country is to abate 1 and have costs 0.5. The condition for symmetric individual rationality is

$$\overline{A} - 0.5\left(\frac{\overline{A}}{n}\right)^2 > (n-1)\frac{\overline{A}}{n} + 1 - 0.5 - X. \tag{5.4}$$

If catastrophic damage X is high enough so that Condition 5.4 is satisfied, the basic game (Equation 5.2) turns into a coordination game. This game has two noncooperative equilibria: All countries abate 1 or $\overline{A}/n$ (we consider only symmetric equilibria). This result implies that the threat of a catastrophe induces the option of countries coordinating on a noncooperative equilibrium; they behave as they would in the cooperative outcome, namely to avoid the catastrophe.

Because the full cooperative outcome is also a noncooperative equilibrium (if Condition 5.4 is satisfied), the grand coalition is stable, provided that we stick to symmetry. If the outsider can commit to lower abatement while conditions are still optimal for the remaining coalition to avoid the catastrophe, the grand coalition is not stable. In contrast, if the remaining coalition can commit, it can force the outsider to higher abatement, in which case the grand coalition is again stable. This is good news for IEAs: Coordinating on a noncooperative equilibrium is easier than cooperating in a prisoners' dilemma game.

However ("however" is common in this literature), if the abatement level $\overline{A}$ that is needed to avoid the catastrophe is uncertain, this property is lost. This result can easily be seen by introducing a cumulative probability density function $F(A)$ so that $1 - F(A)$ denotes the probability that the abatement level A is not sufficiently high to avoid the catastrophe. The basic game (Equation 5.2) changes to (Barrett 2013)

$$\pi_i = A - \big(1 - F(A)\big)X - 0.5a_i^2, A = \sum_{j=1}^{n} a_j, i = 1, 2, \ldots, n, \tag{5.5}$$

which is again a prisoners' dilemma game! Barrett & Dannenberg (2012) report experiments in which they confront the participants with either a certain threshold $\overline{A}$ or a uniform probability distribution over a range of thresholds A. They clearly find that, in the uncertainty case, the participants behave as they would in a prisoners' dilemma game, whereas in the certainty case, they coordinate on the threshold that avoids catastrophe. This is bad news for IEAs because science cannot predict with certainty that an atmospheric CO_2 concentration of 350 parts per million by volume will avoid a climate catastrophe. In the next subsection, we discuss a different IEA design.

5.3. Technology Agreements

We argue above that if countries can share the R&D costs for the development of technology, this additional positive externality increases incentives to cooperate. We can take this scenario one step further and argue that if the new technology spills over to other countries, an extra benefit occurs because these other countries will emit less. In an asymmetric situation, coalition countries may invest even more in R&D than is needed for their own benefit, with the goal of inducing other countries to emit less. Therefore, an international agreement on R&D expenditures for the development of a new environmentally friendly technology may work better because of the option to influence other countries' emissions.

Hoel & de Zeeuw (2014) analyze this possibility in the following way. In the first stage, countries decide whether they want to be a member of the coalition; in the second stage, the coalition decides on the level of R&D expenditures; and in the third stage, each country decides whether to adopt the new technology. The countries differ in their valuation of adopting the new technology, and R&D expenditures lower the costs of adoption. Therefore, the higher are the R&D expenditures, the higher is the number of countries that adopt the new technology. We require that the coalition investing in R&D be stable in the sense of Section 2.3. Therefore, just enough countries, starting with those that have the highest valuation, join the coalition to make R&D investment worthwhile. The most important result of the analysis is that full adoption of the new technology may occur: The coalition may want to invest sufficiently in R&D to induce all the n countries to adopt the new technology because of the lower emissions. This is good news for IEAs. Partial adoption of the new technology occurs in two situations. If some countries have a very low valuation, investing so much in R&D that they will adopt does not pay. Similarly, if just a few countries have a low valuation, investing so much in R&D that they will adopt also does not pay. In a more balanced situation, however, a technology approach to IEAs may work.

6. CONCLUSIONS

Global environmental problems are becoming more urgent, but solving these problems is very challenging because voluntary cooperation between sovereign states is necessary. This article reviews several game-theoretic approaches that give some perspectives but also have some drawbacks. The main issues are how to suppress free-rider behavior and how to make outcomes individually rational. The main theme is that posing threats to the other countries may work but may be hard to implement in the international policy arena. In asymmetric situations, side payments may be needed but are hard to implement; issue linkage may be an option.

Another main theme is the design of IEAs. Top-down general agreements on emission reductions such as the Kyoto Protocol did not work. This review discusses a number of alternatives. One possibility is to design an IEA such that joining the agreement yields more positive externalities. A second approach is that some countries take the lead and show that reciprocal strategies work so that evolution may lead to full conditional cooperation. A third option is to focus the IEA on technology development and make that technology publicly available so that a larger number of countries will adopt the new technology and emit less.

This article does not give a complete overview of the literature, of course. It focuses on some of the fundamental debates and provides ways to progress on research on this important and challenging issue.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I am grateful to Scott Barrett, Carlo Carraro, Herman Cesar, Effrosyni Diamantoudi, Michael Finus, Michael Hoel, Yulia Pavlova, Santiago Rubio, Eftichios Sartzetakis, and Henry Tulkens for many fruitful discussions on the subject of this review.

LITERATURE CITED

d'Aspremont C, Jacquemin A, Gabszewicz J, Weymark J. 1983. On the stability of collusive price leadership. *Can. J. Econ.* 16:17–25

Bargiacchi R. 2006. *Modelling and testing behavior in applications to climate change*. PhD Thesis, Tilburg Univ.

Barrett S. 1994. Self-enforcing international environmental agreements. *Oxf. Econ. Pap.* 46:878–94

Barrett S. 2003. *Environment and Statecraft: The Strategy of Environmental Treaty-Making*. Oxford, UK: Oxford Univ. Press

Barrett S. 2006. Climate treaties and "breakthrough" technologies. *Am. Econ. Rev. Pap. Proc.* 96:22–25

Barrett S. 2013. Climate treaties and approaching catastrophes. *J. Environ. Econ. Manag.* 66:235–50

Barrett S, Dannenberg A. 2012. Climate negotiations under scientific uncertainty. *PNAS* 109:17372–76

Botteon M, Carraro C. 1997. Burden sharing and coalitional stability in environmental negotiations with asymmetric countries. In *International Environmental Negotiations: Strategic Policy Issues*, ed. C Carraro, pp. 26–55. Cheltenham, UK: Edward Elgar

Boyd R, Richerson P. 1988. The evolution of reciprocity in sizable groups. *J. Theor. Biol.* 132:337–56

Breton M, Sbragia L, Zaccour G. 2010. A dynamic model for international agreements. *Environ. Resour. Econ.* 45:25–48

Carraro C, Siniscalco D. 1993. Strategies for the international protection of the environment. *J. Public Econ.* 52:309–28

Cesar H, de Zeeuw A. 1996. Issue linkage in global environmental problems. In *Economic Policy for the Environment and Natural Resources*, ed. A Xepapadeas, pp. 158–73. Cheltenham, UK: Edward Elgar

Chander P, Tulkens H. 1995. A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. *Int. Tax Public Finance* 2:279–93

Chander P, Tulkens H. 1997. The core of an economy with multilateral environmental externalities. *Int. J. Game Theory* 26:379–401

Chwe M. 1994. Farsighted coalitional stability. *J. Econ. Theory* 63:299–325
Diamantoudi E, Sartzetakis E. 2002. *International environmental agreements—the role of foresight*. Work. Pap., Dep. Econ., Univ. Aarhus
Finus M. 2003. Stability and design of international environmental agreements: the case of transboundary pollution. In *The International Yearbook of Environmental and Resource Economics 2003/2004*, ed. H Folmer, T Tietenberg, pp. 82–158. Cheltenham, UK: Edward Elgar
Friedman J. 1986. *Game Theory with Applications to Economics*. Oxford, UK: Oxford Univ. Press
Fuentes-Albero C, Rubio S. 2010. Can international environmental cooperation be bought? *Eur. J. Oper. Res.* 202:255–64
Germain M, Toint P, Tulkens H, de Zeeuw A. 2003. Transfers to sustain dynamic core-theoretic cooperation in international stock pollutant control. *J. Econ. Dyn. Control* 28:79–99
Hoel M. 1992. International environmental conventions: the case of uniform reductions of emissions. *Environ. Resour. Econ.* 2:141–59
Hoel M, de Zeeuw A. 2014. Technology agreements with heterogeneous countries. In *Towards a New Climate Agreement: Conflict, Resolution and Governance*, ed. T Cherry, J Hovi, D McEvoy, pp. 199–219. London/New York: Routledge
Keohane R. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, NJ: Princeton Univ. Press
Kneese A. 1988. *Environmental stress and political conflicts: salinity in the Colorado River*. Work. Pap., Resour. Future
Lenton T, Held H, Kriegler E, Hall J, Lucht W, et al. 2008. Tipping elements in the earth's climate system. *PNAS* 105:1786–93
Mäler K-G. 1989. The acid rain game. In *Valuation Methods and Policy Making in Environmental Economics*, ed. H Folmer, E van Ierland, pp. 231–52. Amsterdam: Elsevier
McGinty M. 2007. International environmental agreements among asymmetric nations. *Oxf. Econ. Pap.* 59:45–62
McGinty M. 2010. International environmental agreements as evolutionary games. *Environ. Resour. Econ.* 45:251–69
Nowak M, Sigmund K. 2006. Five rules for the evolution of cooperation. *Science* 314:1560–63
Ochea M, de Zeeuw A. 2015. Evolution of reciprocity in asymmetric international environmental negotiations. *Environ. Resour. Econ.* In press
Pavlova Y, de Zeeuw A. 2013. Asymmetries in international environmental agreements. *Environ. Dev. Econ.* 18:51–68
Pintissalgo P, Finus M, Lindroos M, Munro M. 2010. Stability and success of regional fisheries management organizations. *Environ. Resour. Econ.* 46:377–402
van der Ploeg F, de Zeeuw A. 1992. International aspects of pollution control. *Environ. Resour. Econ.* 2:117–39
Raiffa H. 1982. *The Art and Science of Negotiation*. Cambridge, MA: Harvard Univ. Press
Sebenius J. 1983. Negotiation arithmetic: adding and subtracting issues and parties. *Int. Organ.* 37:281–316
Taylor M. 1976. *Anarchy and Cooperation*. London: John Wiley & Sons
de Zeeuw A. 2008. Dynamic effects on the stability of international environmental agreements. *J. Environ. Econ. Manag.* 55:163–74