An Article Submitted to

The B.E. Journal of Economic Analysis & Policy

Manuscript 2593

Equity, Heterogeneity and International Environmental Agreements

Charles D. Kolstad*

*University of California - Santa Barbara, kolstad@econ.ucsb.edu

Equity, Heterogeneity and International Environmental Agreements*

Charles D. Kolstad

Abstract

Much of the literature on international environmental agreements (IEA) considers the case of identical countries. There is a much smaller literature concerning the more complex but more realistic case of country heterogeneity. This paper involves modifying the standard static homogeneous country model of international environmental agreements (IEA). In particular, we consider two types of countries, differing in size as well as in marginal damage from pollution. Although the IEA does not have a unique size in this case, we do introduce two equilibrium refinements and explore the implications for coalition size. The two refinements include one based on efficiency and one based on equity.

*Department of Economics and Bren School of Environmental Science & Management, University of California, Santa Barbara; University Fellow, Resources for the Future; Research Associate, National Bureau of Economic Research. Comments from Scott Barrett, Don Fullerton, Ross Mohr, an anonymous referee and participants at an RFF/University of Chicago workshop in January 2010 are much appreciated.

The 2009 stalemate at COP-15 in Copenhagen regarding taking action on climate change illustrates the difficulties in getting nearly 200 countries to agree to act on anything, let alone on an issue as complex and significant as climate change.[1]

The challenge faced by the participants at Copenhagen was to forge an international environmental agreement (IEA) to reduce greenhouse gas emissions leading to climate change. What makes the problem particularly tough is that countries cannot be forced to participate but must do so voluntarily. And because a reduction in greenhouse gases is truly a global public good, while abatement is privately costly, it is far more attractive to free-ride on any agreement than to participate as an abating party. From an incentive point of view, the only stable agreement is one for which every participant finds it individually rational to participate; that is, participation makes the country better off than free-riding.

The issue of the formation of IEAs has also been of concern to the academic community for some time, both in political science (eg, Young 1994) and economics (eg, Barrett 2003). The main question asked by this literature is what characteristics of the problem lead to strong or weak IEAs? A corollary to this question is what structural features can be incorporated into IEAs to improve their performance?

Although the political science literature is more nuanced than the economics literature in generating understanding of IEAs, the economics literature has provided many insights into how IEAs work. One of the first papers in this literature is due to Scott Barrett (1994). In that paper, he develops a simple game-theoretic model and argues that IEAs are unable to improve very much on the status quo of no agreement: either agreements involving very many parties are not stable or, if large agreements can be formed, they do not improve welfare much relative to the case of no agreement. Other questions/issues include the effect of uncertainty on IEA formation[2] and the use of commitment mechanisms to strengthen agreements.[3]

This paper raises the question of the effect of heterogeneity of countries on the formation of agreements. Simply put, is it easier for a meaningful agreement to form when countries are similar or when they are not? Do differences among countries retard or enhance the formation of IEAs? Does heterogeneity increase abatement and thus welfare? Implications are important for improved design of an IEA.

[1] COP-15 is the fifteenth "Conference of the Parties" to the Framework Convention on Climate Change treaty.
[2] On uncertainty, see Na and Shin (1998), Helm (1998), Kolstad (2005), Ulph (2004) and Kolstad and Ulph (2008).
[3] On commitment, see Carraro and Siniscalco (1993) and Barrett (1997b).

Intuition suggests that if countries are very similar, it may be easier to reach agreement to solve a common problem, and thus homogeneity can increase efficiency. It turns out that this intuition is correct.

I. BACKGROUND

In reviewing the literature on International Environmental Agreements (IEAs), we first consider the general question of determinants of stable IEAs. We then turn to the issue of heterogeneity of participants, the issue of concern in this paper. Finally, we move away from pure theory, examining the experimental evidence on IEA formation.

A. IEA Size

International environmental agreements are the subject of a significant economic literature, primarily post-1990.[4] Most of the literature focuses on self-enforcing agreements; i.e., agreements that are structured so that they are effective and cohesive (or stable) without recourse to a larger context of international law for enforcement.[5] The most common, and simplest, notion of stability draws on the cartel stability literature (eg, d'Aspremont et al, 1983; Donsimoni et al, 1986), wherein a stable cartel is defined as a cartel for which no individual members has an incentive to leave nor any outsider to join. This turns out to be a very strong assumption in the sense that many potential cartels fail the test. Chander and Tulkens (1992, 1994) adopt an even stronger assumption that should any individual member of a voluntary agreement choose to leave, the entire agreement would be null and void. Between these two concepts is the notion of "farsighted stability" (Ecchia and Mariotti, 1997; Eyckmans, 2003). The idea here is that an agreement is stable if no country has an incentive to leave or join, but in evaluating those incentives, countries look beyond their act of joining or leaving to the credible additional actions that other countries may take.

Some of the earliest work (Hoel, 1992; Carraro and Siniscalco, 1993; Barrett, 1994) finds that such agreements are either unlikely to consist of very many participants or, if the agreement involves a large number of countries, then

---

[4] Wagner (2001) and Barrett (2003, 2007) provide comprehensive reviews of this literature. See also Finus (2001) and several chapters in the volume edited by Guesnerie and Tulkens (2009).

[5] The literature on IEA uses the term "stable" to refer to coalitions of countries that will tend to stay together and not break up. This is a somewhat unfortunate choice of words, since stability is generally a dynamic term referring to the tendency of an equilibrium or coalition to remain unchanged when conditions are perturbed slightly. That is not the meaning here. In the interests of clarity, we use the standard term "stability" here to describe coalitions that are cohesive, recognizing the less-than-satisfactory nature of the term.

the gains from cooperation must be low.[6] The basic idea is that the incentives for free-riding must be low or else most countries will choose to free-ride and not belong to the agreement. A low incentive to free-ride is the flip-side of a small gain from cooperating.

This conclusion is based on a simple model of N homogeneous countries, each with a marginal cost of abatement c and a marginal damage from pollution b. Each country acts as a payoff maximizer, choosing emissions, e_i, from the interval [0,1], which may be discrete or continuous. In the simplest form of this model, payoffs are linear, and each country either pollutes or abates, resulting in a payoff of Π_i:

$$\Pi_i = ce_i - bE, \text{ where } E=\sum e_i. \tag{1}$$

The countries play a two-stage game: in the first stage, countries announce whether or not they will join the IEA. In the second stage, the IEA and the fringe choose emission levels. In each stage, Nash play is typically assumed. Using the notion of cartel stability mentioned above, it is easy to show that c/b is the unique size of a stable IEA, or more precisely the unique size is the integer rounded up from c/b. The logic is that with an agreement size of c/b, it is optimal for the IEA to abate. But for an agreement size of $c/b-1$, it is optimal to pollute. Thus every member of the IEA views itself as pivotal in that if any single member defects, the entire IEA ceases abating.

Payoffs are normalized so that if all countries abate, the aggregate payoff is zero. With a stable IEA of size c/b (ignoring the integer issues), the aggregate payoff is $-b(N-c/b)^2$. In a noncooperative equilibrium, each country pollutes with an aggregate payoff of $cN-bN^2$. Thus the gain from an IEA is $c(N-c/b)$. This suggests that an IEA does the most good when c/b is small and the least good when c/b is large. But this also means that when an IEA can do the most good, the equilibrium size will be small; and conversely, when the equilibrium size is large, the gain from the IEA is smaller.

The literature varies on the structure just described. Some authors set the problem up as a Stackelberg equilibrium, wherein the IEA is the leader and the fringe members are followers. As mentioned above, the strategies may be discrete as illustrated or they may be continuous. Payoffs may be linear as illustrated or quadratic or even more general. Finally, the model above is fundamentally static – a one-shot game. An obvious extension is to consider a repeated game framework or other dynamic representations. Clearly international agreements on the environment are not reached in a static context; however the

---

[6] Many of the results in the literature rely on simulation models and are thus less in the nature of proofs than illustrations. Rubio and Ulph (2007) provide analytic proofs of some of these early results.

difficulty of developing credible dynamic models has limited the literature to static models for the most part (though exceptions are Finus, 2001, and Barrett, 2002).

The example above illustrates a fundamental issue in this entire literature. How big will an agreement be? What is the size of a "stable" agreement? The answers hinge on what holds an agreement together, or what keeps countries in the agreement. Assumptions can range from complex commitment procedures, to punishments for defecting, to simple self-interest without commitment (as in the example).

B. Heterogeneity

Another issue is the extent to which participating countries are homogeneous or heterogeneous. Most of the results in this literature assume homogeneity of participating countries. But "real" countries are not all the same. Countries may be large or small. Some countries see small marginal damage from climate change (in the sense of small additional damage from additional change in the climate), due to low population levels, large geographic scale or other geographic specifics. Other countries experience high marginal damages. Similarly, the marginal cost of abatement (per ton of greenhouse gases may differ dramatically from one country to another. However, many agreements involve trading mechanisms designed to equalize the marginal cost of abatement among countries; thus marginal costs may not be a major source of heterogeneity. But total costs and total emissions may vary significantly. In comparing the United States to Switzerland, a big difference is that the US emits far more. If the US and Switzerland reduce emissions to a given level of marginal costs, the absolute size of the reduction will be far greater than in Switzerland.

The literature is modest on the significance of heterogeneity. One of the earliest papers on this topic is Barrett (1997a). In that paper he focuses on how countries within a coalition may share the joint payoff from cooperation. He uses Shapley values to divide the joint payoff and examines a world in which there are two types of countries. Although he is able to develop some analytic results, for the most part he uses simulations to derive his results. Botteon and Carraro (2001) extend this framework to five types of countries, again using simulation. McGinty (2006) extends this to 20 countries, also using simulation. To illustrate the difficulty of developing analytic solutions to this problem, it is only recently that an analytic solution has been found for the heterogeneity problem with two types of countries, continuous strategies and quadratic payoffs (Fuentes-Albero and Rubio, 2010). Results of this literature are mixed, though a common conclusion of the simulation literature is that, in some cases, heterogeneity increases the effectiveness of international environmental agreements.

Barrett (2001, 2002) takes a different approach to this problem. Rather than view the problem as a static game, he views it as a repeated game and posits credible punishments that can be built into the agreement. The classic problem with enforcing an agreement with punishment is that it is often not in a country's self-interest to punish. By weakening the abatement undertaken by the coalition, Barrett is able to construct renegotiation-proof agreements that involve credible punishments. Furthermore, weakening the requirements of coalition members tends to expand the size of the coalition. Barrett (2001) specifically shows that heterogeneity of countries can reduce the free-riding problem and thus help support larger coalitions. In his model, heterogeneity facilitates commitment. And commitment is the big problem in self-enforcing agreements. One of his results is that if the countries are substantially different, then only countries with high marginal damage will form a coalition/IEA. But these countries have a collective incentive to bribe the low damage countries to join the IEA, increasing the effectiveness of the IEA. Transfers are a moot issue in the homogeneous country case. But in the heterogeneous country case, the surplus that accrues to some countries is more than enough to "bribe" other countries to participate. The author points out that this is illustrated by the Montreal Protocol. Rich countries stood to gain the most and could afford to pay developing countries to participate.

C. Experimental Evidence

Until recently, virtually all of the economics literature on IEAs has been theoretical in nature. After all, treaties are complex, and it is difficult to do econometrics with them. But an IEA is fundamentally an institution for coordinating voluntary contributions of countries to a global public good. And we know from the literature on voluntary contributions to public goods that theory is often at odds with empirical and experimental evidence (eg, Isaac and Walker, 1988).

Burger and Kolstad (2009) take an experimental approach to validating the theoretical models of IEA formation. Using a structure very similar to the model of Barrett (1994) and others, they find that one of the basic theoretical results does not hold experimentally. As discussed above, a basic theoretical result is that a smaller abatement cost relative to marginal damage from emissions (c/b), results in a smaller stable IEA and thus a lower aggregate level of abatement. Experiments find the opposite result, very much in line with the experimental work on voluntary provision of public goods. In fact, the parallel with the voluntary provision of public goods literature is striking; probably the same underlying forces are at work increasing cooperation relative to the theory. This finding calls into question the validity of theoretical models of IEA formation.

Kosfeld et al (2009) examine the same problem, though they couch it as a public goods provision problem. Furthermore, they only examine the case of four countries/participants. Nevertheless, they confirm the result of Burger and Kolstad (2009) that as the abatement cost-marginal damage ratio increases the size of the stable IEA increases, in contradiction to theory. However, the authors posit inequality aversion as an explanation for the experimental results. This represents a new and innovative direction of research in bringing together theory with experimental results on IEA formation.

II. A MODEL OF AGREEMENTS

We are interested in comparing the situation where all countries are identical to the situation where countries differ. Countries may differ in terms of size, marginal abatement costs, and/or marginal damage. The model we examine is closely related to that of Barrett (2001) and Saha (2007).

The basic structure of the problem is the standard two-stage game model of an IEA, as discussed earlier in this paper.[7] In the first stage, countries decide whether or not to join an IEA or to stay in the fringe (the "membership game"). In the second stage, the IEA acts as a joint payoff maximizer and acts as a single decision-maker in choosing emissions and competing with the individual members of the fringe (the "emissions game"). Equilibrium is a Nash equilibrium in the membership game followed by a Nash equilibrium in the emissions game.

We will introduce an additional variable into this problem, size of the county, and also let the abatement costs and marginal damages vary. Through the size variable we are capturing two things. One is that abatement in large countries is expected to be greater than in small countries. Furthermore, if one thinks of marginal damage as fundamentally per capita, then aggregate marginal damage in a country is also proportional to size. Letting size of county i be S_i, payoff for country i is still as in Eqn. (1), with two modifications:

- e_i is either 0 or S_i (rather than 0 or 1) (2a)
- b_i is proportional to per capita damage (β_i) and S_i: $b_i \approx \beta_i S_i$. (2b)

In other words, marginal damage b_i can vary from country to country, both in per capita terms and in aggregate terms. Furthermore, although the marginal cost of abatement is equalized across countries (due to approaches like the Clean

[7] As discussed in the previous section, there are more complex models of international agreements. The one adopted here is pedagogically attractive and the most common in the literature.

Development Mechanism), the aggregate amount of abatement depends on the size of the country.

To illustrate the richness of this representation, consider a few examples. The US is a large country but one with relatively modest per capita damage from climate change (though this is clearly debatable). Canada and Australia are much smaller countries, though with similar per capita damage. India is a large country with potentially large per capita damage from climate change (though moderated by low per capita income).

To be specific, consider i=1,...,N countries, each of "size" S_i, emitting pollution (e_i) which contributes to the global commons ($E=\sum_j e_j$). For simplicity, assume each country makes a discrete choice regarding how much pollution to emit, which without loss of generality may be restricted to 0 or S_i: to abate (e_i=0) or to pollute ($e_i= S_i$). Marginal damage from aggregate emissions varies from one country to another but the marginal cost of abatement does not. As in Eqn. (1), each country's payoff is represented as a linear function of own emissions and aggregate emissions:

$$\begin{aligned} \Pi_i(e_i,E_{-i}) &\equiv ce_i - b_iE \\ &= ce_i - b_i(e_i + E_{-i}) \qquad (3) \end{aligned}$$

where $E_{-i}=\sum_{j\neq i} e_j$. Thus $\gamma_i \equiv b_i/c$ is the benefit-cost ratio for emissions control – the ratio of own marginal environmental damage from emissions to the marginal cost of emissions control. Clearly we wish to focus on the case of $\gamma_i < 1$ ($b_i < c$), and we make that assumption; otherwise abatement is a dominant strategy for individual countries, and cooperation is unnecessary. Furthermore, the γ_i's cannot be too small; otherwise full cooperation (and efficiency) will not result in any abatement.

A. A Self-Enforcing IEA

As mentioned earlier, we represent the formation of an IEA as a two-stage game, consisting first of a membership game followed by an emissions game. The first-stage membership game is an announcement game in which countries decide whether or not to join the IEA.[8] In the second-stage emissions game, the membership of the IEA is given and countries decide how much to emit. In the emissions game, we assume the members of the IEA decide on emissions jointly, and the non-members (the fringe) decide individually. The coalition acts as a singleton and each member of the fringe acts in a Nash noncooperative manner; a

[8] In the announcement game, each country announces "in" or "out." A Nash equilibrium is a set of announcements for which no country can do better by unilaterally changing its announcement.

Nash equilibrium results. Membership of the coalition cannot change in the emissions game.

We now assume two types of countries in our world, characterized by different sizes and marginal damages: S_j and b_j. It is straightforward to consider the case of M different types, though the presentation is messier. Type 1 countries have size S_1 and marginal damage b_1; type 2 countries have size S_2 and marginal damage b_2. Without loss of generality, assume $b_1/S_1 < b_2/S_2$. Further, assume N_j of type j countries ($N=N_1 + N_2$), but only n_j of type j countries in a coalition/agreement ($n_j \leq N_j$). Thus a coalition (n_1,n_2) consists of n_1 of country 1 and n_2 of country 2. Further, by assumption, full cooperation involves abatement by all countries. As we will see below, this is equivalent to $N_1\gamma_1 + N_2\gamma_2 \geq 1$.

Equilibrium can be determined using backward recursion—first find the equilibrium in the emissions game, conditional on the membership in the IEA, and then find the equilibrium in the membership game. In the emissions game, it is a dominant strategy for the fringe to pollute. Conditional on having (n_1,n_2) members of the coalition, the coalition sees the following coalition aggregate payoff (CAP):

Abate: $$CAP = -n_1b_1[S_1(N_1-n_1)+ S_2(N_2-n_2)] - n_2b_2[S_1(N_1-n_1)+ S_2(N_2-n_2)]$$

$$= -(n_1b_1 + n_2b_2)\,[S_1(N_1-n_1)+ S_2(N_2-n_2)] \qquad (4a)$$

Pollute: $$CAP = (n_1S_1+n_2S_2)c - [S_1N_1 + S_2N_2)](n_1b_1 + n_2b_2) \qquad (4b)$$

⇨ Abate weakly preferred iff $n_1b_1 + n_2b_2 \geq c$ (4c)

In the membership game, it is clear that an equilibrium in the announcement game is associated with $n_1\gamma_1 + n_2\gamma_2$ being as close to 1 as possible without being less than 1 ($n_1b_1 + n_2b_2$ is as close to c as possible without violating the inequality). Then all members of the coalition are pivotal in the sense that if any one defects, the coalition's decision rule on emissions switches from abate to emit, and thus the defector is worse off.

An interesting consequence of the condition in Eqn. (4c) is that size of a country does not matter (from the point of view of emissions abatement). What does matter is damages. In fact, Eqn (4c) can be viewed as a variant of the Samuelson condition for the efficient provision of a public good. The left hand side of the inequality is the aggregate marginal damage from one ton of emissions, and the right hand side is the marginal cost of abating one ton.

We will now make the assumption that in equilibrium, Eqn. (4c) holds with equality. This is approximately true but ignores the integer nature of n_1 and

n_2. We will be interested in comparative statics and welfare, and thus we argue this simplification is reasonable. Thus the equilibrium condition is

$$n_1\gamma_1 + n_2\gamma_2 = 1 \Leftrightarrow \quad n_1 b_1 + n_2 b_2 = c \qquad (5)$$

Note that the problem has multiple equilibria as stated. Many combinations of n_1 and n_2 will satisfy Eqn. (5).

B. Equilibrium Refinement

Multiple equilibria usually means a degree of arbitrariness in narrowing the set of possible equilbria. The issue here is: for all of the possible coalition compositions, which is likely to emerge? Two clear candidates emerge for shrinking this set, selecting a narrower set of equilibria for further consideration – what we call a "refinement." One criterion for refinement is based on efficiency; the other is based on equity (consistent with Kosfeld et al, 2009). The efficiency criterion would involve the coalition that maximizes the joint payoff for coalition members. The equity criterion is a bit more complex.

In fact there are many ways of defining an equity criterion. One could modify the efficiency criterion so that coalition members also take into account the welfare of non-members. As Prof. Scott Barrett has pointed out in discussant comments on this paper, equity could involve the richest countries (per capita) making the greatest sacrifices. Rather than consider all possible ways of characterizing equity, we adopt one of the simplest representations: equating the payoff to each of the two groups of countries within the coalition (there may be other interpretations of what equity means).

Another issue, also raised by Prof. Barrett, is that if one is using an equity refinement for the second stage game (the emissions game), then it would be logical to use an equity criterion for the first stage game – the participation game. This is a valid point; however, we are not expanding the number of equilibria in this game, just choosing among several. Thus we are not changing the fundamental nature of the two-stage game.

The aggregate payoff for each of the two groups in the abating coalition (CAP_j) can easily be computed:

$$CAP_j = -n_j b_j\,[S_1(N_1-n_1) + S_2(N_2-n_2)],\ for\ j=1\ or\ 2 \qquad (6)$$

The two refinement conditions would then be:

Efficiency: Find (n_1,n_2) to max $[CAP_1 + CAP_2]$ (7a)

Equity: Find (n_1,n_2) such that $CAP_1 = CAP_2$ (7b)

In solving for (n_1,n_2) that satisfy either of these two refinement conditions, we must also satisfy Eqn (5), and n_j must lie on the closed interval $[0,N_j]$. There may be more than one solution to each of these two conditions.

It is easy to see that efficiency involves maximizing $[n_1S_1+n_2S_2]$, subject to Eqn. (5). Because the problem is linear, the maximum occurs at a corner when either n_1 or n_2 is as large as possible, depending on which is larger, S_1/b_1 or S_2/b_2. By assumption, $S_1/b_1 > S_2/b_2$, so $n_1 = \min[c/b_1,N_1]$. This implies n_2 is either zero, or whatever is necessary so that Eqn. (5) holds: $n_2=\max[0,(c-N_1b_1)/b_2]$.

For equity, it is easy to see that $n_j = c/(2b_j)$, provided N_1 and N_2 are sufficiently large. If enough type j countries do not satisfy this condition, then the equity refinement is ill-defined.

These results can be summarized in the following proposition:

Prop. 1. The model of linear payoffs described above has many coalitions that satisfy the conditions for stability articulated in Eqn. (5). Two refinements result in unique coalition sizes. With an efficiency refinement (coalition maximizes coalition payoff), the size is

$$\text{Efficiency:} \quad (n_1,n_2) = (\min[c/b_1,N_1], \max[0,(c-N_1b_1)/b_2]). \tag{8a}$$

With an equity refinement, provided N_1 and N_2 are sufficiently large (i.e., $c/(2b_j) \leq N_j$), then

$$\text{Equity:} \quad (n_1,n_2) = (c/(2b_1), c/(2b_2)). \tag{8b}$$

The interpretation of these two results is straightforward. For the efficiency refinement, the countries with the smaller b_j/S_j constitute the bulk of the coalition (in fact it includes all of the coalition, if N_j is large enough). As described earlier, b_j represents the marginal damage from emissions for the entire country; b_j/S_j is closer to the per capita marginal damage.

In some sense, this is counterintuitive. The countries that suffer the greatest damage from emissions have the largest incentive to do something about it, but this result says that they stay in the fringe. With plenty of type 1 countries, then the coalition will consist entirely of type 1 countries, while all type 2 countries remain in the fringe. On the other hand, type 2 countries have stronger incentives to free ride, which leads them to be in the fringe.

In fact, the result is not at all counterintuitive given the structure of the problem. When countries with smaller marginal damage form a coalition, the stable coalition is larger and thus more effective. It is not surprising that on efficiency grounds, we would want as many low damage countries as possible in the coalition.

For the equity refinement, both types of countries will be in the coalition, and participation is independent of size (S_j) of the country. Countries with smaller marginal damage (b_j) will make up a larger proportion of the coalition.

C. Comparative Statics

The next question is what are the implications of heterogeneity? How does the effectiveness of an IEA change as countries become more heterogeneous? To answer this question, we compare the case of $b_1 = b_2$ to $b_1 < b_2$, preserving the expected value of b. We will hold $S_1 = S_2=1$. In particular, define

$$\gamma_1 = \gamma - \delta/N_1 \tag{9a}$$

$$\gamma_2 = \gamma + \delta/N_2 \tag{9b}$$

For any $\delta \geq 0$, the weighted sum of γ_1 and γ_2 will remain equal to γ. With all countries the same ($\delta = 0$), the equilibrium size of an IEA will be $1/\gamma$, and welfare will be W_0:

$$W_0 = - (N\gamma - 1)^2/\gamma \qquad for\ \delta = 0 \tag{10a}$$

and for the case of $\delta > 0$, assuming N_1 is sufficiently large, the equilibrium IEA will consist of only type 1 countries, including $1/(\gamma - \delta/N_1)$ of them, which is a larger IEA. Although that means more abatement, welfare may or may not be larger. Welfare (for $\delta > 0$) will be (following from Eqn. 4):

$$W_+ \quad = [N-1/(\gamma - \delta/N_1)]\ (N\gamma - 1) \tag{10b}$$

To determine whether W_+ is larger or smaller than W_0, for small δ, differentiate Eqn. (10b) with respect to δ and evaluate at $\delta=0$:

$$d\ W_+/\ d\delta = - (N\gamma - 1)\ (\gamma - \delta/N_1)^{-2}/N_1 \tag{11}$$

Clearly the right hand side of Eqn. (11) is negative when evaluated at $\delta=0$. This implies that welfare decreases with small increases in heterogeneity. Consequently, $W_+ < W_0$.

What this says is that heterogeneity decreases welfare, even though the equilibrium size of the IEA may increase. This result is consistent with prior literature on IEAs. As the problem is characterized here, the IEA is dominated by type 1 countries. So the fact that there are type 2 countries is virtually irrelevant. The prior literature tells us that as γ decreases, the size of an IEA increases but the aggregate welfare decreases. Our result is an extension of this.

III. CONCLUSIONS

Although theoretical models of the formation of international environmental agreements are highly stylized and cannot be used to predict outcomes in real treaty negotiations, they provide useful guidance in how incentives play out in settings in which multiple players must agree to coordinate action to improve the provision of environmental goods. It is in this arena that this paper makes a contribution.

In this paper we examine the important question of how heterogeneity among countries affects the development of international environmental agreements. The question is whether heterogeneity (as opposed to homogeneity) leads to larger or smaller IEAs, to higher or lower levels of aggregate welfare and to more or less abatement. Is heterogeneity a hinderance or a help to forming IEAs to solve environmental problems?

We have developed a simple theory here, extending work by Barrett (2001) on a similar topic. We conclude that heterogeneity of countries may increase the size of IEAs, but it weakens them in terms of the aggregate welfare of all participants. Furthermore, the efficiency criterion for formation of an IEA suggests that the agreement would primarily consist of countries with low per capita marginal damage, due to the nature of the equilibrium (small marginal damage supports a larger coalition). On the other hand, the equity equilibrium refinement suggests a coalition made up of a combination of high and low damage countries.

REFERENCES

Barrett, Scott, "Self-Enforcing International Environmental Agreements," *Oxford Economic Papers*, **46**:878-94 (1994).

Barrett, Scott, "Heterogeneous International Environmental Agreements," in Carlo Carraro (Ed), *International Environmental Negotiations: Strategic Policy Issues* (Edward Elgar, Cheltenham, UK, 1997a).

Barrett, Scott, "The Strategy of Trade Sanctions in International Environmental Agreements," *Resource and Energy Economics*, 19:345-61 (1997b).

Barrett, Scott, "International Cooperation for Sale," *European Economic Review*, **45**:1835-50 (2001).

Barrett, Scott, "Consensus Treaties," *J. Institutional and Theoretical Econ*., **158**:529-47 (2002).

Barrett, Scott, *Environment and Statecraft* (Oxford University Press, Oxford, 2003).

Barrett, Scott, *Why Cooperate? The Incentive to Supply global Public Goods* (Oxford University Press, Oxford, 2007).

Botteon, M. and C. Carraro, "Environmental Coalitions with Heterogeneous Countries: Burden Sharing and Carbon Leakage," in A. Ulph (Ed), *Environmental Policy, International Agreements, and International Trade* (Oxford University Press, New York, 2001).

Burger, Nicholas and Charles Kolstad, "Voluntary Public Goods Provision, Coalition Formation, and Uncertainty," National Bureau of Economic Research Working Paper 15543 (Cambridge, Mass., 2009).

Carraro, Carlo and Domenico Siniscalco, "Strategies for the International Protection of the Environment," *J. Public Econ*., **52**:309-28 (1993).

Chander, Parkash and Henri Tulkens, "Theoretical Foundations of Negotiations and Cost-Sharing in Transfrontier Pollution Problems," *Eur. Econ. Rev.,* **36**:288-99 (1992)

Chander, Parkash and Henri Tulkens, "A Core-theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution," *Intl Tax and Public Finance*, **2**:279-93 (1994).

Cooper, Richard N., "International Cooperation in Public Health as a Prologue to Macroeconomic Cooperation," pp 178-254 in R. N. Cooper et al (ed.), *Can Nations Agree?* (Brookings Institution, Washington, DC, 1989).

d'Aspremont, C., A. Jacquemin, J. Jaskold-Gabszewicz and J. Weymark, "On the Stability of Collusive Price Leadership," *Canadian J. Economics*, **16**:17-25 (1983).

Donsimoni, M.-P., N.S. Economides, and H.M. Polemarcharkis, "Stable Cartels," *International Economic Rev*., **27**:317-27 (1986)

Ecchia, Giulio and Marco Mariotti, "The Stability of International Environmental Coaltions with Farsighted Countries: Some Theoretical Observations," Ch. 10 in Carol Carraro (Ed), *International Environmental Negotiations* (Edward Elgar, Cheltenham, UK, 1997).

Eyckmans, Johan, "On the farsighted stability of international climate agreements," University of Leuven working paper (Jan 2003).

Finus, Michael, *Game Theory and International Environmental Cooperation* (Edward Elgar, Cheltenham, UK, 2001).

Finus, Michael, M. Elena Saiz and Eligius M.T. Hendrix, "An Empirical Test of New Developments in Coalition theory for the Design of International Environmental Agreements," *Env. and Dev. Econ*., **14**:117-37 (2009).

Fuentes-Albero, Cristina and Santiago J. Rubio, "Can International Environmental Cooperation be Bought?," *Eur. J. Operations Research*, **202**:255-64 (2010).

Guesnerie, Roger and Henry Tulkens (Eds), *The Design of Climate Policy* (MIT Press, Cambridge, 2009).

Helm, Carsten, "International Cooperation Behind the Veil of Uncertainty," *Environmental and Resource Economics*, **12**:185-201 (1998).

Hoel, Michael, “International Environmental Conventions: The Case of Uniform Reductions of Emissions,” *Environmental and Resource Economics*, **2**:141-59 (1992).

Isaac, R.M. and J. M. Walker, “Group Size Effects in Public Good Provision: The Voluntary Contributions Mechanism,” *Q. J. Econ*., **103**:179-99 (1988).

Kolstad, Charles D., “Piercing the Veil of Uncertainty in Transboundary Pollution Agreements,” *Env. & Res. Econ*., **31**:21-34 (2005).

Kolstad, Charles D. and Alistair Ulph, “Learning and International Environmental Agreements,” Climatic Change, 89:125-41 (2008).

Kosfeld, Michael, Akira Okada and Orno Riedl, “Institution Formation in Public Goods Games,” *Amer Econ. Rev*., 99:1335-55 (2009).

McGinty, Matthew, “International Environmental Agreements Among Asymmetric Nations,” Oxford Economic Papers, 59:45-62 (2007).

Na, Seong-lin, and Hyun Song Shin, “International Environmental Agreements under Uncertainty,” *Oxford Economic Papers*, **50**:173-85 (1998).

Rubio, Santiago and Alistair Ulph, "An Infinite Horizon Model of Dynamic Membership of International Environmental Agreements", J. Env. Econ. & Mgmt., **54**:296–310 (2007).

Saha, Sarani, “Heterogeneity in International Environmental Agreements,” Ch. 1 in *Three Essays in Environmental and Public Economics*, PhD Dissertation, University of California, Santa Barbara (2007).

Ulph, Alistair, “Stable International Environmental Agreements with a Stock Pollutant, Uncertainty and Learning,” *J. Risk & Uncert*., **29**:53-73 (2004).

Wagner, Ulrich J., “The Design of Stable International Environmental Agreements,” *J. Economic Surveys*, 15:377-411 (2001)

Young, Oran, *International Governance: Protecting the Environment in a Stateless Society* (Cornell University Press, Ithaca, NY, 1994)